



COMPUTATIONAL
INTELLIGENCE TECHNIQUES

ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS, DEMOCRACY, AND THE RULE OF LAW

Edited by
Pushan Kumar Dutta, Bhupinder Singh,
Vishal Jain, Christian Kaunert, and Komal Vig



CRC Press
Taylor & Francis Group



COMPUTATIONAL
INTELLIGENCE TECHNIQUES

ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS, DEMOCRACY, AND THE RULE OF LAW

Edited by
Pushan Kumar Dutta, Bhupinder Singh,
Vishal Jain, Christian Kaunert, and Komal Vig



CRC Press
Taylor & Francis Group

Artificial Intelligence and Human Rights, Democracy, and the Rule of Law

This book introduces a comprehensive model for evaluating the effects of artificial intelligence (AI) on human rights, democracy, and the rule of law and an innovative framework that merges Human Rights Impact Assessments (HRIAs) practices with AI-focused algorithmic impact assessment approaches. The book is divided into three subthemes: Understanding AI, Its Concepts and Implications; Challenges and Solutions in AI Systems; and AI Applications, Compliance Issues, and Future Prospects.

Features:

- Emphasizes the importance of integrating HRIAs with AI-centered approaches.
- Evaluates impacts of AI on human rights, democracy, and the rule of law to manage potential risks and impacts of AI.
- Underscores the necessity of a methodology that follows the proportionality principle.
- Highlights the importance of engaging with stakeholders throughout the impact assessment process.
- Provides real-world applications of computational intelligence in the HUDERAF context.

This book is aimed at graduate students and researchers in artificial intelligence, computer science, cybersecurity, and human rights studies.

Computational Intelligence Techniques

Series Editor: Vishal Jain

The objective of this series is to provide researchers a platform to present state of the art innovations, research, and design and implement methodological and algorithmic solutions to data processing problems, designing and analyzing evolving trends in health informatics and computer-aided diagnosis. This series provides support and aid to researchers involved in designing decision support systems that will permit societal acceptance of ambient intelligence. The overall goal of this series is to present the latest snapshot of ongoing research as well as to shed further light on future directions in this space. The series presents novel technical studies as well as position and vision papers comprising hypothetical/speculative scenarios. The book series seeks to compile all aspects of computational intelligence techniques from fundamental principles to current advanced concepts. For this series, we invite researchers, academicians and professionals to contribute, expressing their ideas and research in the application of intelligent techniques to the field of engineering in handbook, reference, or monograph volumes.

Big Data Computing: Advances in Technologies, Methodologies, and Applications

Edited by Tanvir Habib Sardar and Bishwajeet Kumar Pandey

Software Defined Network Frameworks: Security Issues and Use Cases

Edited by Mandeep Kaur, Vishal Jain, Parma Nand and Nitin Rakesh

Machine Learning Hybridization and Optimization for Intelligent Applications

Edited by Tanvir Habib Sardar and Bishwajeet Kumar Pandey

Information Technology Ethics

Hamed Taherdoost

Applications of Blockchain and Computational Intelligence in Environmental Sustainability

Edited by Hamed Taherdoost, Mohsen Saeedi and Aydin Shishegaran

AI and Fintech: Improving the Financial Landscape

Edited by Jaheer Mukthar K. P., Rosario Mercedes Huerta Soto, Vishal Jain and Edwin Hernan Ramirez Asis

Artificial Intelligence and Human Rights, Democracy, and the Rule of Law

Edited by Pushan Kumar Dutta, Bhupinder Singh, Vishal Jain, Christian Kaunert, and Komal Vig

For more information about this series, please visit:

www.routledge.com/Computational-Intelligence-Techniques/book-series/CIT

Artificial Intelligence and Human Rights, Democracy, and the Rule of Law

Edited by

Pushan Kumar Dutta, Bhupinder Singh, Vishal Jain,
Christian Kaunert, and Komal Vig



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Designed cover image: Shutterstock

First edition published 2026

by CRC Press

2385 NW Executive Center Drive, Suite 320, Boca Raton FL 33431

and by CRC Press

4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

© 2026 selection and editorial matter, Pushan Kumar Dutta, Bhupinder Singh, Vishal Jain, Christian Kaunert, and Komal Vig; individual chapters, the contributors

Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, access www.copyright.com or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. For works that are not available on CCC please contact mpkbookspermissions@tandf.co.uk

For Product Safety Concerns and Information please contact our EU representative GPSR@taylorandfrancis.com. Taylor & Francis Verlag GmbH, Kaufingerstraße 24, 80331 München, Germany.

Trademark notice: Product or corporate names may be trademarks or registered trademarks and are used only for identification and explanation without intent to infringe.

ISBN: 9781032796635 (hbk)

ISBN: 9781041132127 (pbk)

ISBN: 9781003668657 (ebk)

DOI: [10.1201/9781003668657](https://doi.org/10.1201/9781003668657)

Typeset in Times
by Newgen Publishing UK

Contents

[Preface](#)

[Editors' Biography](#)

[List of Contributors](#)

Chapter 1 [AI and Human Rights: Future Implications and Ethical Considerations](#)

Kamalesh Ravesangar, Tarun Kumar Kaushik, Ravish, Anurag Singh, Anjali Raghav, Bhupinder Singh, and Saurabh Chandra

Chapter 2 [Assessing AI-Driven Human Rights Violations: A Framework for Accountability](#)

Debashree Chakravarty, Ipseeta Satpathy, Vishal Jain, and B.C.M. Patnaik

Chapter 3 [Computational Intelligence in Human Rights Protection: A Review and Future Directions](#)

Budi Agus Riswandi, Saquib Ahmed, Anjali Raghav, Bhupinder Singh, Atul Sangal, Manmeet Kaur Arora, Sahil Lal, and Saurabh Chandra

Chapter 4 [The Economic Potential of AI: Balancing Growth with Human Rights and Ethics](#)

Kushagra Shukla and Rashmi S

Chapter 5 [AI, Data Privacy, and the Personal Data Protection Act: Implications for Human Rights](#)

Soorya Balendra

- Chapter 6** [AI Ethics and Colonial Legacies: Evaluating UNESCO's Recommendations for Inclusivity](#)
Akash Bag, Anwesha Ghosh, Tejaswini Tripathy, and Piotr Pietrzak
- Chapter 7** [The Role of AI in the Legal System: Implications for Justice and Rights](#)
Geetha Manoharan and M Nivedha
- Chapter 8** [AI in Emotional Intelligence and Leadership: Enhancing Accountability in Business Practices](#)
Gaurav Kulat, Neeta Shah, Kanchan Naidu, Gayathri Band, and Rupesh Pais
- Chapter 9** [AI and Human Rights: Guardians of Freedom or Threats to Democracy?](#)
Prashasti Pritiprada, Ipseeta Satpathy, Christian Kaunert, and B.C.M. Patnaik
- Chapter 10** [Big Data, Privacy, and the Rights of Senior Citizens under the Digital Personal Data Protection Act](#)
Shayut Pavapanunkul, Aditi Srivastava, and Ritu Gautam
- Chapter 11** [AI-Driven Cybersecurity: Protecting Human Rights in the Digital Age](#)
T. Manonmani, S.N. Kumar, Jomin Joy, and Anandhu Venu
- Chapter 12** [AI and Inclusive Democracy: Transforming Governance for Equitable Societies](#)
Kushanthi S. Harasgama, Vaishali Arora, Manvendra Singh, and Tahir Qureshi

Chapter 13 [AI in Law Enforcement and Counter-Terrorism: Supporting SDG 16 for Human Rights](#)

Bhupinder Singh, Christian Kaunert, Saquib Ahmed, Anjali Raghav, and Saurabh Chandra

Chapter 14 [AI in Cybersecurity: Safeguarding Cyberspace and Human Rights](#)

Sakthivel M, Suresh Kumar K, Joshua Bapu J, SwathyPriyadharsini P, Girimurugan R, and Raghavendran N

Chapter 15 [Ethics of AI-Driven Targeted Advertising: Safeguarding Consumer Rights](#)

Gigimon V S and Darsana R S

Chapter 16 [AI-Generated Content and Copyright: Legal and Ethical Challenges in India](#)

Hammouch Hind, Sushma Singh, and Anushka Singh

Chapter 17 [AI's Role in Cybersecurity: Balancing Innovation with Privacy and Rights](#)

Rahul Joshi, Suman Kumari, Krishna Pandey, Laith Farhan, and Amr Badr

Chapter 18 [Transforming Education: The Role of AI-Driven Chatbots in Advancing Equal Access, Accountability, and Empowerment in Educational Systems](#)

Santu Karmakar, Ankur Nandi, Tapash Das, and Tarini Halder

Chapter 19 [AI, Consent, and GDPR: Ensuring Privacy and Rights in the Digital Age](#)

Kittisak Jermsittiparsert, Saurabh Chandra, and Bhumika Muchan

Chapter 20 [AI and IoT in Cybercrime Investigation: Protecting Human Rights in the Digital Sphere](#)

Laeq Razzak Janua, Bhupinder Singh, Saquib Ahmed, Anjali Raghav, Sahil Lal, Manmeet Kaur Arora and Saurabh Chandra

Chapter 21 [Role of Artificial Intelligence in Protection of Human Rights with Special Reference to International Law](#)

Tarun Kumar Kaushik, Saquib Ahmed, Ravish, Anurag Singh, Anjali Raghav, and Saurabh Chandra

Chapter 22 [Facial Recognition, Civil Rights, and AI: Legal and Ethical Considerations](#)

Nilanjan Chakraborty, Hamideza Rabiei Dastjerdi, and Khder Alakkari

Chapter 23 [Impact Analysis of Emotional Intelligence by AI on Sales Team Performance: Case Study of ICICI Group Companies in Nagpur](#)

Gaurav Kulat, Neeta Shah, Kanchan Naidu, and Gayathri Band

Chapter 24 [AI and Effect of Digital Transformation on Banking Audits and Compliance Management](#)

Devendra Kakwani, Kanchan Naidu, and Gayathri Band

[Index](#)

Preface

The 21st century world presents us with unprecedented opportunities and challenges at a unique intersection—the nexus of human rights, democracy, rule of law, and computational intelligence. This preface elaborates the deep repercussions and revolutionary possibilities of embedding frontier computational intelligence into the fundamentals that set up for human civilization yet long ago. Justice, equality, and freedom are built on the foundation of human rights, democracy, and the rule of law. As we move into an age of lightning-fast technological change, the baking in of computational intelligence requires a re-examination. Focusing on steps taken by national sovereign will help to amplify the advocacy reach of human rights: In this area, computational intelligence (e.g., artificial intelligence, AI) can have an impact as much for its potential applications in terms of enhancing surveillance and scrutiny but also reinforcing into guard points; through expanding access information while same time ensuring that machines do not act under certain setting concepts which are divided upon under rule law. This introduction of AI and machine learning (ML) into governance as well as legal systems throws up some ethical in addition to legal issues. As algorithms and automated systems start to shape decisions that have been the realm of human discretion, inherent problems with privacy, accountability, and transparency come into play. These biases that are found in AI systems threaten the equitable application of laws and safeguarding individual rights. Despite this, the possibilities of helping human rights and democracy to develop or turning democracies into states based on law may be so big that we need increasingly sophisticated methods for them all to have a fair chance! AI-driven analytics can be used to reveal human rights abuses, watch over the electoral process to ensure fair outcomes, and help streamline judicial processes, making them more efficient and accessible. In addition, the democratization of access to

information through digital platforms can give citizens jurisdictional participation and transparency. The purpose of this book is to introduce an extended qualitative study that interrogates these themes, advocating for a human rights-respecting approach towards Computational Intelligence (CI). As we continue to navigate this era of transformation, it is vital that we build strong and protective structures so that technology becomes a tool for good, equipped in the hands of everyone with justice and equal opportunity rather than perpetuating injustice, division, or oppression.

Editors' Biography

Pushan Kumar Dutta is an Associate Professor at Amity University Kolkata who has carved a distinctive niche in academia with over 12 years of academic experience. He earned his Ph.D. from Jadavpur University in 2015, followed by postdoctoral work as an Erasmus Mundus Scholar at the University of Oradea in Romania in 2016, which laid the foundation for a career drive for a position in institution of academic excellence. At the forefront of his contributions is an impressive editorial portfolio. Dr. Dutta has meticulously curated and edited over 60 books in the past couple of years for globally recognized publishers such as Springer, Elsevier, IET, Taylor & Francis, Cambridge Scholar Publishers, Degruyter, and IGI Global. His role as the manager of the “Sustainable Industrial Engineering Systems” book series for CRC Press further underscores his influence in shaping scholarly discourse on sustainability and engineering innovation. His editorial expertise spans rapidly evolving domains including artificial intelligence in healthcare, Industry 5.0 technologies, quantum computing applications, aviation and digital transformation strategies. A dedicated mentor and educator, Dr. Dutta has consistently empowered his students to excel and innovate, guiding them to garner international recognition through competitions like the Johnson Control International Whitepaper contest and Ideathon 5.0. His contributions to nurturing young talent have been acknowledged through accolades such as the “Mentor of Change” award by NITI Aayog, along with the prestigious Young Faculty Awards in 2018 and 2024. His keynote speeches at international events, notably at Data Science Europe 2022, reflect his role as a thought leader. Complementing his teaching and editorial endeavors is an extensive research portfolio that includes over 200 Scopus-indexed publications, two pioneering Indian patents, and an Indian copyright. His professional recognitions extend to prestigious memberships and fellowships, including

Senior Membership at the Indian Institute of Engineering, a Threves Fellowship, and a Fellowship at the EnTech Foundation. Dr. Dutta's academic journey also includes significant teaching stints as an Assistant Professor at RGM CET in Nandyal, Andhra Pradesh, and as a lecturer for the Marine Electro Technology course at DG Shipping Taratala. His diverse experience spans classroom instruction in Data Science and AI, comprehensive courses on Electrical, Electronics, Python and coding and mentorship roles that integrates design thinking and hackathon participation and sports.

Bhupinder Singh works as Professor at Sharda School of Law, Sharda University Greater Noida, India. He is also Honorary Professor at University of South Wales UK and Santo Tomas University Tunja, Colombia. He has 3 books, 97 paper publications, 163 paper presentations in international/national conferences and seminars, participated in more than 40 workshops/FDP's/QIP's, 25 courses from international universities of repute, organized more than 59 events with international and national academicians and industry people's, editor-in-chief and co-editor in journals, developed new courses. He has given talks at international universities, resource person in international conferences such as in Nanyang Technological University Singapore, Tashkent State University of Law Uzbekistan; KIMEP University Kazakhstan, All'ah meh Tabatabai University Iran, the Iranian Association of International Criminal law, Iran and Hague Center for International Law and Investment, The Netherlands, Northumbria University Newcastle UK, Taylor's University Malaysia, AFM Krakow University Poland, European Institute for Research and Development Georgia, Business and Technology University Georgia, Texas A&M University US, to name a few. His leadership, teaching, research, and industry experience is of 16 years and 3 months. His research interests are health law, criminal law, research methodology, and emerging multidisciplinary areas such as Blockchain Technology, IoT, Machine Learning, Artificial Intelligence, Genome-editing, Photovoltaic PV Glass, SDG's and many more.

Vishal Jain is presently working as a Professor (CSE) at the Department of Computer Science & Engineering, School of Engineering & Technology, Vivekananda Institute of Professional Studies—Technical Campus, New Delhi. Before that, he worked as Professor at Sharda University, Greater Noida and Associate Professor at Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi. He is associated as a member of the Faculty Board, Board of Planning and Management, Executive Council, Academic Council, Curriculum Development and Review Committee and Academic Audit of various higher education Institutions. He is associated with Kuala Lumpur University of Science and Technology (KLUST) (formerly known as Infrastructure University Kuala Lumpur (IUKL)), Malaysia, Lincoln University College, Malaysia and Research Institute of IoT Cybersecurity, Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Taiwan for research-oriented activities. He has supervised three PhD and two scholars are pursuing. He has more than 3300 research citations with Google Scholar (h-index score 30 and i-10 index 66) and has authored more than 250 research papers in professional journals and conferences. He has authored and edited more than 70 books (most of them are indexed at the Scopus) with various reputed publishers, including Elsevier, Springer, IET, Apple Academic Press, CRC, Taylor and Francis Group, Scrivener, Wiley, Emerald, NOVA Science, River Publishers, IGI-Global, and Bentham Science. He has more than 300 Scopus publications. He is the book series editor of 10 book series with the reputed international publishers. Along with the regular teaching, he had organized various conferences, faculty development programs, seminars, hackathon, faculty and student-oriented activities. His research areas include machine learning, information retrieval, semantic web, ontology engineering, data mining, ad hoc networks, sensor networks, and network security. He received a Young Active Member Award for the year 2012–2013 from the Computer Society of India, and Best Faculty Award for the year 2017 and Best Researcher Award for the year 2019 from BVICAM, New Delhi.

Christian Kaunert is Professor of International Security at Dublin City University, Ireland. He is also Professor of Policing and Security, as well as Director of the International Centre for Policing and Security at the University of South Wales. In addition, he is Jean Monnet Chair, Director of the Jean Monnet Centre of Excellence and Director of the Jean Monnet Network on EU Counter-Terrorism (www.eucter.net). Previously, he served as Academic Director and Professor at the Institute for European Studies, Vrije Universiteit Brussel, a Professor of International Politics, Head of Discipline in Politics, and the Director of the European Institute for Security and Justice, a Jean Monnet Centre for Excellence, at the University of Dundee. Kaunert has researched and taught in many international universities, such as University of Shandong, Jinan, China, IBEI Barcelona, Spain, Diplomatic Academy Yerevan, Armenia—United Nations Development Mission, University of Cairo, Egypt, Collegio Carlo Alberto, Turin, Italy, Ecole nationale d'administration Paris, European Institute of Public Administration Maastricht, Netherlands, European Studies Institute (ESI) at Moscow, University of Porto, Portugal, etc. He was previously Senior Lecturer at the University of Dundee, Marie Curie Senior Research Fellow at the European University Institute Florence, and Senior Lecturer in EU Politics & International Relations, University of Salford.

Komal Vig is working as the Director of Delhi Metropolitan Education, Noida, India. She is LLD in Criminal Law and MBA (HR). Komal Vig has served in the field of academics in different capacities for more than 21 years. She has published more than 95 research papers in International and National Journals (including 23 Scopus), 20+ book chapters, 30+ conference paper publications. She supervised six scholars who have successfully completed PhD under her guidance and many others are pursuing. She has presented more than 65 research papers in national and international seminars/conferences. Her area of interest is Criminal law, Family law, and laws related to Women's rights. Komal Vig is a member of the inspection committees of various governing and accrediting bodies. Komal Vig has chaired numerous sessions in international and national

conferences/seminars. She is active member of Board of Studies of various universities, received Life membership of the Indian law Institute (Deemed University) New Delhi, “Women Empowerment Award—2023” from Council of Research and Sustainable Development, India.

Contributors

Saquib Ahmed

Sharda School of Law

Sharda University

Greater Noida, Uttar Pradesh, India

Khder Alakkari

Faculty of Economics

Tishreen University

Latakia, Syria

Manmeet Kaur Arora

Sharda School of Law

Sharda University

Greater Noida, Uttar Pradesh, India

Vaishali Arora

Sharda School of Law

Sharda University

Greater Noida, Uttar Pradesh, India

Amr Badr

Faculty of Science, School of Science and Technology

University of New England

Armidale, NSW, Australia

Akash Bag

Adamas University

Kolkata, West Bengal, India

Soorya Balendra

Centre for Human Rights and Legal Pluralism, Faculty of Law
McGill University
Montreal, Quebec, Canada

Gayathri Band

Shri Ramdeobaba College of Engineering and Management
Nagpur, Maharashtra, India

Nilanjan Chakraborty

Amity Law School
Amity University Kolkata
Kolkata, West Bengal, India

Debashree Chakravarty

School of Management
KIIT University
Bhubaneswar, Odisha, India
Sharda University
Greater Noida, Uttar Pradesh, India

Saurabh Chandra

School of Law, Bennett University
Greater Noida, Uttar Pradesh, India

Tapash Das

Kazi Nazrul University
Asansol, West Bengal, India

Laith Farhan

Manchester Metropolitan University
England, UK

Ritu Gautam

School of Law, Sharda University
Greater Noida, Uttar Pradesh, India

Anwasha Ghosh

Rajiv Gandhi National University of Law
Patiala, Punjab, India

Tarini Halder

University of Kalyani
Kalyani, West Bengal, India

Kushanthi S. Harasgama

Sri Lanka Institute of Information Technology, Sri Lanka

Hammouch Hind

University Sidi Mohamed Ben Abdellah, Fez, Morocco

Joshua Bapu J

Department of Electronics and Communication Engineering
Dr. Sivanthi Aditanar College of Engineering
Tiruchendur, Tamil Nadu, India

Vishal Jain

Department of Computer Science & Engineering, School of Engineering and
Technology
Vivekananda Institute of Professional Studies—Technical Campus
New Delhi, India

Laeq Razzak Janua

WSB University, Poland

Kittisak Jermsittiparsert

Faculty of Political Science, North Bangkok University
Bangkok, Thailand

Rahul Joshi

Department of Journalism & Mass Communication, School of Media
Studies & Humanities
Manav Rachna International Institute of Research & Studies
Faridabad, Haryana, India

Jomin Joy

Amal Jyothi College of Engineering
Kanjirappally, Kerala, India

Suresh Kumar K

Department of MBA
Panimalar Engineering College
Chennai, Tamil Nadu, India

Devdendra Kakwani

Shri Ramdeobaba College of Engineering and Management
Nagpur, Maharashtra, India

Santu Karmakar

Kazi Nazrul University
Asansol, West Bengal, India

Christian Kaunert

Dublin City University
Dublin, Ireland

Gaurav Kulat

Shri Ramdeobaba College of Engineering and Management
Nagpur, Maharashtra, India

Tarun Kumar Kaushik

Sharda University
Greater Noida, Uttar Pradesh, India

S.N. Kumar

Department of CSE
Mepco Schlenk Engineering College
Sivakasi, Tamil Nadu, India

Suman Kumari

Department of Journalism & Mass Communication,
School of Media Studies & Humanities
Manav Rachna International Institute of Research & Studies
Faridabad, Haryana, India

Sahil Lal

School of Law, Galgotias University
Greater Noida, Uttar Pradesh
India

Sakthivel M

Department of Artificial Intelligence and Data Science
Sri Shanmugha College of Engineering and Technology
Sankari, Tamil Nadu, India

Geetha Manoharan

School of Business
SR University

Warangal, Telangana, India

T. Manonmani

Department of CSE

Mepco Schlenk Engineering College

Sivakasi, Tamil Nadu, India

Bhumika Muchan

Advocate

Delhi High Court

New Delhi, India

Raghavendran N

Department of Computer Science and Engineering

Francis Xavier Engineering College

Tirunelveli, Tamil Nadu, India

Kanchan Naidu

Shri Ramdeobaba College of Engineering and Management

Nagpur, Maharashtra, India

Ankur Nandi

University of Kalyani

Kalyani, West Bengal, India

M Nivedha

School of Creative & Cultural Business

Robert Gordon University

Aberdeen, Scotland, United Kingdom

Swathypriyadharsini P

Department of Computer Science and Engineering

Bannari Amman Institute of Technology
Sathyamangalam, Tamil Nadu, India

Krishna Pandey

Department of Journalism & Mass Communication, School of Media
Studies & Humanities
Manav Rachna International Institute of Research & Studies
Faridabad, Haryana, India

Rupesh Pais

Shri Ramdeobaba College of Engineering and Management
Nagpur, Maharashtra, India

B.C.M. Patnaik

School of Management
KIIT University
Bhubaneswar, Odisha, India

Shayut Pavapanunkul

Universitas Muhammadiyah Sidenreng Rappang
Indonesia

Piotr Pietrzak

Sofia University St. Kliment Ohridski, Sofia, Bulgaria and In Statu Nascendi
Think Tank, Sofia, Bulgaria

Prashasti Pritiprada

KIIT School of Management
KIIT University
Bhubaneswar, Odisha, India

Tahir Qureshi

GN Groups of Institute

Greater Noida, Uttar Pradesh, India

Girimurugan R

Department of Mechanical Engineering
Nandha College of Technology
Perundurai, Tamil Nadu, India

Darsana R S

DPIIT IPR Chair Maharashtra National Law University Mumbai
Mumbai, Maharashtra, India

Hamideza Rabiei Dastjerdi

Faculty of Humanities and Social Sciences
Dublin College University (DCU)
Dublin, Ireland

Anjali Raghav

Sharda School of Law, Sharda University
Greater Noida, Uttar Pradesh, India

Kamalesh Ravesangar

Tunku Abdul Rahman University of Management and Technology
Malaysia

Ravish

Department of BCom LLB
Meerut College
Meerut, Uttar Pradesh, India

Budi Agus Riswandi

Faculty of Law
Universitas Islam Indonesia, Indonesia

Rashmi S

Dayananda Sagar College of Engineering
Bengaluru, Karnataka, India

Atul Sangal

Sharda University
Greater Noida, Uttar Pradesh
India

Ipseeta Satpathy

KIIT School of Management
KIIT University
Bhubaneswar, Odisha, India

Neeta Shah

Shri Ramdeobaba College of Engineering and Management
Nagpur, Maharashtra, India

Kushagra Shukla

Dayananda Sagar College of Engineering
Bengaluru, Karnataka, India
Sharda University
Greater Noida, Uttar Pradesh, India

Anurag Singh

Department of Law
Meerut College
Meerut, Uttar Pradesh, India

Anushka Singh

LLM, Center of Post Graduate Legal Studies
Babasaheb Bhimrao Ambedkar (Central) University

Lucknow, Uttar Pradesh, India

Manvendra Singh

Sharda School of Law

Sharda University

Greater Noida, Uttar Pradesh, India

Sushma Singh

School of Law

Sharda University

Greater Noida, Uttar Pradesh, India

Aditi Srivastava

Sharda University

Greater Noida, Uttar Pradesh, India

Tejaswini Tripathy

Madhusudan Law University

Cuttack, Odisha, India

Gigimon V S

Amity Law School

Amity University Mumbai

Mumbai, Maharashtra, India

Anandhu Venu

Dublin City University

Dublin, Ireland

1 AI and Human Rights

Future Implications and Ethical Considerations

*Kamalesh Ravesangar, Tarun Kumar Kaushik,
Ravish, Anurag Singh, Anjali Raghav,
Bhupinder Singh, and Saurabh Chandra*

DOI: [10.1201/9781003668657-1](https://doi.org/10.1201/9781003668657-1)

1.1 INTRODUCTION

All people have an inherent right to a certain set of liberties, protections, and opportunities guaranteed to them by the society in which they are born and raised, as well as by principles of modernity and international law. No one can sustain their humanity in the absence of human rights, which are so essential to the human condition. A life devoid of human rights is a shadow of what it could be. To put it another way, human rights are too valuable to violate unless there is enough evidence to support doing so. Although it is convenient to think about “artificial intelligence” (AI) in terms of a definition, it should be mentioned that no universally accepted definition has been reached yet. The first of several definitions of AI was proposed in 1956 as part of the Dartmouth Summer Research Project on AI. An early proponent of the field, John McCarthy, deemed “intelligent” any system that could carry out tasks that would be considered intelligent if carried out

by a person. In this view, AI consists of nothing more than a system or computer that can mimic human performance in any given job. As per a recent report from Stanford University, AI is described as “a science and a set of computational technologies that are inspired by but typically operate quite differently from the ways people use their nervous systems and bodies to sense, learn, reason, and take action.” Another way to look at AI is according to Stuart Russell and Peter Norving, who propose the following classifications: First, systems that mimic human thought and behavior; second, systems that think and behave logically; and last, systems that mimic human thought and behavior without acting rationally. The common thread running across all of these definitions is that AI systems exist primarily to make human jobs easier. The chapter uses a qualitative approach, using case studies, reports from international organizations, and legal analyses. It evaluates current AI technologies, their implementation, and their ethical implications by drawing on a variety of interdisciplinary sources, including legal frameworks, academic literature, and expert opinions. The chapter finds that while AI can significantly enhance efficiency in monitoring and enforcing human rights, it also raises concerns about privacy, bias, and accountability. The chapter argues that existing legal structures are insufficient to address the complex issues AI presents, and it calls for enhanced international collaboration to develop comprehensive regulations that ensure AI technologies uphold human rights standards.

In essence, it is impossible to detect the difference between a human-performed procedure and the outcome of an activity executed by an intelligent system. In light of this, the rights of humans must be considered in any interaction between an AI-generated intelligent system and humans. Put simply, the development of AI will unavoidably have an impact on human rights, whether that impact is beneficial or negative.

As a matter of right in the nation or community in which they reside, all human beings are entitled to enjoy certain freedoms, immunities, and advantages according to natural law, contemporary ideals, and international law. Human rights are essential to every person’s ability to be human;

without them, people would be unable to carry on living as humans. Life loses all significance, value, and substance in the absence of human rights. To put it another way, human rights are too valuable to violate unless there is enough evidence to support doing so ([Djenna et al., 2023](#)).

One of the primary goals of the international legal system is to ensure the preservation of human rights. After World War I (WWI), the preservation of human rights jumped to the forefront of global politics. Since 1945, a collection of human rights recognized on a global scale has been established via the adoption of various instruments, including treaties. The international human rights framework encompasses a vast array of concerns, including economic, social, cultural, and political rights in addition to civil and political rights ([Awajan, 2023](#)). States are obligated to uphold these rights according to the United Nations Charter and other fundamental international human rights agreements, human rights as outlined in the Universal Declaration of Human Rights (UDHR) and associated United Nations (UN) system, no less than nine multinational human rights accords that followed, and regional human rights treaties, which need bringing the rights and responsibilities guaranteed by them home. The intricate nature of any system necessitates the participation of governments at different levels. Although the specific reporting obligations may differ between instruments or bodies, they nevertheless share the need for nations to keep tabs on the human rights they guarantee ([Attaran & Deb, 2018](#)).

Consequently, there are significant issues with accountability and enforcement around international human rights legislation. As international law is largely derived from treaties and customs, it is fundamentally based on the principle of consent. According to the prevailing legal positivist view, states are independent entities that are only obligated to adhere to the legal standards and requirements that they voluntarily accept. There is persistent and habitual noncompliance with international law even when nations agree to be bound by it; states both fail to safeguard human rights and infringe upon those rights of persons and communities ([Brotcke, 2022](#)).

1.2 ARTIFICIAL INTELLIGENCE

A simple definition of AI is not achievable. The capacity of a digital computer or computer-controlled robot to carry out activities often linked to intelligent creatures is how AI is defined by Encyclopaedia Britannica. AI is more of a term of art than a specific definable “thing” ([Abebe et al., 2020](#)).

There are four main categories into which AI is usually classified: (1) reactive machines, (2) limited memory, (3) theory of mind, and (4) self-awareness. Both limited memory and reactive machines (where AI can absorb inputs and apply predictive modelling to “learn” from them) account for the vast majority of current AI technology. A characteristic of technology with limited memory is its capacity to take in and improve upon massive volumes of training data ([Elluri et al., 2023](#)).

Different, and frequently overlapping, criteria can be used to classify AI technology. Robotics, computer vision, deep learning, automation, and machine learning (ML) are all examples of technology that fall under these broad categories ([Hagendorff, 2019](#)). ML is described by Oxford Language as “the use and development of computer systems that can learn and adapt without following explicit instructions, by analysing and drawing inferences from patterns in data using algorithms and statistical models” ([Singh & Kaunert, 2024](#)). AI system “learns” by continuously enhancing its performance through the analysis of larger datasets and the identification of additional patterns within that data, which ultimately leads to more accurate automated decision-making ([Thapaliya & Sharma, 2023](#)). Deep learning is a specific branch of ML that uses artificial neural networks to identify patterns and correlations in data. This is achieved by organizing networks into many layers. Deep learning is commonly applied in the fields of image and speech recognition ([Dash et al., 2022](#)).

It is crucial to clearly define the distinction between AI and data right from the beginning. These two concepts are often mistaken or mixed together, but it is crucial to understand the difference between them and their interplay ([Singh & Kaunert, 2024](#)). For AI technologies to be

developed and implemented efficiently, they require access to extensive data sets from which they can learn. Each algorithm possesses an input and output, but, with ML, computers autonomously generate their own programs, so relieving us from the need to do so. For machine education, Prioritizing the data is of utmost importance. The creation of an ML algorithm relies on extensive amounts of data, from which the learning process extracts numerous entities, connections, and clusters ([Barocas et al., 2023](#)).

1.3 AI AND HUMAN RIGHTS

Learning about the many ways AI affects human rights—the obstacles, weaknesses, and damages—have been a major subject of interest. Security of personal information and bias have been two main areas of focus in legal discussions. A particular study examined a range of negative consequences, including but not limited to: absence of transparency in algorithms, vulnerabilities in cyber-security, unfairness, bias, and discrimination, absence of competition, legal recognition of AI as a legal entity, intellectual property concerns, negative impact on workers, issues related to privacy and data protection, liability for damages, and absence of responsibility for the negative consequences. The report offers a valuable synopsis of the many forms of harm that are prevalent at the intersection of AI and human rights ([Narayana & Sreedevi, 2023](#)).

AI and its effects on human rights have been examined by researchers in several sectors. As an illustration, a group of experts in ethics and governance from Harvard's Berkman Klein Centre for Internet and Society conducted a comprehensive analysis of the effects of AI systems on human rights in the following six areas of activity: criminal justice, financial system accessibility, healthcare, online content moderation, human resources, and education ([Mhlanga, 2021](#)).

1.4 SUBMISSIONS OF AI IN HUMAN RIGHTS REGULATORS

In the framework of international human rights law, civil society organizations have mostly utilized AI to keep tabs on, record, and report on the exercise and, more importantly, the infringement of human rights ([Pocher et al., 2023](#)). The international human rights arena is already making use of AI in a number of examples ([Chen et al., 2018](#)). Various civil society organizations have created and implemented technology to monitor and track breaches or progress on the enjoyment of certain rights. These examples demonstrate the approaches ([Tyagi et al., 2020](#)).

1. Measuring the devastation in Darfur villages.
2. Predicting global migration with the use of AI.
3. Keeping tabs on capital cases through media monitoring.
4. AI-based monitoring of deforestation.
5. Monitoring ethnic violence in Myanmar with the use of thermal data.
6. Machine learning for monitoring Twitter harassment of women.
7. AI as a tool for increasing language accessibility: translation services and multilingual chatbots.
8. The use of AI as a tool to reduce the impact of trauma exposure and mitigate mental health consequences in human rights workers ([Alazab et al., 2023](#)).

AI may be utilized in the following ways within the framework of international human rights legislation:

- Improve efficiency and output by rapidly sorting through and making sense of vast volumes of data and information; faster and more effective data collecting, processing, and analysis ([Zhang et al., 2022b](#));
- For the purpose of gathering data and writing reports;
- To improve trend forecasting and prediction, which aids nations and organizations in their strategic planning and resource allocation

initiatives;

- To improve the ability of states and organizations to grasp data and information by facilitating a more thorough comprehension of data across demographic markers and rights ([Singh, 2023b](#));
- To provide improved data monitoring over time;
- For the purpose of data disaggregation, to learn how different communities, groups, and subsets of the population are disproportionately affected by the enjoyment (or violation) of rights ([Lee & Shin, 2020](#));
- To flush out data duplication and find missing pieces of open-source information;
- By comparing different datasets, reports and data may be validated and authenticated.

Both governments and civil society organizations can benefit from using AI to track and report on human rights issues in real time and over extended periods of time. Furthermore, these reports may be utilized in a combative manner to ensure that governments are held responsible for human rights violations ([Bhuvaneshwari & Kaythry, 2023](#)).

1.5 RESTRICTIONS AND THREATS

When it comes to AI, many people are afraid. Michelle Bachelet, the United Nations High Commissioner for Human Rights, delivered a passionate speech to the Committee on Legal Affairs and Human Rights of the Council of Europe in September 2021, urging states to halt the use of AI until adequate protections are established to avoid human rights abuses. The probably inevitable human rights ramifications will be dealt with after the fact, and we cannot afford to keep playing catch-up when it comes to AI. Its use should be regulated or supervised. We cannot deny AI's usefulness to humanity, but we cannot ignore its capacity to fuel massive, covert human rights atrocities either. For our collective benefit, immediate measures are required to establish human rights safeguards for AI. It is always possible

for nations to use the same technology for their own ends, such as increasing human rights breaches or avoiding accountability ([Li et al., 2020](#)).

1.5.1 INVESTIGATING THE LIMITS OF AI IN THE AREA OF INTERNATIONAL HUMAN RIGHTS: THE DATA PROBLEM

Access to data and data completeness are the most apparent limitations of AI and ML in the human rights domain. In many cases, the lack of sufficient or appropriate data is more of a barrier to AI adoption than the development of the AI tool or algorithm itself. The ability of algorithms to provide correct findings and meaningful outputs is dependent on the availability of large amounts of data from which they may “learn”—this is the foundation upon which AI rests. In the case of human rights reporting and monitoring, this issue may be particularly acute because both non-state and state actors may lack easy access to the data needed to train the AI ([Cheng et al., 2021](#)).

State actors who self-report, international human rights non-governmental organizations (NGOs; such as Amnesty International and Human Rights Watch), and international institutions (such as the World Bank and the International Monetary Fund) are the three main sources of data used by the international human rights system. Data accuracy and completeness are problems everywhere. Worldwide, human rights groups have, for the most part, served as the go-to for details about human rights abuses ([Bhuvaneshwari & Kaythry, 2023](#)).

1.5.2 ANALYZING DATA AND MAKING DECISIONS

The capacity of existing AI systems to analyze and understand qualitative data that may provide light on the causes and context of human rights abuses is severely lacking. While AI can provide us with information about what is occurring, it cannot explain why. Many international human rights

legislation treaties include cultural relativism, the view that human rights must account for cultural variations. Thus, AI might not always be the best choice ([Mühlhoff, 2021](#)).

It could be useful to use an example to show where AI's capabilities end and its limitations begin. A mapping project to document and quantify human rights breaches in the Democratic Republic of the Congo (DRC) during a ten-year period (1993–2003) was mandated by the UN Secretary-General in 2010 to be led by the Office of the High Commissioner for Human Rights (OHCHR). But as the case studies showed, these features are already part of current AI technology, and they are doing a great job. Ideally, nations would make use of this capability for rights when determining or interpreting the data is not crucial; they may, therefore, allocate their few resources to reporting rights requiring interpretive judgment ([Saltz et al., 2019](#)).

1.5.3 ASSESSING THE POTENTIAL NEGATIVE CONSEQUENCES OF AI AND THE POSSIBILITY OF EXACERBATING AND PERPETUATING HUMAN RIGHTS VIOLATIONS

1. When states are both the regulator and the user of these technologies, they are more likely to misuse them. State abuse of technology to commit human rights violations or avoid accountability is always a concern. In times of government instability, one party may utilize data from a prior administration to damage another. Consider the Rwandan genocide. The Hutu administration persecuted minority Tutsis. AI technology that monitors ethnicity might have been utilized by the Hutu regime to slaughter people. States are the main human rights offenders; thus, human rights groups should be wary of AI and ML concerns ([Hojtink & Planqué-van Hardeveld, 2022](#)).

Reporting human rights breaches may expose vulnerable populations to further damage. The Darfur village destruction case study shows how AI must limit risk and safeguard vulnerable communities. When considering

migratory patterns, another illustration is useful. State Watch was one of over a dozen groups and individuals that wrote a letter to the EU IT-Flows Consortium in September 2022, requesting that they discontinue using EU-Migra-Tool, an AI application created to monitor migration flows in Europe ([Utami et al., 2022](#)).

2. Bias—AI uses data generated by humans or created by humans. Thus, AI inherits the biases of the people who developed the data or system. Meanwhile, AI ML learns from its input. Thus, AI reinforces prejudices. Scholars studying AI bias summarize this: While algorithms are part of (biased) institutions and structures, they may also magnify or add bias by favoring measurable events and human behavior over those that are hard or impossible to quantify. Biased data and AI systems threaten human rights, and different research is showing its effects ([Zhang et al., 2022](#)). This will be covered in more detail later on. When it comes to the international human rights system, the problems caused by prejudice are much worse as reliable data and models are crucial and when prejudices of this kind would have a negative impact due to the disproportionate number of cases of minority victims of human rights abuses ([Baker & Robinson, 2020](#)).

One such case in point is the Dutch government's utilization of an uncontrolled AI algorithm developed and used by tax officials to identify cases of childcare benefit fraud. An example is AI-based criminal sentence utilizing face recognition. Some facial recognition technologies are biased toward darker-skinned persons, with greater mistake rates for black and minority populations. Human rights monitoring is especially relevant to data bias in algorithmic learning as those at most risk of rights abuses are frequently women, ethnic minorities, and children ([Aziz & Andriansyah, 2023](#)).

3. Privacy and data protection—AI and privacy are linked. Much literature discusses AI's digital spying and data privacy threats. By centralizing huge volumes of sensitive biometric or person-specific data in one location or AI platform, even the most law-abiding, human-rights-respecting, careful regimes risk violating their citizens' rights. Systems

can be infiltrated, data compromised, and sensitive information leaked. This is illustrated by the US leaving critical data in Afghanistan after removing forces in 2021, risking Afghani Taliban access. The US military stored Afghani fingerprints, iris scans, and face photographs during the conflict. The technique was utilized to develop a massive biometric identification system. US forces used the technology to verify allies and track dangers ([Bao et al., 2022](#)). A further concern in this regard is the ease of re-identification and the anonymization of data by AI; when combined with data from other sources, anonymized data can be linked to specific individuals' identities ([McKendrick, 2019](#)). In the context of AI and human rights, the tension between data privacy and data utility poses a formidable obstacle. The flip side of the coin is that human rights organizations may more readily utilize precise data to hold nations responsible for discriminatory actions. While this is primarily a big data issue, automation in this domain raises AI-related risks and highlights the necessity to weigh conflicting priorities when deciding whether and how much AI to use ([Rai & Sahu, 2020](#)).

1.5.4 OUTLINE

How well AI can assist human rights relies on whether the necessary limitations and laws can be implemented to ensure moral usage ([Sharma et al., 2023](#)).

The case studies and analysis of limits and damages underpin this Note's approach for analyzing AI's influence on international human rights monitoring and reporting ([Sharma et al., 2023](#)). This methodology can help practitioners identify human rights issues using cost-benefit analysis ([Varma et al., 2022](#)): (1) domains that can safely use AI; (2) domains where AI can be used with suitable precautions and limitations; (3) high-risk sectors where key costs and limitations may be overcome by potential advantages; and (4) locations where AI is never suitable ([Tippins et al., 2021](#)).

Influences for evaluating impressions:

- Assess the individuals or entities responsible for the development and implementation of the AI ([Lui & Lamb, 2018](#)).

Who creates AI technology and deploys or uses it, and its effects must be assessed. The more influence the user has over people whose rights are in danger, the higher the risk of damage and the greater the prudence needed when using it ([Mijwil et al., 2023](#)).

This analysis can benefit from actor type and believability. This may include whether the developer and user is a civil society organization, regional or international organization, or state. Civil society organizations like Amnesty International and Human Rights Watch have less influence over the parties whose rights are being examined ([Schmitt, 2020](#)). They have less influence; therefore, dangers are reduced, and third-party control may be less necessary ([Siwakoti et al., 2023](#)). Therefore, AI may be more acceptable. Organizational credibility, ties to the government, and experience and competence may also be considered. Individual states directly control and adjudicate rights-at-risk parties. Another element is which third-party commercial firms are helping develop these technologies ([Zhang et al., 2020](#)).

- Find current and potential future applications of AI and assess their use.

As the case studies show, AI may be utilized in many ways, and understanding how it will be used is crucial to assessing its advantages and drawbacks. Simpler tasks are usually safer. Simple data compilation or analysis may be safe ([Nassar & Kamal, 2021](#)).

AI may be worth the expenses or hazards in situations where there is no other method to get or assess data. In scenarios with limited resources or no access to on-the-ground data ([Alruwaili, 2023](#)), AI may be preferable. Where AI is being created and used to make human judgments, we should be most worried ([Grimm et al., 2021](#)). This covers AI-based fact assessments, value judgments, and policy creation ([Dhoni & Kumar, 2023](#)). There are other occasions where a

human element cannot be duplicated or where a person conducting AI job is more advantageous ([Kordzadeh & Ghasemaghaci, 2022](#)).

- Analyze the approach to AI development and design.

AI only works when thoroughly created. As mentioned, AI learns from large data inputs, and the data itself affects its effectiveness, bias, and other damages ([Mijwil & Aljanabi, 2023](#)). Data-related problems for AI design and development include ([Sambasivan, 2021](#)):

1. Are there sufficient data points to train the algorithm?
2. Is there a means to get data in the event that there is insufficient data? An example of this would be crowdsourcing, as shown in the cases of Decode Darfur and Amnesty International on Twitter ([Singh, 2023c](#)).
3. Are all the details correct and comprehensive?
4. Are many sources being used to produce the data? Can you trust these sources?
5. Can biases be introduced into the data or the AI algorithm? Can these biases be reduced or eliminated?

Any cost–benefit analysis worth its salt will answer all of these questions.

It may also be important to evaluate the sort of tool being built, as different AIs have distinct risks and implications. Deep learning involves extra analysis of data inputs—is there enough data for neural networks? Natural language processing may demand specialty data that is hard to get or make ([Singh, 2023a](#)).

- How serious is the threat to individual rights?

It is crucial to assess how AI threatens human rights. Data source and protection must be considered as part of this investigation ([Zhang et al., 2023](#)). As mentioned, AI usage and data privacy are linked. In the Displacement Forecasting example, data came from respectable sources, including the UN International Displacement Monitoring Organization, IMF, and World Bank, and was not personal ([Singh, 2023d](#)). In contrast, the DRC Mapping Exercise systematically

gathered offender and victim information, including locations, organizational links, and more ([Katyal, 2019](#)). The former is less hazardous if the information is credible, public, and controlled to protect victims ([Irfan et al., 2023](#)). This is risky as it contains sensitive information about vulnerable groups ([Selbst & Barocas, 2018](#)).

- Assess the possibility of further damage.

The question of whether AI may be used to cause more harm is crucial ([Chohan et al., 2023](#)). This necessitates reviewing Factor 1 and considering regime type, government stability, data access, and safeguards ([Singh et al., 2023](#)).

Victim harm should also be considered. Would AI interviewing or testifying instead of a human rights professional damage victims? Do human rights workers benefit? This pertains to human rights workers' mental health issues, including post-traumatic stress disorder (PTSD) and depression ([Sowmya et al., 2023](#)).

- Analyze the mechanisms in place to ensure proper utilization of AI.

To reduce damages, establish methods to guarantee AI is utilized responsibly. Do AIs have adequate internal validation? These tests reduce risk and make AI deployment more responsible ([Daugherty & Wilson, 2018](#)). Without these internal validation tests, damage is more likely. Second, can external parties verify AI results? AI may be safer if it leverages open-source data certified by third-party watchdogs or civil society groups ([Świątkowska, 2020](#)). AI that uses closed data sources without external evaluation procedures is less likely to be validated and more dangerous ([Berk et al., 2021](#)). Another factor is whether AI is learning from credible sources like international institutions ([Abbas et al., 2024](#)).

- Cost–benefit analysis

Next, compare these characteristics to decide if AI should be built and implemented. When dangers are minimal, AI can be applied easily, and organizations and nations should not be hesitant to use it.

If resources are available, it may be worthwhile ([Sudhakar & Kumar, 2023](#)).

If the AI is limited, it should not be used until it can be overcome ([Kochhar et al., 2023](#)). These situations are error-prone. These are situations where AI's potential is enormous, yet its limitations make it unprofitable. If the AI cannot create enough output, it may not be worth the initial resource pull to set it up ([Kaushik, 2024](#)).

- Making use of the framework.

Use the Yemen famine, one of the worst human rights problems of our day, to demonstrate how this paradigm might be used ([Liu et al., 2024](#)). Let us assume that a regional human rights organization wants to report on Yemen's civil war-related hunger to the UN ([Rajkumar et al., 2023](#)). The company is considering three AI uses: (1) using satellite imagery and remote sensing to determine the precise locations of food supplies in relation to human settlements; (2) using this data in conjunction with publicly available information to extract meaningful insights along demographic dimensions; and (3) creating a report using ChatGPT ([Nisha & Muthurajkumar, 2023](#)).

Evaluating the players engaged in the production and deployment of the AI begins with concerns about the validity of the organization, the third-party technology business developing it, and any other actors using it ([Shukla & Tiwari, 2024](#)). Factor two involves examining how AI is being utilized now and in the future ([Paddalwar et al., 2023](#)). To identify concerns, evaluate each usage independently and collectively. A comprehensive dive into how the algorithm is produced, how the AI is taught, if it can repair faults, etc. is needed to evaluate how the AI will be constructed and evolved ([Itani et al., 2023](#)). The three applications would need independent analyses with red flags (Is there missing or partial data? The data's source? Possible bias? Concerns about data privacy?). Under factor five, further hazards may be added. Does disclosing group whereabouts expose them to government or insurgency harm? Chances are yes ([Odeh & Abu Taleb, 2023](#)). Last, component six examines ways to reduce

these damages. For instance, other parties might evaluate the algorithm and data, set criteria for sharing information outside the organization, limit the quantity of specified data, or anonymize it in reports. The organization should assess these variables to decide if, how, and how much to deploy AI ([Ortiz-Ruiz et al., 2024](#)).

1.6 CONCLUSION

The application of AI in human rights is growing, and new AI technologies are being introduced every day. Concerns about the potential negative effects of AI on society are common. This Note demonstrates that AI offers fascinating and potentially useful advantages in this domain. How can we utilize AI to ensure that governments are held responsible for human rights abuses and those rights are protected? AI presents both a challenge to our understanding of ethics and law and an opportunity for substantial progress in the advancement of human rights. Whether deciding whether and how to utilize AI, case studies of its current uses, dangers, and repercussions might be helpful.

Participants in this field can benefit from this Note and the evaluation method it uses. The varied and rapidly growing field of AI is beyond the capabilities of current government regulation and enforcement systems. This pace issue is exacerbated on a global scale as there could not be a centralized regulatory or enforcement body to deal with the introduction of AI. The design of this Note allows for self-regulation by both public and private entities. Human rights groups and other prospective AI users might find guidance in this framework. It is important to urge states, civil society organizations, and third-party developers to evaluate the benefits and drawbacks of AI, the investments in resources, and the protections of rights before, during, and after its implementation. This Note establishes a new baseline for soft law that civil society actors may use to hold authorities responsible, regardless of whether they analyze the deployment of AI or not. It is possible that AI may enhance human rights after a thorough evaluation and under the appropriate legal and regulatory constraints.

REFERENCES

- Abbas, S., Bouazzi, I., Ojo, S., Al Hejaili, A., Sampedro, G. A., Almadhor, A., & Gregus, M. (2024). Evaluating deep learning variants for cyber-attacks detection and multi-class classification in IoT networks. *PeerJ Computer Science*, 10, e1793. [↗](#)
- Abebe, R., Barocas, S., Kleinberg, J., Levy, K., Raghavan, M., & Robinson, D. G. (2020, January). Roles for computing in social change. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 252–260). [↗](#)
- Alazab, M., Soman, K. P., Srinivasan, S., Venkatraman, S., & Pham, V. Q. (2023). *Deep learning for cyber security applications: A comprehensive survey*. Authorea Preprints. [↗](#)
- Alruwaili, F. F. (2023). Optimal hybrid deep learning enabled attack detection and classification in IoT environment. *Computers, Materials and Continua.*, 75(1), 99–115. [↗](#)
- Attaran, M., & Deb, P. (2018). Machine learning: The new ‘big thing’ for competitive advantage. *International Journal of Knowledge Engineering and Data Mining*, 5(4), 277–305. [↗](#)
- Awajan, A. (2023). A novel deep learning-based intrusion detection system for IOT networks. *Computers*, 12(2), 34. [↗](#)
- Aziz, L. A. R., & Andriansyah, Y. (2023). The role artificial intelligence in modern banking: An exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. *Reviews of Contemporary Business Analytics*, 6(1), 110–132. [↗](#)
- Baker, D. J., & Robinson, P. H. (Eds.). (2020). *Artificial intelligence and the law: Cybercrime and criminal liability*. Routledge. [↗](#)
- Bao, Y., Hilary, G., & Ke, B. (2022). Artificial intelligence and fraud detection. *Innovative Technology at the Interface of Finance and Operations: I*, 223–247. [↗](#)
- Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT Press. [↗](#)

- Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 50(1), 3–44. [↵](#)
- Bhuvaneshwari, A. J., & Kaythry, P. (2023). A review of deep learning strategies for enhancing cybersecurity in networks: Deep learning strategies for enhancing cybersecurity. *Journal of Scientific & Industrial Research (JSIR)*, 82(12), 1316–1330. [↵](#)
- Brotcke, L. (2022). Time to assess bias in machine learning models for credit decisions. *Journal of Risk and Financial Management*, 15(4), 165. [↵](#)
- Chen, Z., Van Khoa, L. D., Teoh, E. N., Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: A review. *Knowledge and Information Systems*, 57, 245–285. [↵](#)
- Cheng, X., Liu, S., Sun, X., Wang, Z., Zhou, H., Shao, Y., & Shen, H. (2021). Combating emerging financial risks in the big data era: A perspective review. *Fundamental Research*, 1(5), 595–606. [↵](#)
- Chohan, M. N., Haider, U., Ayub, M. Y., Shoukat, H., Bhatia, T. K., & Hassan, M. F. U. (2023). Detection of cyber attacks using machine learning based intrusion detection system for IoT based smart cities. *EAI Endorsed Transactions on Smart Cities*, 7(2), e4–e4. [↵](#)
- Dash, B., Sharma, P., & Ali, A. (2022). Federated learning for privacy-preserving: A review of PII data analysis in Fintech. *International Journal of Software Engineering & Applications (IJSEA)*, 13(4). [↵](#)
- Daugherty, P. R., & Wilson, H. J. (2018). *Human + machine: Reimagining work in the age of AI*. Harvard Business Press. [↵](#)
- Dhoni, P., & Kumar, R. (2023). *Synergizing generative AI and cybersecurity: Roles of generative AI entities, companies, agencies, and government in enhancing cybersecurity*. Authorea Preprints. [↵](#)
- Djenna, A., Barka, E., Benchikh, A., & Khadir, K. (2023). Unmasking cybercrime with artificial-intelligence-driven cybersecurity analytics. *Sensors*, 23(14), 6302. [↵](#)

- Elluri, L., Mandalapu, V., Vyas, P., & Roy, N. (2023). Recent advancements in machine learning for cybercrime prediction. *Journal of Computer Information Systems*, 1, 1–15. [↗](#)
- Grimm, P. W., Grossman, M. R., & Cormack, G. V. (2021). Artificial intelligence as evidence. *Northwestern Journal of Technology and Intellectual Property*, 19, 9. [↗](#)
- Hagendorff, T. (2019). From privacy to anti-discrimination in times of machine learning. *Ethics and Information Technology*, 21(4), 331–343. [↗](#)
- Hojtink, M., & Planqué-van Hardeveld, A. (2022). Machine learning and the platformization of the military: A study of Google's machine learning platform TensorFlow. *International Political Sociology*, 16(2), olab036. [↗](#)
- Irfan, M., Elmogy, M., & El-Sappagh, S. (Eds.). (2023). *The impact of AI innovation on financial sectors in the era of industry 5.0*. IGI Global. [↗](#)
- Itani, M., Basheer, H., & Eddine, F. (2023, July). Attack detection in IoT-based healthcare networks using hybrid federated learning. In *2023 International conference on smart applications, communications and networking (SmartNets)* (pp. 1–7). IEEE. [↗](#)
- Katyal, S. K. (2019). Private accountability in the age of artificial intelligence. *UCLA Law Review*, 66, 54. [↗](#)
- Kaushik, K. (2024). Leveraging deep learning techniques for securing the Internet of Things in the age of big data. In: *Applying artificial intelligence in cybersecurity analytics and cyber threat detection* (p. 309). IGI USA. [↗](#)
- Kochhar, S. K., Bhatia, A., & Tomer, N. (2023). Using deep learning and big data analytics for managing cyber-attacks. In *New approaches to data analytics and Internet of Things through digital twin* (pp. 146–178). IGI Global. [↗](#)
- Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388–409. [↗](#)

- Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, 63(2), 157–170. [↵](#)
- Li, L., Zhao, T., Xie, Y., & Feng, Y. (2020). Interpretable machine learning based on integration of NLP and psychology in peer-to-peer lending risk evaluation. In *Natural language processing and Chinese computing: 9th CCF international conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, proceedings, Part II* 9 (pp. 429–441). Springer International Publishing. [↵](#)
- Liu, C., Chen, B., Shao, W., Zhang, C., Wong, K., & Zhang, Y. (2024). Unraveling attacks in machine learning-based IoT ecosystems: A survey and the open libraries behind them. arXiv preprint arXiv:2401.11723. [↵](#)
- Lui, A., & Lamb, G. W. (2018). Artificial intelligence and augmented intelligence collaboration: Regaining trust and confidence in the financial sector. *Information & Communications Technology Law*, 27(3), 267–283. [↵](#)
- McKendrick, K. (2019). *Artificial intelligence prediction and counterterrorism* (p. 9). London: The Royal Institute of International Affairs-Chatham House. [↵](#)
- Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International Journal of Financial Studies*, 9(3), 39. [↵](#)
- Mijwil, M., & Aljanabi, M. (2023). Towards artificial intelligence-based cybersecurity: The practices and ChatGPT generated ways to combat cybercrime. *Iraqi Journal for Computer Science and Mathematics*, 4(1), 65–70. [↵](#)
- Mijwil, M., Salem, I. E., & Ismaeel, M. M. (2023). The significance of machine learning and deep learning techniques in cybersecurity: A comprehensive review. *Iraqi Journal For Computer Science and Mathematics*, 4(1), 87–101. [↵](#)
- Mühlhoff, R. (2021). Predictive privacy: Towards an applied ethics of data analytics. *Ethics and Information Technology*, 23(4), 675–690. [↵](#)

- Narayana, Y. V., & Sreedevi, M. (2023, October). DDCATF: Deep learning approach for detection of cybercrime activities based on temporal features. In *2023 International conference on self sustainable artificial intelligence systems (ICSSAS)* (pp. 462–469). IEEE. [↗](#)
- Nassar, A., & Kamal, M. (2021). Ethical dilemmas in AI-powered decision-making: A deep dive into big data-driven ethical considerations. *International Journal of Responsible Artificial Intelligence*, 11(8), 1–11. [↗](#)
- Nisha, S. R., & Muthurajkumar, S. (2023). Semantic graph based convolutional neural network for spam e-mail classification in cybercrime applications. *International Journal of Computers Communications & Control*, 18(1), 1–5. [↗](#)
- Odeh, A., & Abu Taleb, A. (2023). Ensemble-based deep learning models for enhancing iot intrusion detection. *Applied Sciences*, 13(21), 11985. [↗](#)
- Ortiz-Ruiz, E., Bermejo, J. R., Sicilia, J. A., & Bermejo, J. (2024). Machine learning techniques for cyberattack prevention in IoT systems: A comparative perspective of cybersecurity and cyberdefense in Colombia. *Electronics*, 13(5), 824. [↗](#)
- Paddalwar, S., Ragha, L., & Mane, V. (2023). Cyber threat mitigation using machine learning, deep learning, artificial intelligence, and blockchain. In *Intelligent approaches to cyber security* (pp. 163–179). Chapman and Hall/CRC. [↗](#)
- Pocher, N., Zichichi, M., Merizzi, F., Shafiq, M. Z., & Ferretti, S. (2023). Detecting anomalous cryptocurrency transactions: An AML/CFT application of machine learning-based forensics. *Electronic Markets*, 33(1), 37. [↗](#)
- Rai, R., & Sahu, C. K. (2020). Driven by data or derived through physics? A review of hybrid physics guided machine learning techniques with cyber-physical system (CPS) focus. *IEEE Access*, 8, 71050–71073. [↗](#)
- Rajkumar, S., Sheeba, S. L., Sivakami, R., Prabu, S., & Selvarani, A. (2023). An IoT-based deep learning approach for online fault detection against cyber-attacks. *SN Computer Science*, 4(4), 393. [↗](#)

- Saltz, J., Skirpan, M., Fiesler, C., Gorelick, M., Yeh, T., Heckman, R., & Beard, N. (2019). Integrating ethics within machine learning courses. *ACM Transactions on Computing Education (TOCE)*, 19(4), 1–26. [↵](#)
- Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., & Prabhakaran, V. (2021, March). Re-imagining algorithmic fairness in India and beyond. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 315–328). ACM. [↵](#)
- Schmitt, M. (2020). Artificial intelligence in business analytics, capturing value with machine learning applications in financial services. [↵](#)
- Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham Law Review*, 87, 1085. [↵](#)
- Sharma, A., Mansotra, V., & Singh, K. (2023). Detection of Mirai Botnet attacks on IoT devices using deep learning. *Journal of Scientific Research and Technology*, 1, 174–187. [↵](#)
- Shukla, R. K., & Tiwari, A. K. (2024). Security analysis of the cyber crime. In *The ethical frontier of AI and data analysis* (pp. 257–271). IGI Global. [↵](#)
- Singh, B. (2023a). Blockchain technology in renovating healthcare: Legal and future perspectives. In *Revolutionizing healthcare through artificial intelligence and Internet of Things applications* (pp. 177–186). IGI Global. [↵](#)
- Singh, B. (2023b). Federated learning for envision future trajectory smart transport system for climate preservation and smart green planet: Insights into global governance and SDG-9 (industry, innovation and infrastructure). *National Journal of Environmental Law*, 6(2), 6–17. [↵](#)
- Singh, B. (2023c). Tele-health monitoring lensing deep neural learning structure: Ambient patient wellness via wearable devices for real-time alerts and interventions. *Indian Journal of Health and Medical Law*, 6(2), 12–16. [↵](#)
- Singh, B. (2023d). Unleashing Alternative Dispute Resolution (ADR) in resolving complex legal-technical issues arising in cyberspace lensing e-commerce and intellectual property: Proliferation of E-commerce digital

- economy. *Revista Brasileira de Alternative Dispute Resolution-Brazilian Journal of Alternative Dispute Resolution-RBADR*, 5(10), 81–105. [↵](#)
- Singh, B., & Kaunert, C. (2024). Integration of cutting-edge technologies such as Internet of Things (IoT) and 5G in health monitoring systems: A comprehensive legal analysis and futuristic outcomes. *GLS Law Journal*, 6(1), 13–20. [↵](#)
- Singh, N. J., Hoque, N., Singh, K. R., & Bhattacharyya, D. K. (2023). Botnet-based IoT network traffic analysis using deep learning. *Security and Privacy*, 1, e355. [↵](#)
- Siwakoti, Y. R., Bhurtel, M., Rawat, D. B., Oest, A., & Johnson, R. C. (2023). Advances in IOT security: Vulnerabilities, enabled criminal services, attacks and countermeasures. *IEEE Internet of Things Journal*, 2, 55–56. [↵](#)
- Sowmya, B. J., Hanumantharaju, R., Kumar, D. P., & Srinivasa, K. G. (2023). Identification of authorship and prevention of fraudulent transactions/cybercrime using efficient high performance machine learning techniques. *International Journal of Business Intelligence and Data Mining*, 22(1–2), 144–169. [↵](#)
- Sudhakar, & Kumar, S. (2023, April). ABBDIoT: Anomaly-based botnet detection using machine learning model in the Internet of Things network. In *International conference on IoT, intelligent computing and security: Select proceedings of IICS 2021* (pp. 235–245). Singapore: Springer Nature Singapore. [↵](#)
- Swiątkowska, J. (2020). Tackling cybercrime to unleash developing countries' digital potential. *Pathways for Prosperity Commission Background Paper Series*, 33, 2020–01. [↵](#)
- Thapaliya, S., & Sharma, P. K. (2023). Optimized deep neuro fuzzy network for cyber forensic investigation in big data-based IoT infrastructures. *International Journal of Information Security and Privacy (IJISP)*, 17(1), 1–22. [↵](#)
- Tippins, N. T., Oswald, F. L., & McPhail, S. M. (2021). Scientific, legal, and ethical concerns about AI-based personnel selection tools: A call to

- action. *Personnel Assessment and Decisions*, 7(2), 1.[↵](#)
- Tyagi, A. K., Aswathy, S. U., & Abraham, A. (2020). Integrating blockchain technology and artificial intelligence: Synergies perspectives challenges and research directions. *Journal of Information Assurance and Security*, 15(5), 1554.[↵](#)
- Utami, S. H., Purnama, A. A., Hidayanto, A. N., Utami, S. H., Purnama, A. A., & Hidayanto, A. N. (2022). Fintech lending in Indonesia: A sentiment analysis, topic modelling, and social network analysis using Twitter data. *International Journal of Applied Engineering & Technology*, 4(1), 50–56.[↵](#)
- Varma, P., Nijjer, S., Sood, K., Grima, S., & Rupeika-Apoga, R. (2022). Thematic analysis of financial technology (fintech) influence on the banking industry. *Risks*, 10(10), 186.[↵](#)
- Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2023). *Dive into deep learning*. Cambridge University Press.[↵](#)
- Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022a). Interpretable machine learning models for crime prediction. *Computers, Environment and Urban Systems*, 94, 101789.[↵](#)
- Zhang, Y., Xiong, F., Xie, Y., Fan, X., & Gu, H. (2020). The impact of artificial intelligence and blockchain on the accounting profession. *IEEE Access*, 8, 110461–110477.[↵](#)
- Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022b). *Explainable artificial intelligence applications in cyber security: State-of-the-art in research*. *IEEE Access*, 10, 93104–93139.[↵](#)

2 Assessing AI-Driven Human Rights Violations

A Framework for Accountability

*Debashree Chakravarty, Ipseeta Satpathy,
Vishal Jain, and B.C.M. Patnaik*

DOI: [10.1201/9781003668657-2](https://doi.org/10.1201/9781003668657-2)

2.1 INTRODUCTION

The concept of creating human-like intelligent entities has been ingrained in human history since classical times. This idea of artificial intelligence (AI) has consistently captivated imaginations, and it has prominently featured in science fiction literature. Approaches to AI are often polarized: one viewpoint asserts that AI could pose an existential threat to humanity (Mulgan, 2016), while another anticipates unprecedented levels of human prosperity ([Kurzweil, 2005](#)). Central to these divergent viewpoints is the understanding of general artificial intelligence (AGI). AGI denotes artificial intelligence that attains intelligence comparable to humans and demonstrates proficiency in abstract reasoning ([Weinbaum & Veitas, 2017](#)). However, true AGI, capable of human-like intelligence, remains elusive. Current AI systems excel in specific tasks within narrow domains, such as recognition of images, recognition of face or facial, emotion detection as

well as natural language processing, sometimes surpassing human capabilities.

Developing a framework to assess human rights violations in the use of AI is crucial to ensure responsible and ethical AI deployment. Such a framework should address various dimensions, including privacy, discrimination, transparency, accountability, and the impact on societal well-being. The framework combines legal, ethical, and technical perspectives like legal compliance, transparency and explainability, privacy and data protection, fairness and non-discrimination, accountability and responsibility, social and economic impact, global collaboration and governance, continuous monitoring and evaluation, ethical consideration and public awareness and education. Implementing and refining such a framework require collaboration among governments, industry stakeholders, academia, and civil society. Regular updates and adjustments should be made to address emerging challenges and technological advancements.

This study aims to explore the intersection of vulnerability and the necessity for a regulatory framework in the context of AI systems. As AI technologies continue to advance, concerns about their potential impact on human rights have become more pronounced. The lens of vulnerability provides a nuanced perspective, emphasizing the heightened susceptibility of certain individuals or groups to the risks associated with AI. By scrutinizing this vulnerability, the chapter seeks to underscore the urgency of establishing a robust regulatory framework to protect against human rights violations.

The advancements in today's AI are primarily driven by machine learning techniques. In this approach, machines learn to perform tasks traditionally considered to require human intelligence by analyzing vast datasets, akin to how children learn through interaction and feedback from their environment. However, despite their proficiency in specific tasks, AI systems do not replicate human thought processes. Instead, machine learning simplifies and formalizes complex cognitive tasks like vision and

language processing, reflecting a computational approach to understanding the world ([Singh et al., 2021](#)).

AI is argued to be that “computational rationality” might be a more fitting term than “artificial intelligence,” emphasizing the use of computational methods to achieve rational behavior ([Russell & Norvig, 2016](#)). Over the past decade, AI systems have proliferated in various aspects of daily life, from household appliances and smartphones to autonomous vehicles and predictive algorithms used in policing and risk assessment ([Mercan & Selçuk, 2024](#)).

The “AI effect” ([McCorduck & Cfe, 2004](#)) refers to the phenomenon where actions performed by AI systems, once considered intelligent, become normalized and no longer categorized as “AI” by humans. This is exemplified historically by calculators, which were initially seen as feats of AI but are now universally regarded as routine tools devoid of intelligence.

Machine learning, a cornerstone of AI, operates by continuously processing vast amounts of data. These systems quietly gather extensive data from their environments, effectively creating a pervasive surveillance network that permeates both individual and societal spheres without conspicuous notice. AI systems analyze and predict human behaviors, influencing beliefs and decisions pre-emptively ([Hildebrandt, 2020](#)).

While humans shape AI systems, these systems reciprocally mold human beliefs, thought patterns, and behaviors, posing challenges to democratic principles rooted in individual autonomy and self-determination. This study critically examines AI’s potential profound impacts on democracy.

Furthermore, the interaction between democracy and AI demands an approach that is interdisciplinary. While our research primarily draws on legal and specifically public law perspectives, it also incorporates insights from the social sciences, humanities, and computer science to a lesser extent. This study provides a thorough analysis of the implications of AI technologies on people’s rights and democracy, examining both the challenges and opportunities they present. Initially, it investigates the AI technologies and their impact on foundational democratic rights and human

rights, such as freedom of expression and participation in elections, the right to vote, privacy, and personal data protection. This examination underscores the potential risks and benefits AI introduces to the process of democracy and the exercise of individual rights. The research delves into the complex ways AI technologies intersect with human rights, discussing their potential positive and negative effects on democracy. It critically evaluates how regulatory frameworks and policies should be designed to safeguard people's rights and uphold the values of democracy in the context of AI deployment. The study emphasizes the importance of legislative measures, ethical guidelines, and international cooperation to effectively regulate AI technologies within the framework of international law.

The primary focus of this study is to explore how and why AI influences democratic processes. It does not confine its investigation to specific national contexts but rather seeks to delineate how AI systems, originally designed for narrow tasks, affect democratic institutions within democratic frameworks. Subsequently, the study evaluates the broader implications of these effects on democratic values, adopting a descriptive rather than prescriptive approach.

Given the absence of universally accepted definitions for "AI" and "democracy" in scholarly discourse, this study begins by establishing comprehensive definitions. A precise definition of AI is crucial for accurately assessing its impact on democracy and formulating effective regulatory frameworks. Narrow definitions might overlook critical aspects of AI's influence, while overly broad definitions could hinder the development of targeted regulations. Similarly, a clear conceptualization of democracy is essential to gauge how AI aligns with democratic ideals and operational realities ([Rodrigues, 2020](#)).

For instance, defining AI solely as technical systems without considering its broader implications would limit meaningful comparisons with the abstract ideals of democracy. Conversely, an incomplete definition of democracy might obscure essential aspects of how AI intersects with democratic governance. Therefore, this study endeavors to provide a

nuanced understanding of both concepts to facilitate a comprehensive analysis of their interplay and implications for democratic societies.

2.1.1 OBJECTIVES OF THE STUDY

- To study AI's multifaceted nature and its impact on human rights and democracy.
- To understand the impact of AI technologies on fundamental rights essential to human rights and democracy, such as freedom of expression, it is crucial to propose a framework for implementing ethical standards and legal regulations across various levels. This ensures that AI technologies contribute positively to society while upholding democratic values and human rights.

2.2 CONCEPTUAL FRAMEWORK: AI'S IMPACT ON DEMOCRACY AND HUMAN RIGHTS

At intervals, leaders of innovation firms and researchers have freely forewarned about the capability of computerized reasoning frameworks to outperform human knowledge and possibly rule society, thus presenting dangers to people and networks ([Tegmark, 2019](#)). These warnings resonate widely due to the implications inherent in the term "AI" ([Russell & Norvig, 2016](#)). The term suggests the development of entities that could one day rival or even exceed human intelligence.

Despite its far-reaching public discernment, the expression "artificial intelligence" comes up short on all-around settled-upon definitions in both scholastic writing and lawful settings. As the objectives and extent AI research develop, so too does the meaning of simulated intelligence itself. Hence, the view of what is "intelligence" and the measures for marking a framework as "intelligence" keep on moving ([Russell & Norvig, 2016](#)). Therefore, AI remains an idea that changes over the long run and across alternate points of view.

2.3 DISTINCTION BETWEEN NARROW AI AND AGI

Present-day AI systems are predominantly engineered for specific tasks within confined domains, commonly categorized as “narrow” or “weak” AI. These systems are programmed to achieve predefined goals, with their capability to fulfil these tasks serving as a gauge of their intelligence.

In contrast, “general” or “strong” AI represents a hypothetical category where AI systems would exhibit human-like intelligence, capable of abstract thinking, and setting their own goals (Vinge, 1993). However, true general AI does not yet exist.

2.4 AI ADVANCEMENTS IN NARROW DOMAINS

AI systems developed for narrow domains have demonstrated superior performance compared to humans in various tasks. Examples include computer vision, where AI systems analyze and interpret images or videos, performing tasks such as classifying images, generating realistic images, detecting human poses, facial recognition, and emotion detection from facial expressions ([Zhang et al., 2021](#)).

Another domain where AI excels is natural language processing, enabling systems to interpret, understand, and generate human language. Natural language processing applications encompass tasks like sentiment analysis, machine translation preserving meaning across languages, and conversational agents or chatbots capable of interacting verbally or textually with users ([Russell & Norvig, 2016](#)).

2.5 IMPACT AND INTEGRATION OF AI

The industrialization and affordability of AI technologies have led to their widespread integration into everyday life, impacting various societal sectors. AI’s accessibility and success in tasks driven by extensive data training have further amplified its societal influence, shaping “directly

usable” applications and affecting diverse aspects of human existence ([Zhang et al., 2021](#)).

2.6 AI IN ACADEMIC DISCOURSE

In the academic discourse, AI is delineated from various angles, predominantly revolving around human intelligence and rationality. Russell and Norvig categorize academic definitions of AI into four principal groups such as systems that act like humans, think like humans, act rationally or think rationally ([Russell & Norvig, 2016](#)). One perspective, aligned with Alan Turing’s criterion, defines AI as systems capable of convincingly mimicking human behavior to the extent of passing the Turing Test, signifying intelligence ([Alpaydin, 2020](#)). However, this approach has been critiqued for emphasizing mimicry rather than genuine intelligence.

On the other hand, the rationality perspective defines AI as systems proficient in identifying and executing optimal actions to achieve specific goals. Russell and Norvig depict AI entities as capable of selecting the best course of action in any given situation, while [Nilsson \(2009\)](#) views AI as enabling entities to operate effectively within their environments through proactive decision-making. This rational agent approach predominates contemporary AI research, where machine learning-based systems are engineered to achieve particular objectives optimally, such as enhancing user engagement through recommendation systems on websites.

Beyond Russell and Norvig’s classification, AI is defined in various ways in the literature. For example, Krizhevsky sees AI as automating intelligence in computer science, while Say focuses on AI’s ability to perform cognitive tasks successfully. Turner extends the definition to include moral decision-making processes by non-natural entities ([Gershgorn, 2018](#); [Turner, 2018](#)).

2.7 LEGAL PERSPECTIVE AND FRAMEWORKS OF AI

Under a legalistic framework, social planners endeavor to interpret how existing laws can be applied to AI and robotic technologies. A key challenge lies in determining whether these technologies align with current legal structures, especially when rules face difficulties during their application. This assignment requires specialized lawyers to examine these matters within particular legal sectors or across multiple fields including product safety, consumer protection, liability, intellectual property (IP), labor law, privacy, civil and criminal liability, legal standing, insurance, and tax law.

For instance, a significant debate within IP law revolves around whether inventions generated by AI can be protected and who holds ownership rights. Copyright law necessitates “originality,” requiring a work to display creativity, reflect the intellectual creation of its author, or involve non-trivial skill and judgment. This scenario draws parallels with cases like a photograph taken by a monkey, which was deemed ineligible for copyright protection. Likewise, patent law requires an invention to exhibit an “inventive step,” indicating it should not be apparent to an individual skilled in the relevant field. This prompts considerations about how to define “non-obviousness” and “skill in the art” when considering AIs capable of “recursive self-improvement” ([Judge & Gervais, 2009](#)).

While the legalistic approach fosters discussions on the societal implications of legal frameworks, it also presents challenges. Lawyers, often distant from technological fields, may speculate on legal issues without a full grasp of the underlying science, leading to flawed hypotheses and the emergence of new legal complexities.

For example, if lawyers mistakenly assume computational solutions could cure Alzheimer’s disease, they might advocate robust patent protections for AI in this realm. However, if the scientific consensus favors biotechnological approaches like digital brain emulation (EM), this could result in misguided investments due to overly stringent patent laws, potentially eroding trust in the patent system.

AI and robotics introduce unpredictable challenges, and over-reliance on existing legal frameworks may create blind spots. Laws are abstract

principles based on specific perspectives of the world and its evolution. Excessive reliance on legal interpretations may impede the development of innovative legal concepts necessary to address emerging technologies.

For instance, if an AI develops a time-travel application enabling users to interact with historical figures, should new laws consider potential psychological impacts upon users' return to reality? Similarly, if AI were to dominate society, would new laws be required to safeguard human rights from intelligent machines? Current laws are founded on the assumption of human uniqueness, positioning humans as superior to machines and animals, thus focusing primarily on regulating human interactions with non-humans.

2.8 CONCEPTUAL FRAMEWORK OF DEMOCRACY

The concept of democracy has been widely invoked across political discourse, ranging from macro political regimes to microenvironments of daily life. It serves as a legitimizing principle universally acknowledged, even among authoritarian leaders who may claim adherence to democratic ideals (Barry, 2004). However, the concept itself lacks consensus, leading to what Sartori termed an “age of conceptual chaos” for democracy.

The diversity in defining democracy arises from its usage on different planes: describing existing democratic regimes and portraying an idealized society ([Susskind, 2018](#)). The literature reflects a spectrum of theoretical approaches to democracy, encompassing etymological, normative, realist, rational, and empirical perspectives (Barry, 2004).

While democracy embodies normative ideals of equality and participation, its empirical manifestation is contingent on real-world conditions. Scholars like Dahl ([Beetham, 1999](#)) emphasize democratic governance as a goal requiring normative pursuit, balanced with an empirical evaluation of its operational realities. This tension between democratic ideals and empirical realities underpins the existential credibility of democratic practices and institutions ([Habermas, 2023](#)).

Thus, understanding democracy necessitates reconciling normative aspirations with empirical conditions, acknowledging that the existence of democracy hinges on citizens' perception and participation in democratic processes ([Habermas, 2023](#)).

2.9 IMPACT OF AI ON FREEDOM OF EXPRESSION

In present digital age, cyberspace serves as a vital platform for political discourse and public information sharing. However, AI-driven recommendation systems shape this discourse by filtering content, creating echo chambers that limit diverse viewpoints and hinder democratic participation ([Erdoğ̃an, 2001](#)). AI surveillance capabilities in cyberspace enable the tracking and analysis of citizens' activities, potentially suppressing dissent and undermining freedom of expression ([Erdoğ̃an, 2001](#)). Online content moderation by AI systems, although aimed at filtering harmful content, risks censoring legitimate expression due to inadequate cultural context understanding. States have international obligations to safeguard freedom of expression, ensuring accessible communication tools and preventing monopolies that restrict plurality ([Habermas, 2023](#)). Effective legal frameworks are essential to uphold these rights amidst AI advancements (Rainey et al., 2021).

In democratic societies, freedom of expression is essential for citizens to openly voice their opinions on dominant ideologies, public authorities, and socio-economic conditions, regardless of how controversial or unsettling these views may be. However, the widespread implementation of AI surveillance systems, both online and offline, enables continuous tracking and analysis of individuals ([Erdoğ̃an, 2001](#)). This surveillance dynamic shifts the role of freedom of expression from empowering citizens to oversee their government through public discourse to a scenario where political authorities can identify and suppress dissenting voices, as seen in practices by countries like China, which employ AI to monitor and predict protests ([Erdoğ̃an, 2001](#)).

Moreover, in cyberspace, social media platforms conduct content moderation activities that pose risks to freedom of expression. AI-driven decisions in these activities often struggle to grasp cultural nuances and satire, resulting in the automatic removal of content that may be offensive or provocative. This not only diminishes the diversity of opinions in cyberspace but also raises concerns about the growing influence of these platforms, akin to becoming modern gatekeepers of digital discourse, a role traditionally held by the press (Sartori, 1987).

The transformation of information intermediaries into influential entities necessitates state intervention to ensure their independence and accountability, as underscored by international human rights obligations ([Habermas, 2023](#)). The European Court of Human Rights (ECHR) has emphasized states' responsibilities to maintain accessible communication tools that promote plurality and safeguard the freedom of journalists. Upholding these rights requires comprehensive legal frameworks that provide judicial oversight and prevent arbitrary removal of online content, thereby preserving public interest and democratic values.

2.10 IMPACT OF AI ON THE RIGHT TO FREE ELECTIONS

AI is increasingly utilized in political campaigns, leveraging computational propaganda through bots and troll accounts to manipulate public opinion ([Susskind, 2018](#)). This threatens the integrity of free and fair elections by influencing voter behavior and spreading misinformation.

Micro-targeting techniques amplify these threats, enabling personalized political messaging that undermines electoral transparency and pluralism ([Woolley & Howard, 2016](#)). Such manipulations challenge democratic principles of fair competition and public trust in electoral processes.

2.11 IMPACT OF AI ON PRIVACY AND DATA PROTECTION

AI's reliance on big data intensifies surveillance practices, collecting and analyzing vast amounts of personal information without adequate consent. Aggregation of this data, including biometric and behavioral data, poses profound threats to privacy and autonomy (Ambitions, 2020).

Sophisticated AI algorithms can identify individuals across various datasets, expanding the definition of personal data and eroding privacy rights ([Yeung, 2018](#)). Commercial applications like Clearview AI exemplify these concerns by enabling extensive surveillance capabilities, challenging traditional notions of privacy in public spaces.

In democratic contexts, the integrity of elections relies on the ability of citizens to freely express their views without undue interference. However, during the 2016 November US presidential election, it was found that there were significant numbers of bot-generated tweets favoring Donald Trump over Hillary Clinton, suggesting an imbalance in public discourse. These bots not only skewed the perception of public opinion but also had the potential to intimidate opposing viewpoints, thereby undermining the free exchange of ideas and the fairness of the electoral process.

Furthermore, AI technologies enable micro-targeting strategies that disseminate disinformation and deepfakes disseminated via bots and troll accounts can sway broad audiences and distort political competition. Such practices not only compromise the democratic principle of fair competition but also suppress the freedom of expression among affected groups.

The influence of AI surveillance systems extends beyond elections, threatening citizens' rights to privacy and autonomy. These systems collect vast amounts of personal data, which are analyzed to create detailed profiles that predict individuals' behaviors and preferences (Ambitions, 2020). Such practices, including online advertising and manipulation techniques like micro-targeting, erode individuals' ability to make autonomous decisions and participate effectively in democratic processes ([Valcke et al., 2021](#)).

Moreover, the extensive use of AI for surveillance purposes, such as facial recognition and emotion analysis, poses significant challenges to privacy and data protection rights (Ambitions, 2020). These technologies enable the identification of individuals from non-traditional data sources,

necessitating broader definitions of personal data ([Yeung, 2018](#)). For instance, applications like Clearview AI collect billions of data points from the internet to compile comprehensive profiles of individuals, thereby infringing on their privacy rights.

AI-driven surveillance risks transforming democracies into “modulated” systems, where individual autonomy and private spaces for free expression are constrained by pervasive monitoring ([Cohen, 2013](#)). This undermines the foundational principles of liberal democracy, hindering rational deliberation and meaningful public engagement ([Hill & Krolik, 2020](#)).

2.12 AI: UPHOLDING EQUALITY AND NON-DISCRIMINATION

AI systems possess the potential to generate discriminatory outcomes that impact specific groups due to factors such as biased training data, system design, feature selection, and underlying classification criteria. These biases can hinder certain individuals’ effective and equitable participation in the processes of democracy. For instance, biased algorithms used in online content moderation may automatically remove or reduce the visibility of certain ideas, thereby compromising the freedom of expression more severely for some individuals in online spaces than others. This unequal access undermines the principle of equality, denying all citizens a fair opportunity for democratic participation online.

Moreover, AI systems deployed as decision-making tools in contexts like criminal justice can perpetuate discriminatory results. For example, research into the compass application, used to assess individual risk in criminal cases, revealed that of non-offending black defendants were labeled as “high risk” twice as often as the non-offending white defendants. Conversely, white defendants with a history of re-offending often received lower risk scores compared to black defendants, exposing evident racial bias in the system’s outcomes.

These instances emphasize the urgent need to identify and rectify biases within AI systems to prevent the perpetuation or exacerbation of

discrimination. Ensuring equity in AI and eradicating discriminatory outcomes are critical for upholding principles of equality and non-discrimination in the digital era. This is vital for protecting democratic rights and fostering inclusive participation for all individuals.

2.13 ASSESSING AI REGULATION: IMPLICATIONS FOR DEMOCRACY AND HUMAN RIGHTS

In a recent research endeavor, [Habermas \(2023\)](#) analyzed the public sphere transforming into digital environments, emphasizing the substantial risks that AI systems pose to democracy and human rights. He underscores the critical importance of strengthening mechanisms aimed at safeguarding human rights.

The Council of Europe's report, explores how AI jeopardizes the effectiveness of human rights frameworks ([Yeung, 2018](#)). According to the report, the complexity and opacity of AI systems present challenges in establishing a direct connection between AI interventions and the resulting violations of human rights ([Yeung, 2018](#)).

If specific job advertisements on a social media platform consistently exclude female users, it could lead to gender-based discrimination in employment opportunities. Affected users often remain unaware of these exclusions, making it difficult to prove discrimination and its impact on their employment rights.

Additionally, human rights frameworks relying on individual consent mechanisms face challenges in the context of AI. Even though individuals have inherent rights such as the right to life and freedom from torture, obtaining personal consent can become a procedural hurdle, especially in situations where individuals might theoretically waive their protections. This issue is exacerbated by power imbalances between data processors and data subjects, where users might relinquish their rights to data protection, freedom of expression, and privacy in exchange for free services, efficiency, or convenience.

At a societal level, AI systems can subtly influence democratic principles and institutions, even if their effects appear minor at the individual level. For instance, a Facebook experiment in 2010 manipulated the news feeds of six crore users, introducing a “click to vote” button to a subset of them. The study revealed that individuals exposed to this manipulation were 0.39% more inclined to engage in voting behavior compared to those who were not, highlighting the potential significant impact on democratic processes. Despite individual awareness of such manipulations, proving their direct influence on voting behavior and mobilizing human rights protections can be formidable challenges.

While the immediate impact of AI on individuals may appear limited, the cumulative effects of widespread AI applications can erode fundamental democratic principles and human rights. These effects pose a threat to societal values and interests that are often overlooked in current discussions focused on individual human rights, thereby weakening both the social and technical conditions essential for upholding democracy and human rights). Addressing these challenges, the Council of Europe’s report proposes proactive measures to safeguard democratic principles and human rights. These include establishing collective complaint mechanisms, formulating appropriate state-level policies, integrating human rights considerations into AI research, creating independent oversight bodies for safeguarding human rights in AI contexts, and defining corporate responsibilities.

In his 2018 work, Nemitz proposes a framework suggesting that AI should follow legal principles similar to those of “law,” ensuring adherence to fundamental rights, the rule of law, democratic principles, and legitimacy throughout its lifecycle of development and deployment. This approach requires strong policies and legal structures capable of accommodating various national strategies for regulating AI. ([Nemitz, 2018](#))

The EU’s Declaration on Digital Rights and Principles, released on January 23, 2023, offers a comprehensive guideline addressing the impacts of AI while prioritizing human rights and democratic values. Although not legally binding, the declaration outlines digital rights and terms for their safeguarding and application. It places human beings at the center of digital

transformation, supports digital literacy and fair labor conditions, upholds freedom of choice in interactions with AI systems, promotes inclusive digital participation, enhances digital safety and empowerment (especially for young people), advocates for sustainability, and guarantees privacy rights and control over personal data. Key provisions include facilitating democratic engagement in the digital realm, protecting freedom of expression, ensuring transparency and accountability on digital platforms, and preventing misinformation and manipulation. The declaration aims to steer international relations and policies related to AI, reflecting the EU's commitment to integrating human rights and democratic values into the evolving digital landscape.

The European Union's proposed Artificial Intelligence Act, introduced in 2021 and approved by the European Parliament in June 2023 pending negotiations among member states, aims to regulate AI systems through a risk-based classification framework. It categorizes AI into four tiers: unacceptable risk (outright prohibition), high-risk (under stringent conditions), limited risk (primarily transparency requirements), and minimal risk (exempt from legal obligations). The Act mandates rigorous pre-market assessments, risk management frameworks, data governance compliance, and human oversight for high-risk AI systems. It also proposes national supervisory authorities, market surveillance bodies, and a European AI Board, though it has faced criticism for perceived gaps in industrial and environmental risk coverage, and the reliance on self-assessment for high-risk systems without adequate avenues for redress by affected individuals.

Amid rising public apprehensions regarding AI, diverse stakeholders are advocating for new ethical frameworks to mitigate its risks and ensure responsible deployment. This includes initiatives by major companies like Google and Microsoft, which have developed their own ethical principles for AI. Furthermore, collaborative endeavors involving private enterprises and a variety of stakeholders, such as the Partnership on AI, the Asilomar Principles, and OpenAI, aim to establish guidelines for ethical practices in AI development and implementation. Civil society and academia have also

intensified their efforts to formulate ethical frameworks for AI. Organizations like Fairness, Accuracy, and Transparency in Machine Learning (fatml.org) have assembled diverse stakeholders to promote standards in machine learning and advocate for governmental and corporate commitments to prevent bias in AI-driven decisions. Stanford University's Human-Centered AI Institute and the Institute of Electrical and Electronics Engineers (IEEE) have similarly mobilized scholars, technologists, and civil society activists through initiatives like the Global Initiative on Ethics of Autonomous and Intelligent Systems.

While commendable, these initiatives each address specific facets of the AI challenge without fully encompassing the entire spectrum of risks to human welfare or addressing the comprehensive governance challenges posed by AI. Crucially, none of these efforts has achieved universal global consensus or broad stakeholder buy-in. Acknowledging the positive steps taken by private firms and nonprofit groups in clarifying their ethical commitments and publishing principles for AI, it remains uncertain how influential these efforts will be beyond the entities directly involved. Ethical guidelines can guide those who adopt them but do not establish a unified governance framework applicable to all. To effectively confront the wide-ranging challenges posed by AI to society, a more systematic and comprehensive approach is imperative. This approach must transcend specific interests and garner global consensus, addressing the roles and responsibilities of both governments and the private sector in ensuring accountability for AI-based decisions. Fortunately, a well-established foundation for this comprehensive and global approach already exists in the form of the universal human-rights framework established under the auspices of the United Nations in 1948. This framework provides a principled basis for regulating AI in a manner that upholds human rights, ensuring that AI developments align with ethical standards and societal values on a global scale.

The Universal Declaration of Human Rights (UDHR), crafted in the aftermath of World War II and adopted by the UN General Assembly on December 10, 1948, established a set of rights applicable universally,

irrespective of geography or regime. Subsequent international treaties collectively known as the International Bill of Human Rights further elucidated these rights, encompassing political, civil, economic, social, and cultural dimensions. Together, these documents form the basis of international human rights law, outlining governments' obligations to protect individuals from rights violations by both state and non-state actors, including private entities.

Applying a human rights-based approach to governing AI can achieve objectives similar to those sought by emerging ethical frameworks. Four key aspects of the human rights framework enable some functions: (1) Placing the human person at the core of AI governance ensures that the impact on individuals remains paramount. (2) Encompassing a wide range of concerns raised by AI, both procedural and substantive. (3) Defining roles and responsibilities for governments and the private sector in safeguarding and respecting human rights. (4) Enjoying broad global consensus, making it universally applicable and advantageous geopolitically. Central to the human rights framework is its foundation in human dignity, affirming humans as the focal point of governance and societal development. This principle ensures that AI development prioritizes support for human welfare over potential harms, particularly crucial as AI assumes roles once held exclusively by humans.

The comprehensive scope of international human rights law addresses contemporary societal apprehensions regarding AI. Articles within the UDHR directly address critical issues such as bias mitigation in data, fairness in automated decision-making, the dangers posed by autonomous weapons, the right to privacy in data-driven societies, and concerns about algorithmic content moderation affecting freedom of expression and democratic processes. Moreover, the framework outlines the obligations of governments and the private sector to uphold and remedy human rights violations. Governments bear the primary responsibility for protecting rights within their jurisdiction, while private firms, under the UN Guiding Principles on Business and Human Rights, are tasked with respecting human rights in their operations and providing redress for violations linked

to their products and services. Given the pivotal role of technology companies in AI development, these principles hold significant normative weight.

In summary, leveraging the human rights framework for AI governance ensures a principled approach that prioritizes human welfare and rights amidst technological advancements. It provides a robust foundation for addressing ethical concerns and promoting accountability in both state and corporate practices related to AI development and deployment. The International Bill of Human Rights and the UN Guiding Principles on Business and Human Rights together offer a robust framework for governing AI technologies and assessing their societal impacts. This framework addresses a fundamental question: What are the respective responsibilities of governments versus tech companies regarding the consequences of new AI technologies? No other proposed AI governance framework matches the human rights framework in defining individual rights, governmental obligations, and corporate responsibilities comprehensively. Moreover, unlike emerging ethical frameworks, the existing human rights framework enjoys unparalleled geopolitical recognition and status under international law. It has been widely adopted and integrated into national constitutions and regulations worldwide, reflecting decades of global consensus and stakeholder negotiations. Non-compliance with these norms risks international condemnation, highlighting their legitimacy despite uneven implementation across states.

Even states that do not fully adhere to human rights principles often frame their actions in human rights terms rhetorically, demonstrating the enduring influence of these norms. In contrast to ad hoc ethical guidelines, universal human rights principles have served as globally recognized governance norms for over seventy years. This framework remains relevant in the digital age, as affirmed by the UN Human Rights Council's resolution emphasizing that rights applicable offline must also be upheld online.

While challenges persist in applying human rights to AI, starting with established global standards provides legitimacy and a shared approach across states and stakeholders. A human rights-based approach offers the

best foundation to mitigate AI's risks while maximizing its potential benefits. As AI capabilities evolve, ongoing efforts to apply international human rights law will be crucial in safeguarding the diverse rights inherent to every individual globally.

2.14 DISCUSSION AND CONCLUSION

In this study, we have explored the profound influence of AI on evaluating the adequacy of human rights and democratic frameworks in addressing these challenges. Our research highlights that AI profoundly disrupts the fundamental principles of democratic institutions.

A precise definition of AI is essential for comprehensively understanding its societal impacts and for implementing effective legal measures to mitigate these effects. In our study, AI is defined not just by its technical components but also as a vast industrial field encompassing significant socio-technical dimensions. Therefore, regulatory efforts aimed at mitigating AI's threats to democracy should encompass not only technical advancements but also the purposes and applications of AI systems. This approach ensures that human decision-making processes are integrated throughout the design and deployment of AI. Given AI's expansive industrial footprint, it is imperative to balance market dynamics with democratic values and address environmental concerns.

The Digital Services Act includes provisions aimed at preserving freedom of expression amidst AI's impact on online democratic participation. These clauses require transparency in recommendation systems, impose restrictions on political advertising that uses personal data, and mandate clear practices for online content moderation. Such measures are essential for enhancing and protecting human rights in the digital realm.

Similarly, the Digital Markets Act targets the disproportionate market power of major online platforms designated as gatekeepers, aiming to foster fair competition and uphold democratic values in digital markets. While still under negotiation, the AI Act Proposal represents a pioneering effort to comprehensively regulate AI's impacts on democracy. It seeks to prohibit

AI systems that pose severe threats to democratic foundations while imposing obligations on systems that may undermine fundamental rights.

A primary concern regarding the efficacy of human rights amidst AI's influence on democracy stems from the nascent stage of current legal regulations. As AI evolves rapidly, presenting new and diverse threats, existing laws intended to mitigate these risks are hindered by bureaucratic inertia. Hildebrandt pessimistically argues that in the digital age, human rights and democracy may struggle to endure without robust legal frameworks that can adapt to rapidly evolving conditions. Applying overarching, abstract rules to the personalized and frequently opaque outputs of AI systems presents formidable challenges.

Democratic governments confront substantial risks until robust legal regulations capable of adequately addressing the inherent threats of AI are universally enacted. To prevent a potential "AI winter," it is crucial for public pressure to integrate principles of democracy, human rights, and the rule of law into the AI domain, which is heavily shaped by user and consumer interests.

REFERENCES

- Alpaydın, E. (2020). *Yapay öğrenme: yeni yapay zeka*. Tellekt. [📄](#)
- Beetham, D. (1999). *Democracy and human rights* (Vol. 249). Cambridge: Polity Press. [📄](#)
- Cohen, J. (2013). What privacy is for (126. *Harvard Law Review*, 1904). *Harvard Law Review* 7. [📄](#)
- Erdoğan, M. (2001). Demokratik Toplumda İfade Özgürlüğü: Özgürlükçü Bir Perspektif. *Liberal Düşünce Dergisi*, 24, 8–13. [📄](#)
- European Parliament—Legislative Observatory. (2023). Artificial Intelligence Act. <https://oeil.secure.europarl.europa.eu/oeil/popups/summary.do?id=1747977&t=e&l=en>

- Gershgorn, D. (2018). The inside story of how AI got good enough to dominate Silicon Valley. *Quartz*. Retrieved, 5, 12–15. [↗](#)
- Habermas, J. (2023). *Kamusalığın yeni bir yapısal dönüşümü ve müzakereci demokrasi* (T. Bora, Çeviri). İstanbul: İletişim Yayınları. [↗](#)
- Hildebrandt, M. (2020). Smart technologies. *Internet Policy Review*, 9(4), 1–16. [↗](#)
- Hill, K., & Krolik, A. (2020). At Talkspace, start-up culture collides with mental health concerns. *The New York Times*. [↗](#)
- Judge, E. F., & Gervais, D. J. (2009). Of silos and constellations: Comparing notions of originality in copyright law. In *Intellectual property protection of fact-based works*. Edward Elgar Publishing. [↗](#)
- Kurzweil, R. (2005). The singularity is near. In *Ethics and emerging technologies* (pp. 393–406). London: Palgrave Macmillan UK. [↗](#)
- McCorduck, P., & Cfe, C. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. AK Peters/CRC Press. [↗](#)
- Mercan, G., & Selçuk, Z. V. (2024). Artificial intelligence (AI) activities in legal practices. *International Journal of Eurasian Education and Culture*, 9(25), 131–144. [↗](#)
- Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180089. [↗](#)
- Nilsson, N. J. (2009). *The quest for artificial intelligence*. Cambridge University Press. [↗](#)
- Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, 4, 100005. [↗](#)
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Pearson. [↗](#)
- Singh, N., Bandyopadhyay, T. K., Sahoo, N., & Tiwari, K. (2021). Intellectual property issues in artificial intelligence: Specific reference to

- the service sector. *International Journal of Technological Learning, Innovation and Development*, 13(1), 82–100. [↗](#)
- Susskind, J. (2018). Chatbots are a danger to democracy. *The New York Times*, 4. [↗](#)
- Tegmark, M. (2019). *Yaşam 3.0 yapay zeka çağında insan olmak*. İstanbul: Pegasus Yayınları. [↗](#)
- Turner, J. (2018). *Robot rules: Regulating artificial intelligence*. Springer. [↗](#)
- Valcke, P., Clifford, D., & Dessers, V. (2021). Constitutional challenges in the emotional AI era. [↗](#)
- Weinbaum, D., & Veitas, V. (2017). Open ended intelligence: The individuation of intelligent agents. *Journal of Experimental & Theoretical Artificial Intelligence*, 29(2), 371–396. [↗](#)
- Woolley, S. C., & Howard, P. N. (2016). Political communication, computational propaganda, and autonomous agents: Introduction. *International journal of Communication*, 10. [↗](#)
- Yeung, K. (2018). A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. *MSI-AUT*, 5, 25–27. [↗](#)
- Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3), 107–115.
- Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., ... & Perrault, R. (2021). The AI index 2021 annual report. arXiv preprint arXiv:2103.06312. [↗](#)

3 Computational Intelligence in Human Rights Protection

A Review and Future Directions

Budi Agus Riswandi, Saquib Ahmed, Anjali Raghav, Bhupinder Singh, Atul Sangal, Manmeet Kaur Arora, Sahil Lal, and Saurabh Chandra

DOI: [10.1201/9781003668657-3](https://doi.org/10.1201/9781003668657-3)

3.1 INTRODUCTION

The integration of human rights and computational intelligence (CI) is a vital field of study in modern society, as progress in artificial intelligence (AI) and machine learning (ML) progressively integrates our social, political, and economic environments ([Brotcke, 2022](#)). As these technologies scale, they open entirely new ways to improve human rights protections and access to services and empower vulnerable groups in entirely new ways ([Abebe et al., 2020](#)). But they are also fraught with challenges and serious risks that could potentially erode fundamental rights and freedoms ([Elluri et al., 2023](#)). Thus, the duality of this factor requires us to conduct a thorough analysis of what is both beneficial and challenging in which CI influences human rights with accompanying requirements for

adequate legal frameworks or other regulatory measures as well ([Hagendorff, 2019](#)).

However, CI encompasses a collection of techniques and methodologies that empower robots to acquire knowledge from data, adjust to novel information, and independently make judgments. Also, this skill has significant ramifications for multiple industries, such as healthcare, law enforcement, education, and governance. For instance, AI systems can examine extensive quantities of data to detect patterns of discrimination, oversee violations of human rights, and assist in ensuring fair treatment under the law. In the field of healthcare, AI tools help to prescribe personalized interventions and predict health hazards to enable patients to receive optimized quality care, resulting in a step more toward Right to Health ([Singh & Kaunert, 2024](#)). Second, if properly supported with data, transparency and accountability in governance can be optimized by CI from both analyzing public information and direct citizen engagement. However, the implementation of AI and ML technologies is also riddled with pitfalls ([Thapaliya & Sharma, 2023](#)). Algorithmic bias can introduce discriminatory practices, especially against minorities. For instance, predictive policing algorithms can lead to unfair profiling of minority communities and further violations of the right not to be discriminated against ([Dash et al., 2022](#)).

Likewise, governments and corporations use AI for surveillance to snoop on people without their permission, which violates privacy rights. Nevertheless, the use of AI and ML technologies comes with inherent drawbacks. The possibility of algorithmic bias gives rise to issues of discrimination, especially toward underrepresented populations ([Barocas et al., 2023](#)). The absence of responsibility can weaken confidence in organizations and corrode the principles of justice, ultimately jeopardizing the fundamental pillars of democratic society ([Narayana & Sreedevi, 2023](#)).

Hence, it is crucial to set unambiguous standards and legislation that guarantee the development and implementation of AI systems in a manner that upholds and advances human rights ([Mhlanga, 2021](#)). The absence of responsibility can weaken confidence in organizations and corrode the principles of justice, ultimately jeopardizing the fundamental pillars of

democratic society ([Pocher et al., 2023](#)). Hence, it is crucial to set unambiguous standards and legislation that guarantee the development and implementation of AI systems in a manner that upholds and advances human rights ([Chen et al., 2018](#)).

3.1.1 DEFINITION OF CI AND ITS RELEVANCE IN CONTEMPORARY SOCIETY

CI is the science that studies the extent to which we can apply principles from biology, neuroscience, and linguistics for developing new computational methodologies ([Tyagi et al., 2020](#)). The aim is to create algorithms and methods that can learn from data, adjust over time by adapting themselves within a variable state environment, and permit collaborative completion of tasks once thought too difficult in ways compatible with human behavior ([Cao et al., 2021](#)). Neural networks, fuzzy systems, and evolutionary computation are the three primary pillars on which CI stands ([Alazab et al., 2023](#)). Neural networks are so-called because they are inspired by the structure of neurons in human brains, which process and complete data similar to a preset example ([Zhang et al., 2022b](#)).

The definition of “AI” was coined by J. McCarthy in the year 1956 ([Singh, 2023a](#)). Also, the initial focus on AI was on defining intelligence as the starting point of human thinking, but this also caused controversy because it is difficult to measure ([Lee & Shin, 2020](#)). After all, there is no definition for generalizing intelligent behavior, and at the same time, technical processes in the ICT and Internet technologies have created a foundation from which to further legally and practically characterize AI ([Bhuvaneshwari & Kaythry, 2023](#)). These days AI as a field encompasses so much under its name, from ML to neural networks to computer vision ([Li et al., 2020](#)). AI in contemporary societies: legal status and definition, implementation in the public sector across various countries ([Cheng et al., 2021](#)).

Fuzzy systems use a collection of fuzzy logic rules to map inputs (which form vectors) to outputs, and they perform approximate reasoning rather than exact calculations ([Bhuvaneshwari & Kaythry, 2023](#)). This means that when creating an output value based on the mapping from input values using these rules, no guarantees can be made for correct decisions ([Mühlhoff, 2021](#)). Evolutionary computation is a family of algorithms for global optimization, inspired by the principles inherent to biological evolution, such as mutation and selection ([Saltz et al., 2019](#)). More recently, CI has been further reinforced with other bio-inspired methods such as swarm intelligence, artificial immune systems, and more into the deep learning phase ([Hojtink & Planqué-van Hardeveld, 2022](#)). These methods allowed for building more complex models of AI capable of handling various practical tasks in image recognition, natural language processing, or robotics ([Utami et al., 2022](#)). The importance of CI in the modern-day world is manifold. But as data has exponentially multiplied in volume and complexity, so have the demands on intelligent systems that can cut through it all to provide meaningful insights and make decisions ([Zhang et al., 2022a](#)). CI-based algorithms are applied in almost all domains of interest, for example, healthcare, finance, smart cities, etc. to automate processes, customize services, and find hidden patterns within data ([Baker & Robinson, 2020](#)). Furthermore, CI is playing an essential role in the evolution of autonomous systems such as self-governing cars and smart personal assistants. Workloads that will be solved by methods in this track require the perception of the environment, reasoning about it, and action selection to determine the correct course of action in real time based on these perceptions ([Aziz & Andriansyah, 2023](#)). In an increasingly more connected and automated world, continuous integration will also be key to keeping such systems safe, efficient, and reliable ([Bao et al., 2022](#)).

3.1.2 OVERVIEW OF HUMAN RIGHTS PRINCIPLES AND THEIR IMPORTANCE IN SAFEGUARDING INDIVIDUAL FREEDOMS AND DIGNITY

Human rights principles constitute the essential norms that help to safeguard fundamental freedom and dignity ([McKendrick, 2019](#)). These principles, which are stated in international documents, that is, Universal Declaration of Human Rights, highlight the natural dignity and rights that every human has, and no one can take them away from him/her because they are inalienable ([Rai & Sahu, 2020](#)). Covering civil, political, economic/social and cultural rights that make sure people are not discriminated against or persecuted. The core principles are relating to equality, non-discrimination, and the right to life and liberty ([Sharma et al., 2023](#)). They ensure freedoms of speech, assembly, and religion, as well as the right to fair trials and privacy. These include such vital services as education, healthcare, and an adequate standard of living ([Sharma et al., 2023](#)). These principles are important as they conserve the freedoms of an individual, bring about justice, and promote human dignity ([Varma et al., 2022](#)). They offer a standard for preventing violations, promoting inclusive societies, and ensuring governments are held to account. Defending human rights is important for living together in peace, social cohesion, and the most fragile populations ([Tippins et al., 2021](#)).

3.1.3 ROLE OF CI IN PROMOTING HUMAN RIGHTS

CI is a form of AI that includes technologies such as neural networks, fuzzy systems, and evolutionary computing ([Lui & Lamb, 2018](#)). These technologies are essential for the support of human rights and enforcement through advanced data analysis and better decision-making processes, or else in policymaking that defends individual freedoms and dignity (Mijwil, 2023). The convergence of CI and human rights is now generating cutting-edge approaches to responding to complex global problems.

3.1.4 IMPROVING DATA ANALYSIS AND MONITORING

Advanced data analysis and monitoring are key functions of CI to enable human rights ([Siwakoti et al., 2023](#)). Using AI tools, one can mine data

from numerous sources (like social networks, government-run databases and satellite imagery) that could be employed to reveal patterns of human rights abuses ([Zhang et al., 2020](#)). By combining aspects of these creation models with ML, algorithmic models can be built to review posts (for example) on social media or in news articles and identify trends where human rights abuses such as hate speech, discrimination, or violence by the majority toward a minority are taking place ([Nassar & Kamal, 2021](#)). In conflict regions, CI can analyze satellite photos to signal military activity or population displacement and raise alarms on a potential human rights crisis ([Alruwaili, 2023](#)). This information allows international institutions, governments, and NGOs to act more quickly and effectively in order not to affect the populations at risk ([Grimm et al., 2021](#)).

3.1.5 ENHANCED DECISION-MAKING AND POLICY EXECUTION

Thus, CI plays its role in enhancing decision-making and policy implementation by providing accurate information at the right time to stakeholders ([Dhoni & Kumar, 2023](#)). AI-driven decision support systems can be used to simulate how different policies would play out in practice, helping governments and organizations pick the best plans for safeguarding human rights ([Kordzadeh & Ghasemaghaei, 2022](#)). This might mean using AI models to help regionally analyze the most effective interventions in a given refugee crisis and coordinate allocation of resources and humanitarian aid with an eye on positive outcomes ([Mijwil & Aljanabi, 2023](#)). Besides, CI can be used for the development of human rights policies. They find the best practices and takeaways based on historical data of what worked (or not) to allow themselves a fact-based approach in designing new policy ([Sambasivan et al., 2021](#)). It is a well-known fact that the use of data enhances effectiveness and efficiency in terms of human rights ([Singh, 2023b](#)).

3.1.6 ENABLING JUSTICE

Justice should be easily accessible to everyone, and if used correctly, CI can play a major role in providing access to legal resources ([Singh, 2023c](#)). Therefore, for the people to get information about their rights and different legal remedies that can be taken by them, AI-powered platforms will provide help in assistance ([Zhang et al., 2023](#)). This allows people to make complaints, seek legal advice from lawyers, and understand the law better so that they can represent themselves in court ([Singh, 2023d](#)). Moreover, CI could also be of quite use to legal professionals in managing cases and performing research tasks ([Katyal, 2019](#)). NLP algorithms can parse legal documents and distill case law and key points of data to streamline time-consuming (and potentially fallible) processes ([Irfan et al., 2023](#)). It can even be leveraged to detect biases in judicial decisions and ensure a more equitable justice system ([Selbst & Barocas, 2018](#)).

3.1.7 TRANSPARENCY AND ACCOUNTABILITY

This aids in maintaining basic human rights, making transparency and accountability an important part of the justice system where CI can be beneficial ([Chohan et al., 2023](#)). With the blockchain technology coming as one of the types, which is also underlined by DLT, it becomes hard to falsify such records like, for example, ownership and election results or established supply chain processes ([Singh et al., 2023](#)). Blockchain creates an indelible ledger of transactions, which can prevent corruption by restricting fraudulent and other practices that destroy human rights ([Sowmya et al., 2023](#)). AI-supported views give insights into anomalies and likely corruption in financial transactions and procurement ([Daugherty & Wilson, 2018](#)). This is particularly important when it comes to monitoring the use of public funds and international development assistance, as this can help make sure that those resources are spent properly on human rights activities ([Świątkowska, 2020](#)).

3.2 RISKS AND CHALLENGES OF CI ON HUMAN RIGHTS

CI offers many opportunities to advance human rights, but it brings with it a number of risks and challenges that should be dealt with very attentively ([Berk et al., 2021](#)). This goes on in many aspects, from the privacy side to bias and accountability issues, passing by potential misuse of technology ([Abbas et al., 2024](#)). It is, therefore, crucial to address these risks and challenges so that CI does make a positive difference for human rights ([Sudhakar & Kumar, 2023](#)). Once we have sound ethical guidelines, regulatory frameworks, and constant supervision in place while using AI, only then will we be able to balance the advantages of this powerful technology with our own dignity as people ([Kochhar et al., 2023](#)).

Privacy Concerns: CI requires tremendous amounts of data collection, which raises privacy issues ([Kaushik, 2024](#)). Many AI systems need access to a lot of personal data to work. When this happens, the data might be easily misused, which can result in illegal surveillance or breaches of private information or privacy rights ([Liu et al., 2024](#)). It is essential to minimize the risks by implementing strong data protection policies, as well as transparent use of data.

Bias and Discrimination: AI systems can—intentionally or otherwise—encode and perpetuate biases, discrimination, and inequalities. The problem is that ML algorithms learn from historical data, which may contain its own biases. As a result, these biases can be present in the decisions and predictions made by AI systems, causing injustices to particular races, genders, or social statuses. To combat this challenge, it is essential that unbiased algorithms be developed and AI systems be audited periodically for discrimination ([Rajkumar et al., 2023](#)).

Transparency and Accountability: The functionality of awaiting is so difficult that even when told where a decision is to be made, AI systems will still have problems with identification because some algorithms are complex and black box, which affects, unlike learn combination sufficiently, transparency ([Nisha & Muthurajkumar, 2023](#)). The AI's black box can guard against efforts to keep the individuals or organizations responsible for human rights abuses.

There is a need for developing AI models that are explainable and can be held accountable under well-defined frameworks so that we can troubleshoot the cause of such errors being made by the systems ([Shukla & Tiwari, 2024](#)).

Misuse of Technology: There is also the substantial risk of CI being misapplied in degenerative political systems or by malign actors. For surveillance and social control as violations of human rights. Deployment of AI technologies may lead to situations in which governments or other actors use them for mass monitoring. For the misuse of AI, this requires strong legislation or comprehensive regulation and respect for regulations across different countries to keep these technologies working properly ([Paddalwar et al., 2023](#)).

3.3 CONSIDERATIONS AND REGULATORY FRAMEWORKS

With the progress of CI technologies and their extensive application in society, it is important to provide detailed ethical guidance and enter robust regulative frameworks allowing these tools to work for us responsibly ([Itani et al., 2023](#)). The restoration of these dimensions is essential to safeguard human rights, rebuild confidence, and foster inclusive development. Finding the right way to regulate and balance such innovations with ethical responsibility is key for protecting human rights and societal well-being while benefiting from all of these advantages that might arise by applying CI. By effectively addressing ethical concerns and setting up proper regulations, we can pave the way to ensure that AI technologies are built and operate in ways that respect principles of fairness, transparency, responsibility, etc., thereby aiding in creating a just and equitable society ([Odeh & Abu Taleb, 2023](#)).

3.3.1 ETHICAL CONSIDERATIONS

Training CI systems on biased data can preserve and magnify those existing biases (similar to use cases) in cybercrime ([Ortiz-Ruiz et al., 2024](#)). This means carefully managing the source and architecture of datasets to be varied without being prejudiced, training an algorithm in a manner that does not reveal bias, and keeping AI systems under check for any discriminatory practices. When we build more ethical AIs, those systems should be explicitly designed to minimize bias and ensure equality for all demographics ([Ali et al., 2024](#)). As the majority of AI models are so-called “black boxes” due to their opacity, this fact gives rise to a whole host of serious ethical issues ([Hamad, 2023](#)). When it comes to AI and modern tech, transparency of how decisions are being made is paramount. If an AI is opaque, stakeholders will have a problem verifying the behavior of that system and correcting errors or biases. Explainable AI (XAI) is a novel discipline aimed at making the decision processes of AI more understandable to human beings ([Latif et al., 2024](#)).

Privacy and Data Protection: AI systems require huge volumes of data, which introduce serious privacy issues ([Anis et al., 2023](#)). The data should be collected, stored, and used in a transparent way with permission from the individual, enabling ethical AI. The use of appropriate techniques for protection, anonymization, and compliance with data privacy laws such as General Data Protection Regulation (GDPR) are essential to avoid any abuse or violation of the rights imposed by law ([Abdalrdha et al., 2023](#)).

Autonomy and Consent: Gaining user autonomy and getting the permission to install forced consent are basic ethical concepts. The problem has received a lot of attention as many AI applications are in sensitive domains where it is crucial for individuals to understand how choices will be made and what effect those decisions could have on their lives (such as healthcare or finance). This includes clear communication and understandable consent procedures ([AL-Dosari et al., 2024](#)).

Accountability: Who is accountable for AI-driven decisions? And this is a major ethical concern that must be addressed. There needs to be more clarity around what the obligations of developers, deployers, and users—when it comes to algos—should look like. Accountability also requires effective redress mechanisms and avenues for challenging or appealing AI acts ([Ambika, 2024](#)).

3.4 REGULATORY FRAMEWORKS

Comprehensive Legislation: The key is for governments to adopt robust regulatory frameworks that encompass all aspects of AI development and deployment. These should include, inter alia, data protection laws, algorithmic transparency and bias avoidance, as well as accountability. In a rapidly changing technological landscape, regulations need to be as flexible for organizations while setting clear guidelines on what compliance looks like ([Aljuhani et al., 2023](#)).

Standards and Coordination at the International Level: This drives full international collaboration and some establishment of global standards as AI tech is a global technology trend. The United Nations, the European Union, and other such organizations have been developing frameworks capable of governing AI ethics to confirm that their design as well as implementation are in tandem with human rights principles. Streamlining regulations by nation starts to reduce the potential for regulatory arbitrage, helping establish consistent digital ethics standards.

Recommended Ethical Guidelines and Best Practices: Professional bodies and industry groups are needed to produce common ethical principles and best practices on AI development. These principles need to be built on interdisciplinary research, including perspectives of ethics, law, and technology. This includes the longstanding work of IEEE through their Ethically Aligned Design, which sits alongside

similar documents such as Europe's *Ethics Guidelines for Trustworthy AI*.

Regulatory Sandboxes: Regulatory sandboxes are environments in which to test new AI technologies under close regulatory oversight. As such, these sandboxes are designed to facilitate innovation while preserving adherence to ethical and legal standards. Consequently, regulators can improve and update frameworks with a view of monitoring these AI applications that are actually entering the market to see where there might be issues.

Public Engagement and Educational Activities: Democratic governance requires that the public be involved in discussions about AI ethics and regulation. Raising general awareness of the capabilities and limitations of AI technologies promotes a more fruitful public debate, guided by an umbrella notion on how these tools should be utilized in our society. You can have public consultations and participatory policy processes that make sure different views are brought into the regulatory decisions ([Lone et al., 2023](#)).

3.5 CONCLUSION AND FUTURE SCOPE

CI is changing the landscape of human rights, enabling new and powerful instruments for data analysis, decision-making processes, and access to justice, making those empowered by transparency and accountability. Using these technologies, it can design more efficient human rights protective and promotional strategies on a global scale. It is important that when we deploy CI, we do so in a way that is ethically guided and respects privacy to avoid misuse, or else these technologies will not serve the interest of human rights.

REFERENCES

- Abbas, S., Bouazzi, I., Ojo, S., Al Hejaili, A., Sampedro, G. A., Almadhor, A., & Gregus, M. (2024). Evaluating deep learning variants for cyber-attacks detection and multi-class classification in IoT networks. *PeerJ Computer Science*, 10, e1793. [↗](#)
- Abdalrdha, Z. K., Al-Bakry, A. M., & Farhan, A. K. (2023). A survey on cybercrime using social media. *Iraqi Journal for Computers and Informatics*, 49(1), 52–65. [↗](#)
- Abebe, R., Barocas, S., Kleinberg, J., Levy, K., Raghavan, M., & Robinson, D. G. (2020). Roles for computing in social change. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 252–260). ACM. [↗](#)
- Alazab, M., Soman, K. P., Srinivasan, S., Venkatraman, S., & Pham, V. Q. (2023). *Deep Learning for Cyber Security Applications: A Comprehensive Survey*. Authorea Preprints. [↗](#)
- AL-Dosari, K., Fetais, N., & Kucukvar, M. (2024). Artificial intelligence and cyber defense system for banking industry: A qualitative study of AI applications and challenges. *Cybernetics and Systems*, 55(2), 302–330. [↗](#)
- Ali, H. H., Naif, J. R., & Humood, W. R. (2024, March). Deep learning algorithms for IoT security (survey). In *AIP conference proceedings* (Vol. 2885, No. 1). AIP Publishing. [↗](#)
- Aljuhani, A., Kumar, P., Alanazi, R., Albalawi, T., Taouali, O., Islam, A. N., ... & Alazab, M. (2023). A deep learning integrated blockchain framework for securing industrial IoT. *IEEE Internet of Things Journal*, 1, 10–15. [↗](#)
- Alruwaili, F. F. (2023). Optimal hybrid deep learning enabled attack detection and classification in IoT environment. *Computers, Materials & Continua*, 75(1), 99–115. [↗](#)
- Ambika, N. (2024). Early detection of darknet traffic in Internet of Things applications. In *Automated Secure Computing for Next-Generation Systems* (pp. 139–153). Wiley. [↗](#)
- Anis, G. M., Aboutabl, A. E., & Galal, A. (2023). Machine learning for detecting cybercrime in the banking sector. *Journal of Southwest Jiaotong*

University, 58(5), 5–7. [↗](#)

- Aziz, L. A. R., & Andriansyah, Y. (2023). The role artificial intelligence in modern banking: An exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. *Reviews of Contemporary Business Analytics*, 6(1), 110–132. [↗](#)
- Baker, D. J., & Robinson, P. H. (Eds.). (2020). *Artificial Intelligence and the Law: Cybercrime and Criminal Liability*. Routledge. [↗](#)
- Bao, Y., Hilary, G., & Ke, B. (2022). Artificial intelligence and fraud detection. In *Innovative Technology at the Interface of Finance and Operations: I* (pp. 223–247). Springer. [↗](#)
- Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press. [↗](#)
- Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 50(1), 3–44. [↗](#)
- Bhuvaneshwari, A. J., & Kaythry, P. (2023). A review of deep learning strategies for enhancing cybersecurity in networks: Deep learning strategies for enhancing cybersecurity. *Journal of Scientific & Industrial Research (JSIR)*, 82(12), 1316–1330. [↗](#)
- Brotcke, L. (2022). Time to assess bias in machine learning models for credit decisions. *Journal of Risk and Financial Management*, 15(4), 165. [↗](#)
- Cao, L., Yang, Q., & Yu, P. S. (2021). Data science and AI in FinTech: An overview. *International Journal of Data Science and Analytics*, 12, 81–99. [↗](#)
- Chen, Z., Van Khoa, L. D., Teoh, E. N., Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: A review. *Knowledge and Information Systems*, 57, 245–285. [↗](#)
- Cheng, X., Liu, S., Sun, X., Wang, Z., Zhou, H., Shao, Y., & Shen, H. (2021). Combating emerging financial risks in the big data era: A perspective review. *Fundamental Research*, 1(5), 595–606. [↗](#)

- Chohan, M. N., Haider, U., Ayub, M. Y., Shoukat, H., Bhatia, T. K., & Hassan, M. F. U. (2023). Detection of cyber attacks using machine learning based intrusion detection system for IoT based smart cities. *EAI Endorsed Transactions on Smart Cities*, 7(2), e4–e4. [↗](#)
- Dash, B., Sharma, P., & Ali, A. (2022). Federated learning for privacy-preserving: A review of PII data analysis in Fintech. *International Journal of Software Engineering & Applications (IJSEA)*, 13(4), 3–7. [↗](#)
- Daugherty, P. R., & Wilson, H. J. (2018). *Human + Machine: Reimagining Work in the Age of AI*. Harvard Business Press. [↗](#)
- Dhoni, P., & Kumar, R. (2023). *Synergizing Generative AI and Cybersecurity: Roles of Generative AI Entities, Companies, Agencies, and Government in Enhancing Cybersecurity*. Authorea Preprints. [↗](#)
- Elluri, L., Mandalapu, V., Vyas, P., & Roy, N. (2023). Recent advancements in machine learning for cybercrime prediction. *Journal of Computer Information Systems*, 1, 1–15. [↗](#)
- Grimm, P. W., Grossman, M. R., & Cormack, G. V. (2021). Artificial intelligence as evidence. *Northwestern Journal of Technology and Intellectual Property*, 19, 9. [↗](#)
- Hagendorff, T. (2019). From privacy to anti-discrimination in times of machine learning. *Ethics and Information Technology*, 21(4), 331–343. [↗](#)
- Hamad, S. I. (2023). A review of deep learning applied to cyber security. *Texas Journal of Engineering and Technology*, 22, 49–57. [↗](#)
- Hoijsink, M., & Planqué-van Hardeveld, A. (2022). Machine learning and the platformization of the military: A study of google's machine learning platform TensorFlow. *International Political Sociology*, 16(2), olab036. [↗](#)
- Irfan, M., Elmogy, M., & El-Sappagh, S. (Eds.). (2023). *The Impact of AI Innovation on Financial Sectors in the Era of Industry 5.0*. IGI Global. [↗](#)
- Itani, M., Basheer, H., & Eddine, F. (2023, July). Attack detection in IoT-based healthcare networks using hybrid federated learning. In *2023 International Conference on Smart Applications, Communications and Networking (SmartNets)* (pp. 1–7). IEEE. [↗](#)

- Katyal, S. K. (2019). Private accountability in the age of artificial intelligence. *UC Berkeley, Public Law Research Paper*, 66, 54.[↵](#)
- Kaushik, K. (2024). Leveraging deep learning techniques for securing the internet of things in the age of big data. *Applying Artificial Intelligence in Cybersecurity Analytics and Cyber Threat Detection*, 309.[↵](#)
- Kochhar, S. K., Bhatia, A., & Tomer, N. (2023). Using deep learning and big data analytics for managing cyber-attacks. In *New Approaches to Data Analytics and Internet of Things through Digital Twin* (pp. 146–178). IGI Global.[↵](#)
- Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388–409.[↵](#)
- Latif, S., Boulila, W., Koubaa, A., Zou, Z., & Ahmad, J. (2024). DTL-IDS: An optimized intrusion detection framework using deep transfer learning and genetic algorithm. *Journal of Network and Computer Applications*, 221, 103784.[↵](#)
- Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, 63(2), 157–170.[↵](#)
- Li, L., Zhao, T., Xie, Y., & Feng, Y. (2020). Interpretable machine learning based on integration of NLP and psychology in peer-to-peer lending risk evaluation. In *Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, Proceedings, Part II* 9 (pp. 429–441). Springer International Publishing.[↵](#)
- Liu, C., Chen, B., Shao, W., Zhang, C., Wong, K., & Zhang, Y. (2024). Unraveling attacks in machine learning-based iot ecosystems: A survey and the open libraries behind them. arXiv preprint arXiv:2401.11723.[↵](#)
- Lone, A. N., Mustajab, S., & Alam, M. (2023). A comprehensive study on cybersecurity challenges and opportunities in the IoT world. *Security and Privacy*, 6(6), e318.[↵](#)
- Lui, A., & Lamb, G. W. (2018). Artificial intelligence and augmented intelligence collaboration: Regaining trust and confidence in the financial

- sector. *Information & Communications Technology Law*, 27(3), 267–283. [↵](#)
- McKendrick, K. (2019). *Artificial Intelligence Prediction and Counterterrorism* (p. 9). London: The Royal Institute of International Affairs-Chatham House. [↵](#)
- Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International Journal of Financial Studies*, 9(3), 39. [↵](#)
- Mijwil, M., & Aljanabi, M. (2023). Towards artificial intelligence-based cybersecurity: The practices and ChatGPT generated ways to combat cybercrime. *Iraqi Journal for Computer Science and Mathematics*, 4(1), 65–70. [↵](#)
- Mijwil, M., Salem, I. E., & Ismaeel, M. M. (2023). The significance of machine learning and deep learning techniques in cybersecurity: A comprehensive review. *Iraqi Journal for Computer Science and Mathematics*, 4(1), 87–101.
- Mühlhoff, R. (2021). Predictive privacy: Towards an applied ethics of data analytics. *Ethics and Information Technology*, 23(4), 675–690. [↵](#)
- Narayana, Y. V., & Sreedevi, M. (2023). DDCATF: Deep learning approach for detection of cybercrime activities based on temporal features. In *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)* (pp. 462–469). IEEE. [↵](#)
- Nassar, A., & Kamal, M. (2021). Ethical dilemmas in AI-powered decision-making: A deep dive into big data-driven ethical considerations. *International Journal of Responsible Artificial Intelligence*, 11(8), 1–11. [↵](#)
- Nisha, S. R., & Muthurajkumar, S. (2023). Semantic graph based convolutional neural network for spam e-mail classification in cybercrime applications. *International Journal of Computers Communications & Control*, 18(1), 1–5. [↵](#)
- Odeh, A., & Abu Taleb, A. (2023). Ensemble-based deep learning models for enhancing iot intrusion detection. *Applied Sciences*, 13(21), 11985. [↵](#)

- Ortiz-Ruiz, E., Bermejo, J. R., Sicilia, J. A., & Bermejo, J. (2024). Machine learning techniques for cyberattack prevention in IoT systems: A comparative perspective of cybersecurity and cyberdefense in Colombia. *Electronics*, 13(5), 824. [↵](#)
- Paddalwar, S., Ragha, L., & Mane, V. (2023). Cyber threat mitigation using machine learning, deep learning, artificial intelligence, and blockchain. In *Intelligent Approaches to Cyber Security* (pp. 163–179). Chapman and Hall/CRC. [↵](#)
- Pocher, N., Zichichi, M., Merizzi, F., Shafiq, M. Z., & Ferretti, S. (2023). Detecting anomalous cryptocurrency transactions: An AML/CFT application of machine learning-based forensics. *Electronic Markets*, 33(1), 37. [↵](#)
- Rai, R., & Sahu, C. K. (2020). Driven by data or derived through physics? A review of hybrid physics guided machine learning techniques with cyber–physical system (CPS) focus. *IEEE Access*, 8, 71050–71073. [↵](#)
- Rajkumar, S., Sheeba, S. L., Sivakami, R., Prabu, S., & Selvarani, A. (2023). An IoT-based deep learning approach for online fault detection against cyber-attacks. *SN Computer Science*, 4(4), 393. [↵](#)
- Saltz, J., Skirpan, M., Fiesler, C., Gorelick, M., Yeh, T., Heckman, R., & Beard, N. (2019). Integrating ethics within machine learning courses. *ACM Transactions on Computing Education (TOCE)*, 19(4), 1–26. [↵](#)
- Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., & Prabhakaran, V. (2021, March). Re-imagining algorithmic fairness in India and beyond. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 315–328). [↵](#)
- Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham Law Review*, 87, 1085. [↵](#)
- Sharma, A., Mansotra, V., & Singh, K. (2023). Detection of Mirai Botnet attacks on IoT devices using deep learning. *Journal of Scientific Research and Technology*, 1, 174–187. [↵](#)
- Shukla, R. K., & Tiwari, A. K. (2024). Security analysis of the cyber crime. In *The Ethical Frontier of AI and Data Analysis* (pp. 257–271). IGI

Global.[↗](#)

- Singh, B. (2023a). Federated learning for envision future trajectory smart transport system for climate preservation and smart green planet: Insights into global governance and SDG-9 (Industry, Innovation and Infrastructure). *National Journal of Environmental Law*, 6(2), 6–17.[↗](#)
- Singh, B. (2023b). Tele-health monitoring lensing deep neural learning structure: Ambient patient wellness via wearable devices for real-time alerts and interventions. *Indian Journal of Health and Medical Law*, 6(2), 12–16.[↗](#)
- Singh, B. (2023c). Blockchain technology in renovating healthcare: Legal and future perspectives. In *Revolutionizing Healthcare Through Artificial Intelligence and Internet of Things Applications* (pp. 177–186). IGI Global.[↗](#)
- Singh, B. (2023d). Unleashing alternative dispute resolution (ADR) in resolving complex legal-technical issues arising in cyberspace lensing E-commerce and intellectual property: Proliferation of E-commerce digital economy. *Revista Brasileira de Alternative Dispute Resolution-Brazilian Journal of Alternative Dispute Resolution-RBADR*, 5(10), 81–105.[↗](#)
- Singh, B., & Kaunert, C. (2024). Integration of cutting-edge technologies such as Internet of Things (IoT) and 5G in health monitoring systems: A comprehensive legal analysis and futuristic outcomes. *GLS Law Journal*, 6(1), 13–20.[↗](#)
- Singh, N. J., Hoque, N., Singh, K. R., & Bhattacharyya, D. K. (2023). Botnet-based IoT network traffic analysis using deep learning. *Security and Privacy*, e355.[↗](#)
- Siwakoti, Y. R., Bhurtel, M., Rawat, D. B., Oest, A., & Johnson, R. C. (2023). Advances in IOT security: Vulnerabilities, enabled criminal services, attacks and countermeasures. *IEEE Internet of Things Journal*, 1, 10–15.[↗](#)
- Sowmya, B. J., Hanumantharaju, R., Kumar, D. P., & Srinivasa, K. G. (2023). Identification of authorship and prevention of fraudulent transactions/cybercrime using efficient high performance machine

- learning techniques. *International Journal of Business Intelligence and Data Mining*, 22(1–2), 144–169. [↵](#)
- Sudhakar, & Kumar, S. (2023). ABBDIoT: Anomaly-based botnet detection using machine learning model in the Internet of Things network. In *International Conference on IoT, Intelligent Computing and Security: Select Proceedings of IICS 2021* (pp. 235–245). Singapore: Springer Nature Singapore. [↵](#)
- Swiątkowska, J. (2020). Tackling cybercrime to unleash developing countries' digital potential. *Pathways for Prosperity Commission Background Paper Series*, 33, 2020–01. [↵](#)
- Thapaliya, S., & Sharma, P. K. (2023). Optimized deep neuro fuzzy network for cyber forensic investigation in big data-based IoT infrastructures. *International Journal of Information Security and Privacy (IJISP)*, 17(1), 1–22. [↵](#)
- Tippins, N. T., Oswald, F. L., & McPhail, S. M. (2021). Scientific, legal, and ethical concerns about AI-based personnel selection tools: A call to action. *Personnel Assessment and Decisions*, 7(2), 1. [↵](#)
- Tyagi, A. K., Aswathy, S. U., & Abraham, A. (2020). Integrating blockchain technology and artificial intelligence: Synergies perspectives challenges and research directions. *Journal of Information Assurance and Security*, 15(5), 1554. [↵](#)
- Utami, S. H., Purnama, A. A., Hidayanto, A. N., Utami, S. H., Purnama, A. A., & Hidayanto, A. N. (2022). Fintech lending in Indonesia: A sentiment analysis, topic modelling, and social network analysis using Twitter data. *International Journal of Applied Engineering & Technology*, 4(1), 50–56. [↵](#)
- Varma, P., Nijjer, S., Sood, K., Grima, S., & Rupeika-Apoga, R. (2022). Thematic analysis of financial technology (Fintech) influence on the banking industry. *Risks*, 10(10), 186. [↵](#)
- Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2023). *Dive into Deep Learning*. Cambridge University Press. [↵](#)

- Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022a). Interpretable machine learning models for crime prediction. *Computers, Environment and Urban Systems*, 94, 101789.[↵](#)
- Zhang, Y., Xiong, F., Xie, Y., Fan, X., & Gu, H. (2020). The impact of artificial intelligence and blockchain on the accounting profession. *IEEE Access*, 8, 110461–110477.[↵](#)
- Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022b). *Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research*. IEEE Access.[↵](#)

4 The Economic Potential of AI

Balancing Growth with Human Rights and Ethics

Kushagra Shukla and Rashmi S

DOI: [10.1201/9781003668657-4](https://doi.org/10.1201/9781003668657-4)

4.1 INTRODUCTION

Artificial intelligence (AI) refers to technologies that enable machines to mimic human abilities, such as problem-solving, learning, and decision-making. Through automation and data analysis, AI can significantly improve productivity and reduce operational costs. For example, AI's predictive analytics help businesses make informed decisions, understand customer behavior more deeply, and plan future strategies. Technologies like machine learning (ML) and natural language processing (NLP), both integral parts of AI, are already transforming industries such as healthcare, finance, manufacturing, and transportation. These innovations make processes more efficient, enhance outcomes, and contribute to overall economic growth. However, the rapid implementation of AI also raises challenges, including job displacement, income inequality, and ethical concerns.

4.2 HOW AI ENHANCES EFFICIENCY

Automation ([Al Rubaie et al., 2024](#)) brought on by AI is changing jobs. It does this by freeing individuals from doing tasks that are repeated and increasing productivity. The prominent development of robotics is the base of the new working style where things operate more efficiently, and this concept results in higher productivity and cost savings.

4.2.1 AUTOMATING REPETITIVE TASKS

One of the main advantages of ML in automation is in completing repetitive tasks that require human time. AI algorithms can be used to automate tasks such as data entry, processing large datasets, and repositioning. These routine tasks are now computerized, releasing human capital to work in more strategic and creative professions. The main aim of automation is to save time and manpower. Firstly, it takes risks concerning human traits by eliminating errors which results in a more precise and accurate yield. Furthermore, it saves time and energy because some tasks are performed automatically. Lastly, automation means that businesses can go about their scaling more efficiently than before, especially if automated processes are in place that do not require more resources to accommodate more work.

4.2.2 APPLICATION OF AUTOMATION

With AI intervention, the response time is reduced, thereby ensuring not only consistent but also accurate customer support throughout. Chatbots can be programmed to handle all kinds of customer queries, from account information queries to troubleshooting technical problems and complex issues. The time that the employees spend interacting with customers can thus be used for more valuable tasks like solving complex issues and nurturing team bonding. Customer relationships are not the only area ([Moch, 2023](#)) where the bots can be applied. They can be used in other areas, including marketing and sales. For instance, bots can be utilized to interact with website visitors, answer product queries, and offer a route toward purchase. Automation of such interactions can be a great advantage

for businesses as it can result in higher conversion rates which in turn increase the revenue.

4.3 DATA-DRIVEN OPTIMIZATION

4.3.1 AI DETECTING INEFFICIENCIES FROM VAST DATASETS

AI-based optimization is an innovative strategy that enables one to analyze the enormous data packet and detect the inefficiencies of the various processes that run any business ([Walas Mateo & Redchuk, 2021](#)). The analysis findings gained serve to guide organizations to operationalize measures that will cut wastage, optimize productivity, and hence bring down operating costs to a significant level. At the core of AI ML-powered optimization processes is the use of computer algorithms that can dig into very large data volumes. Such algorithms have been devised to detect patterns, trends, or anomalies that human analysts may be prone to overlook. AI as a tool in the process of data-based optimization has a big advantage, in that it can spot inefficiencies across various business areas. AI can assist in logistics management, production lines, or even the customer service departments by revealing weak areas through competitive data analysis. Such an approach may consist of pinpointing bottlenecks, minimizing waste, ensuring a rational input–output balance, and making decisions in a time-bound manner.

Process Automation: Through the automation of repetitive tasks and the implementation of workflows, the teams can reduce manual errors, increase efficiency, and provide more opportunities for creative and revenue-generating activities. Organizations can make data-driven forecasts of demand ([Sun et al., 2020](#)), tune up maintenance schedules, and keep inventory levels within minimum and maximum limits.

Dynamic Pricing: AI tools can assess market tendencies, competitor pricing tactics, and shopper behavior to set optimal prices in an instant,

reaching revenue and profit goals.

Resource Allocation: Through the process of undertaking a study on resource exploitation and productivity parameters, organizations can make appropriate resource allocations to guarantee that resources will be distributed properly and appropriately.

Supply Chain Optimization: AI empowers the supply chain with its capability to map the most favorable transport routes, and possible routes of transport, and reduce inventory levels with demand forecasting.

4.4 AI-BACKED DECISION-MAKING

Decision-making is getting more advanced and better with time, with the help of AI. Different sectors are harnessing AI for activities such as supply chain, management, healthcare, finance, marketing, and transportation. Through analyzing vast datasets and identifying patterns, AI helps us with valuable insights, predictions, and recommendations for better decision-making processes. One of the leading consumer goods companies Unilever (Unilever, 2024) uses AI to trace its palm oil supply chain and discover endangerment to deforestation. The company can do tracking by online mobile cell signals and satellite, through the tracing of the slate from the palm oil to different branches of the channels. The companies will be notified when there is any transition of forests in real-time.

Smart port practices now rely on AI-enabled technologies to expedite decision-making, enhance work efficiency, and factor in the consequences of costly environmental pollution. Through the Port of Rotterdam in the Netherlands ([Purdy, 2023](#)), a platform named PortXchange has emerged that integrates data from various sources to provide a real-time dashboard view of all aspects in question. Such platforms are employed by seaports worldwide to facilitate more targeted decision-making and timely port operation planning. When it comes to healthcare, AI may help doctors to lessen pressure by analyzing necessary information, that is, those needed

for effective decision-making. Many other utilizations can be envisioned, including those in areas of business continuity and disaster governance, to mention simply a few. We can also employ this for crisis and risk management in different types of financial investments.

AI is also used to create co-pilots of decisions that can analyze information in changeable circumstances, make considerations and choices, and even complete tasks. A Chicago-based firm named Fusion Risk Management provides a software platform for operational risk management as it is developing a Copilot Resilience assistant that is powered by AI. The Copilot Resilience assistant uses AI to analyze volumes of operational risk information, present to executives the features that matter, and generate executive summaries and insights.

Generative AI is also used as a reputation management tool for organizations that are designed to listen to social talks and help marketing and social media managers track feedback provided online in real time. Reputation, an online review management software vendor from California, employs enterprise-level monitoring of companies. It works with events in real time, shows sentiment indices, identifies crises, and gives its social media managers feedback on how to respond to negative reviews.

One of the biggest potential applications of generative AI is in the checking and testing of ideas, providing a kind of virtual sounding board. By using AI as a virtual sounding board, organizations can test and validate their ideas and decisions, ensuring that they are making the best possible choices.

4.4.1 REAL-LIFE EXAMPLES OF IMPROVED DECISION-MAKING

Airbnb has been actively using AI (LuluQ, 2017) to improve its business operations and enhance the user experience. By using AI, the company used a customer relationship management (CRM) process, making it faster for both agents and customers.

Customer Service: AI technology will be used to enhance the capabilities of customer service. For instance, an agent responsible for navigating through 72 different user policies can be very time-consuming. AI will streamline Airbnb's CRM process, making it faster for the agent and the customer. The goal is for customers to have a seamless experience without realizing that AI is involved in this process.

Review Summaries: AI will help summarize reviews for listings with excessive reviews so guests can get familiar with the property quickly. This will make the listing process more efficient for hosts, allowing them to better understand the amenities and features of the property.

Predictive Analytics: It helps identify which guests are likely to throw house parties before they secure a booking. This helps Airbnb maintain a safe and enjoyable environment for all guests.

Listing Picture Verification: AI is used to verify if listing pictures submitted by hosts match up with satellite images. This helps to remove fake listings from the platform, ensuring that guests have an accurate representation of the property they are booking and are legitimate.

Smart Pricing: AI automatically adjusts rates for all the properties based on demand and the holiday season. This feature makes it easier for guests to find their desired property.

Automated Listing Descriptions: AI algorithms can automatically generate hundreds of unique and optimized descriptions based on information provided by the host such as photos and amenities. By analyzing patterns in language usage and user behavior, these algorithms can make compelling descriptions that highlight the most important features of each property while also ensuring consistency in terms of content and style.

4.5 AI AS A DRIVER OF ECONOMIC GROWTH, PRODUCTIVITY, AND OUTPUT

The latest developments in the field of AI ([Aghion et al., 2017](#)) hold a major promise for several industries, and these could result in the transformation of the economy through the optimization route.

AI is viewed as an exceptional product that is in sharp distinction with commonly known technologies because of its big-scale impact. Nevertheless, the question of whether AI can significantly stimulate economic growth by its interaction with labor markets and output is still valid. Moreover, AI can be considered either as a productivity enhancer or a substitute that will replace the jobs of some workers.

These risks are most evident in areas where there could be increased capital inflow through foreign direct investment, and therefore, an economic policy should consider issues such as income inequality and occupational displacement while promoting economic growth.

4.5.1 AGGREGATE IMPACT OF AI

Economists use this model to measure the potential growth of gross domestic product (GDP) as a function of inputs of capital (K), labor (L) ([Amundi, 2024](#)) as shown in [Figure 4.1](#), and the efficiency of the production process (productivity) [denoted by α], typically expressed as the *Cobb–Douglas production function*:

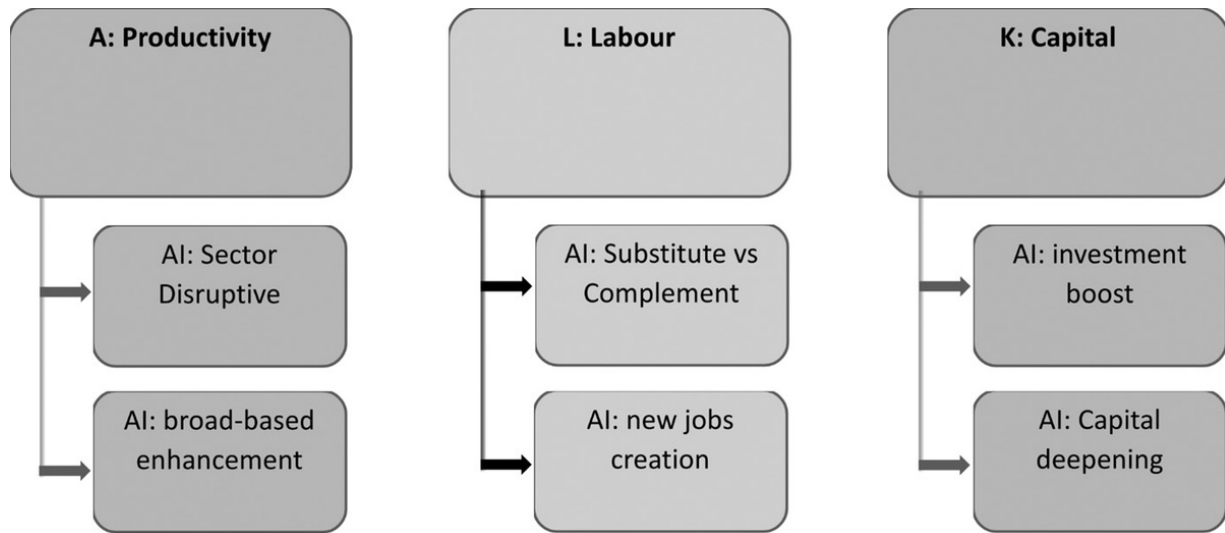


FIGURE 4.1 Projected economic impact of AI on global GDP growth by 2030. [📄](#)

$$Y = AL^{\alpha} K^{1-\alpha}$$

Higher inputs generate higher output, as does higher productivity. AI could impact growth not only through higher productivity but also through its direct impact on labor and capital inputs. Moreover, the impact of AI in the short term may be substantially different from its longer-term impact.

This AI and economic development growth excitement is present on the international stage.

More importantly, we also have to notice the upheavals that AI will bring to the labor market as it has an impact on jobs and workforces. The figures predict a whopping 60% of jobs might be impacted, either by AI automation taking some positions or partially taking over some others.

AI, at the same time, offers massive possibilities for increasing these returns and improving the gross national product. Nevertheless, automation at the same time could reinforce this trend and worsen the existing income disparity unless managed properly.

The government is the part that needs to smarten up with the policy ([Greenspan, 2023](#)). AI would shape and shake industries as well as the labor market and hence we will require programs that can help workers gain the necessary skills for new roles. We should see to it that if firmly

embedded into AI-specific economics, growth and prosperity will be widespread.

Indeed, AI represents a breakthrough and many great economic benefits lie ahead. However, should they fail to conduct a regulated competition, namely one that does not suck the marrow out of the employment sector, the possible benefits could cause more harm than good. Developing such a strategy to preserve the same would be the main problem for government officials in the upcoming years.

4.6 FUELING INNOVATION

AI is gaining ground at the moment, and it stands to bring technologies within reach of innovators who will use them to energize the economy. Increasing the power of artificial imitation driven by AI systems provides companies with an innovative edge to produce, and business models. It allows them to automate tedious jobs, simplify processes, and open up new operational speed and efficiency.

Although ways of finding jobs and doing work are changing fast in the AI era, a wide range of new opportunities is expected to come up in the future in every field. AI is thus comparable to fuel which propels innovation by specialists, and it is expected to drive an AI-driven economic increase in the future.

And not only are the profits of big companies included, but also sustainable approaches of small-scale businesses and marketing entrepreneurs are applied. AI is not just an opportunity; it is also a powerful accelerator of technological breakthroughs and business ventures; therefore, it is a source of constant economic growth and raising all boats ([Mariani et al., 2023](#)).

The productivity brought by AI may produce an upper-digit figure of the GDP in the world economy. People will use these kinds of services and devices more without affecting the budgets of their families. AI evokes a momentum of innovations, which, in turn, helps economic development and financial growth.

In particular, the presence of many obstacles is mainly not the case related to such serious breaches in the norm. However, if we approach closing the loop and developing a system based on the AI-fueled innovation ecosystem, we will reap the benefits. The future of humanity today is captured by the embrace of AI which leads directly to prosperity.

4.6.1 AI ACCELERATES RESEARCH AND DEVELOPMENT

AI technologies are capable of speeding up research and development (R&D) processes at very high speeds for industries from different ranges. AI systems are the major reason for shortening the product life cycle and improving the line of innovation and product development process which depends on innovation pipelines.

Cooperatively with this critical idea, the models of creative AI act like an accelerator in the process of rapid exploration and combination of new ideas. AI helpers can create solutions, images, and plots, which are based on what you intuitively tell them. This pronounced creative power evades cognitive limits leveling to a more sophisticated and deep investigation stage of solution space in contrast with the conventional human approach.

AI technologies also support faster iteration and refinement through automated tests and simulation at the scale of volume, which is another important feature of them. ML algorithms can replace tedious manual tasks and conduct a rapid and accurate validation of product variations in millions of real-world scenarios, user interactions, and situations on the edge of the filing system. Through this stepwise validation method, recognition of defects and optimization issues with utmost visibility is made possible, which would be impossible through manual prototype activity.

AI analytical capacities are also the game changers in providing explanations to complex systems and in overcoming the problem of high-degree non-linear relationships that often warp R&D initiatives' success ([Babina et al., 2024](#)). AI mirrors the interdependent cause-and-effect chains and therefore emerges as a great tool to capture both ripple effects and

essential causes of severe issues hidden beneath the surface. These will be the key contributors in the elimination of some bottlenecks, thereby helping to achieve cost savings to avoid costly oversights.

Moreover, NLP is the one that gives natural AI run across the whole R&D period. NLP models may interpret engineering specifications, generate codes for auto-casing, conduct detailed testing, and continually monitor the performance post-release. This AI-reengineered DevOps pipeline would also upgrade the ability to produce and minimize workflow deflection.

With this AI-enabled process, the time to market for research and products is drastically reduced. Organizations can seamlessly transition from the conception to the tested prototype and back with unprecedented rapidity and thorough quality assurance that is characteristic of additive manufacturing. This development of product launch in rapid time is the source of a considerable competitive strong point.

4.6.2 CREATION OF NEW SOLUTIONS

AI is on the threshold of being an indispensable tool for bolstering competitiveness in almost all sectors ([Lozovsky, 2024](#)). AI technologies in effect augment human creativity to the extent beyond what is expected for the given task by the use of data analysis and other computational methods to overcome our limitations.

AI has a unique ability in data mining from data volumes of different dimensions on a large scale. ML, especially algorithms, may unveil those subceilings and tacit relationships not seen by traditional analysis at all. This intelligent perspective visualizes further issues in the field that can now be approached with creative solutions or surprising solution vectors that can contribute to solving the existing issues.

The generative AI models can then speedily produce ideas that are formed by melting together previously gained insights in various combinations, exceeding a single individual's finite experiential radius.

After the likely solutions are identified, AI provides new finite-generated data for general modeling and stress tests exist for probable implementations of any policy choices on a range of scenarios.

Besides, AI automation replaces many other operations like those of ML which are standard in the development cycle. Natural language input can set requirements for framework synopses, create code frameworks, and automatically generate documentation. Vision machines deliver fast and reliable quality assurance, inspecting products for defects or testing appearance against the design. Besides, AI supports real-time tracking and gradual alteration of installed solutions based on intelligent information mining.

The various advantages of AI become game changers as they lead to the accumulative impact of technology ([Bhandari et al., 2022](#)). Organizations can do multiple things such as thump up the creative speed, reduce the validation period, and increase the thrust of the developer's power. This critical mix-up gives rise to three benefits that help pin down the competitive edge—shorter time to market, higher quality solutions, and agility to acquire and take over rivals with futuristic ideas.

The expanding capabilities of AI for generating innovative copy for specific challenges will drive the sectors of businesses with complex problems and intense rivalries. The enterprise that feeds on AI-spliced innovation systems may still come off as more advanced than those that adhere to the old ways. Now, the day of co-creation between humans and machines is going to dawn, where they can build up radical solutions.

4.7 EMERGING INDUSTRIES

AI is being applied across a wide range of industries to drive innovation, efficiencies, and new business opportunities ([Maheshwari, 2024](#)).

AI in healthcare provides computerized diagnostic imaging interpretation, patient-centered treatment, record analysis, fraud detection, work-as-AI robot assistants, and remote patient monitoring.

AI can help in retail and e-commerce by way of customized shopping, algorithmic pricing, inventory enhancement, the use of chatbots for customer service, and visual search excellence.

The food sector applies AI in all areas of supply chain management, quality checks/inspection, taste/recipe analysis, and individualized restaurant menus.

The banking and financial services sector implemented AI in various forms, ranging from fraud detection, credit scoring panels, chatbots, algorithmic trading and its robo-advisors to money laundering curtailment.

AI unifies these companies' operations through shipment tracking, traffic management, inventory control, and resource allocation, utilizing all the supply chain and logistics companies.

Travel companies have increasingly integrated AI into the travel planning processes, such as giving personalized travel plans, sending real-time recommendations, as well as making bookings and intelligent shopping tools easier to use.

AI is used in real estate for automatic estimating property worth, operation of portfolios as well, and creation of leads.

AI becomes a companion for media and entertainment enterprises in content recommendations for music/film, game design, storytelling, movie editing, etc. ([Espina-Romero et al., 2023](#)).

Manufacturers enjoy AI-driven benefits because of defect detection, administrative control, predictive maintenance, related quality, and process optimization.

The automotive industry utilizes AI to become a foundation of smart traffic, remarkably increased navigation, and personalized customer experiences.

AI helps to provide a personalized learning process, intelligent tutoring systems, auto-grading, learning analytics, and virtual assistants in the field of education.

AI technology allows the fashion industry to forecast trends, perform merchandising analysis, design and product development, make stylish advice, and let people see how clothes will look on them virtually.

For instance, the AI applications in this field assist with deal sourcing, due diligence, risk assessment, portfolio optimization, and predictive analytics. Law offices use AI for contract review, documentation of legal facts, e-discovery, risk and compliance assessment, chatbots, and automated documents.

IT companies may employ AI for services such as chatbot use or support automation, cybersecurity, data analytics, software testing, and IT asset management.

The hospitality sector benefits from AI capabilities like personalized recommendations, revenue management, sentiment analysis, fraud detection, smart room controls, language translation, and energy optimization.

As AI continues to evolve, its application across industries will accelerate the pace of innovation, introduce new business models, and drive economic growth—if implemented responsibly with robust strategies

4.7.1 HOW DOES THIS LEAD TO ECONOMIC EXPANSION

AI employment rate is rising in a variety of industries, which adds impetus to innovation and a spurt in productivity. This is a revolution that is driven by AI to have significant effects through different vectors and which power up the growth of the global economy.

AI helps individuals, organizations, and the wider community through multilateral processes of automation ([Probasco, 2023](#)), data optimization, and objectively depicts process and key performance indicators (KPIs) progress, which catalyze operational efficiency, innovative business models, and expansively accelerated technology based on the available data.

The main AI mode of stimulation of economic activity is manifested in such factors as its ability to promote productivity gains that are supposed to raise the output of sectors. AI plays a key role in manufacturing through predictive maintenance to detect instances of potential failures before the

production line can be shut down for maintenance resulting in less costly downtime and better operational efficiency.

Chains of supply remain AI-driven forecasting of demand and optimization of logistics reduce ineffective warehousing and boost distribution velocity. The savings ultimately translate into profit margins, which in turn further build up revenue. In other words, the cost savings and improved revenue streams contribute to the fueling of profitability.

In addition to its ability to complete difficult and error-inducing tasks at a faster rate ([Ramalingam et al., 2023](#)), AI manifests the impact of money on the economy. In fields such as healthcare AI, time-consuming administrative work is replaced by efforts that focus on tasks with more value by professionals by just removing the hassle. Also, by using AI to build virtual assistants and chatbots customer service is being hastened in the different sectors; this helps consumers become more content while reducing operational costs. With this, human labor is multiplied and you need not spend unnecessary money.

While from the quantitative perspective, AI creates new asset classes, from a qualitative perspective AI is responsible for the initiation of innovative products, services, and business models that result in pioneering new economic effectiveness. AI robotizes drug development in the pharmaceutical industry, where a huge amount of biological data is analyzed in a short time to find the main subject for the creation of new and effective therapies, and therefore, it increases profits.

Additionally, AI enables businesses to translate their databases into useful information, thus providing the sensitivity for data-driven decision-making that uncovers new markets. Categories that were previously unknown have now been formed around the principle of ensuring the development of AI solutions to the platforms and complementary technologies like autonomous vehicles and intelligent robotics.

Innovation powered by AI first takes a course through different industries and their interlinked supply chains, moving them to positive effects in totality. Furthermore, the core benefit of autonomous vehicles is the trend of expanding economic activity in connected industries, such as

automotive manufacturing, transportation services, infrastructure development, and even urban planning.

This is how the positive cycle in the context of microeconomic and macroeconomic scales can be created. From the macroeconomic perspective, AI is forecasted to be a major contributor to an impressive jump-starting of global economic growth. The International Data Corporation is forecasting that AI could add \$15 trillion to the world's GDP in 2030 (PricewaterhouseCoopers, n.d.). It will be a tremendous economic contribution to the world (over 14% economic growth).

Given that the share of prosperity from AI would be concentrated among AI-leader nations in the beginning, its economic benefits are expected to trickle down everywhere when AI adoption spreads completely. Yet, the AI-guided economic expansion is not free of challenges. The fact that AI is disruptive implies that it is quite likely that it will simply replace certain job roles while at the same time emerging new jobs in other high-demand fields.

This dynamic transformation in the labor landscape should include a well-trained workforce with the help of retraining programs and social policy additions that will help ease the change of jobs. Besides that, there is a need to reconcile the rising fear of widening income inequality which could emerge as a result of AI automation that will be favorable to the capitalized group of people compared to the labor.

To achieve this, it is imperative to stimulate AI proliferation with sustainability in mind, while implementing policies that help deal with workforce transitions and the redistribution of wealth. With the help of AI, a new era of growth can be unfolded, bringing with it productivity, innovation, and unexplored economic opportunities across the globe.

There is great potential in this area with careful planning. However, to make the best of this potential and to minimize the risks, public and private entities should join hands strategically. Adoptive policies that promote AI development alongside job training, inclusive growth, cybersecurity, and even AI ethics are essential. Employment of the above measures will ensure

that AI's power to redefine and sustain the modern globally centralized economy is accurately leveraged.

4.7.2 CASE STUDY 1: AMAZON

Amazon is one of the biggest key players for something in stores, electronics, and technology spheres. It started as an online bookstore, but it has now evolved into a mega store carrying all kinds of products available nowadays. Amazon's influence has diversified so that it has hands in everything ranging from clothing and electronics to groceries and even cloud computing services.

In the last few years, Amazon has been fueling its business growth and expansion via widely used advanced technologies such as AI. AI plays a decisive role for organizations like Amazon being a leader in the quest of staying competitive as well as exceeding customers' expectations.

This case study aims to discover how precisely the company has deployed AI and other technologies to enhance its business and make those profits bigger. Among all, we will focus on how Amazon has used AI in various facets of its business, from offering personalized product suggestions to streamlining its complicated supply chain.

4.8 RECOMMENDATION ENGINE

The recommendation engine has become one of the top instruments that Amazon has used to incorporate AI. This AI-driven platform leverages customer data by analyzing browsing activities and purchase patterns and subsequently recommends products that are based on individual preferences.

4.9 SUPPLY CHAIN OPTIMIZATION

Amazon's supply chain is a massive and intricate system comprising an array of warehouses ([Wadhwa et al., 2017](#)), distribution centers, and

distribution channels. For the successful management of this complex system, Amazon has found the power of AI and ML algorithms a good tool.

AI platforms for Amazon improve utilization of inventory, forecasting demand of the products and logistics in a cost-effective way which also boosts delivery times.

4.9.1 INVENTORY MANAGEMENT

Monitoring huge amounts of goods from quite several warehouses is at the same time very difficult. These systems mine tremendous amounts of customer data, market trends, and product sales revenues to forecast which products will be in demand and when. Then with the knowledge of demand, Amazon makes sure that there is neither too much nor too little of each item in the warehouses. This allows customers to get their wanted items without having them out of stock or Amazon facing losses because of having too much inventory.

4.9.2 WAREHOUSE OPERATIONS

Through advanced technology, Amazon can get a hold of its inventory. For product selection, packing, and sorting, the firm utilizes robotic systems, automated guided vehicles (AGVs), and other types of conveyors such as Snake, Overhead, and Accumulation conveyors ([Wu & Baeuml, 2023](#)). Machines with AI algorithms assigned to a warehouse management system (WMS) determine the most appropriate place for goods, reduce picking errors, and augment overall efficiency. Robotics and AI algorithms used with WMS enhance product placement, lower picking mistakes, and boost operational efficiency. Amazon can handle a large number of orders quickly and precisely because of these technologies.

4.9.3 LAST MILE DELIVERY

Amazon makes use of technologically advanced technologies to guarantee effective last-mile delivery. The business plans and optimizes delivery routes for its vast network of delivery drivers using GPS tracking and route optimization algorithms. Amazon has also made investments in drone technology in preparation for future driverless deliveries. Amazon can improve package tracking capabilities, shorten delivery times, and provide consumers with real-time order updates by utilizing technology.

4.9.4 CLOUD COMPUTING AND AI SERVICES (AWS)

Using Amazon Web Services (AWS) (Statista, 2024), Amazon generated revenue which then supported other verticals which are operated by Amazon. AWS is a service that is offered to other companies from their ML-based and cloud requirements.

This way all the AI-based companies are using these systems to automate tasks and administrate efficiently.

4.10 INFERENCES

This case study showed us how AI helps a business in several verticals, thus expanding the economy and creating value for the customers. Through the smooth implementation of AI in different functions such as personalized product recommendations, supply chain improvement, cloud computing, and autonomous delivery services, Amazon has been a trendsetter in bringing the advantages of AI in terms of customer experience, process improvement, and opening up new revenue sources.

It is Amazon's AI-driven transformation that amazes most with its wealth of scale and scope. Since adopting AI for intelligent inventory management to last-mile delivery sites through autonomous drones, Amazon has seen measurable, financial advancements that have contributed to its economic growth and Amazon's dominance in the e-commerce and technology sectors.

Furthermore, it shows that data-driven decision-making is a right and powerful thing to do that gives an impressive insight when humans use ML and advanced analytics. Amazon has used the data obtained by constantly observing a large amount of customer data, market trends, and operational metrics for informed strategic decisions, meeting changing consumer demands, and adapting the services following the trends, thus ensuring its dominance in the industry.

Forging ahead and with tech innovations, Amazon strives to invest in technologies of the future and pioneer novel applications of AI and its impact on economies globally. Thus, it paints a clear picture of the enormous power contained in technology to transform an industry as a whole and to drive economic growth on all continents. On the one side, the side of technology application, AI raises questions of what concerns the ethics and the wider social impact. Therefore, AI raises the issues of privacy, job displacement, and data usage accountability.

4.10.1 CASE STUDY 2: WALMART

Walmart stands for a customer household, so the story of this brand does not need any introduction. It is not only the biggest retailer in the world but also its opportunities extend into groceries, electronics, and clothes as well as many other things.

Recently, Walmart has been channeling into modern tech such as AI to push the moment in its advancement. AI has become the primary material for the likes of Walmart to run their businesses with great speed, make customer service better, and above others stay at the top.

These days AI is sowing the seeds of the retail revolution. It has allowed companies access to big data on customers, merchandise trends, or sales behaviors to identify market demand and therefore to take advantage of opportunistic business strategies.

4.10.2 INTERNAL INVENTORY MANAGEMENT

Managing millions of items in thousands of stores and distribution centers is not a simple task ([Alexander, 2024](#)), but it is Walmart that realizes its secrets in collaboration with those intelligent computer programs. These virtual brains, which process humongous information of customer data, market trends, and past sales, forecast the future best-selling products when the sales are at their peak.

These data will help to make sure that the stores and warehouses will always have the right quantity of each product so that Walmart does not need to buy more or less. This helps the buyers to get the items they want without any frustrating stockouts, and simultaneously, Walmart will not gain extra stock at the increased cost.

The key to Walmart's inventory system is the engine with very complex data analytics which enables monitoring the levels of product availability and quantity in stock in a real-time mode. This is made possible through the seamless integration with Walmart's point-of-sale systems, which do so automatically every time a customer makes a purchase and turns this input into an inventory adjustment. Such a kind of prompt response in stock-level restocking contributes to the fast identification of possible stockouts and replenishment initiation.

4.10.3 VENDOR INVENTORY MANAGEMENT

Walmart's inventory system adopts a vendor-managed inventory (VMI) strategy, which allows the supplier to be directly involved in the management of stock levels of their products within Walmart's distribution system.

Using secure Walmart inventory data access, suppliers can track in real-time stock amounts remaining and make informed decisions as to when and how much product should be shipped to Walmart's fulfillment centers and retail stores. The combined effect of this coordinated method helps the supply chain to provide timely delivery and means—ends of the products to reach their destinations swiftly.

The VMI model is the best choice for both Walmart and its suppliers as it helps the retailer optimize logistics, receive real-time demand information, and increase sales. This way, Walmart can utilize the co-responsibility of suppliers for the process of monitoring and replenishment of their product lines. Such a model would eliminate the need for Walmart to put in place intricate mechanisms of regulating inventories of all the individual suppliers while the cost savings and operation efficiency are increased.

For the suppliers, the VMI model gives the manufacturers better visibility and more control of the availability of their products within the Walmart logistics system. This is achieved by having direct access to real-time inventory data which allows the suppliers to streamline their production and distribution processes, thereby minimizing the risk of stockouts or excessive inventory buildups, both of which can result in financial losses and can make customers unhappy.

4.10.4 DELIVERY STRATEGY

When your order is in its last phase of processing and ready for shipping, Walmart's technology blades go in again to ensure the smooth running of the process. AI effectively plans the shortest and fastest transport routes while considering traffic patterns, detours, and even weather conditions.

GPS tracking allows you to track the package journey in real time because you will always know the exact moment the item will be delivered. Furthermore, Walmart plans to employ emerging delivery methods such as drones and autonomous vehicles to achieve orders that do not require any drivers and where the time taken is less compared to human driver ([Choo, 2024](#); [McKeehan, 2023](#)).

These new technologies being used in conjunction with each other enable consumers to have a flawless shopping experience from the time they add items to their shopping bags to when the package arrives at their front door. It is best practice of fighting logistics problems using technology by boosting efficient operations and handling customers conveniently. The

adoption of AI and other advanced technologies is in the corporate strategy of Walmart and it becomes evident that the future of the retail sector might be through these innovations. These technologies will enable increased operational efficiencies, cost reduction, and growth of the economy in the long run.

By incorporating AI and automation into the different areas of its supply chain as inventory classification and demand forecasting, warehouse functions, and last-mile delivery, Walmart has shown clearly it strives to be at the forefront kind of retailers and face the ever-changing area.

AI is also a notable feature of Walmart's transformation and has resulted in several quantifiable improvements that include, among others, faster inventory turns, lower labor costs, and delivery times. These economic gains not only collectivize Walmart's assets but also escalate consumer contentment, increasing loyalty and enlargement in the future.

Apart from that, Walmart's strategic partnership with suppliers via its VMI model also reflects how AI coupled with data-driven technologies can create synergies. The transparency that Walmart gives suppliers with information about real-time inventory levels and demand patterns has made operations in the supply chain smooth, shortening delays and boosting a smooth flow of products to its large number of retail outlets.

While Walmart continues to invest in breakthrough technologies including autonomous delivery vehicles and robotics, it is in the best position to spearhead the revolution of shopping as the best convenience model, with due consideration to efficiency and sustainability needs.

On the other hand, the ethical issues and impacts on society of the deep-rooted implementation of automation and AI, in addition to the job loss and responsible use of data, must be considered carefully.

4.10.5 CHALLENGES AND CONSIDERATIONS

AI technologies keep rapidly entering the market; thus, the dual problem of job loss and retraining becomes more acute. AI can replace traditional roles

as it is responsible for automation and some tasks are reached by a human being, particularly the tasks that are monotonous and predictable.

With the transition, the key is reskilling and upskilling the workers for their adaptation to the new needs of both the market and their job market. As AI systems grow and start getting used in various industrial fields, the necessity of workforce training takes prominence (Oluwaseyi & Cena, 2024).

Society and individuals alike have to go through continuous learning and skill development and utilize the competitive environment to adapt to technological innovations.

Through education and training, people can gain the required skills for coping with a more dynamic job market and capturing the advantages offered by AI-led automation.

4.10.6 BIAS AND FAIRNESS IN AI

AI algorithm bias due to the data used to train those systems is a major topic that leads to problems of fairness and ethical issues. Along with the increasing prominence of AI technologies in the decision-making processes, it is inevitable to emphasize the responsible development and usage of these systems to help avoid those biases that can lead to inequality.

Transparency, accountability, and justice should be vital components of the design and implementation of AI technology to ensure ethical issues and inclusiveness are considered when working with people ([Ferrara, 2023](#)). Through developing an environment that suspends biases underlying ethical considerations of AI, the stakeholders can contribute to a world where AI supports human abilities while preserving principles of fairness, equity, and social responsibility.

Having a grip on these issues is essential for fully utilizing AI technology while at the same time pre-empting the unintended consequences and fostering an equitable and inclusive society.

4.11 CONCLUSION

In this chapter, we have explored the transformative impact of AI on industries, economies, and the workforce. AI has proven to be a powerful tool for enhancing efficiency, driving innovation, and promoting economic growth by automating repetitive tasks, improving decision-making, and accelerating R&D. Through case studies of Amazon and Walmart, we have seen how these companies have harnessed AI to optimize their supply chains, improve customer service, and streamline operations, ultimately contributing to their economic success. However, while the potential of AI to drive growth is immense, its adoption is not without challenges. Key issues such as job displacement, the widening of income inequality, and ethical concerns related to biased AI algorithms need to be addressed. As AI replaces certain types of jobs, it creates an urgent need for workforce reskilling and upskilling to prepare individuals for new roles in an AI-powered economy. Similarly, the development of ethical frameworks is crucial to ensure that AI systems are fair, transparent, and inclusive, preventing the perpetuation of biases that could exacerbate social inequalities. Looking ahead, AI has the potential to not only increase productivity and economic output but also create new industries and business models. However, to fully realize these benefits, it is essential for policymakers, businesses, and educational institutions to collaborate. They must develop strategies that promote responsible AI adoption while minimizing its negative societal impacts. By investing in ethical AI development and reskilling initiatives, we can ensure that the economic gains from AI are distributed equitably and sustainably. In conclusion, AI is reshaping industries and economies on a global scale, offering both immense opportunities and significant challenges. With thoughtful governance and strategic planning, AI can be a force for positive change, driving economic growth while fostering innovation and creating new opportunities for the workforce of the future.

REFERENCES

- Aghion, P., Jones, B. F., & Jones, C. I. (2017). *Artificial Intelligence and Economic Growth*. Stanford. <https://web.stanford.edu/~chadj/AI.pdf>
- Al Rubaie, H. D., Al Shahri, A. A., & Al Qamashoui, A. A. (2024). *Strategies for Business Transformation- Accelerators for Sustainable Growth*. Zenodo. <https://doi.org/10.5281/zenodo.10432143>
- Alexander, J. (2024, April 17). *Walmart Inventory Management: A Comprehensive Guide*. SellerApp Blog. www.sellerapp.com/blog/walmart-inventory-management/
- Amundi. (2024, January 25). *Will Artificial Intelligence Increase Economic Growth?* Amundi Research Center. <https://research-center.amundi.com/article/will-artificial-intelligence-increase-economic-growth#:~:text=Higher%20inputs%20generate%20higher%20output,from%20its%20longer-term%20impact>
- Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151, 103745. <https://doi.org/10.1016/j.jfineco.2023.103745>
- Bhandari, M., Chang, A., Devenyns, T., Devereson, A., Loche, A., & Van Der Veken, L. (2022, November 16). *How AI Can Accelerate R&D for Cell and Gene Therapies*. McKinsey & Company. www.mckinsey.com/industries/life-sciences/our-insights/how-ai-can-accelerate-r-and-d-for-cell-and-gene-therapies
- Choo, J. (2024, January 5). *Walmart Business Strategy: A Comprehensive Analysis*. The Strategy Journey. <https://strategyjourney.com/walmart-business-strategy-a-comprehensive-analysis/>
- Espina-Romero, L. C., Sánchez, J. G. N., Hurtado, H. G., Conde, H. O. D., Castro, Y. S., Cajo, L. A., & Corredoira, J. R. (2023). Which industrial sectors are affected by artificial intelligence? A bibliometric analysis of trends and perspectives. *Sustainability*, 15(16), 12176. <https://doi.org/10.3390/su151612176>
- Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Science*, 6(1), 3. <https://doi.org/10.3390/sci6010003>

- Greenspan, R. (2023, July 22). *Walmart's Inventory Management*. Panmore Institute. <https://panmore.com/walmart-inventory-management>
- Lozovsky, D. (2024, February 17). How AI is fueling innovation and disruption across industries. *Medium*. <https://medium.com/@daniel.lozovsky/how-ai-is-fueling-innovation-and-disruption-across-industries-3bf64a207037>
- LuluQ. (2017, April 7). *Airbnb: Data Informed Decision Making*. Digital Innovation and Transformation. <https://d3.harvard.edu/platform-digit/submission/airbnb-data-informed-decision-making/>
- Maheshwari, S. (2024). *Artificial Intelligence in Industries*. ResearchGate. www.researchgate.net/publication/378297887_Artificial_Intelligence_in_Industries
- Mariani, M. M., Machado, I., Magrelli, V., & Dwivedi, Y. K. (2023). Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions. *Technovation*, 122, 102623. <https://doi.org/10.1016/j.technovation.2022.102623>
- McKeehan, J. (2023, November 21). *Walmart Introduces Parcel Stations to Speed Up Delivery*. <https://corporate.walmart.com/news/2023/11/21/walmart-introduces-parcel-stations-to-speed-up-delivery>
- Moch, E. (2023). Industrial revolution of artificial intelligence. *International Journal of Public Administration, Management and Economic Development*, 8(2). <https://doi.org/10.60026/ijpamed.v8i2.104>
- Oluwaseyi, J., & Cena, J. (2024b). *Analyzing the Impact of Artificial Intelligence on Job Displacement and Income Inequality*. ResearchGate. www.researchgate.net/publication/377979619_Title_Analyzing_the_Impact_of_Artificial_Intelligence_on_Job_Displacement_and_Income_Inequality
- PricewaterhouseCoopers. (n.d.). *PwC's Global Artificial Intelligence Study: Sizing the Prize*. PwC. www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

- Probasco, J. (2023, December 8). Generative AI and its economic impact: What you need to know. *Investopedia*. www.investopedia.com/economic-impact-of-generative-ai-7976252/
- Purdy, M. (2023, October 26). How AI can help leaders make better decisions under pressure. *Harvard Business Review*. <https://hbr.org/2023/10/how-ai-can-help-leaders-make-better-decisions-under-pressure/>
- Ramalingam, A., Karunamurthy, A., Victoire, T. A., & Pavithra, B. (2023). Impact of artificial intelligence on healthcare: A review of current applications and future possibilities. *Quing: International Journal of Innovative Research in Science and Engineering*, 2(2), 37–49. <https://doi.org/10.54368/qijirse.2.2.0005/>
- Statista. (2024, February 9). *Amazon Web Services Revenue 2023*. www.statista.com/statistics/233725/development-of-amazon-web-services-revenue/
- Sun, S., Cao, Z., Han, Z., & Zhao, J. (2020). A survey of optimization methods from a machine learning perspective. *IEEE Transactions on Cybernetics*, 50(8), 3668–3681. <https://doi.org/10.1109/tcyb.2019.2950779/>
- Unilever. (2024, February 1). *How AI and Digital Help Us Innovate Faster and Smarter*. www.unilever.com/news/news-search/2023/how-ai-and-digital-help-us-innovate-faster-and-smarter/
- Wadhwa, B., Vashisht, A., & Kaur, D. (2017). Business model of Amazon India—a case study. *International Journal of Advanced Research*, 5(8), 1426–1433. <https://doi.org/10.21474/ijar01/5200/>
- Walas Mateo, F., & Redchuk, A. (2021). *Artificial Intelligence and Machine Learning as a Process Optimization Driver under Industry 4.0 Framework, the Role of the People in the Process*. Global Conference on Engineering Research. www.researchgate.net/publication/352544108_Artificial_Intelligence_and_Machine_Learning_as_a_Process_Optimization_driver_under_Industry_40_framework_the_role_of_the_people_in_the_process/

Wu, C., & Baeuml, M. (2023). *AWS Last Mile Solution for Faster Delivery, Lower Costs, and a Better Customer Experience*. Amazon Web Services.
<https://aws.amazon.com/blogs/supply-chain/aws-last-mile-solution-for-faster-delivery-lower-costs-and-a-better-customer-experience/>

5 AI, Data Privacy, and the Personal Data Protection Act

Implications for Human Rights

Soorya Balendra

DOI: [10.1201/9781003668657-5](https://doi.org/10.1201/9781003668657-5)

5.1 INTRODUCTION

The significance of data protection is growing in Sri Lanka due to the rapid expansion of digitization and digital connectivity. According to the World Bank's compilation of development indicators, the estimated mobile cellular subscriptions (per 100 individuals) in Sri Lanka in 2022 were 143. This data was sourced from officially recognized channels. At the onset of 2023, the internet penetration rate in Sri Lanka had reached 66.7%, encompassing a user base of over 14.58 million individuals within the nation. In January 2023, the social media usage in Sri Lanka was seen to be 7.20 million individuals, accounting for around 32.9% of the total population ([Samahon & Liyanagamage, 2021](#)). The ongoing trend of digitalization is leading to an increasing volume of data, thus emphasizing the importance of enacting data protection and privacy legislation.

Nevertheless, Sri Lanka has been among the 21% of countries globally that do not possess comprehensive legislation pertaining to data privacy, as identified by the United Nations Conference on Trade and Development (UNCTAD) among the 107 nations it recognizes in 2019 (The Secretariat of

the United Nations Conference on Trade and Development, 2019). On a global scale, a significant majority of consumers, amounting to 63%, hold the belief that corporations lack transparency when it comes to disclosing their utilization of personal data. Consequently, this lack of transparency has resulted in approximately 48% of consumers ceasing their purchases from these companies ([Komnenic, 2022](#)).

Moreover, privacy concerns have caused 33% of users to leave organizations. This phenomenon has affected social media sites, internet service providers, retailers, credit card companies, and financial institutions. Hackers target internet-connected PCs every 39 seconds ([Mutambik et al., 2022](#)). The preservation of privacy in the realm of cyberspace is a matter of worldwide significance during this era of digital advancements. Technological advances and internet connectivity have transformed daily living. Data breaches, cyber threats, and surveillance fears have increased due to these advancements. Sri Lanka's acknowledgement of privacy as an individual's right to a private life is incomplete. Moreover, the rapid advancement of technology has surpassed the existing legislative safeguards ([Samarakoon & Fonseka, 2020](#)).

Sri Lanka, classified as a developing nation, has had a notable increase in internet usage and e-commerce, resulting in a revenue of \$400 million by the year 2020 ([Abeysekara & Ranasinghe, 2022](#)). It endeavors to assess the extent to which the existing act protects privacy rights in Sri Lanka and explores the emerging privacy challenges posed by advancing technologies in the country. In addition, the study aims to investigate whether the absence of constitutional recognition poses any significant harm. Several key questions are addressed in this study: To what extent does the existing act safeguard privacy rights in Sri Lanka? What are the emerging privacy challenges posed by advancing technologies in the country? Does the lack of constitutional recognition in Sri Lanka's constitution have any noteworthy consequences? These inquiries form the crux of the investigation in the study. By employing a comparative international study, particularly with a focus on the General Data Protection Regulation (GDPR) of the European Union (EU), this chapter aims to uncover and

evaluate Sri Lanka's current stance regarding the protection of privacy rights. The preservation of data is of utmost importance to safeguard information, foster confidence, and mitigate the risks of data compromise, unauthorized access, and data integrity issues. The maintenance of a delicate equilibrium between transparency and privacy holds utmost significance within a democratic framework. To safeguard the rights of individuals and uphold responsibility, it is imperative for legislation to strike a delicate balance ([Samarakoon & Fonseka, 2020](#)).

This chapter undertakes a comparative analysis of the legal framework in Sri Lanka and the United Kingdom (UK), focusing on their respective data protection and privacy legislation. The objective is to identify any deficiencies in the Sri Lankan legal environment and propose potential enhancements. The study herein assesses the efficacy of the recently enacted Personal Data Protection Act No. 09 of 2022 (PDPA) in Sri Lanka, while also providing recommendations for potential enhancements.

5.2 LITERATURE REVIEW

5.2.1 WHAT'S A PRIVACY RIGHT: A LITERATURE REVIEW?

Privacy, as per the Oxford Dictionary's definition, encapsulates the "entitlement to shield oneself from unwarranted intrusion and the capacity to safeguard specific facets of personal life from public scrutiny" ([Law, 2015](#)). MacMenemy, in his comprehensive exploration of the concept, underscores privacy as a foundational element of individual autonomy. This pertains especially to private realms, where individuals can reasonably expect to remain unobserved ([McMenemy, 2016](#)).

In the year 1985, Spiros Simitis, a famous privacy scholar in Germany, was among the first to recognize the potential risks posed by the automation of data for privacy, individual liberties, and democratic processes ([Morozovarchive, 2013](#)). Simitis proposed an intriguing perspective; he asserted that privacy should not be viewed as the ultimate objective in itself,

but rather as a pivotal instrument in nurturing a self-reflective democracy. Within such a democracy, citizens do not unwittingly become mere suppliers of information to all-seeing, all-optimizing technocrats.

It is a well-established fact that data can be employed in undisclosed ways, necessitating proactive measures to safeguard data privacy. According to the present studies, data privacy constitutes an intrinsic right of individuals. This right empowers individuals to maintain the confidentiality of their data and to exercise control over its utilization, particularly concerning governmental entities and corporations engaged in the acquisition and dissemination of personal information.

Radhi and Irina emphasize the fundamental nature of the right to privacy as it underpins various essential rights, including freedom of conscience, personal security, and the right to abstain from self-incrimination ([Romansky & Noninska, 2020](#)). They extend the definition of privacy in the digital age, labeling it as “e-privacy.” They advocate for recognizing the right to privacy as an independent and legally protected right. In their view, the “right to privacy” encompasses safeguarding a personal sphere encompassing elements such as one’s body, residence, possessions, thoughts, emotions, secrets, and identity.

Furthermore, privacy as a critical human right has been acknowledged in diverse regional and international documents. Privacy is intrinsically linked to the duty and necessity of preserving individuals’ private lives, ensuring the confidentiality of personal information throughout its collection, storage, dissemination, and transfer. These requirements give rise to two significant rights: the “right to privacy” and the “right to data protection” ([Romansky & Noninska, 2020](#)).

The Geneva Academy’s meeting report identifies the right to privacy as a liberty right, serving to protect an individual’s autonomy regarding what information to share and with whom. The report also emphasizes that while the right to privacy is not absolute, any limitations imposed on it must meet specific criteria: they must be prescribed by law, necessary to achieve a legitimate objective, and proportionate in their application (Geneva Academy of International Humanitarian Law, 2022).

An examination of the definitions put forth by prominent legal academics reveals diverse perspectives and insights into the concept of privacy rights. Despite their distinct characteristics, these definitions do share some notable commonalities. First and foremost, all of these definitions and elaborations underscore the fundamental nature of privacy as a basic human right. They collectively acknowledge the imperative of safeguarding personal information, particularly in the context of our rapidly evolving digital landscape. Moreover, a central theme in these definitions is the preservation of autonomy and the assertion of control over one's personal data.

However, distinctions can be discerned among these definitions. The Oxford Dictionary offers a broad, encompassing definition of privacy, while MacMenemy delves deeper into the idea of privacy within private spheres and its integral connection to individual autonomy. Spiros Simitis introduces the concept of privacy as a tool to uphold self-critical democracies and protect against data-related risks. Radhi and Irina extend the notion of privacy to encompass e-privacy and advocate for the necessity of legal protection. Lastly, the Geneva Academy's Meeting Report places a particular emphasis on privacy as a liberty right and elucidates the legal boundaries that must be adhered to when limiting this right.

In conclusion, it is evident that these definitions, while shedding light on the fundamental importance of privacy rights and the need to safeguard personal information, do not explicitly delve into the vulnerabilities that arise in the face of emerging technologies such as artificial intelligence (AI), machine learning, the Internet of Things (IoT), blockchain, cryptocurrencies, facial recognition technology, biometric data collection, and data mining. Furthermore, they do not elaborate on the profound impact of digital transformation on the right to privacy, the effectiveness of enforcement mechanisms, or the adherence to international privacy standards. These gaps in the existing literature highlight the pressing need for further research in these areas, as they are integral to understanding and addressing the complexities of privacy rights in our ever-evolving digital world.

5.2.2 SRI LANKA'S LEGISLATIVE FRAMEWORKS ON CYBER PRIVACY

Privacy values are the bedrock principles of a democratic society, safeguarding individual liberties like freedom of speech, expression, association, and personal autonomy. The absence of privacy would make individuals feel stifled, underscoring its profound importance in our lives. Acknowledging privacy as a legal right holds critical significance, particularly in its role as a shield against unwarranted public exposure.

In his comprehensive analysis of privacy-related computer crimes, Marsoof presents various perspectives supporting privacy as a fundamental right, implicitly embedded within Chapter III of Sri Lanka's Constitution ([Masroof, 2007](#)). Marsoof delineates three key dimensions of privacy: Privacy as Expression, Privacy as Movement, and Privacy as an Integral Component of Quality of Life, shedding light on nuanced interpretations. According to him, it is unlikely that the framers of the Constitution deliberately omitted privacy as a fundamental right. Despite the Computer Crimes Act of 2007 addressing certain privacy-related offenses, it falls short in its coverage, necessitating proposed reforms.

Marsoof also scrutinizes the current legal landscape in Sri Lanka concerning the right to privacy, particularly in the context of information communication technology (ICT) advancements. He highlights the potential infringement on privacy through modern technology. It is worth noting that while Sri Lanka's Constitution lacks explicit provisions safeguarding the right to privacy, Marsoof's paper draws a contrast with India's Constitution, which explicitly includes Article 21, addressing privacy rights.¹

Until recently, Sri Lanka lacked a dedicated law governing data protection and privacy, resulting in specific aspects related to data protection and privacy being addressed in various industry-specific legislations like the Intellectual Property Act No. 36 of 2003 and the Computer Crimes Act of 2007 ([Abeysekara & Ranasinghe, 2022](#)). A significant milestone in Sri Lanka's legislative landscape occurred with the

enactment of the PDPA on March 18, 2022, marking Sri Lanka as the first South Asian nation to embrace comprehensive data protection legislation.

Within Sri Lanka, the PDPA's jurisdiction extends territorially to the processing of personal data by entities operating within the country's borders. Furthermore, the act assumes extraterritorial applicability when entities situated outside Sri Lanka provide goods or services to individuals within the country or monitor their online behavior.

While the above literature review provides valuable insights into the evolving landscape of privacy rights in Sri Lanka, it also uncovers several critical areas that remain underexplored. While the scholars have highlighted the PDPA as a groundbreaking development, they have barely scratched the surface. We need a more in-depth look into the PDPA's fundamental details, its provisions, mainly if it is practicable in real life, and whether it truly safeguards individuals' data privacy rights. We also need to see if it stands up to international data protection standards. Thus, this research aims to dive into checking the effectiveness of PDPA while comparing it with two foreign jurisdictions: the UK and Canada.

5.2.3 CHALLENGES IN ENFORCING SRI LANKA'S PRIVACY LAWS

The concept of "Big Brother Is Watching" has long been embedded in the public's collective consciousness, underscoring the widespread presence of surveillance and infringements on privacy. In today's rapidly evolving technological landscape, there is a growing tension between harnessing technology for societal benefits and the potential emergence of new threats to privacy. Privacy, a complex and multifaceted idea, has assumed paramount importance as technological advancements, particularly the proliferation of social media and networking services (SMNS), challenge its very essence. According to Zwick and Dholakia, factors such as social, cultural, political influences, and historical contexts significantly mold privacy patterns, resulting in diverse interpretations across different societies (Zwick & Dholakia, 1999). Privacy is often weighed against other

values, further complicating its implementation. Given the escalating importance of personal data, the development of an effective data protection strategy becomes imperative to mitigate the risks of data loss, theft, and corruption ([Abeysekara & Ranasinghe, 2022](#)).

Although technology transcends geographical borders, it falls upon the government's shoulders to establish robust privacy and data protection laws and policies. It is important to note that mere legislative and judicial acknowledgment of privacy and data protection may render them ineffective ([Koops, 2014](#)). Scholars advocate for immediate measures to enhance information technology standards, particularly in relation to data protection and privacy, in conjunction with the enactment of regulations.

Drawing from the example of the GDPR in the EU, Sri Lanka can glean insights from other countries' experiences in formulating effective data protection regulations. GDPR was introduced to empower individuals in the EU and EEA with control over their data while harmonizing data protection regulations across the EU. Despite Sri Lanka's proactive stance in ratifying international regulations related to ICT law, it faces challenges in effectively implementing and adhering to these regulations. This deficiency poses a significant threat to Sri Lankan businesses engaging with European consumers, as GDPR applies to all European citizens. The absence of comprehensive data privacy laws might even serve as a non-tariff trade barrier, impeding Sri Lanka's benefits from its recently acquired EU's GSP+ status and negatively impacting foreign investment inflows ([Burgess, 2018](#)).

The absence of adequate data protection laws and regulatory authorities poses a tangible threat to the e-Sri Lanka program, aimed at digitizing all government departments and enhancing electronic services. The article brings out Article 14(1) of the Sri Lankan Constitution and two cases, including *Hewamanna v. Attorney General* (1999), relating to the individual privacy where the Supreme Court has highlighted the importance of the individual's right to privacy. The article was written in 2008 and it has done a major analysis of the Sri Lankan legislation till then.

According to the Talking Economics journal, as of 2019, Sri Lanka falls within the 21% of countries lacking well-implemented data protection laws, according to the UNCTAD. The article also discusses emerging technologies such as cloud computing, IoT, 5G networks, and big data analytics in the Fourth Industrial Revolution (4IR) and elucidates the challenges they pose to data protection and privacy. This underscores the urgency for Sri Lanka to establish a comprehensive data protection framework to effectively safeguard individual data and privacy rights.

Furthermore, Sri Lanka grapples with the challenge of striking a balance between protecting privacy rights and ensuring the right to information. The absence of explicit recognition of the right to personal privacy as a fundamental right in the 1978 Constitution exacerbates this situation. Although recent cases in Sri Lanka's Supreme Court have highlighted the importance of an individual's right to privacy, the common law does not fully acknowledge the right to the protection of personal information. This inadequate legal framework contributes to the overall challenges Sri Lanka faces in effectively enforcing privacy laws.

Despite the acknowledged importance of data privacy and the growing recognition of its significance in Sri Lanka, there is a noticeable research gap in understanding the practical obstacles and stakeholder perspectives that hinder the effective enforcement of privacy laws, particularly concerning the protection of personal information in the digital age. Although existing literature highlights legal deficiencies and the absence of comprehensive data protection regulations, there is limited exploration into the effectiveness and practical challenges faced by individuals, businesses, and governmental agencies in adhering to privacy standards and ensuring data protection. A thorough investigation by this study into these practical barriers and stakeholder perceptions by fulfilling this research gap could provide valuable insights into the implementation and enforcement of privacy laws in Sri Lanka and inform policy recommendations for strengthening data protection measures.

5.2.4 THE PROTECTION OF ONLINE PRIVACY RIGHTS IN THE UK

In today's modern world, safeguarding our personal information has become more critical than ever before. Countries like the UK have implemented laws to protect the privacy of their citizens. Even in a developing country like Sri Lanka, there are valuable lessons to be learned from these comprehensive legal frameworks. The rapid advancement of technology and the changing needs of society have led to the creation of global rules for safeguarding personal data. These rules may differ somewhat from one place to another; however, they all draw inspiration from organizations like the Organization for Economic Co-operation and Development (OECD), the Council of Europe, and the EU, especially the GDPR. These rules are meant to establish guidelines for how personal information should be handled, highlighting the growing consensus worldwide about the importance of data protection ([Gyulavári & Menegatti, 2022](#)). Privacy posed apprehensions not only for the government but also for the judiciary. In 1961, there was an endeavor to establish a privacy entitlement in the UK. However, the Lord Chancellor of that era expressed reservations, fearing that judicial intervention in privacy affairs might lead to perceptions of judges as content controllers. This stands in contrast to the contemporary judicial stance, where judges frequently grant legal orders to safeguard privacy ([Leith, 2016](#)).

Concerns regarding the automated processing of personal data in the UK can be traced back as far as 1961. This sparked a series of discussions, official reports, and the development of legislation centered around the issues of privacy and data protection. One particularly noteworthy piece of legislation in this context was the Data Protection Act of 1984. Interestingly, its enforcement unfolded progressively, with the final phase, which pertained to individuals' rights to access their data, commencing in 1987.

In the UK, the legal framework establishes a sharp division between intercepting communication or tracking data and the actual content of these

messages. To conduct an inquiry into an internal private cloud system without infringing upon privacy laws, organizations must obtain internal authorization within their own company ([Taylor et al., 2010](#)). In Canada, privacy holds a status akin to a quasi-constitutional right, meaning it is treated almost like a fundamental right. It receives protection through specific laws such as the Privacy Act for government entities and the Personal Information Protection and Electronic Documents Act (PIPEDA) for businesses. What sets Canada apart is the pivotal role played by privacy commissioners. These officials ensure compliance with privacy laws, address complaints from citizens, and vigilantly monitor any practices that may infringe upon individuals' privacy ([Bronskill & McKie, 2016](#)).

In the UK jurisdiction, the privacy laws that have evolved over time primarily concentrate on the oversight of databases and facilitating access to data. Navigating the complexities of the digital landscape and addressing data breaches continue to present challenges. The fundamental concept of “consent” plays a pivotal role in these laws, underscoring the significance of individuals agreeing to the utilization of their data. However, given the intricacies of the digital era, ongoing discussions about privacy legislation remain imperative in both countries. Thus, this research gap will be addressed by the researcher in this study by thoroughly examining the legislative frameworks of the UK protecting the privacy rights, adhering to address modern challenges.

5.2.5 RESEARCH GAP

Within the existing body of literature on privacy rights, this literature review reveals a notable research gap that this present study seeks to address. While legal scholars have offered various perspectives on privacy rights and their legal frameworks, a critical issue remains largely unexplored.

In the realm of definitions and identification of privacy rights, scholars have generally emphasized their fundamental importance and the necessity of safeguarding personal information. However, a common oversight in this

literature is the failure to explicitly delve into the vulnerabilities that arise in the face of emerging technologies.

Furthermore, when it comes to the examination of legislative policies governing privacy rights, a stark contrast emerges among the jurisdictions under study. Sri Lanka appears to lag behind, while the UK has taken a more proactive stance in addressing the challenges posed by emerging technologies. However, the effectiveness of Sri Lanka's existing legal framework in responding to these emerging challenges has received only superficial attention in the available literature.

What is particularly striking is the absence of comprehensive analyses and comparisons between the legal framework safeguarding privacy rights in Sri Lanka and those in the UK. Equally noteworthy is the lack of literature comparing Sri Lanka with either of these jurisdictions on this critical topic.

These research gaps underscore the significance of this study, which aims to provide a more thorough investigation into the effectiveness of the rules governing privacy rights in Sri Lanka compared to the legal system of the UK. By addressing this pivotal research question, the present study intends to contribute valuable insights into the protection of privacy rights in the digital age. Its primary objective is to fortify the legal framework concerning privacy rights in Sri Lanka to effectively address the challenges brought about by emerging technologies.

In the subsequent chapters, this study will present an in-depth analysis and comparative assessment of the legal systems in Sri Lanka and the UK in the context of privacy rights, aiming to bridge the identified research gap and offer practical recommendations for enhancing the legislative framework governing privacy rights laws.

5.3 LEGAL FRAMEWORKS RELATED TO PERSONAL DATA PROTECTION IN SRI LANKA

Sri Lanka, as a forefront country from South Asia, has passed its Data Protection Act in 2022, incorporating many of the standards from the

GDPR, and globally recognized standards. However, prior to the Act, few other legislations, and constitutional provisions, have embedded protections for personal data at different levels – both directly and indirectly.

5.3.1 THE CONSTITUTION OF SRI LANKA (1978)

Originally, the current Sri Lankan Constitution (1978) did not recognize the rights to information or privacy. The subsequent introduction of the right to information, facilitated by the 19th Amendment to the Constitution, resulted in the enactment of the Right to Information Act (2016). Notably, Article 14A (2) imposes limitations on the right to information, allowing for restrictions prescribed by law to safeguard national security, territorial integrity, public safety, prevent crimes, and protect health, morals, and individual privacy. However, it is crucial to highlight that this provision does not explicitly grant or elaborate upon the right to privacy as a fundamental right for Sri Lankan citizens.

Evidently, Article 14A (1) ensures the right to access information for individuals without concurrently guaranteeing their right to privacy. This situation raises concerns about potential infringements on individuals' privacy ([De Soysa, 2022](#)). In cases where the government or its administrative branches violate citizens' fundamental rights, such actions can typically be challenged by invoking the exclusive jurisdiction of the Supreme Court, as outlined in Article 17, read in conjunction with Article 126 (1). Nevertheless, in the context of privacy rights, this recourse is unavailable, given that the right to privacy is not established as a fundamental right under the Constitution of Sri Lanka.

5.3.2 THE COMPUTER CRIMES ACT, NO. 24 OF 2007

The Computer Crimes Act No. 24 of 2007 plays a significant role in addressing computer-related offenses in Sri Lanka. It describes computer crime detection and prevention. Significant substantive offenses are recognized under this paradigm.

The provision of this act covers a wide array of computer-related crimes such as unlawful computer access,² securing unauthorized access to conduct a crime,³ and causing a computer to perform a function without lawful authority.⁴

Although the Computer Crimes Act of Sri Lanka does not directly address data protection and privacy, it is evident that certain provisions within the Act have some relevance to these issues. However, such rules are not enough to address a country's data protection and privacy needs. The Computer Crimes Act criminalizes unauthorized access to computers, programs, data, and information. It does not explicitly regulate personal data processing and transportation. Therefore, it is often advisable to address specific legal matters with a dedicated piece of legislation or regulations designed explicitly for that purpose rather than relying on multiple laws that do not directly address the issue at hand.

5.3.3 ELECTRONIC TRANSACTIONS ACT, NO. 19 OF 2006 (ETA)

This digital law primarily governs the use of electronic data and digital documents in official and personal transactions. This statute aims to regulate electronic transactions. The Act relates to electronic data or communications; however, it does not define personal data or give data protection and privacy safeguards. Despite Sri Lanka's advanced electronic transaction regulation, data, privacy, and consumer protection are lacking. These gaps hinder the effective enforcement of the law governing electronic transactions.

5.3.4 RIGHT TO INFORMATION ACT, NO. 12 OF 2016 (RTI)

This provides citizens with an unequivocal right to access information, aligning with their constitutional entitlement under Section 3 of the Act. However, this right is subject to certain restrictions outlined in Section 5 of

the RTI Act ([Greenleaf, 2018](#)). Section 5(1)(a) is vital to data privacy and protection. Access to information shall be denied if the disclosure of personal information unrelated to any public activity or interest would unwarrantedly invade an individual's privacy, unless there is a substantial and justifiable public interest or written consent from the concerned individual. This provision clearly aims to strike a balance between the right to information and the principles of data protection and privacy.

5.3.5 BANKING ACT, NO. 30 OF 1998

Financial service providers, both banks and non-bank financial institutions, unquestionably handle substantial volumes of personal data. Section 77 places privacy obligations on individuals such as directors, managers, officers, and other employees working within licensed commercial banks or specialized banks.⁵ Presently, financial institutions face an escalating risk of data and privacy breaches perpetrated by criminals, given the critical nature of the data they manage.

5.3.6 TELECOMMUNICATION ACT, NO. 25 OF 1991

Section 49 outlines various offenses related to telecommunications officers and individuals performing official duties in connection with telecommunication services.⁶

Section 52 of the Act specifies that any person who intrudes without lawful authority into the content of a message or its usage information, with the intention of interfering with the message or unlawfully accessing its content, commits an offense.⁷ Section 54 also identifies the interception and disclosure of message contents by telecommunication officials as an offense. It is evident that these provisions are closely associated with concerns related to data protection and privacy.

5.3.7 INTELLECTUAL PROPERTY ACT, NO. 36 OF 2003

Section 160 of the Act addresses issues related to unfair competition and undisclosed information. According to Section 160(6)(a), unfair competition occurs when industrial or commercial activities disclose undisclosed information without consent from the rightful holder and violate honest commercial practices.

According to Section 160(6)(b), third parties disclosing, acquiring, or using undisclosed information without consent may result in industrial or commercial espionage, breach of contract, breach of confidence, and inducement to commit any of these acts. These provisions appear to have relevance in protecting the data and privacy of both organizations and individuals who possess valuable data.

5.3.8 PERSONAL DATA PROTECTION ACT, NO. 09 OF 2022

The PDPA was introduced by the government with the goal of changing the current data protection system in Sri Lanka. Section 4 of the Act requires data controllers to process personal data. Data controllers must also explain why they process personal data under Section 6(1). In addition, they must guarantee that personal data will only be handled for stated, specified, and lawful reasons. A person's right to correction is guaranteed by Section 15 of the Act.⁸

Data controllers and processors must comply with Part III of the Act. Entities that collect and process personal data are referred to as data controllers and processors. As part of their obligations under this Act, controllers and processors must designate or appoint a data protection officer.⁹

It appears that Sri Lanka's Data Protection Authority is in charge of all matter's ancillary to the protection of personal data, including carrying out the proposed Act's requirements. Another essential duty of the Data Protection Authority is its ability to issue directives to organizations that do not abide by the proposed Act's rules and to levy administrative fines.¹⁰

It is evident that the committee that drafted this Act adhered to best practices that were incorporated into a number of international standards, including the OECD Privacy Guidelines, the APEC Privacy Framework, the EU GDPR, etc. It can be claimed that Sri Lanka urgently needs a data protection law, given the rapid advancement of technology and other related issues such as digital tactics used by the public and corporate sectors.

5.3.9 CRITICISMS AGAINST THE PERSONAL DATA PROTECTION ACT, NO. 09 OF 2022

The efficiency of the PDPA is disputed by certain detractors, notwithstanding the advantages already mentioned. The Act's inclusion of ambiguous clauses has drawn the most criticism. It seems that critics argue that this legal ambiguity can stifle foreign investment into the nation. Next, unlike the GDPR, which permits derogations where data subjects express their explicit approval, this Act does not have a provision that facilitates data transfers with user consent. Additionally, according to Section 28(1) of the PDPA 2022, the Sri Lankan government may establish or nominate anyone to serve as the Data Protection Authority.

However, PDPA 2022 does not prohibit the government from creating an independent Data Protection Authority, but as the government would undoubtedly have a large amount of control over this Authority, its standing as an impartial expert body is likely to be compromised. As an illustration, the fact that the government can direct the Data Protection Authority in connection with the performance of its duties demonstrates the authority's lack of independence. That runs counter to the standards outlined in the EU GDPR as it implies that the supervisory authority established by the member states must be an impartial public body.

The Act's requirements for both data controllers and processors provide another issue. Therefore, data processors are required to abide by the terms of processing outlined in the Bill's five schedules. If the data processors do not follow the Act's requirements, they will be punished. Data processors will not be subject to these fines under the GDPR. Critics claim that these

components violate international norms, that they would make it more difficult for data processors to comply with regulations, and that they will have an adverse effect on foreign investment in Sri Lanka's data processing and outsourcing sector.

The proposed Act also mandates that data controllers perform Data Protection Impact Assessments (DPIAs). When processing data poses a danger to the rights and liberties of the data subjects, DPIAs must be performed. DPIAs should be carried out in connection with sensitive data processing on a broad scale and profiling. This can be viewed as an especially broad requirement that will turn DPIA into a preventative measure that will delay the delivery of innovative goods and services.¹¹ In addition, the PDPA 2022 divides data into two identical categories: special categories of personal data and personal data. It has drawn criticism, nevertheless, because it is fairly problematic.

When enforcing this law, it can be argued that the following factors should also be considered: a specific exception to ensure that the RTI will not be superseded in any case of inconsistency; an independent data protection authority free from government interference; and the removal of financial data and personal information related to offenses, criminal proceedings, and convictions from the list of "special categories of personal data" to ensure that the law is not violated ([Abeysekara & Ranasinghe, 2022](#)).

5.4 LEGAL FRAMEWORKS IN THE UNITED KINGDOM PROTECTING PRIVACY RIGHTS

5.4.1 DATA PROTECTION ACT, 2018 (DPA)

There are six main components to DPA 2018.¹² Under the DPA 2018, everyone must follow the strict "data protection principles" while utilizing personal data. Thus, everyone must utilize information honestly, lawfully, and publicly. Only necessary information should be used. Data should also

be secured to prevent unauthorized processing, access, loss, destruction, or damage.

The 2018 DPA states that individuals have the right to know what data the government and other organizations keep about them. You also have the right to know how your data will be used, correct any inaccurate information, delete data, limit or restrict processing, and object to processing in certain instances.

It is evident that the DPA 2018 only governs how data can be used, processed, and gathered legally in the UK.

5.4.2 UK GENERAL DATA PROTECTION REGULATION

Since the UK GDPR's provisions were developed in response to domestic needs, its enactment can be considered as a progressive step. The GDPR in the UK, following Brexit, is quite comparable to the EU rule and imposes identical requirements on data controllers and processors ([Carey, 2018](#)).

DPA 2018 is an addition to the UK GDPR. The DPA 2018 established data processing regimes for law enforcement and intelligence procedures and applied GDPR principles to public authority processing and other areas outside the regulation. Thus, the UK GDPR and DPA 2018 coexist.

UK GDPR applies to UK and non-UK companies that collect, store, or process personal data of UK residents or observe UK citizens' behavior. UK enterprises processing domestic personal data must obey the DPA 2018 and UK GDPR.

The GDPR grants UK data subjects the right to be informed, access, rectify, erase, object, and more. DPA 2018 and UK GDPR enforcement would be useless without data subjects' rights.

In addition, the UK GDPR enables the transfer of personal data under certain conditions. However, there are reportedly many advantages to GDPR compliance, such as enhancing customer trust, decreasing the risk of data breaches, increasing privacy and information security, etc.

5.4.3 CRITICISMS AGAINST DATA PROTECTION AND PRIVACY LAWS IN UK

The UK's data protection and privacy laws are among the most advanced in the world. As a result, it still has a few outdated components, such as certain gaps and contradictions. The DPA 2018's provisions are being criticized in the first place for giving an unjustified amount of power to amend the GDPR's rules. In addition, DPA 2018 does not offer sufficient protection in relation to the exceptions to the ban outlined in Article 22 of the GDPR or the use of automated decision-making without human involvement.

Moreover, DPA 2018 is not clear about what will happen to people's personal information or what they should do if it is mishandled or if there is not a strong enough legal remedy for data breaches. In addition, DPA 1988 gives intelligence agencies unrestricted authority to transfer personal data across international borders without putting in place the necessary protections. Evidently, not everyone is in favor of DPA 2018; critics claim that this legislation calls for greater accountability and transparency from organizations, as well as stricter laws to prevent data loss and theft and heavy fines for those who intentionally or negligently misuse data.

5.4.4 CASE STUDIES OF THE UK

The legal case of *Douglas v. Hello! Ltd*¹³ in the UK acknowledged privacy as a legal principle derived from the fundamental concept of personal autonomy. The aforementioned stance was upheld in the legal case of *Campbell v. Mirror Group Newspapers Ltd*,¹⁴ when it was established that the concept of privacy is fundamental to the concept of individual freedom within a contemporary society. The preservation of an appropriate level of privacy is crucial for the overall welfare and advancement of an individual.¹⁵ However, this case explains the need to uphold the concept of personal autonomy in Sri Lanka.

Moreover, in the case of *Denco v. Joinson*,¹⁶ the appellant was permitted limited access to the computer system as part of his employment as a metal worker, but he was accused of attempting to access information on the firm's clients, which went beyond the scope of his access rights. The appellant used an ID and password assigned to a separate firm when filing a case alleging unjust dismissal. This case explains that the individual privacy should be upheld in the country.

The Computer Misuse Act 1990, according to *R v. Bignall*,¹⁷ only safeguards against unlawful access to computers and the data they contain, not access to computers and data by authorized persons who use them for unauthorized reasons. This case determines that there should be a mechanism to safeguard against unlawful access to computers and the data-driven processors.

The case of *S and Marper v. the United Kingdom*¹⁸ established that the act of public authorities retaining and storing personal data, regardless of how it was obtained, is considered to have a direct influence on an individual's right to privacy. Thus, it is important to penalize people who are unauthorizedly access others' privacy.

Furthermore, the case, *Johnson v. Eastlight Community Homes Ltd*¹⁹ deals with a different legal matter. The case of EWHC 3069 pertains to the determination of whether claims should be assigned to the High Court or the County Court.

When evaluating the case laws decided in the UK, it is evident that there is a collective landmark judgement relating to data crimes, which is significant for the proper administration of justice. Thus, this series of cases provides the necessity of recognizing the right to privacy in Sri Lanka.

5.5 CONCLUDING REMARKS

This chapter discussion traverses through the comparative study, contributing to a comprehensive understanding of privacy rights in the digital era and the legal frameworks that govern them. The introduction stressed privacy rights and the need for a solid legal framework in Sri

Lanka that set the stage for the study. A research gap was uncovered by a thorough literature evaluation of academic publications, which inspired this study. There is no complete examination of Sri Lanka's privacy protection system in light of technological improvements and related challenges. The analysis and debate chapter stressed the importance of legislative frameworks in protecting privacy rights, especially in the digital age. Global data governance was illuminated by studying international privacy, cross-border data transfers, and data protection issues. The chapter also highlighted Sri Lanka's privacy issues, citing data breaches and cyberattacks. The chapter also critically examined key legal precedents and judgments, shedding light on Sri Lanka, the UK's judicial systems. In conclusion, the inclusion of a comprehensive comparative table facilitated a thorough evaluation of private data security in Sri Lanka, the United Kingdom, and Canada, thereby providing a visual representation that enhanced the research's conclusions.

Also, based on the above study, I hereby suggest a few recommendations to enhance the privacy protections across the digital platforms of Sri Lanka.

They include:

1. Effectively exercising individuals' rights to take legal action against data violations is vital under the Personal Data Protection Act No. 09 of 2022. Without strong enforcement, compliance would be difficult to ensure. Thus, laws would be symbolic and ineffective.
2. The government must emphasize preserving the Data Protection Authority's independence. To avoid conflicts of interest, its members should be appointed by an independent institution distinct from the government.
3. Data protection obligations must apply to companies of all sizes and industries. Data protection and privacy duties should not be categorized by company type, size, or location.
4. Data controllers and processors must implement organizational and technical steps to follow data processing principles. These requirements require these entities to adopt and maintain proper systems and processes.

They improve data processing efficiency. Organizations can use the DPIA to evaluate the privacy risks of certain actions or activities. This study aims to discover and analyze data privacy impacts.

5. Balancing the needs of data subjects, users, and society is crucial. By treating all parties fairly and prioritizing data subjects' rights, this objective can be achieved.
6. Certain requirements apply to the transmission of personal data of Sri Lankan nationals across borders. First, data transfer requires explicit consent from the data subjects. The target country must also have laws protecting personal data and privacy. Insufficient and inefficient data protection and privacy legislation in a country might be problematic.
7. Sri Lanka must recognize the importance of protecting the fundamental right to privacy for individuals. The Right to Information Act No. 12 of 2016 lacks a privacy right, sparking criticism. The significance of data security makes respecting and maintaining privacy even more important. The right to privacy and data protection coexists in certain situations.

Incorporating these suggestions will bolster Sri Lanka's privacy and data protection regulatory frameworks, particularly in effectively enforcing the recently passed PDPA. This enhancement will foster trust and reliability on the internet, crucial from a developing country perspective. It not only promotes increased internet usage but also contributes to the development of e-commerce and attracts direct investments in the near future.

NOTES

1. Article 13 of the Sri Lankan Constitution safeguards "personal liberty." However, it applies only to the arrest and custody of persons. [!\[\]\(48a7667d09d5a06397e047ee4537bb6f_img.jpg\)](#)
2. Section 3. [!\[\]\(3df135a685d1b545c4fa64a5f3516545_img.jpg\)](#)
3. Section 4. [!\[\]\(de62294faded52808857591d246c2e7a_img.jpg\)](#)
4. Section 5. [!\[\]\(632b91ede65784e1fc241c52ebe20c23_img.jpg\)](#)
5. Under this provision, these individuals are required to sign a declaration prior to commencing their duties, pledging to maintain strict

confidentiality (with certain exceptions) concerning all bank transactions, customer information, state of accounts, and related matters.[↗](#)

6. It states that an individual commits an offense if they, without legitimate cause: (a) willfully alter, destroy, or tamper with the content of received messages meant for transmission or delivery, (b) fail to transmit, intercept, or withhold any message, or (c) disclose the content of a message to anyone other than the intended recipient, except as directed by a court.[↗](#)
7. This includes the willful interception of telecommunication transmissions and the disclosure of message contents.[↗](#)
8. This section requires data controllers to correct or complete erroneous or incomplete data. Additionally, under certain conditions.[↗](#)
9. Section 22(1) PDPA 2022—data controllers must also take appropriate organizational and technical measures to maintain the security and confidentiality of personal data. Additionally, they must constantly take into account the Act’s emphasis on openness requirements.[↗](#)
10. Section 32 of PDPA 2022.[↗](#)
11. Section 24(1) PDPA 2022.[↗](#)
12. (a) General processing. (b) Processing by law enforcement. (c) Processed by intelligence services. (d) The UK’s data supervisory authority. (e) Information Commissioner’s Office (ICO), which is enforcement, as well as (g) supplemental and concluding measures.[↗](#)
13. [2005] EWCA Civ 595.[↗](#)
14. [2004] UKHL 22.[↗](#)
15. Campbell v. Mirror Group Newspapers Ltd [2004] UKHL 22.[↗](#)
16. [1991] 1 WLR 330.[↗](#)
17. [1998] 1 Cr App R 1.[↗](#)
18. [2008] ECHR 1581.[↗](#)
19. [2021] EWHC 3069.[↗](#)

REFERENCES

- Abeysekara, T., & Ranasinghe, A. (2022). Holistic Approach in Introducing Proper Legal Framework to Regulate Data Protection and Privacy in Sri Lanka. *Vidyodaya Journal of Management*, 8(1), 169–200. [↵](#)
- Bronskill, J., & McKie, D. (2016). *Your Right to Privacy: Minimize Your Digital Footprint*. Self-Counsel Press. [↵](#)
- Burgess, M. (2018, May 24). What Is GDPR? The Summary Guide to GDPR Compliance in the UK. *Wired*. Retrieved from www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018 [↵](#)
- Carey, P. (2018). *Data Protection: A Practical Guide to UK and EU Law*. Oxford University Press. [↵](#)
- De Soysa, S. (2022, March 22). The Right to Privacy and a Data Protection Act: Need of the Hour. *Daily FT*. Retrieved from www.ft.lk/article/606874/The-right-to-privacy-and-a-data-protection-act:- [↵](#)
- Geneva Academy of International Humanitarian Law. (2022). *The Right to Privacy in the Digital Age: Meeting Report*. Geneva.
- Greenleaf, G. (2018, 27). *Privacy in South Asian (SAARC) States: Reasons for Optimism*. Privacy Laws & Business International Report. [↵](#)
- Gyulavári, T., & Menegatti, E. (Eds.). (2022). *Decent Work in the Digital Age: European and Comparative Perspectives*. Bloomsbury Publishing. [↵](#)
- Komnienic, M. (2022, December 12). What Is Data Transparency? Q&A With Privacy Experts. *Termly*. Retrieved from <https://termly.io/resources/articles/data-transparency/> [↵](#)
- Koops, B. J. (2014). The Trouble with European Data Protection Law. *International Data Privacy Law*, 4(4), 250–261. [↵](#)
- Law, J. (2015). *Oxford Dictionary of Law*. Oxford University Press. [↵](#)
- Leith, P. (2016). *Privacy in the Information Society: Volume II*. Routledge. [↵](#)
- Masroof, A. (2007). Privacy Related Computer Crimes; A Critical Review of the Computer Crimes Act of Sri Lanka. *Sri Lanka L.C.L Rev.* [↵](#)
- McMenemy, D. (2016, August 13–19). *Rights to Privacy and Freedom of Expression in Public Libraries: Squaring the Circle*. 82nd IFLA General

Conference and Assembly, Columbus, USA.↵

Morozovarchive, E. (2013, October 22). The Real Privacy Problem. *MIT Technology Review*. Retrieved from www.technologyreview.com/2013/10/22/112778/the-real-privacy-problem/↵

Mutambik, I., Lee, J., Almuqrin, A., Halboob, W., Omar, T., & Floos, A. (2022). User Concerns Regarding Information Sharing on Social Networking Sites: The User's Perspective in the Context of National Culture. *Plos One*, 17(1), e0263157.↵

Romansky, R. P., & Noninska, I. S. (2020). Challenges of the Digital Age for Privacy and Personal Data Protection. *Mathematical Biosciences and Engineering*, 17(5), 5288–5303.↵

Samahon, A., & Liyanagamage, V. (2021, March 25). A Study on the Implementation of Data Protection Laws in Sri Lanka. *Colombo Telegraph*. Retrieved from www.colombotelegraph.com/index.php/a-study-on-the-implementation-of-data-protection-laws-in-sri-lanka/↵

Samarakoon, C., & Fonseka, B. (2020). *Right to Privacy in Sri Lanka*. Center for Policy Alternatives.↵

Taylor, M., Haggerty, J., Gresty, D., & Hegarty, R. (2010). Digital Evidence in Cloud Computing Systems. *Computer Law & Security Review*, 26(3), 304–308.↵

The Secretariat of the United Nations Conference on Trade and Development. (2019). *Trade and Development Report 2019*.

6 AI Ethics and Colonial Legacies

Evaluating UNESCO's Recommendations for Inclusivity

*Akash Bag, Anwesha Ghosh, Tejaswini
Tripathy, and Piotr Pietrzak*

DOI: [10.1201/9781003668657-6](https://doi.org/10.1201/9781003668657-6)

6.1 INTRODUCTION

Artificial intelligence (AI) still remains an essential source of severe ethical challenges in a number of industries, especially when it is important to study such principles as social justice, explainability, bias, and privacy ([Carvalho et al., 2022](#)). The two forms of idea analysis include descriptive idea analysis, which provides an overview of the United Nations Educational, Scientific and Cultural Organization (UNESCO) recommendations' extent and purpose, and critical idea analysis, which examines the potential application and the potential beneficiaries of these recommendations. The conversation revolves around the assumption that the recommendations promote an open universalism deeply rooted in the ethic of human rights. This approach corresponds with postcolonial theoretical frameworks that emphasize the continuing differences in power and voice between the various global communities. In the medical field, the application of AI technologies necessitates strict adherence to ethical

requirements to preserve patient confidentiality and maintain the purity of medical informatics ([Civaner et al., 2022](#)). Similarly, software companies are more concerned with data security and the development of proper ethical standards due to AI advancements ([Vakkuri et al., 2022](#)). Other components of AI's ethical context include issues such as who is responsible for AI system errors and the ethical implications of machines' behavior ([Hermansyah et al., 2023](#)). Various industries and organizations, such as the Catholic Church, underline the need to create ethical discussions of digital society that will help present the multifaceted effects of AI ([Čerbová, 2021](#)). Furthermore, we must develop ethical principles for the formation of AI systems ([Leikas et al., 2019](#)).

To solve the ethical and philosophical problems associated with AI technology, there is increasing interest in the educational industry in integrating ethics into AI courses ([Burton et al., 2017](#)). This is why it is crucial to ensure that there are official protocols for ethical reasoning in AI to ensure that ethical concerns are systematically considered in the AI processes ([Almpani et al., 2023](#)). Compliance with ethical principles in the creation and use of AI systems is also critical to improving customer satisfaction and trust ([Dolganova, 2021](#)).

Legislators have developed numerous proposals for ethical charters, recommendations, and legal acts to regulate the ethical and legal aspects of AI development and use ([Matyuk, 2022](#)). As previously discussed, the UNESCO Recommendation on the Ethics of Artificial Intelligence provides guidelines for member states to manage AI ethically, demonstrating the diversity of global perspectives and priorities in AI ethics ([Jobin et al., 2019](#); [Wang & Zhang, 2022](#)). UNESCO's recommendation on the Ethics of Artificial Intelligence is a critical attempt to establish ethical principles in the design and use of AI systems, with special regard to human rights protection. This is especially important in fields like medical education that require AI's ethical incorporation to solve various issues ([Civaner et al., 2022](#)). Furthermore, the inclusion of other cultures from all over the world, such as the African Ubuntu worldview, broadens the discussion and provides different perspectives on the concept of human rights about AI

([van Norren, 2022](#)). The term “moral diplomacy” can then be considered a more accurate representation of the current ethical situation in AI and its development, especially when considering the topics discussed during the 2021 UNESCO Intergovernmental Meeting of Experts ([Vică et al., 2021](#)). This corresponds with the identification of five primary AI ethical principles: human-centricity, agency, justice, and intelligibility, which are in harmony with UNESCO’s recommendations ([Dolganova, 2021](#)). The United Nations has also discussed AI human rights issues and pointed out major issues such as freedom of speech and the possibility of discrimination, which require strict ethical standards ([Hunkenschroer & Kriebitz, 2023](#)). The ethical, legal, and social implications extend into the financial sector, where the deployment of AI technologies must be critically evaluated to avoid ethical pitfalls ([Ahmed, 2022](#)).

The effort to establish ethical guidelines for the application of AI responds to critical concerns regarding the technology’s alignment with ideological constructs, particularly whiteness. [Katz \(2020, p. 155\)](#) provocatively describes AI as a tool that perpetuates colonial power structures and upholds white supremacy. This perspective sets the stage for scrutinizing whether UNESCO’s Ethical Recommendations on AI, established in 2021, are crafted in an inclusive manner that adequately acknowledges and addresses the historical dominance of whiteness and its implications for racial and ethnic discrimination today.

Thus, the paper’s main question is to assess the efficacy of these ethical recommendations as a blueprint for global leaders and organizations to protect the rights of historically marginalized populations. This includes an analysis of the document’s fundamental human rights provisions, as well as the values and standards it supports. The analysis centers on the recommendations’ targeted audience, such as AI developers and other stakeholders, to assess if the guidelines adequately acknowledge and mitigate the long-lasting impacts of colonial influence. The purpose of the paper is to critically examine the effectiveness of the UNESCO document in addressing the power dynamics inherent in technology. Therefore, the paper aims to shed light on the limitations of the AI recommendations in

terms of ethical framing and directives on the intersectionality of technology, race, and historical subjugation. This method emphasizes how important it is to engage deeply and critically with AI ethics to make sure that developments in technology do not perpetuate present-day injustices and discriminatory behaviors, but rather call into question them. The main focus of the paper will be how well UNESCO's ethical AI guidelines handle the difficulties in promoting inclusive global governance and guaranteeing equitable human rights protections in the era of AI.

6.2 LITERATURE REVIEW

In her analysis, researcher van Norren examines how AI is perceived and connected to the Ubuntu ideology in Africa, as well as UNESCO's ethical guidelines in this area. According to the Ubuntu ideology, the community is the biggest and most significant component of society ([van Norren, 2022](#)). Although UNESCO's recommendations are a global document, the author contends that applications and AI are not contextualized to African ethics; rather, they are typically adapted to English-speaking countries and are in English. Additionally, the author raises the issue that the majority of AI researchers in this field are Westerners. This typically makes AI apps challenging for users that adhere to certain ideologies, like Ubuntu. It causes problems with cybersecurity, which puts up a barrier. AI is always a risk, especially for exposed and vulnerable people and groups ([Mohamed et al., 2020](#)). A critical review of AI research is needed both to ensure that development and to make it more inclusive. The use of AI and science is guided by norms, which must also be morally relevant and contextual ([Mohamed et al., 2020](#), p. 661).

To avoid exclusion, a new methodology must be written and communication between the AI developers and impacted parties must occur ([Mohamed et al., 2020](#), p. 678). The study demonstrates the necessity for robots and AI systems to foster civil rights and human trust. The development of intelligent systems should give ethical considerations more attention in future research investigations. Dignum argues that there are

various ways in which ethical considerations connect to AI. It is necessary to have both an ethical viewpoint on how the AI is designed and responsibility. The persons in charge of the algorithm must be. Engineering methodologies must also take ethics into account. Thus, to integrate social structures with ethical content, analysis and evaluation are required. Moreover, ethical deliberation is required in AI's code of behavior to protect users' and developers' privacy (Dignum, 2018, p. 2).

There are major challenges with the AI systems that have been developed in recent years and the technology has an impact on people's basic human rights ([Aizenberg & van den Hoven, 2020](#)). Furthermore, it is believed that a proactive strategy is needed to deal with AI. It is necessary to change one's appearance based on the sociopolitical climate. Considering the study by [Aizenberg and van den Hoven \(2020\)](#), it is appropriate to discuss the necessity of an AI normative framework parallel to the UN's Universal Declaration of Human Rights. It was the opinion of [Aizenberg and van den Hoven \(2020\)](#) that it would require an even more abstract approach to be implemented on a national scale. AI ethics is lacking in several places and it lacks further consequences if deviations occur ([Hagendorff, 2020](#)). The researcher further believes that the existing ethical guidelines do not affect AI developers' decision-making in practice. Furthermore, it has been argued that the researcher's and society's ignorance of ethical aspects around AI means that there will be no consequences if they violate fundamental rights ([Hagendorff, 2020](#), p. 114).

The ethics of AI has emerged as a critical area of focus in the digital age, prompting a structured approach to integrating ethical principles into AI systems. UNESCO's (2021) recommendation on the Ethics of Artificial Intelligence establishes a comprehensive framework for member states, aiming to guide the governance of AI through clearly defined values, principles, policy actions, and mechanisms for monitoring and evaluation ([Wang & Zhang, 2022](#)). This framework is a response to the increasing recognition of the need for ethical considerations in AI development. The interaction between AI and human rights forms a complex landscape that requires meticulous attention. It emphasizes that different human rights

organizations must deal with the issues that technology brings, underscoring the need to address them to protect human rights ([Nirenberg, 2014](#)). Also, the use of AI in such areas as criminal justice systems poses major ethical issues, including the reinforcement of current bias and discrimination ([Malek, 2022](#)). These discussions stress the need to integrate ethical values, starting from the design process up to the deployment of AI solutions. Furthermore, the creation of reliable and trustworthy AI solutions in the healthcare and education domains emphasizes the importance of compliance with ethical standards and professional codes of conduct to minimize possible threats that may affect society ([Bærøe et al., 2020](#); [Tapalova & Zhiyenbayeva, 2022](#)).

AI's ethical issues extend beyond its technical aspects, encompassing social issues like privacy, equality, and environmental concerns ([Bryndin, 2020](#); [Hermansyah et al., 2023](#); [Ruban, 2022](#)). The expanding body of writing and the international work of expert groups to outline ethical AI are evidence of the global interest in developing solutions to the ethical issues that accompany AI technologies ([Cervantes et al., 2020](#); [Jobin et al., 2019](#)). These debates cut across various fields: psychology, law, philosophy, and business management, thus underscoring the multifaceted nature of ethical issues in AI ([Cath, 2018](#); [Iphofen & Kritikos, 2019](#); [Luccioni & Bengio, 2020](#)). UNESCO's (2021) recommendation on the Ethics of Artificial Intelligence is a crucial global treaty that outlines the principles, values, policy measures, and assessment tools for member states to govern AI ([Huang et al., 2023](#); [Wang & Zhang, 2022](#)). Numerous ethical questions arise in various fields, including healthcare, where it is crucial to guarantee the reliability of AI ([Hermansyah et al., 2023](#)).

Other ethical issues of AI include explainability, bias, and the effects on employment and the military, which further highlights the challenges of creating AI systems that are ethical and aligned with society's norms ([Carvalho et al., 2022](#)). In addition, defining and building ethical competencies in AI systems is crucial to preserving these norms ([Bryndin, 2020](#)). Other than the technical and application-related factors, ethical AI development also has implications for other social and humanitarian

aspects. This entails the implementation of health equity, racial justice, and human rights into the development of AI, a feature that UNESCO noted as essential ([Dankwa-Mullan et al., 2021](#)). The application of postcolonial and feminist theories provides insights into addressing structural inequities and understanding the historical and colonial impacts on racial stratification ([Anderson et al., 2007](#); [Go, 2018](#)). In terms of health technology, ensuring equity involves balancing efficiency with fair access and distribution, which is critical to ethical decision-making in healthcare ([Cookson & Mirelman, 2017](#)).

This chapter will examine the universal ethical recommendations regarding AI. Additionally it will examine any potential issues around the document's universality. The goal of the work is to add an understanding of vulnerable populations, and AI, and how to include them in discussions around AI. As demonstrated above, numerous documents have been the subject of earlier studies on AI and ethics. The study has also shown that when creating AI, programmers must have an ethical stance.

6.3 METHODOLOGY

The paper under review engages with the UNESCO document titled "*Recommendation on the Ethics of Artificial Intelligence*," which was adopted on November 23, 2021. The document, comprising 134 paragraphs, serves as a foundational text, proposing ethical guidelines and principles for AI usage to safeguard human rights. Additionally, the analysis includes Gabriela Ramos's perspective on the practical implications of these guidelines within the workforce, particularly concerning the intersection of AI, employment, skills, and future work dynamics. The analytical approach adopted in the paper is bifurcated into descriptive and critical idea analysis. Initially, the paper offers a descriptive account, which not only elucidates the content of the recommendations but also provides a basis for their validation and further scrutiny. This phase is instrumental in clarifying the theoretical underpinnings and ideological constructs within the document, thereby setting the stage for a more nuanced evaluation.

6.4 THEORETICAL UNDERPINNINGS

The paper delineates its theoretical framework and analytic strategy by explicating the conceptual underpinnings that inform its scrutiny of human rights, globalization, and agency. The discussion is anchored in universalism, as interpreted through the lens of Elena Namli's ([Namli, 2014](#)) conceptualization, while also engaging with postcolonial perspectives, primarily drawing from Ania Loomba's work ([Loomba, 2015](#)), and phenomenological approaches to agency, based on Sara Ahmed's ([Ahmed, 2007](#)) definitions. Universalism in this context refers to the notion that human rights, as articulated in the UN's Universal Declaration of Human Rights, are inherently applicable across diverse global contexts. This Declaration suggests human dignity as foundational to freedom, justice, and peace, asserting that all individuals are born free and equal in dignity and rights. Such rights are endowed upon individuals by their capacity for reason and conscience, compelling them to interact within a spirit of fraternity.

However, the application and interpretation of these universal norms are subjects of considerable debate. Although these rights are supposed to be the standard for all countries, the application of these rights at the national level presents many difficulties. These challenges are partly because the document is quite general, which means that its provisions can be interpreted in various ways depending on the legal and cultural environment. [Namli \(2014, p. 35\)](#) has expressed his concern about the idea of universality of human rights laid in the Declaration, stating that this means that human rights are implemented in the same manner all over the world. She also posits that there is now a general understanding of the concept of human rights as a principle, but the implementation of these rights requires that each independent nation-state uses its power to protect them within its territory. Moreover, the discussion of human rights themselves is built primarily on the principles of Western culture, which means that it is based on the premise that there is a global agreement on such values. This assumption fails to consider that there are great

ideological and cultural differences that affect the perception and application of human rights standards all over the world.

Elena Namli ([Namli, 2014](#)) has suggested a concept regarding human rights norms, which she terms “open universalism.” In her view, many legal and political human rights documents of the global human rights regime underplay the moral aspects of human rights. [Namli \(2014\)](#) asserts that the structure of human rights relies on three fundamental aspects: moral, legal, and political. [Namli \(2014\)](#), in his concept of open universalism, also supports the moral aspect of protecting human rights efficiently. She argues that society and people’s perceptions of right and wrong in various cultural and national settings construct moral norms, not fixing them ([Namli, 2014](#), p. 20). This perspective challenges the notion of a universally applicable set of human rights, asserting that these rights should be sufficiently diverse to accommodate diverse moral and cultural realities.

This theoretical approach is particularly relevant in the analysis of global human rights documents, such as the UN’s Universal Declaration of Human Rights. This declaration, which emerged during the Second World War, sought to establish the foundation of global recognition and acceptance of rights. Nevertheless, following [Namli’s \(2014\)](#) theoretical framework, the application and understanding of these rights can differ greatly from one society to another due to the differences in legal, political, and moral systems. In addition, UNESCO works by the Charter’s principles and acts toward the achievement of peace and cooperation through the means of education, science, and culture. It has introduced human rights norms into diverse fields of science and technology, such as bioethics and AI, demonstrating their incorporation and application.

In addressing the question of the universal applicability and interpretation of human rights, the paper employs the concept of universalism to scrutinize how the document under consideration portrays human rights and places its recommendations within the context of universality. The paper aims to identify the implicit postulates of the human rights enshrined in the document and assess the extent to which various societies portray these suggestions as globally applicable. Transitioning to

postcolonial studies, the paper includes a reference and discussion of Ania Loomba's critique and elaboration ([Loomba, 2015](#)) to elucidate the current colonial influence on socio-economic and cultural processes. Postcolonial studies, as described by Loomba, focus not merely on the historical aspects of colonialism but also on its persistent effects in shaping global relations and power structures today. [Loomba \(2015, pp. 31, 245\)](#) emphasizes the enduring stereotypes and economic disparities that have arisen from colonial histories, suggesting that terms such as "developing countries" reflect these ongoing imbalances and biases.

Furthermore, [Loomba \(2015\)](#) raises concerns about the selective focus within academic discourse on cultural issues at the expense of economic exploitation, which she argues is a critical aspect of the colonial legacy that continues to affect former colonies ([Loomba, 2015, p. 2](#)). This critique extends into an analysis of globalization, which [Loomba \(2015\)](#) views as a phenomenon that, while facilitating technological access and potential economic growth, also exacerbates existing inequalities and reinforces historical power hierarchies ([Loomba, 2015, p. 247](#)). The paper posits that understanding the concept of "world order" as a framework for managing global events and political stability is crucial for analyzing how human rights recommendations perceive and address global disparities. It raises pertinent questions about the differences in recovery and development between Western countries and those historically subjected to colonial rule, as well as the underlying values that shape the current global order.

[Ahmed \(2007\)](#) conceptualizes whiteness as an invisible normative force that essentially controls how bodies orient themselves within societal spaces. According to her, whiteness functions as a sort of gravitational force, establishing what is central and peripheral in social environments ([Ahmed, 2007](#)). Her theory extends to the notion that everyday interactions and spatial negotiations are influenced by histories of whiteness, which prioritize certain bodies over others, thereby determining how individuals and groups occupy and move through various spaces. [Ahmed's \(2007\)](#) discussion of orientation builds upon Franz Fanon's phenomenological exploration of the body schema, which highlights the lived experience of

bodies in space, conditioned by historical and racial contexts ([Romdenh-Romluc, 2024](#)). Fanon suggests that societal norms and structures are inherently configured around a “white” body schema, which positions white bodies as naturally at home within these spaces, while others are marked as alien or foreign ([Ahmed, 2007](#)). The concept of the body schema is integral as it shapes the perceptual lens through which individuals see the world and interact with it.

Interaction with the environment constantly updates this representation, keeping the existing world order based on whiteness intact. This inherited and predominantly white worldview defines what is considered familiar and foreign, accessible and beyond reach, within society’s inherited social order and personal experiences. We must critically examine this orientation’s impact on agency and inclusion, scrutinizing the action and interaction processes within these structured environments raises questions about the possibility of being an actor in a white-oriented order and how it influences one’s perception of inclusion and exclusion in societal contexts. This analysis not only questions the systemic alignments that favor certain groups but also seeks to uncover the implications of these alignments on collective and individual agency in the social world.

6.5 ANALYSIS

The following part will begin by analyzing the recommendations’ definitions of rights, norms, and agency based on the analytical questions posed in the theory and then problematizing the document’s reasoning based on the theoretical starting points presented earlier.

6.5.1 RIGHTS

The document under discussion delineates an encompassing approach toward AI that integrates ethical considerations with the development and implementation of AI technologies. It emphasizes the dual impact of AI—both positive and negative—on society, highlighting its potential to

transform human cognition and societal structures through innovations (UNESCO, 2021, p. 5). However, it also acknowledges the emergence of new ethical dilemmas fostered by AI, particularly in realms such as decision-making, social interactions, and fundamental human rights. Central to the recommendations is the principle that AI should enhance human rights, fundamental freedoms, and human dignity. This principle is elaborated across four targeted domains: fostering peaceful societies, just, and well-connected; addressing environmental and climate concerns; promoting diversity; and encouraging broad participation in AI development and implementation. This document assumes a dogmatic theory that places AI in the context of legal and ethical systems. This framework considers the relationship between human beings, their surrounding environment, and society in general (UNESCO, 2021, p. 6). We identify the design of AI, its use, and its eventual disposal as a key stage at which AI should have a net positive impact on society and develop respect for human rights.

Furthermore, it underlines the need to protect cultural and contextual differences in AI-related solutions. This makes certain that AI technologies respect the diverse needs and rights of people in various categories to achieve principles of equity and access (UNESCO, 2021, p. 20). Regarding policy implementation, the recommendations suggest that there is a need for context analysis to address the possible clash between the espoused values and the actual usage of AI technologies. It is important to avoid violating the rights of persons and fundamental freedoms while applying AI in compliance with international law (UNESCO, 2021, p. 18). Paragraph 28 reads:

AI actors should respect and enforce social justice and the non-discrimination policy of any type under international law. This means that everyone should have an equal chance to reap the benefits of AI technologies, and we must not overlook the requirements of diverse age groups, cultural systems, language groups, individuals with disabilities, girls and women, and those who are disadvantaged,

marginalized, vulnerable, or in vulnerable situations. It also ensures that individuals have equal opportunities to benefit from AI technologies.

UNESCO (2021, p. 20)

The passage examines a turning point in the text that enunciates the claim that everyone should be entitled to AI. Particularly, it claims that AI should be provided to every person, the differences in origin and community should not affect this. This sentiment is summed up in the following statement: “*Member States should strive to ensure that all, including local communities, have access to AI*” (UNESCO, 2021, p. 20), which places the principle of inclusion as a given that needs to be followed. Paragraph 105 reads:

Member States should ensure girls and women, people of different ethnicity and culture, disabled people, people on the margins and in the vulnerable positions or situations, minorities and all those who do not have an equal opportunity to engage in digital society, should be encouraged to lead and participate in AI education programs at all levels, as well as to share and exchange the experiences and the best practices in this regard with other Member States.

UNESCO (2021, p. 34)

The document supports the idea that “AI development is a commonwealth of various parties.” It emphasizes the importance of knowledge dissemination and AI benefit distribution to various groups of people, including women, culturally diverse people, and those without access to digital devices. This strategy operates on the principle that AI should benefit everyone equally, necessitating the incorporation of these groups’ requirements and circumstances into AI innovation and usage. It fails to explain the specifics of how to handle possible drawbacks of AI use, such as prejudice and does not offer guidance for those who may not wish to engage with AI technologies. In addition, it does not offer detailed

strategies for achieving the goal of raising AI's cultural relevance and local adaptation, despite stressing their necessity. The document takes a broad, relatively passive posture, requiring everyone to use AI and participate in its development without clearly outlining how this should be done in practice.

[Ramos \(2022\)](#) elucidates the transformative influence of AI on human learning processes, information assimilation, and interpersonal relations. While AI enhances access to information, it concurrently presents risks, such as the propagation of disinformation and potential infringements on human rights and democratic principles ([Ramos, 2022](#), p. 2). These complexities necessitate the implementation of policy measures and ethical guidelines to prevent misuse. The document further indicates the need for comprehensive education to empower the widespread and ethical application of AI (UNESCO, 2021, p. 35). It refrains from prescribing explicit directives on how member states should utilize the recommendations or develop AI technologies, emphasizing instead a role for UNESCO in monitoring and supporting the implementation of these ethical guidelines through evaluations and capacity-building initiatives. This framework positions UNESCO not as a directive force but as a facilitative body focused on consequence management and support. It provides guidelines for ethical consequences and approaches, as well as guidelines for training decision-makers on the ethical impact of AI, thereby promoting the culture of ethical AI as a shared responsibility of all stakeholders.

At the heart of such an approach lies a general framework for legislation and policy regulating the use of AI, which must be compliant with international law. The recommendations raise awareness of the operation of AI, the type of data it uses, and how these factors affect human rights and freedoms. The recommendations should ensure the avoidance of risks like discrimination and inequality. Additionally, it establishes ethical guidelines for AI use that align with the principles of the Universal Declaration of Human Rights. It advocates for an internationally recognized ethical standard for the use of AI and AI technologies, emphasizing the importance

of education and access to AI, particularly for the world's less privileged, to bridge the gap (UNESCO, 2021, p. 6). The recommendations aim to lay the groundwork for AI, not just as an instrument but as the building block of a new world order based on ethical values and human rights.

This critique aims to delve into the internal contradictions of a UNESCO document that asserts AI as a universal right, implying a global demand and necessity for AI, irrespective of discrimination or apathy. This perspective carries profound implications regarding the incorporation of diversity and the recognition of people's right to choose what to adopt in their lives. Firstly, the document's insistence on establishing AI as an inherent human right fails to consider the choice of people or groups who do not want to engage in AI technology because of past experiences or prejudice. This oversight illustrates the conflict between full inclusion and society's and history's heterogeneity. [Namli \(2014, p. 13\)](#) says that these kinds of universalistic claims may always have colonial roots. This means that they may create values and norms that can be used all over the world, but they are not culturally or politically sensitive to non-Western socio-political realities. This makes these kinds of claims problematic.

Furthermore, the document promotes "open universalism," a moral and normative approach to human rights that should guide social and political structures. Although the document does not mandate the standardization of AI use, it provides guidelines that member states can interpret differently. One could view this flexibility positively, as it allows for variations in the interpretation of human rights across different jurisdictions. However, the idea of training everyone in AI, especially focusing on minorities and other vulnerable groups, as done with good intentions, may also serve to impose a normalizing expectation on people's interactions with technology. This universalism threatens to impose one vision that does not necessarily recognize or respect the history or cultural differences of the world's peoples.

The second problem under discussion pertains to the potential for discrimination against specific groups or individuals based on universal human rights, rather than group concepts. [Namli \(2014, p. 13\)](#) noted that

while individual rights may be good for the majority, group rights are important in protecting the minority. The recommendations, by requiring training for all and the implementation of AI, do not sufficiently address the call for critical interaction with the history and global injustice that currently govern technology use and access. It is also noteworthy that there is no discussion of historical grievances and contemporary world structure in the document, which can affect the approach to creating AI technologies and their use.

6.5.2 WORLD ORDER

According to UNESCO (2021), certain recommendations stress the importance of a normative framework that can help the member states apply and regulate AI technologies. This framework not only aims to properly implement AI systems in a society but also ensures their acceptance by the population. Most importantly, the document underscores the need to include low-income countries, social groups, and least-developed nations in the AI discourse. Their participation is critical to avoid cultural globalization and to ensure that AI's advantages and underpinning technologies do not only mirror the perspectives of other developed countries. These areas are characterized by a heightened focus on capacity development, especially concerning media and information literacy education. Such measures are crucial in preparing these countries to effectively compete in the digital economy, thereby preventing a widening of the technology divide that would only worsen the divide in global society.

Furthermore, it enunciates that technologically developed countries are obligated to spread the advantages of AI technologies to less developed countries. This is considered a way to advance toward a fairer world in which the opportunities and advantages of AI not only reproduce inequalities but also promote greater social, economic, and political stability through, for example, better communication, culture, education, and research. In essence, through international cooperation and targeted support, including financial aid and infrastructural investments, these

recommendations aim to diminish inequality and address development challenges more effectively. This collaborative approach aims not only to democratize the development of AI but also to ensure its ethical implementation and the equitable distribution of its benefits (UNESCO, 2021, pp. 10, 6, 20).

Further can be read in paragraph 80:

Member States should work through international organizations to provide platforms for international cooperation on AI for development, including by contributing expertise, funding, data, domain knowledge, infrastructure, and facilitating multi-stakeholder collaboration to tackle challenging development problems, especially for LMICs, in particular LDCs, LLDCs, and SIDS.

UNESCO (2021, p. 30)

In paragraph 83, it also states that:

Member States should encourage international cooperation and collaboration in the field of AI to bridge geo-technological lines. Technological exchanges and consultations should take place between Member States and their populations, between the public and private sectors, and between and among the most and least technologically advanced countries in full respect of international law.

UNESCO (2021, p. 32)

The discourse encapsulated in the document reflects a juxtaposition of universalism with a nuanced recognition that developed countries must actively support their less developed counterparts, specifically in educating them about AI. It advocates for international cooperation while respecting national sovereignty to foster genuine participation in AI development. [Ramos \(2022, p. 10\)](#) stresses the necessity of crafting regulations to safeguard against market dominance abuses, particularly to protect vulnerable developing countries from exploitation. Concerns about diversity within AI research and development are prominent, highlighting

the sector's male dominance which could perpetuate ethical issues and further inequality. To combat this, there is an imperative to enhance gender equality by training women in AI use and development (UNESCO, 2021, p. 35). The document also suggests that AI ethics should be developed considering national differences, and in multiple languages, to ensure inclusivity across various societal strata.

[Young \(2003, p. 4\)](#) critiques the colonial overtones within such global initiatives, suggesting that they continue to perpetuate a narrative where less developed countries are seen as needing guidance from developed nations. This perspective can echo colonial attitudes where the West positioned itself as a caretaker for the ostensibly incapable colonized states. The document's focus on developing AI primarily in the least developed countries (LDCs), landlocked developing countries (LLDCs), and small island developing states (SIDS) could inadvertently reinforce these outdated power dynamics (UNESCO, 2021, p. 20). [Butler \(2009, p. 37\)](#) raises a pertinent concern regarding the democratization process, which may mirror the issues with AI imposition. She argues that forcing democracy or AI upon states can create significant power imbalances and potentially lead to undemocratic outcomes. The document suggests that AI dissemination aims to foster a more democratic world, yet this could paradoxically result in greater dependence of less developed countries on their more advanced counterparts, thus exacerbating global inequalities rather than alleviating them.

Paragraph 51 (UNESCO, 2021, p. 21) acknowledges the potential for AI to widen the gap between the rich and poor and to deepen digital divides. Nevertheless, the strategy lacks a plan to prevent such outcomes. If we expect the developed nations to provide funds and training as and when necessary, this strategy may perpetuate a dependency that shifts power dynamics in their favor, rather than seeking to subvert them. Therefore, although this document outlines a future where AI is a tool for fair global cooperation, it does not explain the challenges and possible negative effects of such an agenda, which would further entrench historical injustices and generate new forms of dependence and division. The paper argues that the

world order created based on AI recommendations has critical issues in its proactive approach and enforcement mechanisms. Thus, it states that although the recommendations call for peaceful, inclusive, and just AI throughout the AI life cycle, they do not specify procedures or responsible parties (UNESCO, 2021, p. 20). This makes the document passive and can weaken human rights protection, as it is up to individual member states and AI actors to decide how to interpret and enforce it.

Further complicating the landscape is the document's inadequate address of how AI globalization might exacerbate economic disparities and social divisions. [Loomba \(2015, p. 246\)](#) suggests that the modern global power dynamics do not simply revolve around physical invasions but rather involve embedding smaller nations within an economic network that enhances their dependency on more powerful nations. This could be paralleled in the AI sphere, where less technologically developed countries might become dependent on more advanced nations for AI technology and infrastructure. The role of AI in maintaining existing societal structures is also critically examined. Yarden Katz identifies AI as a technology that perpetuates the normativity of whiteness, thus potentially reinforcing racial inequalities and social hierarchies ([Katz, 2020, p. 2](#)). This critique is complemented by data indicating a lack of diversity within the AI research community itself, which is predominantly white and Asian, thus raising concerns about the inclusivity and representativity of AI advancements ([Zhang et al., 2021, pp. 10, 12](#)).

The recommendations state that AI should respect human rights principles and be available to everybody. This article advocates for the inclusion of individuals from diverse countries to avoid prejudice, yet it fails to guide how to prevent power dynamics from distorting the dominance of large American companies in the process. Western corporations currently control AI development, raising several questions about the terms and conditions of its implementation. According to [Rajagopal \(2021, p. 15\)](#), the structural consequences of globalization are the violation of human rights as poor and vulnerable populations receive the norms of the West. This critique highlights a flaw in the recommendations,

as they fail to specify the methods for scrutinizing large corporations to prevent human rights violations and hold them accountable for their actions. Furthermore, while the recommendations mandate member states to establish processes for assigning ethical and legal responsibilities throughout the AI system's life cycle, they do not specify the precise implementation of these responsibilities (UNESCO, 2021, p. 9). This exclusion presents significant challenges in enforcing responsibility within entities, particularly in the context of AI globalization. Incorrect data feeding into AI increases the risk of misuse or rights violations, and the dominance of large tech companies in AI could lead to inadequate accountability.

6.5.3 ACTORSHIP

UNESCO guidelines describe the distribution of tasks between AI actors and Member States to ensure ethical considerations and human rights protection throughout the AI life cycle. AI stakeholders, defined as any participant including but not limited to researchers, engineers, universities, and companies involved in any phase of AI systems' development, should consider diverse groups to avoid prejudice (UNESCO, 2021, p. 10). We expect them to ensure the promotion of justice and non-violence, as well as the development of an interconnected world (UNESCO, 2021, p. 22). UNESCO emphasizes in paragraph 29 that AI actors bear a great responsibility to prevent the reproduction of discriminatory results, guarantee fairness, and provide redress against bias in AI systems (UNESCO, 2021, p. 20). Still, these recommendations are rather vague on how the AI actors can adequately guarantee respect for human rights and freedom or how they can adequately prevent the abuse of AI technologies. This implies that the intention is not to compromise state sovereignty, and consequently, the details of implementation and compliance remain almost within national jurisdiction.

Furthermore, the recommendations state that accountability should be given at all levels of an AI system to make people liable if AI operations

cause damage or human rights abuses. Yet, the document does not specify how these responsibilities should be applied or enforced, reflecting a broader issue of accountability in AI governance. Addressing the complexities of AI agency, AI can never be value-neutral, as it is ultimately controlled by humans, which can lead to inherent biases. The concept of defining who an AI actor is becomes essential to avoid discrimination and ensure that the AI systems work as required. The self-learning nature of AI, which allows it to act autonomously, exacerbates this issue, as it lacks a moral conscience, rendering it incapable of moral accountability. The critique offered here pertains to some major issues with the notion of actorship in the domain of AI, specifically the problem of the Westernization of AI regulation and direction. However, it is evident that the majority of AI researchers, including white Americans, are based in Europe and the United States, as per the recommendations in the AI actors' definition. This demographic dominance raises critical questions about the inclusivity and universality of AI governance.

The recommendations assign substantial responsibility to these actors to ensure human rights are respected in the development and application of AI. However, they fall short in delineating who will oversee the overseers, that is, ensuring that these AI actors themselves do not perpetuate human rights violations. This oversight gap could lead to discriminatory outcomes if the data input into AI systems reflects the biases of their predominantly Western creators. Additionally, the recommendations advocate for increased representation from underrepresented countries in the AI discourse, yet they lack a clear mechanism for integrating these new actors into the AI landscape, potentially perpetuating a skewed world order where AI development remains a Western-centric endeavor. The dilemma extends into AI ethics education, as outlined in the recommendations that propose developing educational materials, including online courses, in local languages to accommodate diversity, including those with functional variations (UNESCO, 2021, p. 35). The challenge, as [Young_\(2003\)](#) highlights, lies in the translation process which is inherently non-neutral and subject to cultural contexts. The choice of translators and the cultural

nuances of language translation can significantly influence the local interpretation and acceptance of AI ethics, potentially reinforcing existing power dynamics.

Paragraph 69 of the recommendations expands the definition of actors to include a broad spectrum of societal roles, from civil society representatives to engineers and users. However, the predominance of white individuals within these roles, as noted by Ahmed, suggests a normativity that reflects and perpetuates a Western, white worldview. This homogeneity risks sidelining diverse perspectives and experiences, particularly those from “*non-traditional*” AI regions like island nations or developing countries, which the recommendations themselves identify as focal points for AI education and development.

6.6 RESULTS

The paper scrutinizes the enduring influence of colonial legacies within UNESCO’s (2021) ethical recommendations on AI. It explores the document’s foundations in human rights norms, the values it promotes, and its targeted audience concerning developers and AI actors. The central query of the paper concerns the portrayal and implications of human rights, world order, and agency in the recommendations and identifies problematic aspects therein. The analysis reveals that while human rights are deemed integral throughout the AI lifecycle, the rapid evolution of AI poses risks to human dignity due to the autonomous nature of learning algorithms. The document itself, despite having a general framework, assumes that the measures are universal and do not require specific guidelines, assigning the protection of these rights to the Member States and AI actors. This passive approach leads to a very liberal scope of interpretation for the Member States, which can reduce the impact of the guidelines in various circumstances. The recommendations mirror the current power structures in international relations and do not sufficiently address past injustices.

Incorporating low-income groups and other minorities into AI development lacks a proper strategy, thereby perpetuating the existing

structures of white supremacy. White male AI researchers and developers from the United States promise to achieve this by serving as the primary disseminators of AI knowledge and structures in less developed regions. This dynamic has the potential to establish economic dependencies akin to colonial power, as less technologically developed countries may require assistance from the United States for AI development. In a very broad way, the recommendations describe actors as anyone who participates in any part of the AI life cycle, from data creation to use. The self-learning nature of AI systems makes it challenging to identify these actors and hold them accountable, despite their critical role in protecting human rights. Major companies primarily establish themselves in Western countries, thereby excluding non-Western perspectives and maintaining a structure that adheres to white Western norms.

6.7 CONCLUSION AND SUGGESTIONS

The paper examines UNESCO's ethical recommendations for AI, asserting that, despite their universal intent to safeguard human rights, their practical implementation is fundamentally reactive. This passivity shifts a significant deal of work to member states and AI actors—mainly developers—to guarantee the protection of human rights when using AI at every stage of the process. The recommendations are anchored in the principles of the UN's Universal Declaration of Human Rights, emphasizing that the right to access and use AI should be universal. The paper critiques the document for not actively directing member states on how to implement these guidelines, thereby making them less effective, especially in postcolonial countries where implementing such broad directives can be challenging. The identified world order within the recommendations suggests that the most technologically advanced countries, predominantly Western, have a responsibility to assist less developed nations in AI management and development. This notion perpetuates a Western-centric approach to AI ethics and development, potentially sidelining other cultural perspectives in the global AI discourse.

Furthermore, the paper problematizes the inclusivity of the “actors” defined by the recommendations. While the document broadens the definition of actors to include anyone involved in the AI lifecycle, from creation to application, it assigns them a significant role in promoting human rights and peaceful societies. Yet, the lack of specific accountability measures for these actors is seen as a deficiency, potentially allowing the prevailing “white, Western” world order to dominate global AI practices. This orientation reflects a Western normativity in both development and education about AI, which may not resonate with or include all global actors, particularly those from non-Western backgrounds. The paper calls for further research into the practical application of these ethical recommendations globally, to better understand their effectiveness and inclusivity in safeguarding human rights in the rapidly evolving domain of AI. This further research could provide insights into how different regions interpret and implement the recommendations and whether they are effective in preventing the exploitation and discrimination of vulnerable groups.

REFERENCES

- Ahmed, F. (2022). Ethical aspects of artificial intelligence in banking. *Journal of Research in Economics and Finance Management*, 1, 55–63. <https://doi.org/10.56596/jrefm.v1i2.7>
- Ahmed, S. (2007). A phenomenology of whiteness. *Feminist Theory*, 8(2), 149–168. <https://doi.org/10.1177/1464700107078139>
- Aizenberg, E., & van den Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society*, 7(2), 2053951720949566. <https://doi.org/10.1177/2053951720949566>
- Almpani, S., Stefaneas, P., & Frangos, P. (2023). Formalization of ethical decision making: Implementation in the data privacy of wearable robots. *International Journal of Extreme Automation and Connectivity in*

- Anderson, J., Reimer Kirkham, S., Browne, A., & Lynam, M. (2007). Continuing the dialogue: Postcolonial feminist scholarship and Bourdieu—discourses of culture and points of connection. *Nursing Inquiry*, 14(3), 178–188. <https://doi.org/10.1111/j.1440-1800.2007.00367.x>
- Bærøe, K., Miyata-Sturm, A., & Henden, E. (2020). How to achieve trustworthy artificial intelligence for health. *Bulletin of the World Health Organization*, 98(4), 257–262. <https://doi.org/10.2471/BLT.19.237289>
- Bryndin, E. (2020). Implementation of competencies by smart ethical artificial intelligence in different environments. *Software Engineering*, 8, 24. <https://doi.org/10.11648/j.se.20200804.11>
- Burton, E., Goldsmith, J., Koenig, S., Kuipers, B., Mattei, N., & Walsh, T. (2017). Ethical considerations in artificial intelligence courses. *AI Magazine*, 38(2), Article 2. <https://doi.org/10.1609/aimag.v38i2.2731>
- Butler, J. (2009). *Frames of War: When Is Life Grievable?* Verso.
- Carvalho, L. P., Murakami, L., Suzano, J. A., Oliveira, J., Revoredo, K., & Santoro, F. M. (2022). Ethics: What is the research scenario in the Brazilian Conference BRACIS? *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, 624–635. <https://doi.org/10.5753/eniac.2022.227590>
- Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 376(2133), 20180080. <https://doi.org/10.1098/rsta.2018.0080>
- Čerbová, V. Č. (2021). Artificial intelligence, ethics and catholic church. In *Human Rights—From Reality to the Virtual World*. www.academia.edu/83123902/Artificial_Intelligence_Ethics_and_Catholic_Church
- Cervantes, S., López, S., & Cervantes, J.-A. (2020). Toward ethical cognitive architectures for the development of artificial moral agents.

Cognitive Systems Research, 64, 117–125.
<https://doi.org/10.1016/j.cogsys.2020.08.010>

- Civaner, M. M., Uncu, Y., Bulut, F., Chalil, E. G., & Tatli, A. (2022). Artificial intelligence in medical education: A cross-sectional needs assessment. *BMC Medical Education*, 22(1), 772. <https://doi.org/10.1186/s12909-022-03852-3>
- Cookson, R., & Mirelman, A. J. (2017). Equity in HTA: What doesn't get measured, gets marginalised. *Israel Journal of Health Policy Research*, 6(1), 38. <https://doi.org/10.1186/s13584-017-0162-3>
- Dankwa-Mullan, I., Scheufele, E. L., Matheny, M. E., Quintana, Y., Chapman, W. W., Jackson, G., & South, B. R. (2021). A proposed framework on integrating health equity and racial justice into the artificial intelligence development lifecycle. *Journal of Health Care for the Poor and Underserved*, 32(2), 300–317.
- Dolganova, O. (2021). Improving customer experience with artificial intelligence by adhering to ethical principles. *Business Informatics*, 15, 34–46. <https://doi.org/10.17323/2587-814X.2021.2.34.46>
- Go, J. (2018). Postcolonial possibilities for the sociology of race. *Sociology of Race and Ethnicity*, 4(4), 439–451. <https://doi.org/10.1177/2332649218793982>
- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99–120. <https://doi.org/10.1007/s11023-020-09517-8>
- Hermansyah, M., Najib, A., Farida, A., Sacipto, R., & Rintyarna, B. (2023). Artificial intelligence and ethics: Building an artificial intelligence system that ensures privacy and social justice. *International Journal of Science and Society*, 5, 154–168. <https://doi.org/10.54783/ijssoc.v5i1.644>
- Huang, C., Zhang, Z., Mao, B., & Yao, X. (2023). An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence*, 4(4), 799–819. <https://doi.org/10.1109/TAI.2022.3194503>
- Hunkenschroer, A. L., & Kriebitz, A. (2023). Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring. *AI and Ethics*, 3(1),

- 199–213. <https://doi.org/10.1007/s43681-022-00166-4>
- Iphofen, R., & Kritikos, M. (2019). Regulating artificial intelligence and robotics: Ethics by design in a digital society. *Contemporary Social Science*, 16, 1–15. <https://doi.org/10.1080/21582041.2018.1563803>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Katz, Y. (2020). Artificial whiteness: Politics and ideology in artificial intelligence. In *Artificial Whiteness*. Columbia University Press. <https://doi.org/10.7312/katz19490>
- Leikas, J., Koivisto, R., & Gotcheva, N. (2019). Ethical framework for designing autonomous intelligent systems. *Journal of Open Innovation: Technology, Market, and Complexity*, 5(1), Article 1. <https://doi.org/10.3390/joitmc5010018>
- Loomba, A. (2015). *Colonialism/Postcolonialism*. Routledge.
- Luccioni, A., & Bengio, Y. (2020). On the morality of artificial intelligence [commentary]. *IEEE Technology and Society Magazine*, 39(1), 16–25. <https://doi.org/10.1109/MTS.2020.2967486>
- Malek, Md. A. (2022). Criminal courts' artificial intelligence: The way it reinforces bias and discrimination. *AI and Ethics*, 2, 1–13. <https://doi.org/10.1007/s43681-022-00137-9>
- Matyuk, Y. S. (2022). Ethical and legal aspects of development and implementation of artificial intelligence systems. In *AmurCon 2021: International Scientific Conference, European Proceedings of Social and Behavioural Sciences*. <https://doi.org/10.15405/epsbs.2022.06.76>
- Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33(4), 659–684. <https://doi.org/10.1007/s13347-020-00405-8>
- Namli, E. (2014). *Human Rights as Ethics, Politics, and Law*. Acta Universitatis Upsaliensis. <https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-230689>

- Nirenberg, J. (2014). A review of human rights and information communication technologies: Trends and consequences of use. *International Journal of Applied Management and Technology*. www.academia.edu/47936933/A_Review_of_Human_Rights_and_Information_Communication_Technologies_Trends_and_Consequences_of_Use ↵
- Rajagopal, B. (2021). *Development Theories, Old and New and Their Implications for Human Rights* (pp. 10–20). <https://doi.org/10.4337/9781781005972.00008> ↵
- Ramos, G. (2022). A.I.'s impact on jobs, skills, and the future of work: The UNESCO perspective on key policy issues and the ethical debate. *New England Journal of Public Policy*, 34(1). <https://scholarworks.umb.edu/nejpp/vol34/iss1/3> ↵
- Romdenh-Romluc, K. (2024). Fanon, the body schema, and white solipsism. *Southern Journal of Philosophy*, 62(1), 110–123. <https://doi.org/10.1111/sjp.12556> ↵
- Ruban, D. A. (2022). Analytical review of conjugation of the ethical bases of artificial intelligence implementation and ecologization in corporate governance. *Journal of Applied Economic Research*, 21(2), 390–418. ↵
- Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial intelligence in education: AIED for personalised learning pathways. *Electronic Journal of E-Learning*, 20(5), Article 5. <https://doi.org/10.34190/ejel.20.5.2597> ↵
- UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. United Nations Educational, Scientific and Cultural Organization. www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence
- Vakkuri, V., Kemell, K.-K., Tolvanen, J., Jantunen, M., Halme, E., & Abrahamsson, P. (2022). *How Do Software Companies Deal with Artificial Intelligence Ethics? A Gap Analysis* (pp. 100–109). EASE '22: Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering. <https://doi.org/10.1145/3530019.3530030> ↵

- Valentine, L., D'Alfonso, S., & Lederman, R. (2023). Recommender systems for mental health apps: Advantages and ethical challenges. *AI & Society*, 38(4), 1627–1638. <https://doi.org/10.1007/s00146-021-01322-w>
- van Norren, D. (2022). The ethics of artificial intelligence, UNESCO and the African Ubuntu perspective. *Journal of Information, Communication and Ethics in Society*, 21, 112–128. <https://doi.org/10.1108/JICES-04-2022-0037>
- Vică, C., Voinea, C., & Uszkai, R. (2021). The emperor is naked. *Információs Társadalom*, 21, 83. <https://doi.org/10.22503/inftars.XXI.2021.2.6>
- Wang, Z., & Zhang, Y. (2022). A study of ethics on intelligent nonlinear prediction creative design. *Scientific Programming*, 2022, e8616308. <https://doi.org/10.1155/2022/8616308>
- Young, R. J. C. (2003). *Postcolonialism: A Very Short Introduction*. Oxford University Press.
- Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J., Sellitto, M., Shoham, Y., Clark, J., & Perrault, R. (2021). *The AI Index 2021 Annual Report*. Stanford University.

7 The Role of AI in the Legal System

Implications for Justice and Rights

Geetha Manoharan and M Nivedha

DOI: [10.1201/9781003668657-7](https://doi.org/10.1201/9781003668657-7)

7.1 INTRODUCTION

Artificial intelligence (AI) is the term for technology designed to do activities that often call for human intelligence. It centers on developing intelligent applications that can mimic cognitive processes such as learning, reasoning, problem-solving, and perception that are similar to those of humans. To analyze information and make wise judgments, AI systems use data and algorithms ([Manoharan, Ashtikar, Rao, Natarajan, et al., 2024](#)). AI may be roughly divided into two types:

1. *Narrow or Weak AI*: These AI systems operate in their assigned domains and are suited for certain jobs; they do not carry out duties outside their areas of expertise. Alexa and Siri are two examples of voice assistants.
2. *General or Strong AI*: This is a higher-level AI that can potentially perform any mental work that a human can. Research and discussion on strong AI are still underway.

According to the essay by [Ferdousi \(2024\)](#), there are many issues facing the international community in academic society today. Professionals and

academics now have new options due to technological advancements like AI which have raised concerns about plagiarism, cheating, and fabrication, which have also made it more difficult to uphold academic integrity. Researchers and academics must ensure that their use of AI aligns with ethical norms and academic integrity. AI is a set of moral standards and ideals that guide behavior in academic and research settings. We are still figuring out how the industry will change in this specific environment, what safety precautions will be required, and how everyone-including the governments-needs to get ready. Examining AI's impact on legal studies and the legal profession in the context of ethical standards is the aim of this research study. We want to find out what the pros and cons of using AI in legal study might be, as well as how it might affect lawyers, customers, and the legal system as a whole. Effectively sorting through huge amounts of data, AI systems can cut costs and time in half while retaining all important information. AI is also capable of analyzing case laws and precedents, providing historical information to support lawyers in developing litigation strategies and making well-informed decisions. Legal professionals are using the AI facility for a variety of purposes, including conducting legal research; summarizing legal narratives; reviewing legal documents; drafting/templating legal contracts; conducting due diligence; reviewing discovery; and drafting/templating communications (such as memos, emails, correspondence to opposing counsel, etc.). Furthermore, AI helps with contract management by automatically extracting important words, clauses, and conditions from contracts; detecting possible hazards; guaranteeing compliance; and creating customized contracts based on specified inputs. AI improves the efficiency of procedures and processes, which may occasionally be quite drawn out and repetitive, by automating repetitive and time-consuming jobs. Legal professionals are using the AI facility for a variety of purposes, including conducting legal research; summarizing legal narratives; reviewing legal documents; drafting/templating legal contracts; conducting due diligence; reviewing discovery; and drafting/templating communications (such as memos, emails, correspondence to opposing counsel, etc.).

Robots are slowly taking over jobs that used to be done by people because they can do many tasks faster and better. Automation poses a danger to routine and administrative work, but new research indicates that it might also affect sectors that demand more nuanced human labor. An excellent example of this is the legal profession, where developing arguments based on facts and, of course, critical thinking are essential activities that entail outlining a legal strategy. After all, robot attorneys with AI capabilities have long been fashionable, particularly in recent years.

AI will also change how legal services are provided at the practice level. AI automation ([Deviprasad et al., 2023](#)) and augmentation at scale will revolutionize law practices that often handle similar sorts of legal work. AI will significantly reduce the inefficiencies and redundancies at the heart of these jobs, from preparing briefs to examining evidence and developing legal strategy, freeing up lawyers to concentrate on the strategic work that best fits their positions. Opportunities to grow, or raise income faster than costs, will now be possible thanks to this adjustment. AI will also have an impact on legal research. Several law schools and instructors aggressively integrate AI ethics into the curriculum, despite some academics' worries that AI for law students encourages cheating ([Unver, 2023](#)). For example, some instructors let or even mandate that students experiment with generative AI tools such as ChatGPT; however, they usually stipulate that the responsibility for any wrong information produced by the technology rests with the student if the student fails to fix it ([Yu, 2023](#)). [Jaiswal \(2024\)](#) notes that academics recognize the continuing influence of (AI on the legal field and that prospective attorneys must be conversant in the technology and aware of its appropriate and inappropriate applications.

[Fried \(2024\)](#) reports that although attorneys have always needed to be able to tackle disorganized, complex situations, this ability is becoming increasingly crucial. On the one hand, a lot of the other work that was formerly thought to constitute the foundation of the professional practice is being encroached upon by technology and alternative legal service providers. At the same time, there is a growing feeling of urgency to solve the access to justice gap and wider structural injustice, leading to an

increase in ill-structured problems due to the unique legal concerns brought by new technologies. Law schools should instruct students in a process I refer to as “unstructured,” which involves seeking a deeper understanding of the specifics and context of an ostensibly well-defined problem to find new ways to view it. This will prepare students for this type of professional practice.

7.2 AI IN THE LEGAL PROFESSION

Nearly all professions and industries are being disrupted by AI, but some more quickly and deeply than others. By replacing muscles with hydraulic pistons and diesel engines, the Industrial Revolution mechanized physical labor; the AI-powered revolution, on the other hand, automates cerebral workloads. While certain blue-collar occupations may only be optimized by AI, many white-collar jobs that were previously assumed to be immune to automation are seeing a more profound transformation. According to [Savira et al. \(2023\)](#), AI is enabling professionals in certain fields to perform tasks that were previously unattainable, therefore complementing and even partially substituting their human counterparts in professional settings. This is one way that AI is significantly changing the legal profession. Shortly, AI will probably assist rather than replace lawyers, but it is already being used for contract reviews, legal research, and locating pertinent documents during the discovery phase ([Nirmala et al., 2024](#)). Recently, AI has started to be utilized in contract drafting, case prediction, and even court recommendations about bail and sentences.

AI has the potential to be beneficial to law. Legal productivity can rise, and expensive errors can be prevented. Research and decision-making may move more quickly when it greases the wheels of justice. In the legal profession, AI is still not ready to fully replace human judgment. Before using AI in some legal situations, it is necessary to address the danger of bias in the data that powers the technology ([Singh et al., 2024](#)) and the incapacity to fully understand the reasoning behind choices made by AI in a form that is intelligible to humans (i.e., explainability). Every area of

business is impacted by the law. Companies use legally binding contracts for almost everything they do, including partnerships, sales, purchases, mergers, and reorganizations. If there were no robust intellectual property laws, innovation would come to a complete stop. All of us function against the backdrop of our legal system and the constant threat of lawsuits, whether we realize it or not.

In the upcoming years, expect this to alter. AI, more than any previous technology, is going to drastically change the legal profession. Yes, this procedure has already started. The implementation of AI and machine learning (ML) ([Tripathi et al., 2023](#)) is greatly aided by the law in many respects. Law and ML work on remarkably similar premises: they both use precedent to conclude how rules should be applied in different contexts. Legal theory may be the most similar to a formal logic system among the social sciences. Legal decisions, to put it too simply, are based on a series of presumptions drawn from precedent, which are then applied to the specific circumstances at hand, and a conclusion is drawn from them. Machine intelligence may be effectively used for precisely this kind of task, which follows a logic-oriented technique. When it comes to the legal profession using AI, contract review, contract analytics, litigation prediction, and legal research are the four most crucial areas.

Recent research by [Bishnoi et al. \(2024\)](#) has focused a great deal of emphasis on AI and how it intersects with the law. However, it is crucial to explore the true nature of AI and how it applies to legal management and application. About AI and its legal ramifications, this article seeks to give a thorough yet approachable introduction. For readers who are not technical, it aims to be both sophisticated and clear. Before diving into AI's particular uses in the legal field, for people and companies navigating legal systems, and for government agencies in charge of law enforcement, the conversation first discusses AI in a more generalized sense. This sets it apart from talks that focus mostly on hypothetical future consequences of AI in law. Such theoretical discussions are valuable, but it is important to recognize that they frequently rest on assumptions about the direction that technology will go, some of which may not be supported. The current legal

and policy issues raised by AI technology are sometimes overlooked in these debates about the future.

The field of AI development has significantly changed a new age in digital technology, social economics, human needs, and professional conduct, according to a study by [Masithoh et al. \(2023\)](#). Over time, further facets of the legal profession will be automated by AI. AI further enables the legal industry to automate routine operations, freeing up time for in-depth research and client interactions. The purpose of the study is to determine how literate legal professionals are with AI. Normative–empirical legal research approaches were used in this work. In addition, the research uses secondary data—that is, information from books, journals, and other relevant legal sources—as well as primary data, which is gathered through questionnaire surveys. In the meantime, descriptive qualitative analysis will be used in the research. Studies indicate that law enforcement personnel have a medium degree of literacy when it comes to using AI. Statistics reveal that 75% of legal professionals have favorable things to say about the use of AI in their line of work. In contrast, other data indicates an 8% high percentage and a 17% low percentage.

7.3 AI RESHAPING LEGAL PROFESSION

Although AI has been around for decades, some legal practitioners are unsure about whether they may use AI-powered technologies in their legal practices in an ethical manner, given the current developments in the field. Legal professionals who do not understand and utilize AI may soon be viewed as lacking in legal representation, even though some may be wary about AI's application in law and legal practice, as well as the possible hazards involved in doing so. While AI can improve productivity, streamline processes, and help attorneys plan, there are also substantial hazards associated with it. For instance, AI may assist attorneys in generating first drafts of documents that used to take a lot of time, such as contracts, settlement agreements, motions, and legal briefs. Yet, the application of AI in law also presents problems with algorithmic bias (bias

in the algorithm itself, or bias in the data used to train an otherwise unbiased algorithm), hallucinations (a phenomenon whereby AI chatbots may confidently respond with false information), inaccuracies, and confidentiality issues.

This piece goes into great detail about the current state of AI in the legal field. It talks about the tools and technologies that lawyers can use right now, the pros and cons of using AI to find legal answers, and the future of AI. Continue reading to learn more about what AI means for the legal field and how lawyers can use it to make their work easier and more efficient in a way that is both fair and smart. A study by [Mahadik \(2024\)](#) says that the law business is one of the biggest in the world.

The legal industry is expected to have less of an electronic footprint. While adopting new technology, the legal profession is hesitant and approaches its job in a rather conventional manner. India's judicial system might go through major advancements and improvements due to AI. There are several areas where AI can be useful, even if technology is still in its infancy within the Indian legal system. Lawyers will benefit from increased productivity, cost savings, and a more strategic approach to their profession as AI advances. AI technology can improve legal research and analysis, speed up procedures, boost access to justice, and improve the practice of law generally, despite concerns that it could put lawyers in risk. The advantages of using AI technology in the legal industry are examined in this study, along with the perspectives of legal practitioners on this use of technology. The 106 attorneys in the City of Mumbai who provided primary data for the study were the study's participants. Using the Mann–Whitney test, the Kruskal–Wallis test, percentages, and basic frequency, the analysis was completed using SPSS software. The results indicated that there is a substantial correlation between age and the view of attorneys about AI, but not between gender and this concept. Based on research by [Farrukh et al. \(2024\)](#), lawyers should not ignore the impact that AI will have on their field. First, this piece looks at how far AI research and use have come so far to determine how it will affect the law field. AI is thought of as a tool that could make people smarter and help lawyers do their jobs better. People

often give high-quality work to AI, like answers to legal questions or work on legal tasks. Despite this, this study shows that AI systems do not always get research, analysis, and writing right. The study also looks at the different factors and situations that affect how well AI systems work and makes the case for why law enforcement, human control, and moral standards are needed to make sure that justice is done fairly and correctly.

- *Superhuman Lawyers:* To review contracts more quickly and accurately, lawyers are already using AI, especially ML. They can find mistakes and problems that human lawyers would have missed this way. Discovery, the process by which relevant papers from a suspect in a case are found, has used algorithms for a long time. In this effort these days, ML is also being used.

Trying to come up with a list of every possible method a topic might be mentioned or portrayed is one of the difficulties in seeking and finding every pertinent document. However, some papers are shielded from public inspection, and the court or counsel may try to restrict the search's reach to avoid burdening the producing party excessively. By utilizing both supervised and unsupervised learning, ML threads this needle ([Lourens et al., 2023](#)). Legal research is another field in which AI is now heavily utilized in the practice of law. Due to AI's seamless integration into many research services, practicing attorneys might not even be aware that they are utilizing it in this capacity. Semantic search has added functionality to the keyword or Boolean search strategy that was the service's defining feature for many years. This indicates that rather than merely matching words to keywords, the ML algorithms are attempting to comprehend the meaning of the phrases. Even the instances when a referenced decision has been subtly reversed can be identified by Quick Check.

- *Automated Legal Scholars:* [Šavelka and Ashley \(2023\)](#) hope that the release of ChatGPT has made people more aware of the broad range of uses for large language models (LLMs), including creating documents, asking questions, and summarizing, even in the law field.

Although recent studies have shown how well the technology works in several semantic annotation jobs involving legal texts, more research needs to be done on newer, more powerful (GPT-4) or cheaper (GPT-3.5-turbo) models. That is what this work is about recent improvements in LLMs' ability to properly annotate legal texts in zero-shot learning settings. As generative AI systems become more developed, we compare the performance of GPT-4 and GPT-3.5-turbo(-16k) models to the previous generation of GPT models on three legal text annotation tasks involving different types of documents, such as court decisions, contract terms, and legal provisions. To further grasp the trade-offs, we also examine the cost and performance of the models. Our findings indicate that, on two of the three tests, the GPT-4 model performs much better than the GPT-3.5 models. While the text-davinci-003 model costs 20× more, its performance is equal to that of the more affordable GPT-3.5-turbo. As the batch size grows, speed suffers even if one may annotate numerous data points in a single prompt. For several real-world uses (like contract reviews) and academic endeavors (like empirical legal studies), this study offers pertinent insightful information. These results might be used as a reference for legal scholars and practicing attorneys when making judgments on how to include LLMs in various processes that involve contextual annotation of legal texts. Both creation and analysis of material are possible with AI. Generative AI does not have to be flawless all the time, unlike AI used in self-driving cars, where errors can be lethal. The intriguing thing about AI-generated art is, in reality, the strange and unexpected objects that are connected to it. Because AI views creativity fundamentally differently from humans, it occasionally produces unexpected results or takes unexpected paths. It is known as “emergent behavior” in the context of AI. New techniques to win games, the discovery of pharmaceuticals, or even just creative ways to communicate ideas are examples of emergent behavior. Writing material still requires human authors to oversee the creative process

and decide which of the numerous AI-generated sentences or variants to use. Whether a token is audio or text, the generative model GPT-3 can anticipate what will come next in a sequence. As a transformer, GPT-3 adds context to data sequences, such as sentences, by emphasizing the most pertinent parts. This allows the work to be expanded in a way that is consistent, logical, and natural. Using around 200 billion parameters and half a trillion words to train, GPT-3 is a massive pre-trained model, which sets it apart from other models. Several generating text tools, some based on GPT-3, are already available for usage in the fields of journalism and creative writing, where this method has already been applied. You can anticipate a tale, essay, or report from an AI writer given a brief prompt, but do not expect flawless results.

Occasionally, the AI tool will discuss arbitrary subjects or concepts; also, because AI is devoid of human experience, it could contain odd allusions or factual errors. In Russia, the media environment is constantly changing due to complex political and regulatory frameworks, rapid technological progress, and ensuing societal changes. Within the journalistic community, these modifications upend long-standing roles, conventions, attitudes, and practices. Within the journalistic community, these modifications upend long-standing roles, conventions, attitudes, and practices. Beyond the reasons listed above, the slow infiltration of AI technology into routine journalistic operations is another aspect that has sparked the conversation about professional journalism standards. Research on the professional identity of journalists, the limits of the profession, and the evolution of professional standards was carried out by the author against this backdrop. It was based on eleven semi-structured interviews with representatives of the professional community. Notwithstanding significant differences in opinions, the survey revealed that professionals stress the critical significance of precisely stated “rules of engagement” that are understandable to the media community and its readership. Although the fundamental principles

are essentially constant, practical considerations like the standards for formatting and appropriate journalistic conduct show flexibility in response to changing socio-political, economic, and cultural environments. In contrast, practical and technological standards are more adaptable and sensitive to changing social norms. This is a noteworthy observation on values that are shared by traditional and new media spaces. The study conducted by [Litvinova \(2024\)](#) indicates that the particular standards that editorial teams adopt ultimately depend on the media format and niche that they have within the business.

- *Drafting Legal Contracts:* AI will require legal law training before it can create contracts, for example. To do this, the AI's developer must "label," or gather, legal performance data on different contract wording variations. The AI is then trained on how to create a suitable contract using this labeled data. Nonetheless, a contract's legal execution is frequently context-specific, contingent upon the jurisdiction, and subject to a constantly evolving body of law. Furthermore, as most contracts are never used in court, their terms are untested and kept confidential between the parties. AI generative systems that are trained on contracts may inadvertently magnify both excellent and poor legal work. These factors make it doubtful that AI contract writers will improve significantly very soon. AI technologies are just not precise enough in language or have the necessary subject knowledge to be left on their own. Although these tools can help draft language, the result still has to be reviewed by human specialists before being put to use.
- *Judge Bots:* Predicting legal outcomes is another innovative application of AI. Determining the probability of a lawsuit's victory accurately might be quite beneficial. It gives a lawyer the ability to choose whether to take a case on a contingency fee basis, how much to spend on experts, and whether to counsel clients to settle. The application of AI to counsel judges on bail and punishment is a more worrisome practice. Correctional Offender Management Profiling for

Alternative Sanctions (COMPAS) is one such application. Criminal courts in numerous jurisdictions use COMPAS and related AI techniques to determine a defendant's or convicted person's likelihood of recidivism when making judgments about pre-trial detention, punishment, or early release. The accuracy and impartiality of these methods are hotly contested. Nevertheless, applying AI in this situation could highlight or even reflect the innate prejudice present in the criminal justice system's data. The real arrest and conviction rates that were utilized to train the ML models may have been biased against particular demographics. As a result, it may perpetuate historical injustices or, worse, disingenuously mask them in the mask of computer-generated neutrality.

7.4 AI IS READY TO PRACTICE LAW

There are a lot of questions that AI brings up in the law field:

- Will it ever be considered malpractice to not use AI in some parts of the law, like discoveries? It might be a professional duty to use AI in discovery if, for example, not using it slows down the process or leads to incomplete revelations.
- If AI tools can help a criminal's case, should they be allowed to be used by them?
- Should lawyers say when they use AI in a case? Should they share the training data or other information that was used to set up the ML models if that is the case?
- If ML models are not completely clear, does that mean they should not be used in some legal situations?
- What can we do to make sure there is no discrimination based on sex or race?

AI is a useful tool in the pursuit of justice as it may expedite the legal process and make it more error-avoidant. In addition to improving

predictability and avoiding moral hazard, AI could be a more effective means of resolving civil claims. The employment of AI to substitute human judgment—particularly in the context of criminal law—becomes more difficult. For several reasons, AI is not prepared for this. In the training data, bias might be present, and the resulting ML models will only serve to reinforce and institutionalize it. Eliminating bias from our training data could help us identify and address some of the systemic racism and misogyny that exists in our judicial system.

The lack of transparency and explainability when utilizing AI is another issue with the current method. At least not yet! It is not possible to cross-examine a deep learning AI. AI is a reflection of humanity; it exposes some of our faults. Understanding the truth and limitations of human justifications for actions may improve as a result of unraveling the reasoning behind an AI's advice. However, maybe more significantly, it may just be intolerable to let algorithms make judgments that violate people's rights.

As with the science fiction film *Minority Report* ([Berry 2024](#)), it is not implausible that ML algorithms will start to confidently anticipate when an individual is likely to commit a crime in the future. Judges, attorneys, and society at large could place excessive faith in these algorithms, which is another strong argument against using AI in criminal cases. Automation bias is the term describing the tendency for people to become too dependent on suggestions from technology, even in situations where humans still have the final say over decisions ([Manoharan, Durai, et al., 2024](#)). As the true capabilities of AI technologies could not be as “intelligent” as they seem, this mistaken confidence could be particularly problematic.

7.5 BENEFITS OF AI IN THE LEGAL PROFESS

1. *Due Diligence Review*: According to [Saxena \(2022\)](#), a due diligence study is an audit ([Garg et al., 2023](#)) of an investor, investment, or product to confirm facts, including financial ([Bhavya et al., 2023](#)) and legal records among other documents deemed relevant to analyzing the

potential risks that might hurt the other Party participating in the transaction.

2. *Predicting Outcomes Even for Judicial Use:* In legal analytics, AI may even forecast a judge's potential ruling based on the arguments made. These choices are feasible because computers consider other conventional criteria as well, such as the quality of the evidence, comparable past cases that the decision is likely to be based on, and even the judge's past decision-making history. Apart from analysis, these robots can provide their suggestions on wording and previous instances that attorneys can use as a foundation for their argument. Recommendations, however, may also be utilized in court to inform decisions. A man found guilty in 2015 received a 6-year sentence from the Wisconsin Supreme Court based on a software suggestion. In the meanwhile, others are trying to devise a system that would eliminate the need for judicial intervention.

3. *Contract Formation, Review, and Management:* AI systems can write contracts. The lawyer and client can both use the same platform to build up contracts. User-service contract drafting is available on the platform. The contract's salient provisions can be selected by the parties after deciding on the kind of agreement that is required. The automated contract-making process uses legal criteria as input to establish norms. In some situations, legal firms can tailor their engagement. In cases where a client requests a nonstandard agreement, lawyers have the option to exclusively examine and handle contracts of a certain type. Not only can AI draft contracts, but it can also verify that the terms are compliant with existing agreements and within the legal bounds of the law. Furthermore, when every contract has been finished and approved, it is important to store these records and add any pertinent additional information, such as the date of the notice of renewal or the contract's termination. Those were the days when a spreadsheet had all these facts. The approach has changed with AI to the point that, aside from program setup and frequent fine-tuning, the work is now accomplished without the need for human interaction.

4. *Consultancy Services through Legal Bots:* AI bots are web-based instruments created to communicate with a user base and offer personalized answers to known scenarios. Bots assist clients, both present and potential, in law firms with legal matters according to their facts and situations. Additional organizations ([Manoharan, Durai, et al., 2023](#)) are creating pro bono legal bots to help those who would not otherwise be able to access the legal system.

Rather than eliminating the need for human resources, it appears that AI takes on laborious duties that are nevertheless crucial to the development of a lawyer's campaign. To free up lawyers' time for other, more productive activities, their time is being optimized. This 24/7 work yields extremely accurate outcomes while reducing operating costs through job reduction.

5. *The Verdict:* However, the paper did point out that certain jobs—such as those involving physical labor in “predictable environments,” data collection and analysis, etc.—are more vulnerable than others. In this instance, those who perform the majority of the research work are not on the safe side: paralegals, junior attorneys, and other support staff members. Furthermore, the sudden appearance of these technological developments makes the idea of computers taking over critical-thinking occupations plausible. Therefore, all of them present a strong case for an adoption strategy that decision-makers in industry and government should consider. Alongside addressing job issues, this policy should embrace the benefits that automation may provide. As history has shown, those who do not use technology to advance will be left behind.

Nonetheless, governments everywhere are motivated to pursue automation in all sectors of the economy to increase productivity and meet their development objectives. They will have to deal with labor deployment, though, which is emerging as one of the most significant societal issues brought on by automation. Ultimately, the ability of an economy to expand is largely dependent on the productivity of its labor force. To reach a consensus on the needs, the government might initiate dialogue with

relevant parties on the matter. [Murphy \(2023\)](#) states that the public and private sectors may collaborate to guarantee the execution of these measures. Lawmakers, especially those who observe a shortage of candidates in fields like data science and business translation, might also create legislation and campaigns to assist in educating and training individuals to develop skills appropriate for certain automated professions.

Naturally, economies that resist these contemporary changes will suffer a slowdown or perhaps reverse in the growth of jobs. In addition, attorneys might not be immune to the dreaded human labor shortage.

7.6 CHALLENGES OF AI

According to [Herdiyanti et al. \(2023\)](#), there is a lot of discussion about the ethical issues of legal practice in the digital age as information technology advances. This article examines social media's effects, data privacy and security issues, and AI's place in legal decision-making. This article explores the consequences of using technology, especially AI, in legal practice and emphasizes the significance of taking ethical principles into account. It does this through a review of the literature and a combined qualitative and quantitative research technique. According to [Singh et al. \(2023\)](#), AI is redefining the legal profession through process automation, enhanced decision-making, and changes to the boundaries of legal practice. This research highlights the revolutionary potential and ethical conundrums of AI as it explores the legal system's numerous implications. This article explores the level of acceptance of AI technology among legal professionals, investigates privacy concerns brought up by networked AI bots, and speculates on the possibility of the creation of autonomous AI attorneys. The study paper highlights the need for just legislation, moral AI development, and the preservation of core legal norms in light of technological advancements.

7.6.1 PRIVACY AND DATA PROTECTION

It is very important to keep private data safe when AI systems are used in the law area. So that client information and case-related data stay private, lawyers who use AI must be very careful with security and privacy. The data of AI systems can still be stolen and used for bad things. Laws need to be able to look over AI programs that watch and collect data. While these systems make choices, they need to think about things like national security and other factors as well. Attorneys must handle worries about data ownership and consent to make sure that people keep control of their data and give their permission for it to be used. Legal professionals are very helpful for AI writers who want to make sure that their apps follow the law.

7.6.2 ETHICAL CHALLENGES

There are a lot of moral questions that come up when AI is used in the law field. One of the biggest is how biased AI programs can be. Bias can happen for many reasons, including skewed training data, bad algorithms, biased human code that is not meant to be biased, and systems that cannot be trusted. When AI systems are taught on biased past data, they may incorporate these biases, which could lead to unfair legal outcomes and bad decisions. To make sure that AI-powered court systems are fair, bias must be dealt with, and AI systems must be held accountable for what they do.

7.6.3 LIABILITY

Determining who is responsible when AI systems make bad law decisions is still an important issue. The legal system and groups that make laws need to be clearer about who is responsible for liability: the user, the person who made the AI, or both. The implication is that the people who made the AI hardware or software are not responsible for any harm as long as it was made without any bugs. Having said that, the owner and/or licensee may be responsible if AI production goes wrong or if the licensee changes AI in a way that causes harm.

7.6.4 INTELLECTUAL PROPERTY RIGHTS

Intellectual property rights (IPRs) issues must be thought about when AI makes works, innovations, or materials. This is one of the most important issues in AI and IPR: who owns the things that AI makes? Who owns the rights to anything made by AI is one of the most important issues. Someone who made the work usually has the copyright. People or organizations are often in charge of AI-made content.

On the other hand, there is not a clear legal framework on this matter. The uniqueness of the content produced by AI is another problem. Works must frequently be unique and the product of human imagination to be protected by copyright laws. In the realm of AI, there is an ongoing discussion on what uniqueness looks like. It would be fascinating to know what rulings courts in other countries have made on this matter. For an invention produced by AI to be eligible for patent protection, it must satisfy three requirements: novelty, non-obviousness, and usefulness. Similar to copyright, ownership is a problem in patent law as well. It can be difficult to ascertain who is the rightful owner of an innovation created by AI, particularly. Will AI change the legal industry? We contend that the research by [Kapoor et al. \(2024\)](#) does not support this assertion with the available data. We explore the growing roles of AI in three categories of legal jobs: information processing, tasks requiring creativity, reasoning, or judgment, and future prediction. Based on the degree of difficulty in determining the right response and the observability of information pertinent to the work at hand, we discover that the degree of difficulty in evaluating legal applications differs significantly throughout legal tasks. Because they are more difficult to assess, tasks that would have the most impact on the legal profession are also the ones that are most vulnerable to overconfidence about AI's capabilities. We offer suggestions for improved AI assessment and use in legal settings.

7.7 DISCUSSION ON HOW CAN LAWYER AI BENEFIT THE FIRM AND THE CLIENT

AI makes it easier for lawyers to do their jobs. More generally, AI cuts down on the time needed to do things by hand, making more time available for activities that build relationships and focus on clients.

Law companies can get a lot of benefits for their customers and their pocketbooks, including:

- *Increase Productivity:* The company as a whole gains efficiency by using AI to generate ideas and automate repetitive manual operations. To increase productivity, labor-intensive and time-consuming tasks such as finding a contract, carrying out due diligence, or generating an invoice are eliminated by AI-driven procedures. Attorneys can spend more time with their clients and billable hours when they increase their efficiency.
- *Improve Access to Justice:* Justice-related obstacles, most notably the high expense of obtaining legal counsel, may be lessened by AI and ML. Lawyers can lower estimates and prices for their clients by eliminating time-consuming, manual legal labor. If research that took twenty hours to complete now only takes two, for instance, attorneys might pass those savings forward to clients. Furthermore, attorneys might help more clients with the time saved from completing tiresome research. The potential for applying AI in the legal sector is considerable, even though the benefits are not yet completely realized.
- *Provide a Better Client-Centered Experience:* AI in law companies offers a single, primary benefit that frees up more time for attorneys and other legal professionals. Lawyers may now spend more time building meaningful relationships with their clients by utilizing AI-driven solutions that automate tasks and save time. Prosecutors should be able to assist their clients with more than simply legal matters. With additional time on their hands, attorneys may better get to know their clients and comprehend the circumstances surrounding their legal needs. Your reputation will precede you if you develop into a reliable adviser who spends the necessary time getting to know

your customers and providing prompt, effective service. More referrals and improved online reviews will result in gaining your clients' confidence and trust. Your legal practice may eventually see an increase in clients and income as a result of this strategy.

7.8 CONCLUSION

With its potential to completely change the practice of law, AI is starting to find its way into India's legal system. AI is meant to supplement attorneys, not replace them; hence, cooperation between legal experts and AI systems is essential to using AI in the legal sector. Intellectual property rules face new problems from AI, which means that existing frameworks must change to handle things like ownership of material created by AI. Given the importance of legal ethics in the context of AI, law schools must provide hands-on instruction in the use of AI tools for legal research, contract analysis, document review, due diligence, and other purposes. To achieve the goals of justice, AI has the potential to improve the effectiveness and accessibility of legal services. The legal community must lead this change in the age of AI while maintaining the values of justice, accountability, and openness that guide the legal profession. Typical corporate functions include talent, HR, finance ([Almahairah et al., 2023](#)), accounting, marketing, sales, customer success, and legal. Corporate software businesses ([Manoharan, Razak, et al., 2024](#); [Manoharan, Nithya, et al., 2024](#)) worth billions of dollars have enhanced procedures and efficiency in almost all of these areas throughout the previous 20 years. LinkedIn, Gusto, Workday, NetSuite, HubSpot, Salesforce, and Zendesk are a few examples. It is lawful to provide an unambiguous exception. Little innovation has recently occurred in the legal sector as entrepreneurs disregard it and view it as a cost center. Microsoft Word and email are still widely used in law firms. Given the magnitude of the legal market, this presents an enormous opportunity for value creation. There will be a lot of opportunities to update and change the legislation because of AI and natural language processing.

REFERENCES

- Almahairah, M. S. Z., Goswami, S., Karri, P. N., Krishna, I. M., Aarif, M., & Manoharan, G. (2023, December). Application of Internet of Things and Big Data in Improving Supply Chain Financial Risk Management System. In *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)* (Vol. 10, pp. 276–280). IEEE. [↗](#)
- Beebeejaun, A., & Gunpath, R. (2023). A study of the influence of artificial intelligence and its challenges: The impact on employees of the legal sector of Mauritius. *Global Business Review*, 1, 45–47.
- Berry, S. (2024). AI and the legal profession: Transforming the future of law. Edited by Alex Davies 2023, published by Globe Law and Business. *Legal Information Management*, 24, 61–63. [↗](#)
- Bhavya, K. R., Dharmaraj, A., Pareek, P., Sathe, M. A., Tiwari, M., & Manoharan, G. (2023, September). Application of Soft Computing Techniques for Predictive Analytics in Financial Markets. In *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)* (Vol. 6, pp. 1756–1760). IEEE. [↗](#)
- Bishnoi, R., Kaur, R., Kaur, N., Kaur, H., Kaur, A., & Johal, A. (2024). *Utilizing Artificial Intelligence in Legal System with Special Reference of India. Educational Administration: Theory and Practice*, 30(5). [↗](#)
- Deviprasad, S., Madhumithaa, N., Vikas, I. W., Yadav, A., & Manoharan, G. (2023). The Machine Learning-Based Task Automation Framework for Human Resource Management in MNC Companies. *Engineering Proceedings*, 59(1), 63. [↗](#)
- Farrukh, T., Qureshi, F.N., & Abbasi, S. (2024). Artificial Intelligence (AI) in Legal System. *Journal of Independent Studies and Research Computing*, 1, 12–15. [↗](#)
- Ferdousi, N. (2024). Artificial Intelligence in Legal Research and Practice: An Exploration of Ethics and Policy. *Bangladesh Journal of Bioethics*, 1, 70–71. [↗](#)

- Fried, A. (2024). Unstructuring for Insight: The Legal Profession in an Age of AI and Social Change. *The Law Teacher*, 58, 74–91. [↵](#)
- Garg, S., Theth, H. P., Saxena, A., Manoharan, G., Al-Chilibi, H., & Alazzam, M. B. (2023, December). The Connection among Informatics and Audit Standards is Mediated by Organizational Culture. In *2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG)* (pp. 1–5). IEEE. [↵](#)
- Herdiyanti, S.H., Kurniati, H.Y., & Ras, H.H. (2023). Ethical Challenges in the Practice of the Legal Profession in the Digital Era. *Formosa Journal of Social Sciences (FJSS)*, 1, 12–15. [↵](#)
- Jaiswal, K. (2024). The Impact of Artificial Intelligence on Legal Practice and Ethics. *International Journal for Research in Applied Science and Engineering Technology*, 1, 23–24. [↵](#)
- Kapoor, S., Henderson, P., & Narayanan, A. (2024). Promises and Pitfalls of Artificial Intelligence for Legal Applications. *arXiv*, abs/2402.01656. [↵](#)
- Litvinova, A.I. (2024). *The Opinion of Russian Journalists on Professional Standards in Traditional and Digital Media*. *Communicology*. [↵](#)
- Lourens, M., Sharma, S., Pulugu, R., Gehlot, A., Manoharan, G., & Kapila, D. (2023, May). Machine Learning-Based Predictive Analytics and Big Data in the Automotive Sector. In *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 1043–1048). IEEE. [↵](#)
- Mahadik, S.S. (2024). A Study on Perception of Lawyers about the Impact of Artificial Intelligence on the Legal Profession. *International Journal for Multidisciplinary Research*, 1, 88–89. [↵](#)
- Manoharan, G., Ashtikar, S. P., Rao, C. G., & Nivedha, M. (2024). A Pedagogical Landscape: Narrative Study on Emerging Artificial Intelligence Trends in Education. In *Inclusive Educational Practices and Technologies for Promoting Sustainability* (pp. 22–41). IGI Global.
- Manoharan, G., Ashtikar, S. P., Rao, C. G., Natarajan, S., & Nivedha, M. (2024). Masterminds and Machines: Harnessing AI in Strategic

- Leadership. In *Impact of New Technology on Next-Generation Leadership* (pp. 107–119). IGI Global. [↗](#)
- Manoharan, G., Durai, S., Ashtikar, S. P., & Kumari, N. (2024). Artificial Intelligence in Marketing Applications. In *Artificial Intelligence for Business* (pp. 40–70). Productivity Press. [↗](#)
- Manoharan, G., Durai, S., Rajesh, G. A., Razak, A., Rao, C. B., & Ashtikar, S. P. (2023). A Study on the Perceptions of Officials on Their Duties and Responsibilities at Various Levels of the Organizational Structure to Accomplish Artificial Intelligence-Based Smart City Implementation. In *Artificial Intelligence and Machine Learning in Smart City Planning* (pp. 1–10). Elsevier. [↗](#)
- Manoharan, G., Nithya, G., Rajchandar, K., Razak, A., Gupta, S., Durai, S., & Ashtikar, S. P. (2024). AI in Finance and Banking: The Act of Gyration. In *Revolutionizing Customer-Centric Banking through ICT* (pp. 1–28). IGI Global. [↗](#)
- Manoharan, G., Razak, A., Rajchandar, K., Nithya, G., Durai, S., & Ashtikar, S. P. (2024). Digital Learning for Professional Development in Varied Fields of Service Sectors: Embracing Technological Advancements. In *Embracing Technological Advancements for Lifelong Learning* (pp. 111–137). IGI Global. [↗](#)
- Masithoh, M.Q., Hakim, H.A., Praja, C.B., & Iswanto, B.T. (2023). *AI in Law: How Artificial Intelligence Is Transforming the Legal Profession in Indonesia*. Justitia Jurnal Hukum. [↗](#)
- Murphy, J. (2023). AI in the Legal Sector—An Overview for Information Professionals. *Legal Information Management*, 23, 150–153. [↗](#)
- Nirmala, O., Rao, C., & Manoharan, G. (2024, June). Imperatives of Artificial Intelligence in Armed Forces. In *AIP Conference Proceedings* (Vol. 2971, No. 1). AIP Publishing. [↗](#)
- Šavelka, J., & Ashley, K.D. (2023). The Unreasonable Effectiveness of Large Language Models in Zero-Shot Semantic Annotation of Legal Texts. *Frontiers in Artificial Intelligence*, 6, 10–11. [↗](#)

- Savira, J., Djaja, B., & Sudirman, M. (2023). *Polemic of Artificial Intelligence (AI) Technology on the Development of Human Resources, Especially in Legal Practitioners*. Edunity Kajian Ilmu Sosial dan Pendidikan.[📄](#)
- Saxena, N. (2022). *Artificial Intelligence in Legal Profession: Pros, Cons and Challenges*. Haridra.[📄](#)
- Singh, A., Virmani, D., Kulkarni, N., Tiwari, M., Manoharan, G., & Alanya-Beltran, J. (2024). The Emerging Role and Application of Adopting Revolutionary Internet of Things (IoT) for Future Technology Sustainability. In *Computer Science Engineering and Emerging Technologies: Proceedings of ICCS 2022* (p. 103).[📄](#)
- Singh, L., Joshi, K.A., Koranga, R.S., Pant, S.C., & Mathur, P. (2023). Artificial Intelligence in the Legal Profession: A review on its Transformative Potential and Ethical Challenges. In *Proceedings of the 5th International Conference on Information Management & Machine Intelligence*. ACM.[📄](#)
- Tripathi, M. A., Tripathi, R., Effendy, F., Manoharan, G., Paul, M. J., & Aarif, M. (2023, January). An In-Depth Analysis of the Role that ML and Big Data Play in Driving Digital Marketing's Paradigm Shift. In *2023 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1–6). IEEE.[📄](#)
- Unver, M. B. (2023). Rebuilding 'Ethics' to Govern AI: How to Re-Set the Boundaries for the Legal Sector? In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*.[📄](#)
- Yu, Y. (2023). Discussion on the Reform of Higher Legal Education in China Based on the Application and Limitation of Artificial Intelligence in Law Represented by ChatGPT. *Journal of Education, Humanities and Social Sciences*, 1, 103–106.[📄](#)

8 AI in Emotional Intelligence and Leadership

Enhancing Accountability in Business Practices

*Gaurav Kulat, Neeta Shah, Kanchan Naidu,
Gayathri Band, and Rupesh Pais*

DOI: [10.1201/9781003668657-8](https://doi.org/10.1201/9781003668657-8)

8.1 INTRODUCTION

In this fiercely competitive business environment, it is important for organizations to understand the various components that impact sales performance in order to thrive ([AlDosiry et al., 2016](#)). Over the past few years, leadership types as well as emotional intelligence (EI) have been identified as having significant influence on sales results. EI, also known as the emotional quotient (EQ), plays a vital role in having excellent relationship management or client interaction by using skills like self-awareness, empathy, and emotional control ([Karatepe, 2010](#)). Likewise, the ways sales tactics and team dynamics are handled with different leadership styles (from transactional to transformational approaches; [Schutte et al., 1998](#)).

Advancements in data analytics and machine learning (ML) have helped provide means to (better) measure these variables and thus their correlation

with revenue generation ([Pearman, 2011](#)). ML algorithms learn from giant datasets in order to analyze, recognize patterns, and predict better; then, by using these methods, businesses can then discover more about the impact of leadership styles and EI traits on sales outcomes ([Mayer et al., 2004](#)).

This study will utilize ML techniques for this void in knowledge and is anticipated to help us understand the relationship between leadership styles, EI, and sales effectiveness ([Lopes et al., 2003](#)). A case in point, it explores how various EI functionalities—like self-regulation and empathy—and leadership methodologies—such as transformational and transactional styles—affect sales achievements and collapses ([Lopes et al., 2003](#)). In an effort to improve this data, sales team data is measured across many different industries and assessed in order to identify what is working for successful sales teams. By including ML techniques along with this investigation, we gain insight into these connections and receive predictive methods to help advise key decision points ([Goleman et al., 2002](#)). Businesses that wish to achieve superior sales performance and a competitive advantage may therefore find this study highly enlightening, paving the way for EI and leadership process ([Haakonstad & Håvold, 2011](#)).

8.2 LITERATURE REVIEW

Recent research also has emphasized the importance of EI in sales effectiveness. It found that EI was a significant predictor of relationship management and stress tolerance in high-pressure sales situations ([Honeycutt Sigler & Shepherd, 2023](#)). According to the study, salespeople with higher-than-average EQs are better able to determine just what it is their customers want.

In subsequent studies Patel and colleagues also showed that there was evidence from such social skills and empathy also fall under the umbrella of EI and are integral for building rapport with customers and in turn earning trust. More sales due to this trust and happy customers ([Law et al., 2004](#)). They validated these findings from a cross-section of businesses using

survey data to show that salespeople with high EI are better at both being communicators and managing client relationships ([Mortan et al., 2014](#)).

Leadership styles and sales success: This is another area that has been heavily researched. Leading sales performance is an area where transformational leadership rises into the spotlight ([Mehlhorn, 2020](#)). This type of leadership consists of those who inspire and encourage the staff. Sales performance is positively related to transformational leadership as well, in that it facilitates an empowering and supportive climate (Judge & Piccolo, 2019, meta-analysis). This method incubates what is needed the most to achieve desired sales—innovation and dedication.

In contrast, sales performance, which is more sluggish and stable, has been linked to styles of transactional leadership that revolve around reward and punishment based on productivity. Transactional leadership might be great at managing everyday sales tasks, but it is not always inspiring and motivating for producing top sales results (Ralston et al., 2020). One study found that how sales managers react to necessary changes is likely a mix of the costs themselves—and provided several varying combinations of transformational or transactional strategy with positive effects in different situations.

The recent studies have shown that by adopting ML methods, sales performance could be better understood. By integrating customer information, sales touchpoints, and evolving market signals, Zhang and colleagues highlight additional ways machine learning can be used to forecast business performance. Their results show that these models can accurately anticipate sales outcomes, helping organizations shape more effective future strategies.

Secondly, machine learning can be applied to predict sales volume by integrating data on emotional intelligence and leadership styles. Combining these human-focused factors with ML techniques provides a richer, more complete understanding of the drivers behind sales success. The researchers found that it was possible to provide an even richer picture of the factors affecting sales performance. This way, companies are able to identify and

analyze the data well enough to better plan their sales into specific target groups and reduce the chance of failure.

These findings are promising, but for my intent to combine EI and leadership styles using ML, there is still work to be done. Data privacy and algorithmic bias are two key ethical issues that need careful attention. To ensure sales performance analysis remains fair and transparent, machine learning must be implemented correctly and responsibly.

Future research should focus on enhancing these ML models to be explainable, focusing on how they can study EI and leadership dynamics' influence on sales success in the long term. Organizations will still have to be mindful of how EI and leadership styles impact new-age tools as they continue to adopt in the new-age technology era and must tread carefully when it comes to optimizing sales tactics and ensuring long-term success.

8.3 OBJECTIVES OF THE STUDY

- To assess the impact of EI on sales performance.
- To examine the influence of leadership styles on sales performance.
- To analyze the interaction between EI and leadership styles.

8.4 HYPOTHESIS OF THE STUDY

H₀ (Null Hypothesis): There is no significant relationship between EI and sales performance. EI does not significantly affect sales outcomes.

H₁ (Alternative Hypothesis): There is a significant positive relationship between EI and sales performance.

8.5 RESEARCH METHODOLOGY

The current study follows a mixed-methods strategy: The researchers use a quantitative approach to investigate how EI predicts sales performance. The

quantitative aspect of the study involves a survey of salespeople at various companies to gather and analyze numerical data.

The survey also encompasses success metrics on sales goal achievement and outcome measures, as well as trusted EI assessments such as the Emotional Quotient Inventory (EQ-i). Statistical approaches such as multiple regression and correlation analysis are used to evaluate the relationship of EI measures with sales performance outcomes.

The qualitative data further enable better understanding of what EI entails while determining success in new sales. The first set of interviews are semi-structured in nature and focus on understanding how EI of salespeople and managers impacts the way they sell to the customers and make their sales initiative a success. Interview data is then categorized, and the commonalities/trends are identified using thematic analysis.

And, if you want to get even more technical, we plug in ML algorithms that predict sales success based on EI qualities and identify hidden insights in the dataset. This is the part of the strategy where classification and regression models are built to forecast and validate the relationships found through quantitative research.

The main purposes of the investigation were to clarify EI as a predictor of sales success and to provide practical advice aimed at enhancing sales results using EI.

8.6 DATA ANALYSIS AND DISCUSSION

TABLE 8.1
Descriptive Statistics

Variable	Mean	Median	Mode	Standard Deviation	Range
Age (years)	34.5	33	30	8.2	22–55
Years of sales experience	8.7	8	6	4.5	1–20
EI score (out of 100)	75.2	76	80	12.3	50–95

Variable	Mean	Median	Mode	Standard Deviation	Range
Monthly sales revenue (Rs.)	12,500	12,000	10,000	2,800	6,000–20,000
Sales targets achieved (%)	85.0	87	90	7.8	60–100
Number of client interactions per month	30	28	25	6.5	15–45

Descriptive data provide an overview of the demographics, sales performance, and EI of 175 workers. Staff ages range from 22 to 55 years, SD = 8.2 years, M = 34.5 years old (see [Table 8.1](#)). Age diversity may lead to lots of different views and experiences on the sales teams, so be prepared!

Although some of the staff may have been at work for shorter periods, most others are sales experts. Great, the average experience is 8.7 years, the median is 8, and the mean is 6. This range also shows the varied edition, and it goes with the sales professionals' various tenure from 1 to 20 years.

The standard deviation for EQ scores is 12.3, and the mean result is 75.2%. It would appear that the squad possesses a moderate degree of EI—scores averaged from 50 to 95 over all tidiness areas. The high number of employees scoring toward the upper limit, indicated by a mean of 80 and median of 76, might help explain why these can drive sales success.

There are levels of uncertainty in performance, as the sales income mean here is 12,500 rupees per month with a standard deviation of 2,800 rupees. Income amounts are everywhere, ranging from Rs. 6,000 to Rs. 20,000, suggesting the majority of the workers make an average level of what we would call as revenue/earnings in their books.

Sales targets were typically achieved, having a standard deviation of 7.8% and a typical attainment of 85%. This means that workers meet or exceed their goals an average of 69% of the time, and it ranges from 60% to 100%. Because most of the workers achieve so close to their target goals, they perfectly hid a small fraction who either do extremely great or, in contrast, very bad.

Finally, staff interact with consumers a mean of 30 times per month and spectral density of 6.5. This distribution reflects the diversity of approaches customers take to get engaged with hyper-personalization personal engagement strategies and the varying amounts of heavy lifting that salespeople are required to perform, from 15 times a month to 45. This could impact their sales effectiveness and how they leverage EI in customer relationships.

The descriptive data revealed that the observed team is made up of seasoned and diverse professional experience with EI and sales performance, both of which could be specifically targeted for development.

8.7 HYPOTHESIS TESTING

TABLE 8.2

Pearson Correlation Coefficient between EI and Sales Performance [↗](#)

Variable	Pearson Correlation Coefficient (r)	Significance (p-Value)
EI score and monthly sales revenue	0.68	0.0003
EI score and sales targets achieved (%)	0.72	0.0001

The Pearson correlation analysis results showed that there is a significant relationship in a positive direction between EI and sales success indicators (see [Table 8.2](#)). The positive relationship is a significant one, which means those who score highly in EI tend to earn more money per month (as indicated by an impressive correlation coefficient of 0.68 shared between them). This association is so strong that it is very unlikely to have occurred by chance, and consequently, this p-value (0.0003) appears well below the standard significance criterion of 0.05 ([Farahani & Zare, 2020](#)).

However, the impact is even stronger in relation to sales targets achieved (percentage)—with a 0.72 correlation value, EI has a considerably large effect on sales targeting for the average worker. It would appear that on the face of it, greater EI results in better performance, either meeting or

exceeding their sales goals. With p-value of 0.0001 which is <0.05 , it refers to the fact that this result is statistically significant.

Cumulatively, therefore, our results lend support to the suggestion that salespeople do significantly better when they have high EI according to H1 (which states an alternative hypothesis effect). Both are very highly correlated with sales efficiencies and performance outcomes, so the impact of EI on those results is often tangibly revealed.

8.8 EI: CONCEPT AND THEORETICAL FOUNDATION

Since the late 1990s, EI has increasingly been seen as a key construct in explaining human behavior, including leadership and organizational contexts. EI, which is defined as the ability to perceive, understand, manage, and use emotions effectively, includes skills that enable us to ease a connection with others and improve our decisions ([Bande et al., 2015](#)). EI, which was introduced years ago in the scientific literature, used as a synonym for EI has long been a topic of several psychological theories, and models have evolved over time since it was theoretically first conceptualized ([Baig & Nesa, 2022](#)).

The notion of EI went mainstream in the mid-1990s because of *Emotional Intelligence: Why It Can Matter More Than IQ*, a book by Daniel Goleman that mixed conversations around psychology with feelings about business practice. According to Goleman, EI is not just an inherited characteristic, but it is a collection of abilities that can be learned and improved as life goes on ([Cherniss & Goleman, 2001](#)). The Mayer et al. model states that EI can be categorized into these five fundamental areas: self-awareness, self-regulation, motivation, empathy, and social skills. The domain that each has a role to play in how we relate to others and how our emotional landscapes get developed.

8.9 KEY INGREDIENTS FOR EI

Self-awareness is the ability to recognize one's own emotions and their effects on thoughts and behavior ([Farahani & Zare, 2020](#)). People with high self-awareness know their strengths and weaknesses.

Impulse/Self-Regulation: Ability to manage emotions and impulses. It provides the set of tools to help manage overall response—especially for the stress-triggered kind.

Background: Intrinsic motivation energizes the individual to set goals and sustain their goal-directed behavior ([AlDosiry et al., 2016](#)), while patients' involvement in health care and their ability to manage chronic diseases is highly dependent on cognition about motivational factors for long-term adherence ([Baig & Nesa, 2022](#)). People who are emotionally intelligent typically have a high need for accomplishment, in its purest sense.

Empathy: The ability to understand and share the feelings of another. It is a very necessary part of communicating and building interaction with other people.

Soft Skills: Skills that are important for handling people, relationships, and building. They can use their social skills to influence, inspire, and effectively lead others.

8.10 EI MODELS

There have been a number of models proposed to measure EI.

Cognitive Ability Model: Originally developed by Peter Salovey and John D. Mayer, this model frames emotional intelligence as a form of intelligence used to interpret and work with emotional information. It highlights four core abilities: recognizing emotions, harnessing emotions to support thinking, making sense of emotions, and regulating emotions effectively.

Including personality avg to balance abilities:

Goleman Model: This model suggests that EI combines cognitive (e.g., problem-solving) and non-cognitive (motivation, for example) traits.

Bar-On Model: Reuven Bar-On did not develop a competency model in the strictest sense of the word, but he consolidated emotional and social competencies into a single framework and promoted their importance for the first time as correlates (components?) of human ability to cope with life and excel (Goleman, 1998).

8.11 EI TRAININGS

Evidence shows that EI drives leadership capability and organizational success. Those with high EI also tend to be more patient and able to deal with stress, develop positive workplace culture, and improve team dynamics. They are better at working through challenging social circumstances than those with low EI. Also, individuals who are emotionally intelligent often excel in sales jobs, as they can read their customer, feel them out on a more emotional level, and then react accordingly to that information. This kind of relationship builds a necessary rapport and loyalty, which are crucial for effective sales tactics.

8.12 ML TECHNIQUES IN HUMAN BEHAVIOR AND PERFORMANCE PREDICTION

A lot of interesting information about predicting and analyzing human behavior and performance is also published using ML. Using big data and sophisticated algorithms, researchers are able to find patterns and revelations that had previously been obscured. Some of the major ML methods we are applying in this field are as follows.

8.12.1 DEEP LEARNING MODELS FOR PREDICTING HUMAN DECISION-MAKING

This paper explores how deep neural networks (DNNs) can be employed to predict human decision-making patterns. The research builds on earlier models by using complex patterns to predict behavior in specific environments, such as sales and leadership interactions.

8.12.2 SUPERVISED LEARNING FOR BEHAVIORAL PREDICTIONS

Supervised learning models are applied to labeled datasets to predict human behavior. This study focuses on the use of support vector machines (SVMs) and random forests to predict the impact of EI on sales performance.

8.12.3 UNSUPERVISED LEARNING FOR DISCOVERING HIDDEN PATTERNS

The paper presents unsupervised learning algorithms for uncovering hidden patterns in human behavior datasets. It highlights how these methods help reveal overlooked dimensions of emotional intelligence and the influence of leadership.

8.12.4 FEATURE ENGINEERING FOR PREDICTING PERFORMANCE

The paper on feature engineering highlights the importance of selecting the right data points for ML models. By focusing on key EI traits, the study demonstrates how predictive accuracy can be improved in performance models.

8.12.5 MULTIMODAL FUSION FOR COMPREHENSIVE BEHAVIORAL PREDICTION

This work discusses the fusion of different data modalities, such as physiological, text, and video data, to create more comprehensive predictive

models. The study applies multimodal ML models to forecast performance in leadership and sales ([Wisker & Poulis, 2015](#)).

8.12.6 ADVANCED PREDICTIVE MODELS FOR LEADERSHIP IMPACT ON SALES

This section highlights research on the application of advanced ML models to predict how leadership styles influence sales performance. The study investigates different leadership strategies and their correlation with sales outcomes.

8.13 EI AND LEADERSHIP ASSESSMENT TOOLS

EI is one of the important contributing factors that affect sales success, and research also has effectively supported the favorable and statistically significant relationship between these two. The study also found a strong correlation from the Pearson results, which suggests a high level of association between EI test scores and monthly income, as well as the percentage of sales goals met. The data suggests people with a high EQ will tend to do the best job, earn much more, and meet or exceed their goals.

The report has a large amount of practical application that is offered to companies that want to sell more by explaining the role that EI plays in higher sale outcomes. Sales teams with a greater focus on developing sales staff EI skills may improve the sales efficiency and overall performance of firms. These results are consistent with the literature and suggest that sales effectiveness might be improved by implementing programs designed to teach and develop EI.

The assessment of EI is now considered an essential part of evaluating leadership. Organizations are now just beginning to understand the significance of EI in supporting good leadership, and a plethora of assessment tools has been created as well to measure and improve these skills. In this article we discuss some of the key EI assessment tools, which are used extensively to measure leadership competency.

8.13.1 TRADES AND SERVICES INDUSTRY SIX SECONDS EI ASSESSMENT(B) BUDGET—\$293.00 PER PERSON SIX SECONDS EI ASSESSMENT

Six Seconds EI Assessment (SEI): The premier tool to measure EI and support development. Built on validated psychometric foundations, it is used worldwide by a range of organizations—from FedEx to the United Nations. The SEI has two forms: one measures EI competencies, and the other examines current success factors, including effectiveness and relationships. This dual method not only deepens our understanding of where EI can be improved and better applied against tangible outcomes but is especially compelling for leaders wanting to advance their ability to connect effectively with people.

8.13.2 EQ INVENTORY (EQ-I 2.0)

EQ-i 2.0 is one of the most well-known EI tests that exists; this model was created by Reuven Bar-On. The tool covers 15 emotional and social skills, making it less of a gross measure of valuing oneself talking about emotions (as self-esteem essentially is) than a more complex profile of an individual's relative strengths and weaknesses in these two areas. This can be particularly useful in leadership development programs, where, for example, using the EQ-i 2.0, we help a group of emerging leaders learn more about how well their emotional skills are serving them at work. The form of the evaluation allows individuals to evaluate and pinpoint areas in which they would specifically like to improve, thus enabling focused development.

8.13.3 MAYER-SALOVEY-CARUSO EI TEST

A performance measure of EI in the Mayer–Salovey–Caruso EI Test (MSCEIT) takes a series of parts and assesses how an individual can process emotional information, which will ultimately help people think, understand emotion, and reason with emotion. Unlike self-report measures,

the MSCEIT is an ability-based test in which people perform a set of tasks designed to simulate real-world emotional problems that are relevant to the typical environment/tests taking place—that is, this performance will map into their everyday living as well and not just castles in the air. This method is a unique way for leaders to get information on their ability to process emotions that may help them make better decisions and manage relationships.

8.13.4 GENOS EI TESTS

The Genos suite of assessments, by contrast, is aimed squarely at the behaviors of leadership associated with EI. This includes self-assessments and 180° or 360° multi-rater feedback options for leaders to gain a holistic view of behavior related to EI in the workplace. The assessments provided by Genos not only measure EI, but they also provide actionable development tips, which are designed to align to the individual and make them a highly practical tool for leaders looking to improve their performance.

8.13.5 EMOTIONAL AND SOCIAL COMPETENCY INVENTORY

Another sense is the Emotional and Social Competence Inventory (ESCI); this assessment of EI is based on feedback from others who know you. The multi-rater approach enables accurate measurement of leader social-emotional competencies rather than just self-awareness or awareness as perceived by those who report to them. The ESCI helps to identify strengths and areas of development, as well as offers awareness regarding the impact of these competencies on leadership quality.

8.13.6 INFLUENCE OF EI ON SALES PERFORMANCE

Sales performance is greatly improved by enabling sales professionals to connect with customers on a more personal level, which can be facilitated using EI. You do not get ahead in sales based only on knowing your product or being good with tools, but by understanding and managing emotions—both yours and your customer's. Studies say that emotionally intelligent salespeople sell more and build better customer relationship translating into better SALES RESULTS. An example is that 90% of top performers have high EI, and they earn an average of \$29k more than their lower-scoring peers.

Of all the ways EI affects sales, one of the most important is the power to respond to objections. Top salespeople, who are emotionally intelligent enough to see these for what they really are, view them as opportunities for dialogue instead of threats. Employ empathy to connect with customer pain points—this will enable you to provide a solution more responsibly. This technique not only assists in overcoming prospects but also grants the trust and relationship with clients that is a long-term foundation.

Also, EI helps with dealing with stress as well as staying professional when faced with difficult situations. Sales also tends to be a high-pressure environment with lots of rejection, but high EI helps people deal with that better. They are still driven and stubborn, viewing rejection as feedback of incorrect choices they made rather than a feeling of being let down. And sales professionals with high EI will thrive in this kind of environment, where they can deliver a personalized customer experience. By being more empathetic to consumer feelings, companies can frame their communication strategies with specific needs—which results in higher levels of customer satisfaction and engagement. This type of relationship-building not only improves short-term sales but also has a long-term effect on customer retention. In conclusion—with a proven impact on those client interactions, managing objections and keeping people on the right side of bright even when things are tough, it is really clear to see how important EI is as far as sales performance goes. As EI becomes an “it” factor for organizations and we outright demand from ourselves and our people that they have emotional, social, and interpersonal competence, those sales effectiveness

improvements likely to come if you invest in developing it will be dwarfed only by the amount of business success you can still create.

8.13.7 LEADERSHIP STYLES AND THEIR IMPACT ON SALES PERFORMANCE

The way in which leaders manage their teams is the starting point for impact on sales performance. How leaders behave can significantly help or seriously hinder sales professionals in hitting their numbers and driving revenue. Organizations that want to base their sales strategies on sound business knowledge are driven toward the understanding of how different leadership styles impact sales outcomes. Among the many styles of leadership, one of the most successful is transformational leadership because it works well in improving sales performances. Transformational leaders empower and energize their teams with a real sense of vision combined with creating innovation/personal-development-supportive environments. BambooHR recognizes the value in giving salespeople autonomy, an outlet for creativity, and a general ethos of lifelong learning. This does help job satisfaction but also allows every team member to feel responsible, which in turn helps drive a culture of productivity and top-line sales. According to studies done about transformational leadership, teams led by these leaders usually surpass their quotas and work even harder for the company as a whole because of the positive working environment created.

Transactional—Mainly focuses on tasks and provides clear instructions and goals. Transactional leaders focus on rewards and punishments according to performance outcomes, which can be appropriate for contexts where tasks are routine and well understood. However, this approach can be constraining in terms of creativity and drive for modern salespeople. Meaning, while it can drive behavior change in the short term using incentives, behavioral traits such as the adaptability that characterizes modern markets resist being pushed/rewarded and hence it is a less efficient way of achieving long-term engagement or desired performance levels.

Transactional leadership over sales teams and their performance: Such a type of leader would normally be effective in delivering on specific targets but not in the development of innovation or adopting new obstacles.

Another important style that focuses on putting the needs of team members first is servant leadership. Servant leaders concentrate on developing rapport, trust, and levels of support within their sales teams. This promotes job satisfaction and, if followed, results in a better performance. Servant leaders put a particular focus on the personal and professional development of their team members, creating an environment where sales people know they are valued and inspired to do their best. Nurtures customer relationships: Empowered staff deliver better services, which helps improve the relationship management of clients that are not tag-alongs. With today's business changes, adaptive leadership is essential. Firms who adopt this style are agile and adaptive to change, thereby allowing their team to ride the storm more favorably as markets fluctuate. Adaptive leaders inspire and empower a learning, resilient culture, helping sales people to adapt to dynamic shifts in customer requirements and competitive demands with skill sets. Both of these moves not only drive higher individual performance but also add to the flexibility of the sales org as a whole. In fact, situational leadership theory accepts that distinct situations demand unique behaviors. Leaders who can identify what their team members need individually and adapt to support them in the most effective way will unlock the potential of each individual as well as drive superior results. Situational leaders can drive substantial improvement in behavior and shared understanding on teams when they adjust their style of leadership to the context. Leadership styles, in summary, have a big impact on sales performance with their ability to form team dynamics, high motivation level, and an effective all-round approach. Sales personnel are more engaged and innovative when led by transformational or servant styles of leadership, though transactional leadership often can produce short-term results at the sacrifice of long-term adaptability. In addition to the sales leaders themselves, both adaptive and situational leadership styles play critical roles in making otherwise top-performing sales teams more

resilient and responsive within a competitive environment. This mutual understanding can help organizations produce successful sales leadership practices that deliver long-term results in business growth.

8.14 INSIGHTS ON ENHANCING SALES PERFORMANCE THROUGH TAILORED LEADERSHIP APPROACHES

Sales leadership directly related to team performance and ultimately organizational goals is the effectiveness of sales leadership. Although leadership is situational and based on context, a one-size-fits-all approach generally does not work. Leaders can adapt by understanding what makes their sales teams tick and creating an environment where individual and collective sales performance hits the high notes—specific to them.

One important responsibility of a tailored leader is seeing the different ways people are motivated and what their various strengths are. The individualized consideration of transformational leadership is excellent at empowering and inspiring the team. These leaders provide tailored coaching, mentorship, and support to help salespeople grow into their full potential and exceed their goals. Transformative leadership was found to predict sales volume, distributor recruitment, and overall team performance in a positive way.

Conversely, transactional leadership is all about setting expectations for performance levels, establishing goals, and providing rewards or punishments depending on perceived output. This well-defined method works when you need repetitive results—for example, to reach your monthly sales quota. However, just transactional leadership may not be enough for addressing complex customer needs and growing market conditions. Servant leadership, on the other hand, focuses on the growth and well-being of team members. Servant leaders help to forge a trustworthy, loyal, and purpose-driven team with a safe and collaborative work environment. This way, one can guarantee not only higher job satisfaction and motivation but also eventually more output, especially in customer-oriented sectors. Today, the business landscape is so dynamic that

adaptive leadership has become a necessity. Leaders who can evaluate shifting realities rapidly and create new courses of action enable their teams to grow during uncertain times. By promoting flexibility, risk-taking, and ongoing learning, adaptive leadership will prepare their sales teams to meet new challenges with the skills and tools necessary to succeed.

Before a sales manager can employ personalized leadership strategies, they must first intimately know their team's strengths/weaknesses and key drivers. Having a quick conversation one-on-one at intervals, performance reviews, and debrief sessions can give you, for example, insights into what people want and need personally. Match the style of leadership with team member characteristics so that they can operate best to achieve their mutual aims. Also, the best sales leaders understand that they, along with their teams, need to grow and learn. Training programs, workshops, and coaching can sharpen the skills and knowledge of sales professionals to provide more fruitful customer journeys and grow your revenue. Prioritizing professional development is one way leaders show they care about the success of their team and value continuous improvement.

So, in total, we can see that sales performance improvement through leadership is all about flexibility, EI, and coaching individual development and team development. When sales leaders recognize the special needs and attributes of their teams, they can apply leadership styles that are transnormal, transactional, servant, or adaptive as required to help each person perform at his best. With customized coaching, transparent communication, and an emphasis on never-ending learning, sales leaders can help unlock the latent potential of their teams and create sustainable results for their respective organizations.

8.15 CONCLUSION

EI is one of the important contributing factors that affect sales success, and research also has effectively supported the favorable and statistically significant relationship between these two. The study also found a strong correlation from the Pearson results, which suggests a high level of

association between EI test scores and monthly income, as well as the percentage of sales goals met. The data suggests people with a high EQ will tend to do the best job, earn much more, and meet or exceed their goals.

The report has a large amount of practical application that is offered to companies that want to sell more by explaining the role that EI plays in higher sale outcomes. Sales teams with a greater focus on developing sales staff EI skills may improve the sales efficiency and overall performance of firms. These results are consistent with the literature and suggest that sales effectiveness might be improved by implementing programs designed to teach and develop EI.

REFERENCES

- AlDosiry, K.S., Alkhadher, O.H., AlAqraa', E.M., & Anderson, N. (2016). Relationships between emotional intelligence and sales performance in Kuwait. *Journal of Business Research*, 69(1), 123–128. <https://doi.org/10.1016/j.jbusres.2016.07.021>
- Baig, D.M.R., & Nesa, D.H. (2022). Emotional intelligence and entrepreneurship ability among youths: A correlational study. *KAAV International Journal of Arts, Humanities & Social Science*, 9(1), 1–7. <https://doi.org/10.52458/23484349.2022.v9.iss1.kp.a3>
- Bande, B., Fernández-Ferrín, P., Varela, J., & Jaramillo, F. (2015). The impact of emotional intelligence on sales performance: A meta-analysis. *Journal of Business Research*, 68(5), 1036–1045. <https://doi.org/10.1016/j.jbusres.2014.11.008>
- Cherniss, C., & Goleman, D. (2001). *The emotionally intelligent workplace: How to select for, measure, and improve emotional intelligence in individuals, groups, and organizations*. Jossey-Bass.
- Farahani, M., & Zare, H. (2020). The relationship between emotional intelligence and sales performance: A case study of Iranian salespeople in the pharmaceutical industry. *International Journal of Pharmaceutical and Healthcare Marketing*, 14(2), 145–162.

- Goleman, D., Boyatzis, R., & McKee, A. (2002). *Primal leadership: Learning to lead with emotional intelligence*. Harvard Business Review Press. [↵](#)
- Haakonstad, A.-K., & Håvold, J.-I. (2011). Emotional intelligence and sales performance: A study of technology sales professionals in the United States. *International Journal of Business and Management*, 6(9), 75–82. [↵](#)
- Honeycutt Sigler, T., & Shepherd, C.D. (2023). Emotional intelligence training and professional salespeople: Best practices for developing effective training programs for sales professionals. *Atlantic Marketing Journal*, 12(2), Article 7. [↵](#)
- Karatepe, O.M. (2010). The effects of emotional intelligence on frontline employees' outcomes in the hotel industry. *International Journal of Contemporary Hospitality Management*, 22(1), 69–84. [↵](#)
- Law, K.S., Wong, C.S., & Song, L.J. (2004). The construct and criterion validity of emotional intelligence and its relation to workplace performance outcomes. *Journal of Applied Psychology*, 89(3), 483–496. [↵](#)
- Lopes, P.N., Salovey, P., & Straus, R. (2003). Emotional intelligence and social interaction. *Personality and Social Psychology Bulletin*, 29(12), 1470–1480. [↵](#)
- Mayer, J.D., Salovey, P., & Caruso, D.R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry*, 15(3), 197–215. [↵](#)
- Mehlhorn, T. (2020). The influence of emotional intelligence on sales performance of companies—with a particular focus on salespersons. *Journal of Business Research*, 112(1), 123–130. [↵](#)
- Mortan, C., Ripoll, P., Carvalho, C., & Bernal, J.-A. (2014). The role of emotional intelligence in the relationship between leadership styles and job satisfaction in the workplace: A study of Spanish employees in various sectors of activity. *Journal of Work and Organizational Psychology*, 30(2), 83–90. [↵](#)
- Pearman, T.A. (2011). The role of emotional intelligence in employee performance evaluations: A review of the literature. *Journal of Business*

Research, 64(7), 747–753. [↗](#)

Schutte, N.S., Malouff, J.M., Hall, L.E., Haggerty, D.J., Cooper, J.T., Golden, C.J., & Dornheim, L. (1998). Development and validation of a scale for measuring emotional intelligence. *Personality and Individual Differences*, 25(2), 167–177. [↗](#)

Wisker, Z.L., & Poulis, A. (2015). Emotional intelligence and sales performance. A myth or reality? *The International Journal of Business and Society*, 16(2), 185–200. [↗](#)

9 AI and Human Rights

Guardians of Freedom or Threats to Democracy?

*Prashasti Pritiprada, Ipseeta Satpathy,
Christian Kaunert, and B.C.M. Patnaik*

DOI: [10.1201/9781003668657-9](https://doi.org/10.1201/9781003668657-9)

9.1 INTRODUCTION

Artificially intelligent technology is a notion which has been an integral component of popular conversation for many years, frequently represented in futuristic movies or discussion regarding how clever computers could triumph overrun the entire globe, reducing the human species to a dull subservient condition in favor of an artificially intelligent regime. While this depiction is a cartoon of artificial intelligence (AI), the truth resides that it is something that already this facade, and numerous of us grapple with the phenomenon on an ongoing basis. The use of AI is not anymore the domain of sceptics; it is already a crucial part of several companies' operational strategies as well as a critical tactical ingredient in the future directions of numerous domains of industry, health, and governance on a worldwide level ([Dwivedi et al., 2021](#)). Such revolutionary influence of AI has sparked enormous attention from academia, with fresh research focusing on the technological developments, repercussions, and outcomes in lieu of its

efficiency and ramifications, which appear having remained the primary study subject for a period of time. Discoveries in machine learning have sparked widespread interest among both academics and professionals, opening potentially an extensive spectrum of favorable potential for AI use in the nongovernmental domain. There has to be an increasing appetite for a comprehensive knowledge of the scope and effect of based on AI uses, as well as the difficulties they provide. Yet, certain studies examine applications of AI along with difficulties in solitude especially fragmentarily. Considering the scarcity of an in-depth analysis of applications based on AI and difficulties for the wider public market, certain theoretical methods analyze and aggregate important conclusions from academic research for providing an integrated picture of applications for AI and associated issues. Few identify AI application categories, detailing how they provide utility alongside operation, in addition to particular general use instances. Going ahead, technological advancement will play a growing significance within energy production. Nonetheless, it remains unclear whether upcoming technological advances will be most important throughout the power industry's change ([Wirtz et al., 2019](#)).

AI is a revolutionary innovation in the modern digital era. Debates regarding what it really is, which it is capable of now, as well as what it could eventually achieve span technological advances, psychological research, political economy, fantasy and science fiction, legislation, and philosophy. AI has been the topic of several conversations and publications, ranging from writings on technological developments to sensational concerns pertaining to its ramifications. While the argument keeps going, the innovations that enable AI advance, allowing for implementations that range from identification of faces in cellphones to mobile apps for consumers that utilize AI algorithms to identify illnesses such as hypertension and diabetes with greater precision. Although AI is often associated with fiction about science and robotics, there are several real-world uses that are affecting every day of our existence. AI has the potential to provide significant advantages and destabilize industries, according to preliminary data. AI's reliance on an internet-based base, along with the

reality that the technology frequently requires training on distinct information, means that enterprises have no alternatives (Lyu & Lin, 2021). Organizations must accelerate their technological excursions, incorporating AI. Deployment statistics show an increasing disparity amongst initial AI users and others as well. The rights of humans are the one and only generally recognized set of current values that, over the last fifty decades, every nation have steadily evolved and articulated within a complete universal regulatory structure. The globe's rights of people system are inextricably linked to global security and peace, growth, including a worldwide push for democratic society, accountable government, including the supremacy of justice. Worldwide, humane and felony laws are now regarded subsets of global rights of people law, which has grown in complexity and has been challenging to monitor since the conclusion of the Cold War. Robotic intelligence is establishing itself as an innovation that is influencing an array of ethical concerns, including prejudice and supplier network verification. Considering firms' growing ethical commitments and the escalating debate on intelligence and rights for humans, many studies have provided insight into multinational entities' obligations for the fulfilment of ethical practices while creating and employing AI ([Chui et al., 2018](#)). What consequences do equitable treatment duties exert on corporations researching and employing AI? Several pieces examined if AI fundamentally clashes against liberty and independence, as well as whether AI could be potentially aligned to the beneficial action criteria of AI principles and implemented in terms of human liberties-related domains.

9.2 UNDERSTANDING AI

The ideas in Alan Turing's 1950 work *Computing Machinery and Intelligence* posed an essential issue, "Can computers anticipate?" Turing later developed a simulation sport, referred to as the Turing test, which is a procedure in which if a machine can hold a dialogue which is unrecognizable coming from one involving a person, it is acceptable to conclude that the device is smart. The Turing examination became an

inaugural demonstration designed to assess the abilities of computers. The initial “AI era” started after the Dartmouth meeting in the year 1956, when AI was given a title as well as an aim. McCarthy developed the phrase called “artificial intelligence,” that subsequently evolved into the trademark for the research discipline. The most significant seminar claim was the belief that: Each component of any additional characteristic associated with education or intellect must be precisely identified for the computer could imitate itself. Individuals felt ecstatic as, for a brief duration, machines appeared to solve queries in the same way as folks. The more general investigation into AI sector was initially optimistic, expressing ambitious projections while gaining ground. The machine learning methods involve a means of developing autonomous robots with resembling behaviors as humans ([Nowak, 2021](#)). The swift development into this field has piqued the interest of plenty of technicians and academics worldwide, and numerous firms spanning multiple disciplines are eager to investigate its possibilities. The use of AI is constantly transforming the surroundings of organizations, and even individual and collective duties, as breakthroughs in academia in the area continue. Considering any discipline that has exploded in prominence in such a brief space of time, programmers working on AI must research its progress, advances, usage, and upcoming aspects of enrichment to gain a deeper understanding of the subject.

AI systems mimic the structure of the human neurological framework in AI challenges by considering computational elements like neuron. The aforementioned grandiose ambition of neuroscience is to produce intelligent machines by designing computers with design that mimics the calculations of the nervous system of humans. The computing capacity of the greatest modern computers pales in comparison to that of the brain of a person, making such a challenging undertaking (Kriebitz & Lutge, 2020). These networks of neurons emerged right after the debut of machines during the 1950s and 1960s. Frank Rosenblatt’s perceptron algorithms were regarded as the essential foundation of neural network technology, sparking initial optimism around the potential of machines with intelligence. Nevertheless, following the initial enthusiasm, an era of disillusionment followed

whereby both thirst for information and technically costly character of artificial brains were viewed as a barrier to their future usefulness. Around the turning point of the decade, improved accessibility to information and processing capacity resulted in enhanced profitability for AI systems, and the discipline was revived according to the recently established moniker of the term “deep learning.”

While AI spreads into crucial sectors, moral issues emerge, revealing an intricate relationship among technology progress as well as cultural ideals. Healthcare is one of the numerous furnaces in which moral quandary emerges. Applications of AI in healthcare diagnosis and therapy management pose worries regarding confidentiality of information, customer permission, especially the accuracy of computations in critical situations. Establishing the right equilibrium among creativity and moral principles is critical for building confidence across doctors and nurses and everyone else ([Raj & Kos, 2023](#)). In a comparable manner in the judicial system, the application of AI toward criminal justice prediction, including punishments analytics, raises many ethical issues. Challenges regarding equity, transparency, and the likelihood of intensification of currently existing inequities necessitate a thorough assessment of ethics. Throughout higher academics, the use of AI-powered technologies enabling customized instruction poses moral issues about privacy of learners, the technological gap, including the possible amplification of currently present disparities in learning. Because autonomous devices acquire essential players in the learning environment, moral dilemmas must drive their creation and implementation to guarantee that they improve student achievement while not increasing inequality within society.

9.3 THE IMPACT OF AI ON HUMAN RIGHTS

Some studies take an empirical viewpoint to rights for people, rather than an ethical or emotional one. Some understanding of rights for humans is based on the global Declaration of Human Rights, which includes three vital agreements. The system of legislation has evolved by way of the

enactment of newly concluded arrangements, the consideration for publication of the overall criticism, as well the application of the terms of agreements by globalization and native legal bodies in particular circumstances ([Aggarwal, 2018](#)). The anonymity encompassing computerized assessments of risk created by entrepreneurs may jeopardize unlawful defendants' constitutional entitlements for an open and transparent trial prior to a tribunal that is impartial and independent, as well as access to necessary defense assurances. Establishing thorough laws and regulations for AI for medical purposes may promote an atmosphere which prioritizes rights for humans while maximizing its possible goodness. This strategy emphasizes sustainable and moral deployment of machine learning, benefiting the public while safeguarding freedoms of individuals. Findings revealed worldwide consolidation of trends across many AI concepts, particularly for the good of society. This integration creates a significant push for international cooperation and norm-setting in this arena. Based on the studied materials, AI is the most appealing community advantages are medical care, educational opportunities, and mobility. A lot of nations concentrated upon their distinct circumstances ([Singh & Joshi, 2024](#)). The significance of AI for benefiting society along with designing with people in mind is well recognized. The pursuit of sustainable growth is one way to look at such attempts.

Much research addressed far-reaching socioeconomic as well as constitutional ramifications. These influence a wide range of essential human rights concepts, including safeguarding information, egalitarianism, liberties, freedom of choice and independence, respect for others, security for humans, consent that is informed, honesty, equitable treatment, fairness, confidentiality, and decision-making. The outcomes of a fiscal evaluation of impact conducted as part of the SIENNA program raised similar concerns. Several broad lawsuits include the absence of machine learning-related restrictions, a shortage of fresh governing bodies in which current ones tumble curly the effectiveness of present laws related to nations, an absence of an explanation on how to use of the current legislation, a shortage of ethical intellectual discussions in specific nations, an absence of legal

recourse experience and expertise, and ambiguity in the legally binding nature of autonomous devices ([Raso et al., 2018](#)).

9.4 PRIVACY AND SURVEILLANCE

Confidentiality with morality has become essential in the framework of machine learning due to the rapid advancement of technological advances, which crosses the borders dividing safeguarding private information versus moral quandaries. Within this instance, integrity applies toward the conscientious along with balanced execution of machine learning to avert damage as well as promote social harmony, whereas anonymity relates to people's basic entitlement to manage their own personal data, and to make sure algorithms have no impact on their personal privacy. The criteria are significant as they safeguard toward analytical prejudices; compromises of information, and the potential overuse of deep learning for surveillance or the selection process. Precisely, AI becomes increasingly integrated throughout everyday activities, a solid basis for security and morality is required to achieve an appropriate equilibrium among creativity and preserving rights for humans, as well as promote confidence in AI systems and guarantee that they enrich as opposed to damage culture. Such concepts are really important. Most importantly, these features are critical for protecting private information against intrusions and maintaining integrity, especially in the modern era of substantial amounts of information and interconnected networks. Furthermore, algorithms serve a significant part in rectifying computational prejudices which may contribute to systemic discrimination and unfair treatment ([Pashkov et al., 2023](#)). Additionally, security and integrity are crucial for developing audience trust and belief in machines powered by AI, as well as to guarantee AI is utilized to benefit humankind instead of causing damage. The expansion of AI spying is unchecked. The application of this tool by authoritarian governments for pursuing certain groups has raised concerns. Despite nations having solid systems democratic customs, AI raises questions about morality. Specialists are concerned regarding a substantial number of detection of faces errors or

misleading results among communities. The general population is becoming more conscious of biased algorithms in AI records, which can negatively affect crime prevention as well as judicial analytics.

AI devices involve surroundings, including healthcare monitoring systems, namely identification tracking systems. These monitoring devices help government agencies execute the rules by tracing people's activities along with extending statistical information to track and anticipate societal behavior. As a result, organizations are viewed as possibly successful instruments in responding to humanitarian emergencies, especially those caused by violence and medical epidemics. A three-dimensional framework might be proposed for assessing these possibilities or risks ([Shackelford et al., 2022](#)). An operational aspect evaluates the kind, quantity, and quality of data that the equipment requires to perform properly. The authorization component examines the individual's capacity to be properly notified regarding and refuse any kind of monitoring, challenging if permission is possible and if an individual is entitled to completely autonomous/independent decisions. Ultimately, there is a sociocultural feature that outlines susceptibility along with the effects of greater authority of entrenched authority structures via novel strategies of sterilization rooted in informational tracking. Many studies address moral and legal issues about AI in promotional activities, including prejudice, confidentiality, even responsibility, through synthesizing data taken from numerous places and proposing remedies like as accountability, moral requirements, discrimination detection implements, as well as confidentiality-enhancing technology. A handful of studies focus on ethical AI deployment to build confidence among customers and comply with law, emphasizing the significance of ongoing investigation and interaction in tackling these growing concerns ([Rodrigues, 2020](#)).

9.5 BIAS AND DISCRIMINATION

People of all kinds face the risk of having their online identities commodified while they engage in a digital space. The surroundings

capture and quantify perplexing ideas and behaviors to categorize, classify, suggest, and arrive at judgments concerning the lifestyles of individuals. Although numerous organizations aim to use data responsibly, prejudices associated with technology can nevertheless have negative consequences, despite their purpose. Hazardous results, especially if unintentional, pose serious hurdles to faith toward AI. AI presents unique challenges which necessitate innovative ways to ensure its safety and security. Machine learning algorithms are frequently used in environments where their application might have a substantial influence. The issue of whether AI has a positive or negative influence is central to the concept of reliable and appropriate AI. AI's negative effects extend beyond the person or business scale and can have far-reaching consequences for the community. AI and predictive algorithms may cause considerable harm spanning disciplines and sectors, necessitating collaborative efforts to mitigate the impact ([Taneja et al., 2024](#)).

The theory of economics has always modelled people's behaviors like may be the case one were statistically optimal. Critics of financial abstraction and its presumptions of sensible forecasting say that many aspects of social behavior are incapable of being formalized. The identical complaint is leveled against machinery that makes decisions. This approach provides the theory of economics—a perfect fit for creating and assessing decision-making programs. Accounting has a strong statistical toolset focused on deductive reasoning. Practical methodologies can enhance conventional science-based techniques for addressing biases within algorithms ([Feldstein, 2019](#)).

To address skew within AI forecasting algorithms, it is crucial to recognize along with mitigate its underlying causes at all stages of creation and deployment. Nevertheless, the present awareness of discrimination in AI is only a portion of a vast glacier. Participants, such as physicians, scientists studying AI, advocates for patients' categories, medical equity researchers, lawmakers, and the business world, ought to work together to improve illustration in AI proforma. This will help guarantee the fact that

AI innovations and instruments improve medical services as opposed to worsen it ([Fontes et al., 2022](#)).

9.6 FREEDOM OF EXPRESSION

While federal scrutiny upon big internet startups grows, corporations alongside politicians alike are looking for scientific responses to complex interface oversight problems like xenophobia and disinformation. Computerized tag recognizing and prognostic algorithms for learning, known as computational censorship structures, are currently being utilized at mass with large players media outlets including Instagram, YouTube, Facebook, and X (formerly known as Twitter). Many articles mention about a simple scientific initial regarding the way computational caution operates; investigate a few currently remaining computerized methods employed on prominent social grounds to cope with plagiarism, violence, and unsafe opinions; and point to necessary ideological as well as moral dilemmas for such mechanisms as their adoption advances. Some study examines the three problems pertaining to confidence in society, openness of data, and accountability that online social networking sites confront because of their posting filtering rules and procedures, and also the consequences for morality and legislation. Such quandaries, which are solvable only via the development and roll out of intelligent technology, represent the complicated interaction amongst online networking sites, subscribers, along with authorities. As social and political issues evolve, so does this connection ([Kumar & Suthar, 2024](#)). In terms of tidying up internet-based material, few articles believe that regulators, internet network corporations, and people ought all to engage an integral part. The program seems advantageous to strengthen interaction among governments and networking companies in addressing relevant issues, while public and private sector influence ought to remain limited. Members of the online community additionally have an obligation to preserve their social networking cleanliness, as well as being conscious of the necessity to restrict and balance the reins of legislatures and corporations.

9.7 RIGHT TO FAIR TRIAL

Hazard evaluations take place at several stages of legal proceedings, involving parole, punishment, and release from prison, in addition to deciding prolonged monitoring and permanent custody decrees for defendants with elevated risk. These threat evaluations are usually carried out by court officials using their own personal discernment and instinct, utilizing medical evaluations, defined by laws and general principles of law, thus containing the notion of personalized fairness. However, as law enforcement becomes more technologically advanced, empirical, driven by data assessments of risk are becoming more common ([Schwartz et al., 2022](#)). Traditional legal evaluation duties are rapidly being supplemented by a variety of computational, complex, algorithmic learning, and AI methods that claim to give precise planning abilities and impartial, reliable risks being assessed. The explosive rise of machine learning is resulting in major developments in many disciplines, notably the courtroom. As AI capabilities grow and spread, bringing them inside legal proceedings brings both potential and concerns. Investigators are interested in how systems of law might be adapted to cope with the rise of AI, notably in the field of justice ([Cowgill & Tucker, 2019](#)). The worldwide problem of robotics entering the legal system demands a thorough analysis and understanding of its consequences, encompassing both possible advantages and threats. By investigating the use of machine learning in judicial procedures, investigators want to uncover methodologies and answers to various issues surrounding integration of AI.

The investigation of legal theory as well as technological advances is inherently contradictory in a component of, whereas innovation in technology occurs swiftly and encompasses concepts including internationalization and globalization, conventional legislation, for the vast majority, takes time to respond to emerging technologies and is primarily restricted to national boundaries. Yet the concept of the supremacy of legislation contradicts the idea that regulation is restricted by countries rather obtains universal respect (Nazer et al., 2023). Still, a severe challenge

to the application of statute looms in the manner of a strike by technical advancements in intelligent machines. As significant advances have been achieved in the educational arena of AI, this AI is beginning to find ways to integrate into virtual governance platforms, effectively displacing the role of humans. A well-known instance for such advancement involves the incorporation of AI to aid attorneys in formulating rulings in courts. Still, in numerous instances, this sort of technology is considered the ‘black box’ owing to its intricacy and legal protection. The dearth of candor and limited understanding of the working of such technologies, which are growing in popularity by governing institutions, is calling into question conventional ideas of the supremacy of justice. That is particularly important for principles closely related to the authority of statute, notably openness, integrity, and elucidation.

9.8 SOCIOECONOMIC RIGHTS

Over the past decade, technical advances in machine learning have resulted in enormous joblessness as well as a variety of injustices. Experts discovered that wealth disparity rose because innovation advances along with investigated whether equipment especially AI impacts in some way people’ occupations along with processes. Several studies use computerization information along with coefficients of Gini to show how the link amongst AI or wealth disparity is typically deemed unfavorable, which is precisely the present investigation found, meaning it possesses immediate effects upon the dispersion of employment and earnings. The digitization is reducing cheap with medium-skilled occupations while growing joblessness and widening the pay disparity among intermediate along with high-skilled labor. Inequality values of emerging economies have larger values compared to those of industrialized ones, showing overall disparity in wealth is greater in emerging economies. AI is starting to create novel educational and instructional alternatives, which are currently being evaluated across a variety of scenarios. AI necessitates cutting-edge facilities plus a vibrant innovative environment, however how

does it address the urgent needs of underdeveloped nations? Do their descendants need to remain patient until the privilege of AI? Alternatively, robotics should be addressed promptly as feasible to lessen cyberspace along with societal divides! The aforementioned represents a few of the concerns raised with esteem; this important conversation has to start having an acute awareness what is occurring and how things may be accomplished. Researchers investigate the way AI may be utilized to boost academic performance within instructional frameworks, leveraging knowledge to promote scholastic access and excellence in poor countries. The growing application of machine learning within governance presents multiple possibilities to administrations globally. The mainstreaming including AI into administrative procedures particularly people-centric environments has an opportunity to dramatically transform conventional processes of assistance supply, legislation, and compliance ([Gorwa et al., 2020](#)). Authorities, for instance, may employ AI to enhance the caliber of services provided to the public, build confidence among netizens, and overall boost their operational productivity and efficacy. Agencies can utilize AI for better forecasting or mimic complicated structures, while allowing for legislative tinkering. Advantage might be generated throughout a variety of administrative areas of operation, including advisory services, conveyance, healthcare for the public, even prosecution.

9.9 TOWARD ETHICAL AI DEVELOPMENT

Machine learning has begun to have a significant influence in civilization. The main concerns surround why, exactly where, and which individuals are going to feel the influence of AI. Several organizations have launched projects to define moral standards for implementing constructive technology. A large number of recommended concepts might be intimidating and misleading. Based on the outcomes, the numerous collections of moral requirements underlying AI could be identical, causing unwarranted overlap and duplication, or drastically different, resulting in uncertainty as well as misunderstanding. The most dire scenario is the

‘place toward values’ in which participants are enticed to ‘shop’ over something particularly enticing. The manner in which may the scenario about ‘theory dissemination’ be considered addressed.

AI is utilized in an extensive number of ways, with the majority of them employing statistical algorithms to identify similarities in massive quantities of information and then generate recommendations that depend on such trends. Because AI becomes more prevalent in extremely dangerous domains, there is an increasing need to create and regulate machine learning in ways which are reliable, liable, truthful, yet accessible. This serves as a fundamental concern highlighted via the researchers, and provides a comprehensive review regarding the moral, law administrations, including technological problems associated with building management frameworks with AI. AI tools have already been used to handle complicated, risky procedures related to sentence giving, health diagnosis, especially investment management ([Tong, 2022](#)). It creates novel obstacles, such as responsibility problems alongside autonomous automobiles, the bounds of present legislation for coping using “large information distinct influence” for hindering computational damage, equitable problems arising from integrating prosecution or bettering society, and digital utilization of media. Regarding widespread influence of AI, several important problems ought to be answered utilizing a comprehensive viewpoint.

To enhance the welfare of all people, AI must be designed and deployed without harming basic principles of humanity. Global rights for people offer a strong and worldwide representation of these principles. AI practitioners require “tough” regulations, technological norms, customs in society, and monetary rewards to integrate fundamental rights throughout their company’s operations, simulations, procedures, and routines. Legal rights advocates must actively participate with AI oversight and accountability efforts. Whenever required, policymakers ought to be prepared to use a system based on human rights to scrutinize how businesses and governments build and use technology. Collaboration between public communities and AI researchers is crucial for identifying risks and anticipating the requirements of populations in danger. Organizations must

collaborate beyond barriers to prevent injury to AI and promote rights for individuals ([Mckay, 2020](#)).

9.10 CONCLUSION

To summarize, the complex link involving machine learning and rights for humans requires our undivided devotion and purposeful action. To traverse the intricacies of a future AI-powered society, it grows clearer component of although AI has enormous promise to improve human rights, nonetheless it presents substantial obstacles and hazards. AI has far-reaching and diverse consequences for basic rights, ranging from breaching privacy to analytical prejudice. Nevertheless, these issues present a possibility for civic engagement and creativity. Society may steer studies on AI onto an era wherein human rights are protected and secured by adopting ethical guidelines, encouraging multidisciplinary partnership, and implementing strong frameworks for regulation. This endeavor will need some cooperation from authorities, executives in industries, scholars, and grassroots groups. It calls for an altar to openness, responsibility, and inclusion in AI development and operation. While we work to capitalize on the revolutionary potential of AI for the sake of mankind, we must stay diligent in safeguarding and supporting each person's innate worth and liberties. By pursuing AI research with modesty, kindness, and a firm commitment to people-centric principles, we may create a future in which AI acts as a protector of the rights of people instead of a danger. This ideal is within easy attract, because it is down to humanity to work together to create a world where technological advancement and dignity live peacefully for the betterment of everyone.

REFERENCES

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and

- agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. [↵](#)
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—applications and challenges. *International Journal of Public Administration*, 42(7), 596–615. [↵](#)
- Lyu, W., & Liu, J. (2021). Artificial intelligence and emerging digital technologies in the energy sector. *Applied Energy*, 303, 117615.
- Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). Notes from the AI frontier: Insights from hundreds of use cases. *McKinsey Global Institute*, 2, 267. [↵](#)
- Nowak, M. (2021). *Introduction to the international human rights regime* (Vol. 14). Brill. [↵](#)
- Kriebitz, A., & Lütge, C. (2020). Artificial intelligence and human rights: a business ethical assessment. *Business and Human Rights Journal*, 5(1), 84–104.
- Raj, R., & Kos, A. (2023). Artificial intelligence: Evolution, developments, applications, and future scope. *Przegląd Elektrotechniczny*, 99(2). [↵](#)
- Aggarwal, C. C. (2018). *Neural networks and deep learning* (Vol. 10, No. 978, p. 3). Cham: Springer. [↵](#)
- Singh, B. P., & Joshi, A. (2024). Ethical considerations in AI development. In *The ethical frontier of AI and data analysis* (pp. 156–179). IGI Global. [↵](#)
- Raso, F. A., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Kim, L. (2018). *Artificial intelligence & human rights: Opportunities & risks*. Berkman Klein Center Research Publication. [↵](#)
- Pashkov, V. M., Harkusha, A. O., & Soloviov, O. S. (2023). The impact of the introduction of artificial intelligence technologies on the current human rights and freedoms concept. *Polski Merkurusz Lekarski: Organ Polskiego Towarzystwa Lekarskiego*, 51(6), 646–653. [↵](#)
- Shackelford, S. J., Asare, I. N., Dockery, R., Raymond, A. H., & Sergueeva, A. (2022). Should we trust a black box to safeguard human rights?: A

- comparative analysis of AI governance. *Journal of International Law and Foreign Affairs*, 26, 35. [↗](#)
- Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, 4, 100005. [↗](#)
- Taneja, S., Shukla, R. P., & Singh, A. (2024). Navigating the intersection of ethics and privacy in the AI era. In *Ethical Marketing through Data Governance Standards and Effective Technology* (pp. 154–166). IGI Global. [↗](#)
- Feldstein, S. (2019). *The global expansion of AI surveillance* (Vol. 17, No. 9). Washington, DC: Carnegie Endowment for International Peace. [↗](#)
- Fontes, C., Hohma, E., Corrigan, C. C., & Lütge, C. (2022). AI-powered public surveillance systems: Why we (might) need them and how we want them. *Technology in Society*, 71, 102137. [↗](#)
- Kumar, D., & Suthar, N. (2024). Ethical and legal challenges of AI in marketing: An exploration of solutions. *Journal of Information, Communication and Ethics in Society*, 1, 45–46. [↗](#)
- Schwartz, R., Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence* (Vol. 3, pp. 12–15). US Department of Commerce, National Institute of Standards and Technology. [↗](#)
- Cowgill, B., & Tucker, C. E. (2019). Economics, fairness and algorithmic bias. *Journal of Economic Perspectives*, 21, 44–45. [↗](#)
- Nazer, L. H., Zatarah, R., Waldrip, S., Ke, J. X. C., Moukheiber, M., Khanna, A. K., ... & Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digital Health*, 2(6), e0000278.
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 2053951719897945. [↗](#)
- Tong, J. (2022). Artificial intelligence is not a panacea: policing content on social media platforms, three dilemmas and their ethical and legal

implications. In *Artificial intelligence and the media* (pp. 123–147). Edward Elgar Publishing. [↗](#)

McKay, C. (2020). Predicting risk in criminal procedure: Actuarial tools, algorithms, AI and judicial decision-making. *Current Issues in Criminal Justice*, 32(1), 22–39. [↗](#)

10 Big Data, Privacy, and the Rights of Senior Citizens under the Digital Personal Data Protection Act

Shayut Pavapanunkul, Aditi Srivastava, and Ritu Gautam

DOI: [10.1201/9781003668657-10](https://doi.org/10.1201/9781003668657-10)

10.1 INTRODUCTION

In the Data Breach Report of IBM Security, the Vice President (Technology) at IBM India and South Asia, with the above quote highlighted, states that phishing is the most common form of attack in India. In 2023, the cost of average breaches reached approximately ₹17.9 crore, representing a substantial 28% increase since 2020 and setting a new high record. This report sheds light on a worldwide contrast in how businesses plan to manage the increasing expenses and frequency of data breaches. While 95% of the surveyed organizations faced multiple breaches, the study highlighted different approaches 57% sought to pass on incident costs to consumers, while 51% aimed to boost investments in security measures. During this timeframe, there was a notable 45% surge in expenses associated with detection and escalation, which constituted the largest portion of breach costs. This significant uptick indicates a shift toward more intricate breach investigations, as researched in the report. Concerning the Indian scenario, phishing emerged as the primary form of

attack, contributing to almost 22% of breaches, closely followed by incidents involving stolen or compromised credentials at 16%.

“In 2021 Cybercrime with a significant surge of 24.4% (65,893 cases) rose from 52,974. Where predominantly, fraud were 64.8% of total, followed by extortion 5.5% and sexual exploitation 5.2%.” The crime rate category escalated from 3.9% to 4.8% in 2022 (National Crime Records Bureau, 2022). The report has singled out “social engineering” to be the causation factor, even though expensive to commit these breaches. The average cost of ₹19.1 crore was lost due to malicious threats, due to adoption of artificial intelligence (AI) and automation for security, a marked improvement was observed in identifying and containing breaches at the earliest. Nowadays, companies in India are extensively using these technologies as a 153-day trial, to decrease the data breaches lifecycle of 225 days versus 378 days (Cost of a Data Breach Report, 2022). Notably, firms globally have extensively integrated AI automation security, reducing nearly ₹95 million less in data breach costs as compared to those not utilizing these private cloud and on-premises environments. This showcases attackers’ ability to compromise multiple platforms stealthily ([Gleiss et al., 2023](#)). The breach of data across these varied environments took an average of 327 days to detect ([Yamout et al., 2023](#)) and contain tools, showcasing their significant impact on reducing breach-related expenses. Approximately 28% of investigated data breaches in India resulted in data loss across diverse settings like public cloud, “Big data ([Espinoza & Aronczyk, 2021](#)) analytics involves using advanced analysis methods on vast and varied datasets that include structured, semi-structured and unstructured data from diverse sources such as collected by artificial intelligence (AI), mobile devices, social media and the Internet of Things, audiovisual content, networks, logs, transactions, web, real-time data at an immense scale. This dynamic space offers both opportunities and challenges for organizations aiming to leverage data for insights and decision-making” (IBM Data Breach Report, 2023). The Digital Personal Data Protection Act not only enhances the protection of personal data but also provides a legal mechanism to balance individual privacy rights with

state interests in ensuring security and transparency in the cyber domain. The paper further explores how emerging technologies such as blockchain, AI, and machine learning can be employed to protect the personal data of seniors.

The updated cybersecurity statistics for 2023 portray a concerning landscape with an alarming 2,200 cyberattacks occurring every 39 seconds in a day. The huge financial impact of data breaches in the United States has taken a toll of \$9.44 million approximately per incident; the forecast prediction is that this number will rise exponentially, estimating a staggering \$8 trillion by the year 2023. Escalating financial ramifications associated with cybersecurity breaches, urging businesses and individuals to prioritize robust cybersecurity measures. Although the term “data breach” is sometimes used interchangeably with “cyber-attack,” it is important to note the distinction. Not every data breach is a cyber-attack and vice versa. Data breaches specifically refer to incidents where the confidentiality of information is compromised. For instance, a ransomware attack aimed at obtaining confidential company data for blackmail purposes would qualify as a security breach. When data on seniors is obtained by third parties for analytics to understand them as a segment, it becomes imperative to identify the major challenges (Saleous et al., 2023) of collecting and managing big data for implementing analytics and forecasting the trends, need to address key issues such as related to introducing special techniques, tools to manage enormous volume of data’s storage, processing and analyze as traditional data methods may be insufficient for handling such diverse datasets due to variety of formats which can be differentiated as “structured,” “semi-structured,” and “unstructured” integrating diverse speeds in a real-time influx of data and to make the outcome to be reliable or accurate and consistency, which is very time-consuming and resource-intensive. Data security and privacy ensure that sensitive information is protected and that collection practices comply with regulations as per Cyber laws ([Trein & Varone, 2023](#)).

The digital infrastructure to handle the growing demands of big data requires the application of scalable technologies to accommodate expanding

data from multiple sources to create a unified and comprehensive quality. Thus, governance policies for ensuring security are an important aspect, as little inconsistencies can lead to graver misuse and increase the cost of processing by organizations, which have a different challenge of data ownership, control, and access of departments or external sources. These challenges must be addressed with a combination of technology, adherence to best practices in big data analytics, and robust data management strategies. Organizations must stay always adaptive to evolving technologies and methodologies to derive meaningful insights from their data. The most positive forecast is that the elderly population in India is anticipated to constitute over 36% of the total population, affirmed by the findings of the World Population Prospects 2022 report, which predicts that the proportion of senior citizens will surge to 36.1% of the Indian population, encompassing a populace exceeding 550 million people.

The share of senior citizens aged 60 years and above will escalate to 20.8% of the population, approximately 347 million individuals. This marks a substantial increase from the 2022 demo Figure, which accounted for 149 million elderly persons, constituting about 10.5% of the country's population.

India Ageing Report, 2023

And, this report also indicates that the elderly population shall increase and simultaneously, the population of the young in India shall decline by 2050, the demoFigure ic size of the elderly in India will surpass that of children aged 14 and below, Senior citizens worldwide will emerge as vulnerable demoFigure ic to cyber victimization but, also “*Silver Economy*” (Ministry of Social Justice and Empowerment, 2021) and as a segment having huge potential to be explored but, only hurdle with this steady growth is a notable upsurge in incidents of their cyber victimization in recent times. As it is becoming imperative to not inclusively participate in digitalization, a new way of life has its adverse effects, both financially and socially. The rising prevalence of cyber threats targeting seniors prompts an urgent need for

heightened awareness, protective measures, software-supported devices specifically designed for aged consumers, and community support systems to safeguard this susceptible group from the detrimental consequences of online exploitation ([Ghosh & Banerji, 2024](#)). As the aging population continues to grow, addressing the unique challenges faced by seniors in the digital realm becomes increasingly crucial to mitigate the far-reaching impacts of cyber victimization on their financial well-being, as the data gathered for analytics infringes the privacy and leads to theft to commit cybercrimes and hinder broader societal dynamics.

10.1.1 UNFOLDING OF DIGITAL PERSONAL DATA PROTECTION ACT, 2023 WITHIN CYBER FRAMEWORK

The Act represents a substantial advancement in safeguarding digital personal data, aiming to strike a delicate balance between protecting individual rights and enabling lawful data processing. At its core, it underscores the responsibilities of those managing data, termed as “Data Fiduciaries,” and the rights of individuals, referred to as “Data Principals,” with financial penalties for breaches. This legislation aims to introduce a strong data protection law with minimal disruption while fostering an environment conducive to both living and doing business effortlessly. Fundamentally, the Act revolves around seven key principles, each crucial in ensuring the careful handling of personal data. It begins with the foundation of consent, legality, and transparency in data usage, progressing to emphasize the significance of limiting data use to specific purposes, minimizing collected data, ensuring accuracy, restricting storage duration, implementing reasonable security measures, and establishing accountability for breaches. These principles form the framework for responsible data handling, placing a premium on individual privacy and expecting ethical conduct from entities managing personal information. Furthermore, by criminalizing the anonymization of data, it aims to safeguard the privacy of individuals and prevent attempts to identify them from datasets lacking personal identifiers. This initiative upholds privacy rights and ensures the

reliability of anonymized data by outlining different levels of protection for various categories of personal information (Bengaluru Bureau, 2030).

The Digital Personal Data Protection Act of 2023 grants vibrant discretionary authority to the central government in unspecified situations. This nuanced evolution suggests reduced responsibilities for businesses and fewer specific safeguards for consumers. Although the recalibration aims for a more efficient approach, it also introduces certain uncertainties that might lead to broader consequences, which have to be taken into account in amendments. The acronym SARAL: Simple, Accessible, Rational, and Actionable Law summarizes it. Using clear, gender-inclusive language without complicated provisions or excessive cross-references, bestows a range of rights upon individuals such as *Access to Information, Rectification and Erasure, Grievance Redressal, Designating a Representative* in cases of incapacity or demise, granting them authority in their dealings with Data Fiduciary and, if dissatisfied, escalate their issues to the Data Protection Board for resolution, overseeing regulations, compliance, levy penalties, ensuring accountability, unidentified datasets, criminalizing actions linked to de-anonymizing individuals even to non-citizens as a more inclusive approach (Charter of Fundamental Rights, 2007).

The Honorable Supreme Court of India observed concerning *section 57 of the Aadhaar Act* that regarding the capabilities of metadata, on storage rules, metadata risks were overlooked, allowing the Indian government to collect and process personal data through Aadhaar without acknowledging metadata's dangers. In Act's "**Sec 2(h)** 'data' means a representation of information, facts, concepts, opinions or instructions in a manner suitable for communication, interpretation or processing by human beings or by automated means" and **2(x)** "processing" about data, where "wholly or partly automated operation" or "set of operations performed on digital personal data" and "all operations such as collection, recording, structuring, storage, retrieval, using, combination, indexing, sharing, disclosure by transmission, dissemination or otherwise making available, all kinds of restriction, erasure or destruction." The obligations on "*significant data*

fiduciaries,” adhere to additional measures such as to appoint “data auditor” for regular “Data Protection Impact Assessments,” and compulsory registration within India’s jurisdiction to ensure heightened data protection. This Act is motivated to safeguard children’s data, making it a necessity to explicitly take parental consent to prevent endangerment of children’s well-being, such as tracking, behavioral monitoring, or targeted advertising.

While maintaining a stringent stance, the Act offers exemptions for specific circumstances, including matters related to security, research, start-ups, enforcement of legal rights, judicial functions, crime prevention, processing data of non-residents who are binded due to foreign contracts, mergers or demergers. These exemptions aim to strike a balance between data protection and societal needs, recognizing the complexities inherent in various situations, together with handling their information responsibly and ethically. The Act imposes stringent penalties up to Rs. 2.5 billion (*approximately USD 30 million*) for non-compliance, for example, to implement reasonable security measures for preventing data breaches. However, there is an absence of a provision for compensating “data subjects,” raising concerns regarding redressal. The vast spectrum of Indian industries falls under its purview due to digital storage and payment information, but imposes fewer obligations as compared to the General Data Protection Regulation (GDPR), a factor that businesses subject to this law perceive positively (Digital Personal Data Protection Act, 2023).

On “*data localization*” the government’s potential to limit data flows to specific countries via notifications, though not explicitly outlined, is for national security reasons. Additionally, the law specifies that these regulations will not hinder actions by sector-specific agencies that have already been imposed or may impose data localization rules, for example, Reserve Bank of India’s existing data localization requirements will maintain their legal validity despite the broader changes brought by the new law. Two specific types of data “sensitive” and “critical” will need stricter protocols for specific data subsets. “*Big data*” is using information for objectives unrelated to its initial collection, analyzing and processing data to deliver personalized services, for example, surveying shopping habits in

a specific region, credit evaluation, healthcare, personal consumption, work activities, tracking movements, online behavior, electronic transactions, smartphone usage, and monitoring communications or videos on applications. Thus, the principles of the United Nations System Organization (Trendmicro, 2016) “*fair, ethical and legitimate processing,*” “*handling of big or personal data,*” “*consent of data subject,*” “*explicit consent might not be available,*” *precision and transparency* to define why data is collected and how it is used? “*Legitimate uses,*” “*Collection limitation principle*” (based on curbing malicious and illegal uses of individual data), *Specific Purpose* clear, *Proportionality/Necessity*, “*Correspondence*” ensure a balance between the business organization’s mandates and the rights or interests of the individuals.

European Court of Human Rights in *Tietosuojavaltuutettu v. Satakunnan Markkinapörssi and Satamedia*, stressed the importance of finding an equilibrium between “freedom of speech” and “privacy.” In *S. and Marper*, retaining personal data for an excessive duration is inconsistent with the characteristics of a democratic society guided by the rule of law. In *Netherlands v. Parliament*, the necessity of upholding fundamental rights as “human dignity.” In *the Digital Rights Ireland case*, necessary data should not be stored for longer than strictly necessary. In 1983, the *Population Census Decision* acknowledged individuals rights to control the flow of their data, intertwined – privacy rights and protection of data needs harmonizing rules and eliminating current fragmentation in national systems, and reducing unnecessary administrative complexities.

10.1.1.1 Right to Privacy of Senior Citizens

Invasion of privacy is a self-defense against deception.

Richard Posner

In 2017, the Honorable Supreme Court ruling in *Justice K.S. Puttaswamy and Anr. v. Union of India and Ors.* Acting as a catalyst to bring forth the “*right to privacy*” contained within Article 21 as an innate component of

the fundamental right to life in India. Emphasizing the critical role of “informational privacy ([Dixon, 2023](#)),” the judgment laid the foundational groundwork for safeguarding such right within the nation’s legal structure, despite acknowledging that the court refrained from prescribing specific parameters for protecting informational privacy. This spurred discussions to explicitly address data protection concerns and the gap between rights to privacy and a concrete infrastructure. In 2018, a bill proposed by the Justice B.N. Sri Krishna committee under the Ministry of Electronics & Information Technology and Honorable Prime Minister Shri Narendra Modi established data protection standards, heavily relying on the globally accepted European Union’s GDPR. The predecessor bill “*Personal Data Protection Bill of 2019*” delayed due to the pandemic was retracted by the government, and replaced with a revised new draft “*Digital Personal Data Protection Bill of 2022*,” which amalgamates elements of the revised draft and evolving perspectives, pivotal to present data security issues.

The government’s subsequent withdrawals reflect a responsive approach, to address potential shortcomings found in earlier versions, keeping strict adherence to the Indian Constitution *Article 19*, which bestows the freedom to individuals as a right to determine their preferences in alignment, *Article 21*, which grants liberty, the freedom to choose across life’s facets – food, attire, faith and privately one’s thoughts and *Article 25* provides freedom of religion inherently includes the right to choose and express faith. Privacy is a cornerstone for freedom, crucial in exercising individual liberties, involving provisions enabling the government to request “*data fiduciaries or processors*” to provide “*anonymized personal data*” or “*other non-personal data*.” This data access is aimed at enhancing governance and optimizing citizen welfare services by facilitating better targeting and administration.

The evolution of human rights in response to the rapid progress of technology and intelligence marks an intricate journey such as the evolution of a new class of rights termed “*Epistemic Rights*” (*Michel Foucault, 1994*) drawn from the Greek notion of “*understanding*” aimed at fortifying existing human rights as “*Fourth Generation Rights*” to safeguard the roles

and responsibilities of those involved in the creation and dissemination of knowledge “*Digital life Worlds 2.0 and 3.0*” borrowed from the German word called “*Lebenswelt*,” commonly used in phenomenology, refer to the contemporary experiential landscapes shaped by electronic devices, interconnected systems and the pervasive use of numerically encoded data. These life worlds are characterized by the dominance of digital technology, where interactions, information, immediate experiences, activities, and relationships collectively shape an individual’s perception of the world and societal structures revolving around digital interfaces, networks, and data-driven processes. This concept transcends conventional rights, emphasizing the protection and recognition of the unique cognitive faculties intrinsic to humans. Nowadays, human beings are intertwined with advanced machines and big data ([Espinoza & Aronczyk, 2021](#)) in such intricate ways to define our present reality of unexplored domains. About 83% of companies have faced multiple data breach incidents with an average cost of around 4.35 million dollars lost due to a data breach.

10.1.1.2 Recent Data Theft Cases in India

As per the Universal Declaration on Human Rights’ definition, “*Data protection refers to the practices, safeguards and binding rules put in place to protect your personal information and ensure that you remain in control of it.*”

A disturbing trend is observed in India, which is the most targeted nation, in comparison to other nations globally. Among 829 million cyber-attacks detected or thwarted worldwide in 2022, nearly 59% targeted India, specifically ransomware (*Indus Face, a Tata Capital-Backed SaaS Security Firm, 2023*). Prevalent and damaging data breaches not only pose immediate risks to individuals but also erode trust in institutions responsible for safeguarding this sensitive data in various sectors. Data breaches affect organizations across various sectors, including governmental, private, and non-profit entities of different sizes. However, the sectors most profoundly affected are finance, healthcare, and public services. The most notable data breaches in India from 2019 were the Police recruitment exam, Just Pay,

cyber-attacks on Government websites, Air India, Common Admission Test, Upstox, Delhi State Health Mission database, Domino's Pizza, and Aadhaar.

In the fiscal year 2022–23, as per the annual report of the Reserve Bank of India (RBI), the digital payment sector experienced a concerning rise in data fraud within banks. The reported cases of data fraud amounted to 13,530, significantly surpassing the 9,097 cases recorded in the preceding fiscal year 2021–22.

Reserve Bank of India

Recent cases due to Agent Tesla Crypt Ransomware strain, an insidious software possessing multifaceted capability, infiltrates via various exploit kits crafted to identify and capitalize on vulnerabilities present in web servers. Through the deployment of this virus, the targeted systems compromise the security of user's workstations, capitalizing on multiple avenues of entry to maximize its reach and potential impact to adapt to a range of vulnerabilities across different platforms. It is in the lineage of Crypto Locker, to compromise the security landscape, posing a significant threat to the users' digital environments with the primary objective of surreptitiously acquiring personal data and transmitting it to a designated Command and Control (C2) server from web browsers, email clients and FTP servers, providing a broad spectrum of potential data points to exploit. It also captures screen and video recordings and documents a user's digital activities, thereby compromising sensitive information and privacy. Malware can also intercept and record data residing in the clipboard, magnifying its scope for capturing confidential snippets, and adeptly records form values, further exacerbating the compromise of private information ([Tikk-Ringas, 2016](#)).

This malware has emerged as a formidable threat in the realm of cyber threats, employing a sophisticated arsenal to transmit personal data, potentially causing severe privacy breaches and security compromises for unsuspecting users. “Deepfakes” is a recent weapon of cybercriminals, with

78% of attacks overall, 66% of cybersecurity professionals have given positive assent to deepfake attacks with a 13% increase in crime, and email being the predominant methodology of committing the crime. Utilizing advanced AI and machine learning, “*deepfake technology*” allows the creation of highly realistic video, audio, images, or text, mimicking the appearance and voice of real individuals.

This sophisticated capability poses a significant risk, enabling cybercriminals to engage in “social engineering” tactics, tricking individuals into divulging sensitive information, participating in financial scams or gaining unauthorized access to corporate networks. The Federal Bureau of Investigation has categorized these attacks as “*business identity compromise*” as a distinct category in cybercrime. Recent incidents employ deepfake holograms on Zoom calls to deceive cryptocurrency executives and utilize real-time voice cloning to orchestrate a \$35 million bank transfer scam. Another tactic involves creating false personas, using deepfake technology and stealing personal information to infiltrate companies by posing as remote employees. To mitigate these risks, the Federal Bureau of Investigation recommends proactive education for both personnel and clients to recognize and respond to deepfake threats. Training should include identifying suspicious behaviors and ceasing information sharing in cloned voices or videos ([Confessore, 2018](#)).

The US Department of Homeland Security suggests specific indicators to determine the authenticity of images, videos, or audio content. As deepfake usage continues to grow, addressing these emerging risks becomes paramount to avoiding potential financial and operational consequences for targeted firms. Highlighting the menace posed by AI-generated deepfakes, cyber criminals produce a video utilizing the likeness and voice of a former IPS officer from the Uttar Pradesh police to coerce a senior citizen. Fearing legal consequences, he succumbed to the scammers’ pressure and made multiple payments under the mistaken belief that the offender was in a position of power. The scammers exploited the victim’s vulnerability by issuing threats of police intervention, leveraging an image they possessed that falsely portrayed the senior citizen as engaging in illicit activities. This

manipulative tactic coerced the elderly man into compliance, ultimately leading to financial extortion through a fabricated narrative orchestrated by the criminals.

10.2 METHODOLOGY

The data is collected by using a Google form as a questionnaire, circulated among senior citizens as a sample population from Delhi and the National Capital Region. The quantitative approach relates to the positivist approach, so that this exploratory paper can be combined with other multidisciplinary approaches. The research questions of this paper explore the challenges faced by senior citizens with digital migration, their participation online, and infringement of privacy, data theft, due to analytics used by third parties, and exposing that data to cyber offenders ([Shahriar et al., 2023](#)). Systematic Literature Review is adopted as the method of collecting data, identifying the highlighted points, and critical appraisal of research. Due to this being a Social science, the overall effect can be determined in later research, taking an estimated population.

10.2.1 THE SURVEY DELVED INTO THE FOLLOWING RESULTS

The survey form's data from senior respondents, as in [Figure 10.1](#) shows that 33.9% of respondents have less than 10,000 monthly income, 30.4% more than 50,000, and remaining 17.9% between 11,000 and 50,000 income.

56 responses

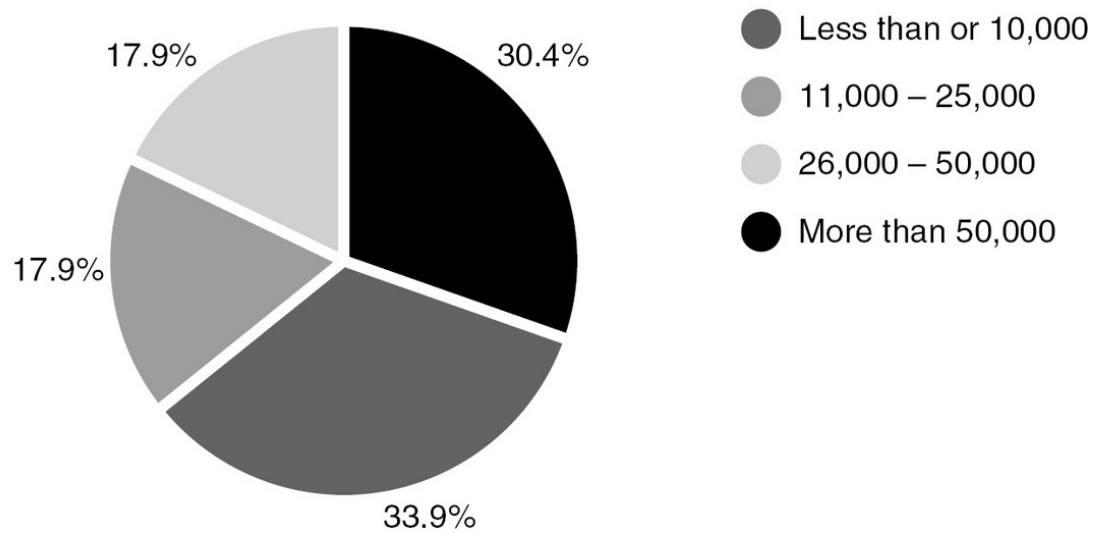


FIGURE 10.1 The financial backup of seniors. [↗](#)

The survey form's data from senior respondents, as in [Figure 10.2](#) represents that at least 66.1% of seniors are independently handling digital transactions, 26.6% are dependent on family members or servants and the remaining take help from bank staff or strangers based on good faith.

56 responses

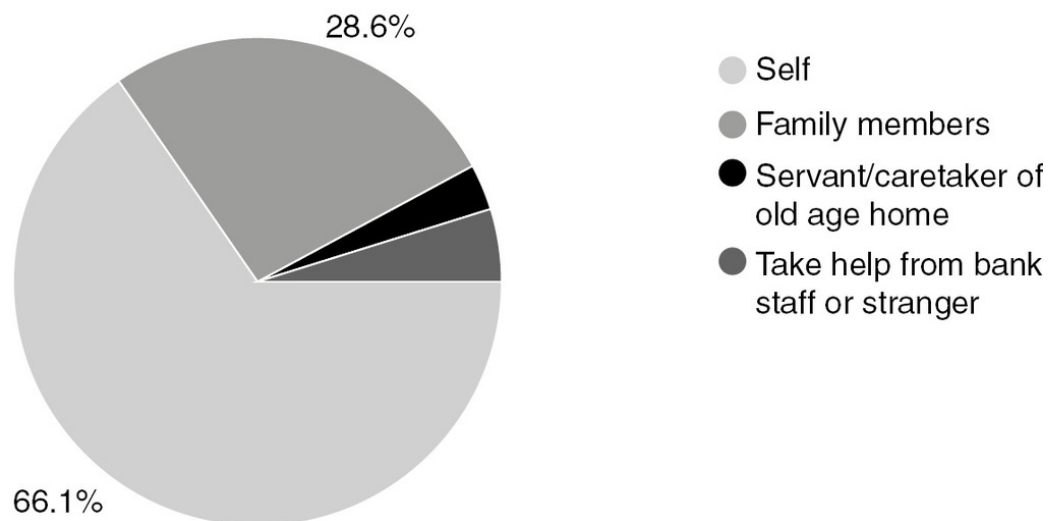


FIGURE 10.2 Digital transaction by seniors. [↗](#)

The survey form's data from senior respondents as in [Figure 10.3](#) represents that 62.5% use WhatsApp/Telegram, 37.5% use Instagram, 35.7% Face book/Messenger and only 1.8% Tik Tok or their data is exposed to cyber space via these applications.

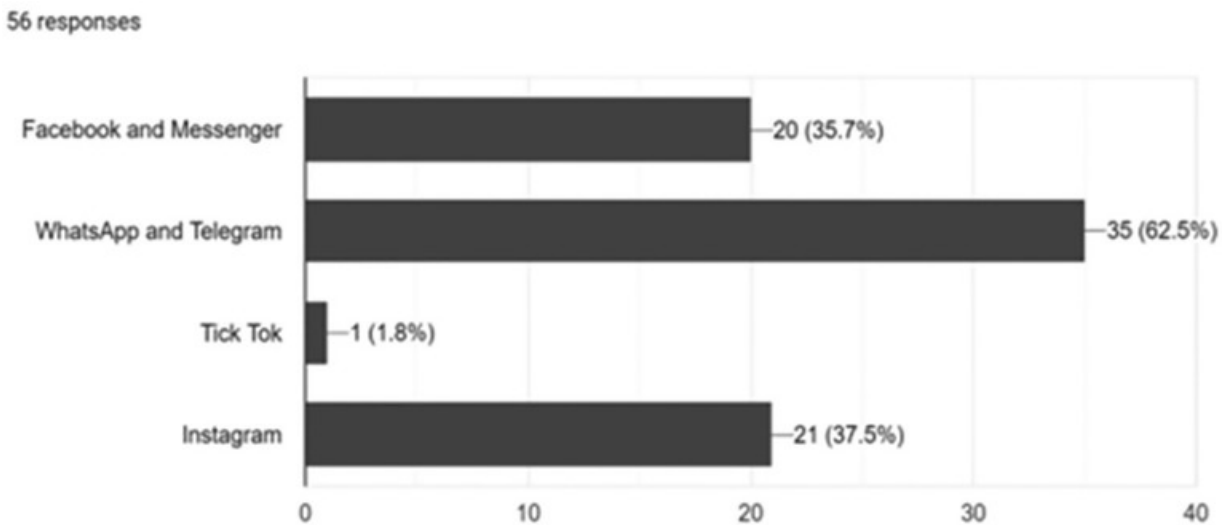


FIGURE 10.3 Most used social media platform by seniors. [↗](#)

The survey form's data from senior respondents as in [Figure 10.4](#) represents that at least 37.5% almost every day, received unknown calls, pop-ups of ads and cookies for requesting acceptance from websites when they were online.

56 responses

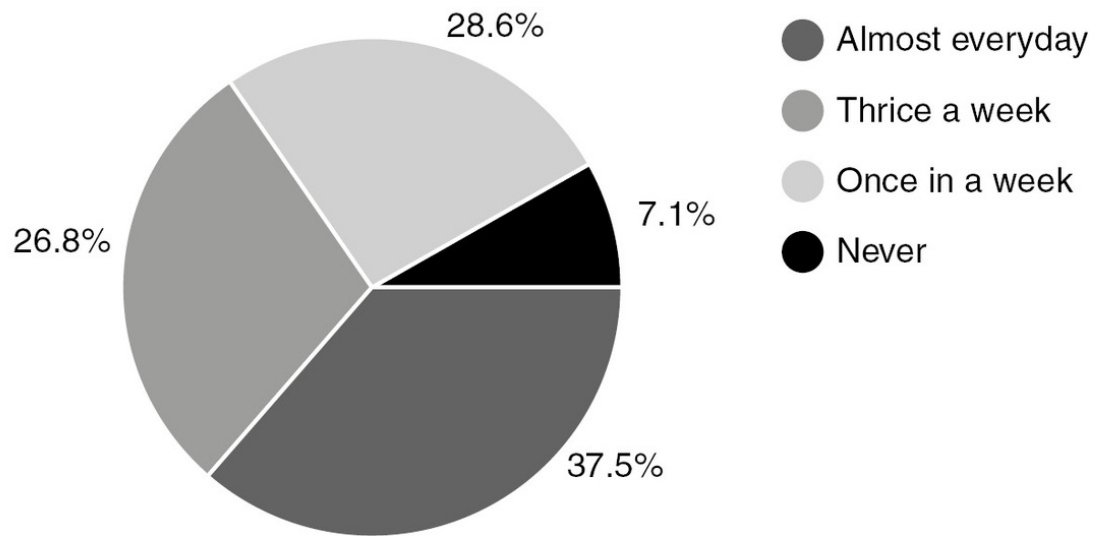


FIGURE 10.4 Compromised data of seniors. [↗](#)

The survey form's data from senior respondents as in [Figure 10.5](#) represents that at least 60.7% were comfortable with the Patym (Pay through Mobile) application, 44.6% used UPI (Unified Payment Interface) and 50% used Gpay (Google Pay), 25% did Net Banking and 3.6% used none

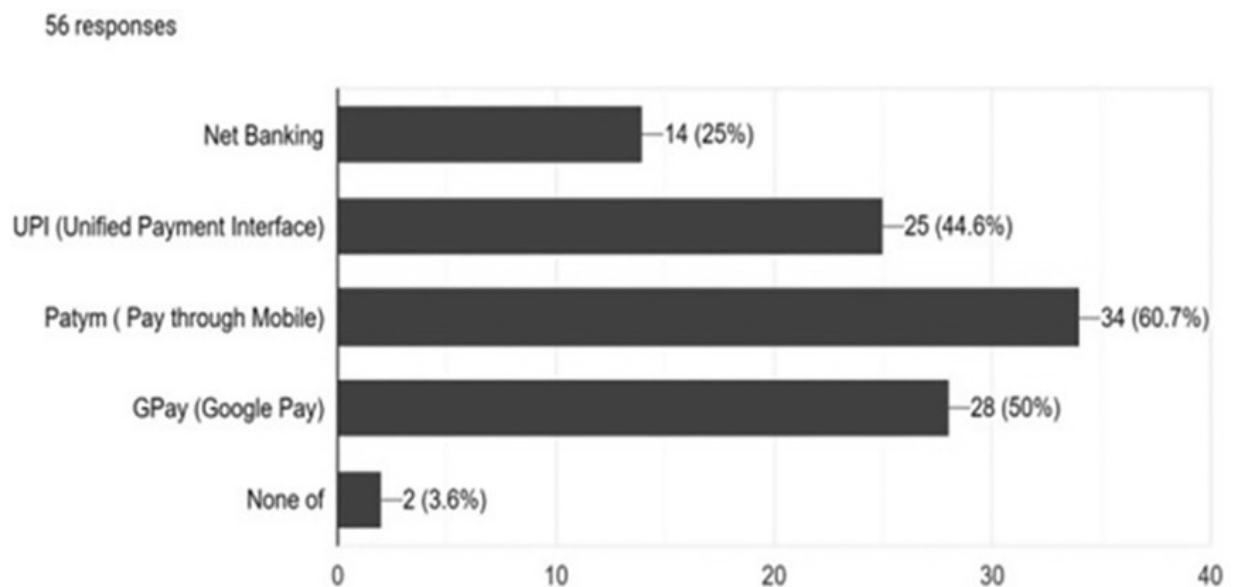


FIGURE 10.5 Convenient financial platforms for seniors. [↗](#)

The survey form's data from senior respondents as in [Figure 10.6](#) represents that at least 78.6% have received calls from individuals misrepresenting as Government or bank employees to commit some fraud or cybercrime such as Grandparent's scam.

56 responses

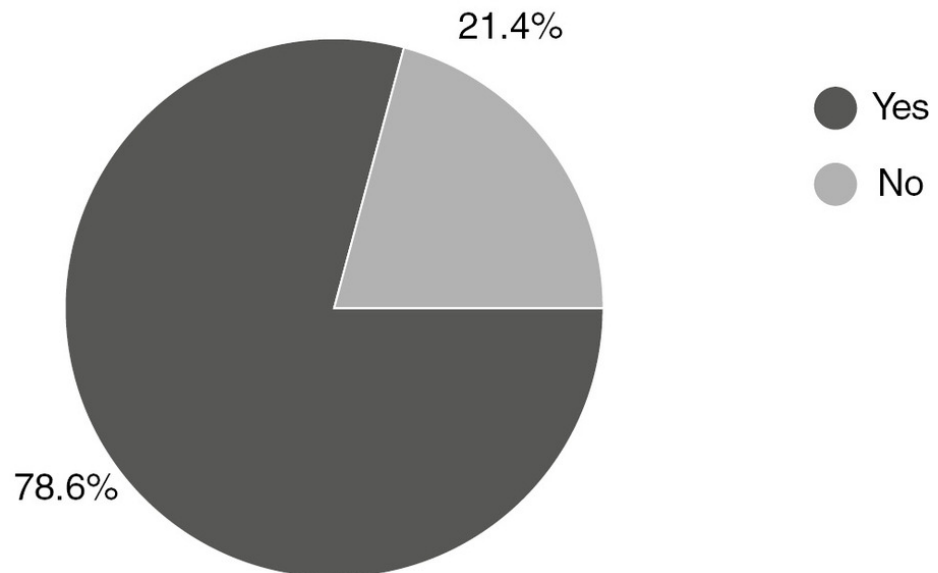


FIGURE 10.6 Invasion of privacy of seniors and data theft. [↗](#)

The survey form's data from senior respondents as in [Figure 10.7](#) represents that at least 82.1% are aware of the Right to Privacy, Digital Personal Data Protection Act, 2023, Digital Signature Consent "By a Click" and 17.9% were not aware.

56 responses

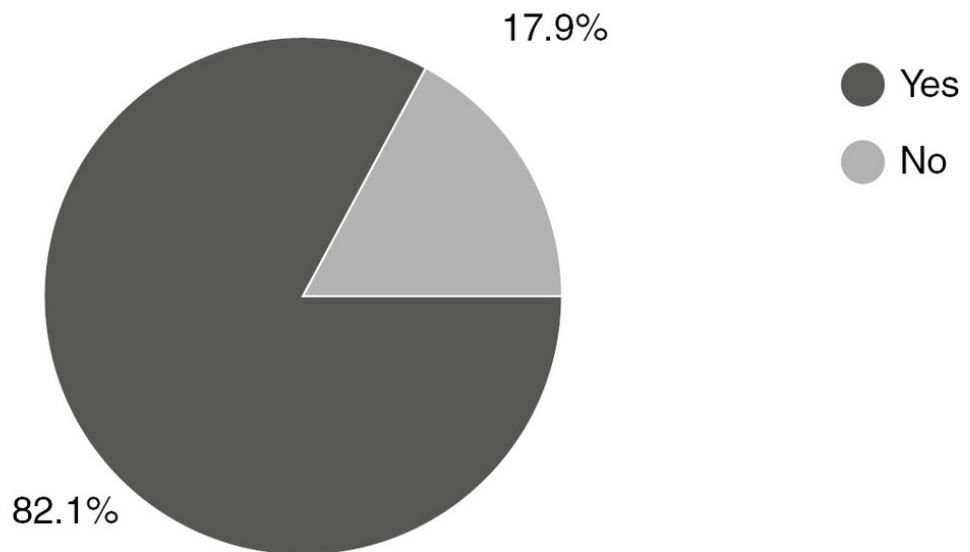


FIGURE 10.7 Digital awareness among seniors. [🔗](#)

The survey form's data from senior respondents represents that 75% just aware of legal laws such as the Information Technology Act, 2008 and Digital Personal Data Protection Act, 2023 without basic knowledge of what is written in them, to protect their digital rights or how to seek a remedy in court, and 25% were unaware ([Figure 10.8](#)).

56 responses

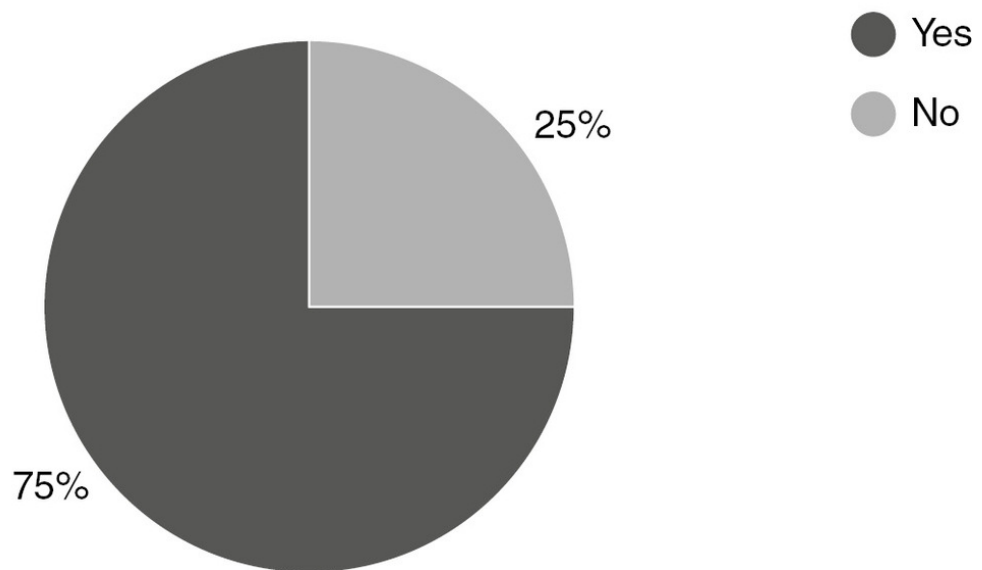


FIGURE 10.8 Legal awareness of senior citizens to protect from cybercrimes. [📄](#)

10.3 DISCUSSION ON FINDINGS

The research paper is set to investigate the challenges faced by seniors' of the Delhi NCR region concerning digital literacy, the scope of victimization and infringement of the right to privacy due to data theft as a drawback of big data analysis and understanding the legal import, social inclusion. It also addresses financial independence, accessibility, skill sets and awareness, legal capital, and experiences in the use of digital technology. The focus on computers, smartphones/tablets, mobile phones, and internet accessibility skills in knowledge society but also due to the importance of this study, to highlight what translates to maximum social inclusion or not in the society, is beneficial to the seniors, who have limited mobility.

The data shows that there is no significant disparity between both the genders, although females were not equally sampled yet, respondents were within the age group of 55–80 as valid sample of 56 senior citizens. Most of the respondents were graduates or held postgraduate education, with 33.9%

having less than 10,000 monthly income, 30.4% more than 50,000 and 17.9% having income between 11,000 and 50,000. Raising the possibilities of acquiring skills needed for use of digital media. Some males among them were retired private, government, and civil servants, while the majority of females among them were professionals or homemakers. The seniors are in moderate health conditions with declining memory power. Fragility does not depend on age alone but on all factors of an improved lifestyle, to affect their capacity for learning digital skills and remember passwords to secure their data. The 37.5% of seniors have reported receiving calls almost every day, some unknown calls requesting acceptance of insurance, loans, medicine purchases, consumer services and goods, and website pop-ups of ads and cookies as they go online. At least 60.7% were comfortable with the Patym (Pay through Mobile) application, 44.6% used Unified Payment Interface (UPI) and 50% used Google Pay (Gpay).

Thus, these platforms must accommodate more senior users in the future and make their interface suitable not just for the young generation but for migrating seniors with easy use of secured AI technology. Invasion of privacy is felt by 78.6% as they received calls from cyber criminals misrepresenting Government or bank employees to commit cybercrime. And, even though 82.1% were aware of “Information Technology Act, 2008,” Right to Privacy, “Digital Personal Data Protection Act, 2023,” Digital Signature and withholding Consent “By a Click” but 42.9% were not aware of the legal recourse even though 46.4% were victims of Phishing Financial and Customer Service Frauds, 37.5% of identity theft, 30.4% received obscene messages and 21.4–26.8% were victims of other cybercrimes. Seniors have reported in the central government’s cybercrime portal and Delhi cyber cell but, due to limited personnel, over burden of cases or less sum of money stolen (as per cyber police) their complaint, is still pending for recovery and they had very little knowledge of pursuing the case in law of court and feared the judicial process’s time of final adjudication in such matters. Senior citizens use digital media reasonably, accessing to a smartphone, internet but still lack confidence and trust in online banking, as they are aware of financial fraud but feel helpless due to

limited knowledge to reaffirm their position post cyber fraud victimization as technology usage, digital inequality.

Despite being educated and retired from a life of employment, respondents as majority did not extensively use digital technology in their “professional and private activities” and felt less adaptable digital media and its gadgets, due to generational identity where a more planned, without “trial and error” approach would give solutions to this challenge. Due to memory loss, seniors forget how to use digital tools for participation in cyberspace. Family and friends of the seniors can influence and help in the adoption of digital media technologies and stay engaged and socially, the rate of adoption and use of media depends on the need the users want to satisfy after retirement. However, seniors are yet to maximize the potentiality of the online media for business transactions as part of digital economy.

10.4 CONCLUSION

The second research question is understood by studying the various aspects of the Act, such as “Privacy and Digital Rights” can be secured by recent advancements within cybersecurity by introducing the Digital Personal Data Protection Act, 2023. Regarding “*legitimate interests*,” “*unconditional*,” Act does not align with the European Act ([Chaturvedi & Sinha, 2017](#)). Aspects like “*specified purposes*” is without guidance, on what constitutes to be a “legitimate processing.” The exemptions for government bodies in the Act’s applicability will necessitate judicial intervention to set reasonable criteria. The English language is dominating digital space, and other languages pose potential challenges for data fiduciaries in regard to notice. The Act’s has a limited application to foreign personal data processed in India, and this could hinder India’s efforts to obtain adequacy rulings from entities like the European Union and impacting cross-border data flow assurances.

The Act’s stringent data breach notification requirements, for example, short 6-hour timeline for notifying CERT-In, could be impractical in real-

world scenarios, posing challenges for compliance, limited powers delegated to the Data Protection Board of India lacks authority for delegated legislation, restricting its role mainly to adjudication rather than facilitating dynamic policymaking and guidance necessary in this evolving landscape of data privacy and technology, amendments in Act will impact “*Right to Information Act*” “shifting the focus on personal data over public interest, tilting the balance in favour of data protection at the expense of transparency.” This Act needs refinement to ensure effective implementation and alignment with global standards. Also, due to the fast evolution of cyberspace a new concept of “Brain wave data” various aspects of brain function, including attention, memory, emotion regulation, and cognitive workload, which has applications in neuroscience research, clinical diagnostics, neurofeedback therapy and brain-computer interface technology.

These bio signal sensing devices offer promising opportunities for healthcare, research and technology applications, it will become crucial to address data privacy concerns through amendment in Data Act, clear consent mechanisms, ethical considerations, compliance with relevant regulations as “Biometric Data” ([Foley, 2023](#)) is highly sensitive as it can reveal intimate details about an individual’s mental state, emotions and even health conditions, posing a risk that “unauthorized access” to this data could lead to “privacy breaches and potential misuse.” A stringent cybersecurity measures must protect this data from unauthorized access ([Calzada, 2021](#)), tampering or breaches such as encryption, storage protocols, and access controls. To mitigate privacy risks, bio signal data must be de-identified, meaning before being shared or used for research purposes. Depending on the jurisdiction where data is being collected, processed, and stored by third parties, there may be specific regulations governing these steps of bio signal data, for example, EEG data. Compliance now will be as per Digital Personal Data Protection Act of 2023, GDPR of Europe and for ethical considerations consent, autonomy, discrimination, or stigmatization based on the analysis of such data is covered within this Act.

Expansion of Digital Earth applications has surged the advancement of AI systems and their ability to handle vast volumes of geospatial data. This synergy has unlocked opportunities for various sectors such as urban planning and environmental monitoring, to disaster management and precision agriculture. The integration of AI into Digital Earth frameworks has facilitated the automation of complex tasks, enabling more efficient data collection, analysis, and interpretation ([Sharon, 2021](#)). However, amidst the optimism surrounding these technological advancements, there is a growing recognition of the challenges and risks associated with the widespread deployment of AI systems within the Digital Earth domain such as ethical implications of AI's decision-making processes, as AI algorithms become increasingly sophisticated, there is a risk of bias, discrimination and opacity in the outcomes they produce ([Yamout et al., 2023](#)). Issues such as algorithmic fairness, accountability, and transparency have come to the forefront, prompting calls for greater scrutiny and regulation of AI technologies. This growth of AI systems and their potential societal impacts, particularly in the context of power asymmetries within a “datafied society” (Computer Law & Security Review, 2020) will conflict with the current Act, which must be either amended to accommodate this unequal distribution of data, disparities in access, opportunities or decision-making processes, concentration of AI power in the hands of a few dominant actors, such as tech giants and government agencies, raises questions about data governance, privacy and surveillance ([Anand, 2023](#)). These challenges ([da Silva Cezar & Maçada, 2023](#)) are growing imperative to adopt a more critical and reflexive approach to the development and deployment of AI technologies within the Digital Earth ecosystem and a complex exhaustive “Digital law” specifically for India upholding ethics ([Huang, 2023](#)), human-centered principles into the design, implementation and governance of AI systems can validity accept the vulnerable group's interests ([Benavides & Licholai, 2022](#)). Initiatives such as the development of AI ethics frameworks, responsible AI guidelines, and regulatory mechanisms are essential steps toward mitigating the risks associated with AI-driven Digital Earth applications and also help users such as senior citizens who need

assistance due to age and other factors fostering greater transparency, accountability and engagement ([Bokolo, 2023](#)), which will serve the public interest and promote societal well-being of senior users and protect them from cybercrimes. Fostering a more inclusive and participatory approach to AI development and deployment, the complex challenges can be turned into assets ([Foley, 2023](#)).

REFERENCES

- Anand, A. (2023). Critical Success Factors for Enabling a National Data Exchange Platform. In *Technology, Policy, and Inclusion* (1st ed.). Routledge India. Retrieved from Indian School of Business. [↗](#)
- Benavides, X., & Licholai, G. (2022). *Healthcare Data Ownership and Privacy: A Perspective for Digital Therapeutics*. Chapman and Hall/CRC, pp. 2–10. [↗](#)
- Bokolo, A. J. (2023). Inclusive and safe mobility needs of senior citizens: Implications for age-friendly cities and communities. *Urban Science*, 7(4), 103. <https://doi.org/10.3390/urbansci7040103> [↗](#)
- Calzada, I. (2021). Data co-operatives through data sovereignty. *Smart Cities*, 4(3), 1158–1172. <https://doi.org/10.3390/smartcities4030062> [↗](#)
- Chaturvedi, A., & Sinha, A. (2017). *GDPR and India*. The Centre for Internet and Society, India. Retrieved from <https://cis-india.org/internet-governance/files/gdpr-and-india> [↗](#)
- Computer Law & Security Review. (2020). Information privacy, impact assessment, and the place of ethics. *Computer Law & Security Review*, 37, 105404. <https://doi.org/10.1016/j.clsr.2020.105404> [↗](#)
- Confessore, N. (2018). Cambridge Analytica and Facebook: The scandal and the fallout so far. *The New York Times*. Retrieved from www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html [↗](#)
- da Silva Cezar, B. G., & Maçada, A. C. G. (2023). Cognitive overload, anxiety, cognitive fatigue, avoidance behavior and data literacy in big data

- environments. *Information Processing & Management*, 60(6), 103482. [↗](#)
- Ministry of Electronics and Information Technology (MeitY). (2023). Digital Personal Data Protection Act, 2023. www.meity.gov.in/static/uploads/2024/06/2bflf0e9f04e6fb4f8fef35e82c42aa5.pdf
- Dixon, S. M. (2023). Exploratory analysis of future privacy issues and policy implications for consumer privacy in the cloud. Doctoral dissertation, Wilmington University. ProQuest Dissertations Publishing. Accession No. 30317426. [↗](#)
- Espinoza, M. I., & Aronczyk, M. (2021). Big data for climate action or climate action for big data? *Big Data & Society*, 8(1), 1–5. <https://doi.org/10.1177/2053951720982032> [↗](#)
- Foley, J. (2023, July 27). Apple's wild new patent could turn its AirPods into a health device. Retrieved from www.yahoo.com/lifestyle/apples-wild-patent-could-turn-200037372.html [↗](#)
- Ghosh, J., & Banerji, O. (2024). The socio-legal challenges of privacy issues in smart healthcare 4.0 ecosystem. *International Journal of Technology Policy and Management*, 24(2), 145–168. <https://doi.org/10.1504/IJTPM.2024.137824> [↗](#)
- Gleiss, A., Degen, K., & Pousttchi, K. (2023). Identifying the patterns: Towards a systematic approach to digital platform regulation. *Journal of Information Technology*, 38(2), 180–201. <https://doi.org/10.1177/02683962221146803> [↗](#)
- Huang, L. (2023). Ethics of artificial intelligence in education: Student privacy and data protection. *Science Insights Education Frontiers*, 16(2), 2577–2587. <https://doi.org/10.15354/sief.23.re202> [↗](#)
- IBM (2023). Data Breach. www.ibm.com/reports/data-breach
- Key4Biz. (2022). Cost of a Data Breach Report, 2022. www.key4biz.it/wp-content/uploads/2022/07/Cost-of-a-Data-Breach-Full-Report-2022.pdf
- Ministry of Social Justice and Empowerment. (2021) Home page of Department of Social Justice and Empowerment. <https://socialjustice.gov.in/>

- National Crime Records Bureau. (2022). Crime in India. ncrb.gov.in/uploads/nationalcrimerecordsbureau/custom/1701607577CrimeinIndia2022Book1.pdf
- National Cyber Security Centre. (n.d.). Yahoo data breach: NCSC response. Retrieved from www.ncsc.gov.uk/news/yahoo-data-breach-ncsc-response
- Shahriar, S., Allana, S., Hazratifard, S. M., & Dara, R. (2023). A survey of privacy risks and mitigation strategies in the artificial intelligence life cycle. *IEEE Access*, 11, 61829–61854. <https://doi.org/10.1109/ACCESS.2023.3287195>
- Sharon, T. (2021). Blind-sided by privacy? Digital contact tracing, the Apple/Google API and big tech's newfound role as global health policy makers. *Ethics and Information Technology*, 23(Suppl 1), 45–57. <https://doi.org/10.1007/s10676-020-09547-x>
- Tikk-Ringas, E. (Ed.). (2016). *Evolution of the Cyber Domain: The Implications for National and Global Security* (1st ed.). Routledge. <https://doi.org/10.4324/9781003422549>
- Trein, P., & Varone, F. (2023). Citizens' agreement to share personal data for public policies: Trust and issue importance. *Journal of European Public Policy*, 1, 1–26. <https://doi.org/10.1080/13501763.2023.2205434>
- Trendmicro. (2016.). EU General Data Protection Regulation (GDPR). Retrieved from www.trendmicro.com/vinfo/in/security/definition/eu-general-data-protection-regulation-gdpr
- Yamout, Y., Yeasar, T. S., Iqbal, S., & Zulkernine, M. (2023). Beyond smart homes: An in-depth analysis of smart aging care system security. *ACM Computing Surveys*, 56(2), 1–35. [https://doi.org/10.1145/3540000](#)

11 AI-Driven Cybersecurity

Protecting Human Rights in the Digital Age

*T. Manonmani, S.N. Kumar, Jomin Joy, and
Anandhu Venu*

DOI: [10.1201/9781003668657-11](https://doi.org/10.1201/9781003668657-11)

11.1 INTRODUCTION

Organizations in the internet age mostly depend on their *information technology* (IT) infrastructure to protect them against cyberattacks. The threat of cybercrime is growing dramatically due to the widespread adoption of digital transformation, highlighting the need for cybersecurity. Strong policies and infrastructure work together to strengthen computer systems and networks against unauthorized access or assaults, and cybersecurity emerges as the steadfast defender. Industries, people, and governments are investing a lot of money to take advantage of cybersecurity's benefits in defending their assets and data from cybercriminals. By utilizing artificial intelligence (AI)-driven tools, such as intrusion detection systems (IDSs) and machine learning algorithms, cybersecurity measures can be enhanced to detect, prevent, and mitigate cyberattacks that threaten democratic institutions and individual freedoms. The capacity of an organization to survive in the current, extremely fierce

competitive marketplace depends on its ability to implement the right cybersecurity strategy and solutions to address them. In the current digital scenario, the role of cybersecurity has a significant impact. It seeks to traverse the complex and volatile field of cybersecurity by looking at its history, the factors that continue to propel it forward, the major obstacles it faces, and its anticipated future. The writers will examine various threat actors and cyber threats in detail before diving into the technologies and instruments that have been developed to address these issues. In addition, several cybersecurity techniques will be examined and contrasted to evaluate their efficiency and consequences. This chapter will analyze real-world case studies of notable cyberattacks to provide an insightful analysis of the lessons gained and their implications for the cybersecurity environment [1].

Technological developments in AI have led to a tremendous surge in automation and creativity. These AI structures have many advantages, but they can also be abused. Attacks that are extremely focused and elusive in innocuous carrier applications—like Deep Locker—have shown how AI may be used maliciously. Threat actors are continually refining and enhancing their attack strategies, focusing special focus on the use of AI-driven tactics, also known as AI-based cyberattacks, which may be combined with traditional attack methods to bring about more harm. Despite several studies on AI and security, scholars have not effectively characterized AI-based cyberattacks to comprehend the tactics used by the adversary and to create successful countermeasures. With the goal of mapping previous research on AI-based cyberattacks onto a recommended framework, this study aims to shed light on potential new threats. Our system distinguishes many aspects of harmful AI usage during the cyberattack life cycle and offers a foundation for detecting these uses to anticipate potential risks. Furthermore, we describe how to use this approach to investigate AI-driven cyberattacks in a highly unlikely scenario of vital smart grid infrastructure.

With a special focus on the financial and insurance sectors, the abstract analyzes the ABACUS survey [2], which was carried out on behalf of the

Australian Institute of Criminology to assess computer security occurrences in Australian firms. It emphasizes the survey's methodology, sample size, and main findings while highlighting the frequency, kind, prices, and effects of cyber threats that Australian businesses must deal with. To expose the effect of cybersecurity events against Australian companies and to figure out its frequency, expenses, and effects of vulnerabilities, ABACUS conducted a survey with 221 participants from various finance sectors such as insurance and banks. The results infer the potential risks in the insurance and financial industries and expose the challenges and risks related to cybersecurity in these sectors. The ABACUS survey also highlights the importance of understanding the threat frequency, expenses, and consequences of various security incidents, which helps to devise tactics to reduce industry risks.

Cybersecurity continues to be a massive cause of concern in this digital era. News on ID theft, data breaches, and cracking captchas are too many to record and have their effects on millions of people and organizations. Designing sufficient controls and processes and effectively implementing them with the utmost accuracy to fight cyberattacks and crimes will continue to be a difficult task. With the great help of AI, cybercrime can be done now with a lot of ease. It has been used in practically all technical and scientific fields. AI has caused a revolution in fields like robotics and healthcare. Cybercriminals were unable to resist this fireball, and as a result, "usual" cyberattacks have developed into "intelligent" ones. The topic of cybersecurity has been significantly affected by AI due to the introduction of "intelligent" cyberattacks, which are more and more complex and difficult to identify and prevent. certain AI techniques, like machine learning algorithms for identifying suspicious activity and predictive analytics for spotting possible attacks, are found to be promising for improving cybersecurity measures. AI applications in cybersecurity demonstrated the potential to improve cyber-attack detection and response, hence improving an individual's or organization's overall security posture [3].

In the digital age of today, enterprises need strong cybersecurity solutions for several reasons. Cyberattacks are growing more prevalent and sophisticated. Hackers target companies of all sizes to extort money, disrupt operations, or steal data. Multiple layers of resistance are provided by robust solutions, making it more difficult for attackers to succeed. Financial data, intellectual property, and sensitive information are essential to businesses. Strong cybersecurity guards safeguard those assets from breaches, leaks, and illegal access. This can avert substantial financial losses, harm to one's reputation, and legal implications. Customers want businesses to take safeguarding their information seriously and are becoming more aware of the potential risks associated with cybersecurity [4]. Sturdy solutions show a dedication to security that builds client loyalty and trust. Cyberattacks have the potential to halt operations entirely, costing businesses money and precious time. Sturdy solutions ensure business continuity and save overall security costs by thwarting assaults and minimizing harm when they do occur. Restrictions requiring specific cybersecurity measures are in place in many businesses. Strong solutions assist companies in adhering to these rules and averting costly penalties or legal action. Having robust cybersecurity is a competitive advantage in today's data-driven environment. Companies that can show strong security are more appealing to investors and partners [5]. Organizations have to establish a defense strategy to safeguard their funds, data, and operations in the face of a constantly shifting cyber threat scenario.

For a holistic approach, an all-inclusive strategy encompasses stacked layers of security measures such as perimeter, network, endpoint, data, role-based access and identity control, application security, security awareness and training, and business continuity planning. Organizations are able to effectively handle risks and lessen the impact of potential security incidents by adhering to best practices, which include doing risk assessments, prioritizing expenditures, updating policies, monitoring and responding to incidents, and evaluating security measures. A well-implemented defense plan promotes a security-aware culture, which helps the company succeed and remain resilient over an extended period of time [6].

Nowadays, for operational efficiency and accurate control, there is a predominant use of digital instrumentation and control systems in nuclear power plants (NPPs). However, because industrial control systems are becoming increasingly targeted by sophisticated and frequent cyberattacks, this change poses cybersecurity challenges. We suggest an extensive cybersecurity solution platform for NPPs to solve this. Systems for data collecting, cyberattack detection, cause analysis, reaction, and primary control room appearance are all part of this platform. It improves cybersecurity by merging process data with conventional host systems and network data through integrating IT and OT skills. Our suggested architecture exhibits efficacy in stopping, identifying, and handling cyberattacks through simulations, in line with NPP safety and security protocols [7].

The safety of organizations faces new issues because of the confluence of cyber risks and digitization, which is happening at a rapid pace in the digital realm. This essay offers a thorough framework for influences. The integration of resilience and capabilities is the main area of focus. Organizations can enhance their capacity to anticipate, mitigate, respond to, and recover from cyber disasters by incorporating all these elements into their cybersecurity procedures. To achieve cyber resilience, this essay highlights the significance of corporate leadership, responsibility, and innovation [8]. [Figure 11.1](#) shows the diagrammatic representation of types of cyberattacks.

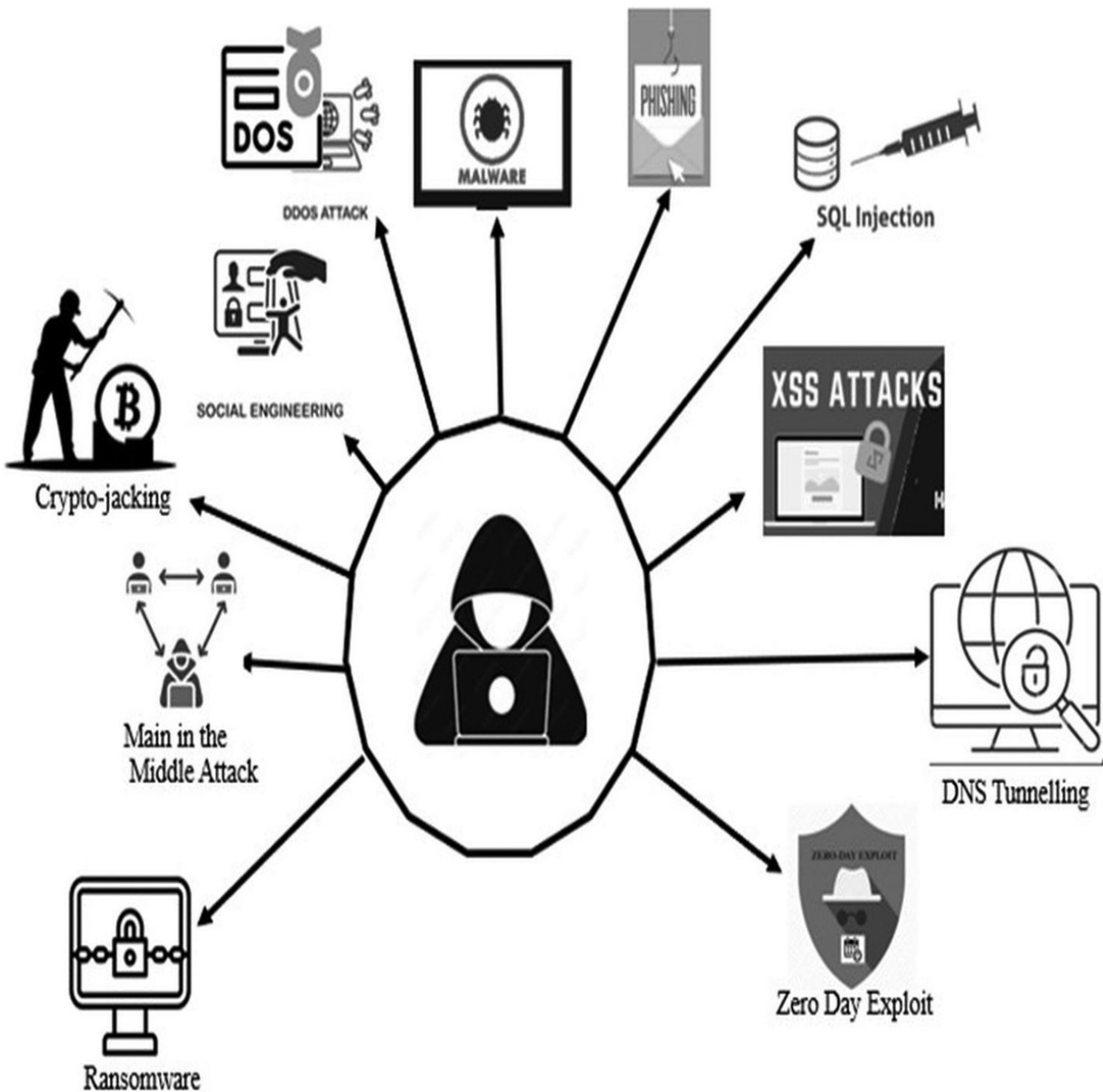


FIGURE 11.1 Types of attacks in cybersecurity. [📄](#)

Sturdy digital immune systems are essential in the dynamic digital age because cybersecurity attacks are an ever-present concern. It seeks to summarize the key ideas of digitally immune systems, as well as the essential elements and tactical methods covered in the paper, to provide a comprehensive understanding of them. Digital immune systems are fundamentally similar to human immune systems. These defenses are fashioned after the way our bodies' defenses against infections anticipate,

stop, identify, and neutralize cyberattacks. The need for these solutions grows more urgent by the fact that cybercrime is growing more and more commonplace globally, putting reputation, financial stability, sensitive data, honesty, and regulatory compliance at risk. A digital immune system's primary parts are two-factor authentication (2FA), firewalls, antivirus software, IDSs, and encryption techniques. Each of them cooperates with each other to produce cyber immunity. Firewall acts as a primary defense wall of the network to prevent illegal access. Questionable activity of the intruders is detected by the IDS. Removal of malicious software, applications, and viruses is carried out by the "Cyber antibodies" the anti-virus software. Encryption protects the data by converting it into unintelligible formats and keeps their knowledge safe from interception. In addition to password protection, 2FA enhances the security part of the access control mechanism. Scheduled backup and maintenance, recovery procedures, security awareness campaigns, and systematic software updates are all part of the cyber immune system. Like boosters against unique virus strains, cyber threats need constant updates for systems to keep them updated. These systems are valued for being capable of detecting and eliminating cyber threats before they have a chance for massive destruction. These systems help us to create a clean digital environment for the future. The digital immune system acts as a barrier against the changing array of cyberattacks by delivering all protection mechanisms. Adopting this system is not only urged but also a vital need for persisting in the digital era [9].

11.2 THE ROLE OF AI IN CYBERSECURITY

Being an imperative asset in the field of cybersecurity, AI introduces innovative methods to tackle the emergent landscape of cyber threats. Let us see the comprehensive breakdown of AI's role in developing cybersecurity:

11.2.1 DATA ANALYSIS AND PATTERN RECOGNITION

The prominent feature of AI is its ability to quickly and efficiently analyze and process voluminous data. Hence, AI becomes an integral part of cybersecurity with a vital feature in detecting anomalies or vulnerability patterns well in advance. With huge data of network logs, user behavior patterns, and system events, to flag off the actions that are questionable in nature and could have bypassed common security controls. Usage of multiple machine learning models for the analysis of cybersecurity data was discussed in ref. [[10](#)].

11.2.2 PREDICTIVE ANALYTICS

Based on historical data and ongoing trends, AI-powered cybersecurity systems leverage machine learning algorithms to predict and anticipate cyber threats. AI can forecast the vulnerable points by scrutinizing patterns from the attack history and thus empowers the organizations to proactively use defensive procedures and reinforce their protections. The integration of analytical modeling methodology in machine learning models enhances the intrusion detection in cybersecurity [[11](#)].

11.2.3 ENHANCED IDSS

Predefined rules and signatures are used in conventional IDS to detect suspicious activities in a network. However, they often struggle to keep pace with the evolving strategies of the attackers. On the other hand, AI-driven IDS utilize innovative machine learning algorithms to constantly acquire and adapt to real-time threats. AI-augmented IDS dynamically adjusts its detection criteria based on the emerging attack patterns and alleviates sophisticated cyber-attacks, including zero-day exploits and polymorphic malware. Authors of ref. [[12](#)] discussed the improvised intrusion detection using AI techniques in cybersecurity.

11.2.4 BEHAVIORAL ANALYSIS AND ANOMALY DETECTION

By monitoring and analyzing the user activity logs and user behavior, AI-powered cybersecurity solutions excel in behavioral analysis to detect the deviations from normal patterns. Suspicious anomalies can be detected by establishing baseline behaviors for users and corresponding devices. AI algorithms can easily pinpoint suspicious activities of potential insider threats, compromised accounts, or unauthorized login attempts by continuously monitoring and comparing them with the baseline data. This will help the organizations to get an early alarm on security incidents effectively, and hence minimize the impact of the attacks. Evangelou and Adams [13] proposed a Quantile Regression Forests model as a baseline algorithm for anomaly detection.

11.2.5 AUTOMATED INCIDENT RESPONSE

Besides threat detection, AI can streamline incident response processes by automating routine tasks and decision-making procedures. AI-driven security orchestration and automation platforms can automatically prioritize and triage security alerts, initiate remediation actions, and coordinate responses across various security tools and systems. By reducing response times and human intervention, AI-driven incident response mechanisms help organizations mitigate the impact of security breaches and minimize downtime [14].

11.3 AI IN VULNERABILITY MANAGEMENT

Vulnerability management is an ongoing process of identifying, categorizing, mitigating, and monitoring vulnerabilities in software applications, networks, and computer systems. A vulnerability is what it is designed to secure any number of digital assets of an organization. The major steps are:

- *Identification:* The initial step here is to point out the vulnerabilities. This is done most often through the use of vulnerability scanning

tools that can check differences in system configurations and software versions against a database of vulnerabilities in discovered and known systems.

- *Categorization:* Once vulnerabilities are detected, effective risk assessment requires categorizing and prioritizing them based on their criticality and likelihood of exploitation.
- *Remediation:* Organizations then go on to plan and strategize the implementation of remediation, which most often involves applying patches, configuring systems, or updating software. Collaboration needs to be with system administrators, network engineers, and application developers.
- *Continuous Monitoring:* Processes for continuous monitoring should be established, to ensure regular vulnerability scanning, IDSs, log analysis, and threat intelligence feeds to detect new vulnerabilities as and when they emerge.

In wrong hands, one single vulnerability may mean an entire system being thrown open to catastrophic data breaches. Effective vulnerability management ensures not only proactive management of risks but also works as a shield against potential attackers. By adopting strong vulnerability management, organizations can:

- guard sensitive data,
- bolster customer trust,
- preserve organization's reputation, and
- be cost-effective.

11.4 LIMITATIONS OF CONVENTIONAL VULNERABILITY MANAGEMENT SYSTEM

- *Manual Effort:* Manual processes are error-prone, unreliable, inefficient, and time-consuming.

- *Limited Scale and Speed:* Traditional approaches are not able to cope with the IT environments growing complexity.

Proactive measures play a vital role in robust security features, with the main objective as patching the vulnerabilities before an attacker exploits them (Beagle Security, [2]).

11.4.1 ADVANCED ANALYTICS AND THREAT MODELING TECHNIQUES

- *Vulnerability Analytics Using AI:* Vast amount of data from dataset comprising vulnerability reports, threat intelligence, and threat actor behaviors is easily analyzed by AI. This capability empowers it to identify patterns and predict potential vulnerabilities even before they are loaded in public databases.
- *Threat Modeling with AI:* Threat models can be created by AI to simulate attacker behavior in an environment. These models provide valuable insights for securing teams by prioritizing vulnerabilities based on their exploitability. Basically, vulnerability management has been a reactive process. Vulnerability scanners identify the potential threats in software, and security personnel patch them before attackers exploit them. There are several limitations in implementing this approach.
- *Data Deluge:* A constant stream of data, often containing false positives, is generated by vulnerability scanners for analyzing the data. Sifting through this data to identify vital threats is a time-consuming and vague process.
- *Zero-Day Exploits:* As the conventional methods often miss “zero-day” vulnerabilities—those unknown to security vendors and for which no prevention mechanisms exist will create a critical window of opportunity for intruders.
- *Patching Lag:* Due to resource constraints, platform compatibility issues, patching the identified vulnerabilities will become a slow

process, and this lag opens the door for intrusion.

Because of these limitations, data breaches, financial loss, and work disruptions can cripple the organizations as the consequences are severe. Taking proactive measures to mitigate the vulnerabilities before exploitation is crucial for robust cybersecurity. Vulnerability assessments can be automated and streamlined by the following methods:

- *Continuous Scanning*: Manual tracking and minimizing the attacker's window of opportunity was overcome by the AI-powered tools, which can automatically detect the vulnerability.
- *Reduced False Positives*: Prevalence of false positives is a common challenge with conventional vulnerability scanners,

Prioritizing patches becomes crucial with a constant invasion of vulnerabilities, AI enables organizations to:

- *Risk-Based Prioritization*: Potential impact on critical systems is calculated by analyzing the severity of vulnerabilities and their exploitability. Based on this, the most critical vulnerabilities are prioritized for patching first.
- *Contextual Awareness*: When prioritizing patches, AI can consider factors like system dependencies and ongoing business processes. This consideration reduces the disruption to ongoing operations while patching the critical vulnerabilities promptly.

11.5 LEVERAGING AI FOR THREAT INTELLIGENCE

Threat intelligence supplies crucial information to defend the cyber-attacks [15, 16]. Valuable insights about the identity, motive, and tactics of the cyber adversaries are identified by collecting and analyzing the data on potential threats. By leveraging this threat intelligence [17], organizations empower proactively to mitigate attacks and strengthen their cybersecurity posture.

11.5.1 LIFE CYCLE PROCESS OF THREAT INTELLIGENCE

The cyber threat intelligence lifecycle [18] consists of six vital parts ([Figure 11.2](#)) to offer a complete comprehension of possible vulnerabilities and direct proactive defense tactics.

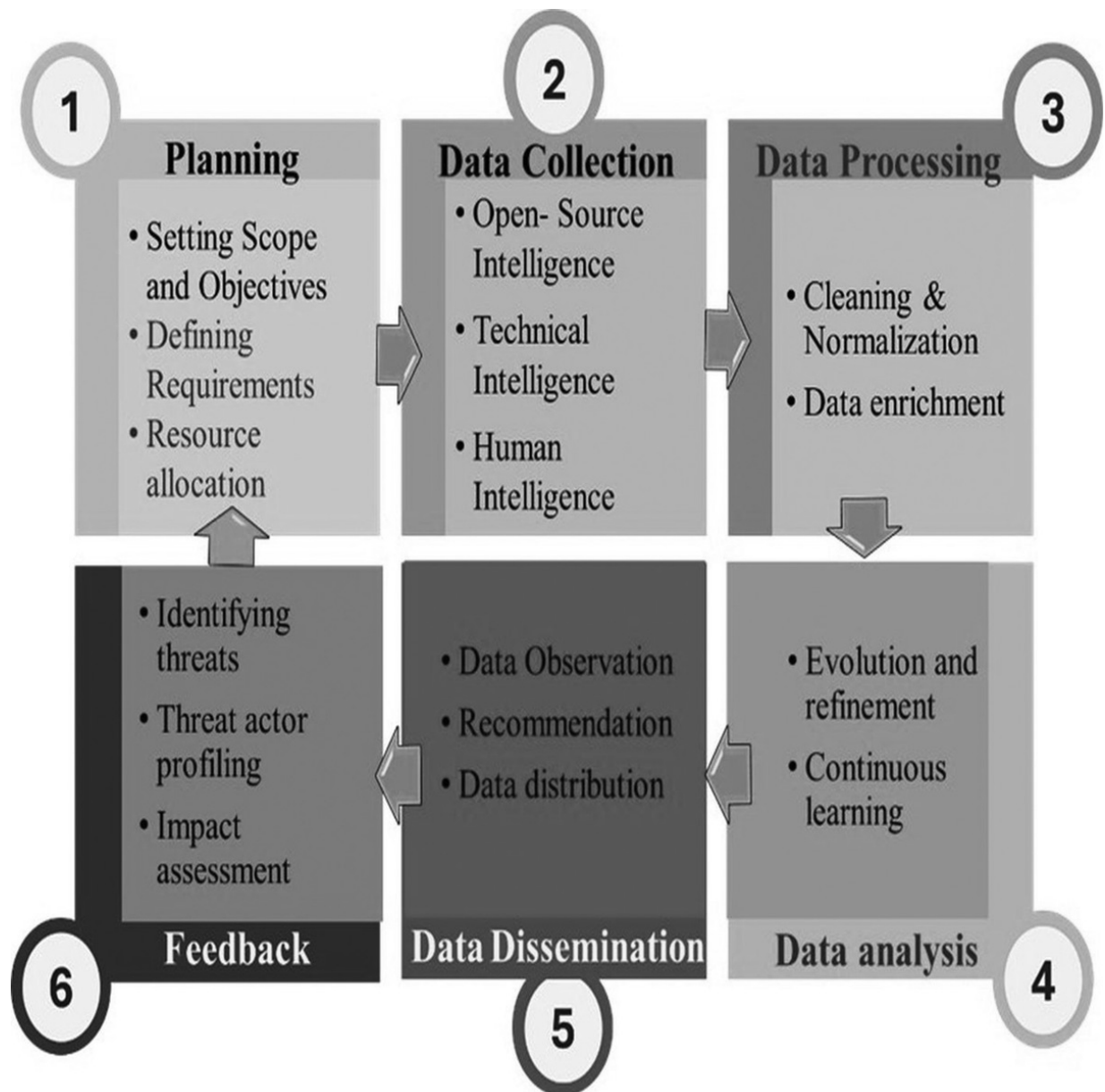


FIGURE 11.2 Life cycle of cyber threat intelligence. [↗](#)

Gathering threat intelligence by conventional approaches needs manual collection of data from various resources, which can be an ineffective and time-consuming process. AI automates this process by utilizing various machine learning algorithms [19] and has made a remarkable change in the process of collecting and processing a huge volume of real-time data. This includes open-source intelligence (OSINT) feeds, threat actor forums, and network traffic logs, thereby providing a clear understanding of the attack landscape. AI can easily predict the future attack trends and identifies the subtle threats missed by humans. Major difference between traditional threat intelligence and AI threat intelligence in terms of data collection, data processing, analysis, speed of processing, accuracy, threat detection, predictive capabilities, insights, efficiency, and overall approach in analyzing the potential threats are given in [Table 11.1](#).

TABLE 11.1
Traditional versus AI Threat Intelligence

Feature	Traditional Threat Intelligence	AI Threat Intelligence
Data collection	Manual collection from various sources (OSINT, HUMINT, TECHINT)	Automated collection from various sources with AI and machine learning
Data processing	Manual cleaning, formatting, and organization	Automated cleaning, formatting, and organization
Analysis	Human analysts identify trends and patterns	AI algorithms identify trends, patterns, and anomalies
Speed	Time-consuming and labor-intensive	Real-time processing of large datasets
Accuracy	Prone to human error	Improved accuracy due to AI pattern recognition
Threat detection	Relies on analysts to identify potential threats	Faster detection of subtle patterns and anomalies indicating threats

Feature	Traditional Threat Intelligence	AI Threat Intelligence
Predictive capabilities	Limited	Predicts future attack trends and tactics
Insights	Requires interpretation of complex data	Generates clear and actionable insights
Efficiency	Requires significant manpower	Frees analysts for strategic tasks
Overall approach	Reactive, focused on identifying existing threats	Proactive, focused on anticipating and preventing threats

AI algorithms [20] collect information from various sources such as security feeds, user and network logs, and data from external threat intelligence platforms. Efficacy of threat intelligence depends on the fundamental capability to collect, analyze, and interpret these huge volumes of data retrieved from diverse sources. This will ensure that the organizations are having the most up-to-date threat data and all-encompassing threat data accessible [21]. To demonstrate this process, implementation of cyber security framework in financial organization is discussed below.

1. *Network Logs:*

- *Source:* Internal network devices such as firewalls, routers, and switches.
- *Data Collected:* Information about both internal and external communications are collected from the network traffic logs, access logs, and event logs.
- *Purpose:* Main objective is to monitor mishaps or any unusual behaviors in the network which represent a network breach or attempted breach.

2. *Security Feeds:*

- *Source:* Third-party security vendors and OSINT platforms.
- *Data Collected:* This includes information on recent trends in cybersecurity, new vulnerabilities, and updates on ongoing cyberattacks.
- *Purpose:* By collecting the security feeds as a continuous process, the organization can stay informed about the recent threat trends and confirm the organization's defense is intact to counter them.

3. *External Threat Intelligence Platforms:*

- *Source:* Subscription-based threat intelligence services.
- *Data Collected:* Detailed threat intelligence reports, indicators of compromise (IoCs), and tactical advice on mitigating specific threats.
- *Purpose:* Threat intelligence platforms are utilized to gain up-to-date knowledge on potential attacks that target the financial organization. This will aid the organization to take proactive defense measures.

4. *Customer Interaction Points:*

- *Source:* Online banking platforms, mobile apps, and customer service portals.
- *Data Collected:* User access patterns, transaction logs, and records of customer service interactions.
- *Purpose:* Data from customer interaction points of a financial organization helps to analyze the deviation of customer behavior to identify the presence of any unwanted activities in financial data transactions. This approach helps in ensuring the safety and integrity of the organization's customer accounts and transactions.

5. *Data Analysis:* By using various machine learning algorithms, AI can easily analyze the huge volume of data. This will help to identify the relationship between complex threat patterns that are usually not

discovered by manual analysis methods. For instance, AI can identify unusual network traffic patterns or predict the known malicious IP addresses present in the collected data. Cyber-attacks that take place immediately after a threat is found and before any preventive measures are put in place are known as Zero-day exploits [22, 23]. Due to their unique characteristics and absence of their established patterns in traditional security tools this attacks pose a significant challenge in cybersecurity domain.

11.5.2 ROLE OF AI—ZERO-DAY ATTACK

- *Data Acquisition:* The AI system continuously acquires data from various sources, such as network logs, security feeds, application logs, and external threat intelligence platforms, to infer the presence of threats.
- *Pattern Recognition:* AI starts analyzing the acquired data by using various supervised and unsupervised machine learning algorithms. It identifies the baseline behaviors within the network and systems and coined a definition for “normal” activity.
- *Detection of Anomalies:* After studying the patterns of various threats and abnormalities, AI becomes adept at detecting the baseline deviations from the established baseline behaviors. For example, it can easily identify the data surge from a specific user account or the routing of network traffic to an unfamiliar IP address. These unusual behaviors indicate the occurrence of a zero-day exploit being utilized against the network.
- *Identification of Threats:* After identifying the threats, the AI system uses the deep learning algorithms to analyze the features of the threat patterns to understand its characteristics. It simply compares the identified threat pattern and its malicious behaviors with the known threat patterns. AI can detect the potential vulnerability based on the context of the threat occurrence, even if the zero-day exploit does not match any existing signatures.

- *Alerting and Response:* Security alerts are given to the cybersecurity personnel to initiate the predefined responses to mitigate the attacks. Mitigation of serious attacks is achieved by blocking the suspicious IP address, isolating the affected system, and comprehending the breach by initiating forensic analysis.

By analyzing the severity, likelihood, and other predominant factors, AI has the capability to differentiate the attacks and offer precise and actionable insights to the cybersecurity team.

11.5.3 BENEFITS OF AI-POWERED DATA INTERPRETATION

- *Automated Threat Detection:* By continuously analyzing voluminous data, AI has the capability to faster identification of potential risks. This capability enables AI to response quickly and more efficiently for emergency situations.
- *Improved Accuracy:* Being equipped with various deep learning and natural language processing algorithms, AI poses the intent to identify the patterns of anomalies in an organization that may go undetected in human analysis.
- *Predictive Capabilities:* By understanding the past data, AI can learn and predict future attack trends of the cybercriminals. This predictive capability will help organizations to proactively strengthen their defenses, mitigating threats before they come into the picture. By ensuring the security and integrity of their systems and data, AI helps organizations to stay ahead of cyber threats.

11.5.4 AUTOMATION OF THREAT INTELLIGENCE PROCESSES WITH AI

AI's role in threat intelligence extends beyond data analysis; it can automate the identification of possible vulnerabilities [\[24\]](#) by continuously

monitoring the organizations networks and systems. Figure 11.3 depicts the process of automating the cyber threat intelligence process. This automation will help to reduce the likelihood of human errors and speed up the response to threats. AI systems remain agile by continuously learning from real-time new data and improving its detection and response capabilities.

- *Automated Response and Feeds:* AI automates the response to various attacks by automatically quarantining infected devices or applying patches to vulnerable systems, to minimize the potential damage.
- *Automated Threat Scoring:* Threat prioritization can be done based on their severity, potential impact, and exploitability
- *Self-Learning and Model Retraining:* AI models are trained with mechanisms for continuous self-learning on new threat data as it becomes available, to ensure their adaptability to new evolving threats.
- *Continuous Learning:* AI poses the capability to learn quickly from the past. By analyzing the historical data, machine learning algorithms improve their capability of risk detection and response over time. This will help the AI systems to easily adapt to the new attack vectors and threats with an effective threat intelligence process.

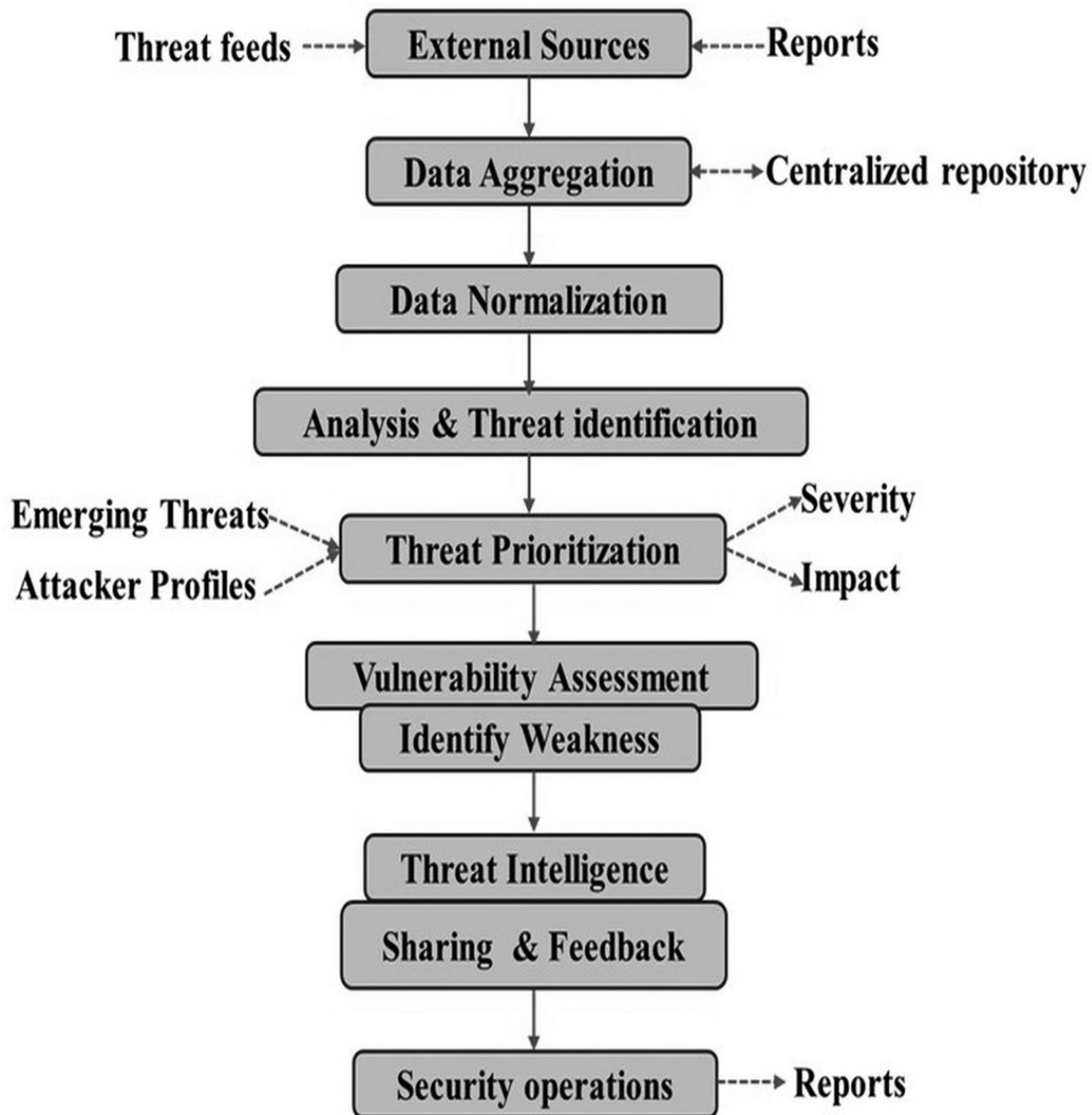


FIGURE 11.3 Automation of cyber threat intelligence process.

For example, AI-powered IDSs can automatically analyze network traffic to identify and block potential threats, such as malware or unauthorized access attempts.

11.6 PREDICTING ATTACK TRENDS AND PROVIDING ACTIONABLE INSIGHTS

AI assists the organizations to proactively prepare and mitigate attacks by analyzing historical data and current threat vectors. They can easily identify the common patterns of the threats by analyzing the past data. For example, it can easily infer the increase in phishing attacks targeting a specific domain or the usage of a new type of malware by cyberattackers. Using this predictive analysis, organizations can anticipate the future attacks and strengthen their defenses accordingly. Based on the current data, AI's predictive modelling capabilities can predict the trends of future attack vectors. For instance, the likelihood of a specific type of attack occurring in a certain time duration is predicted by machine learning models will help the organizations to take pre-emptive measures.

AI can also provide actionable insights that guide security strategies. For example, if AI predicts an increase in ransomware attacks or phishing attacks targeting a particular industry, organizations within that sector can take pre-emptive measures to secure their systems.

[Figure 11.4](#) depicts a scenario where an organization is monitoring the dark web for mentions of its brand. A dark web investigation [\[25\]](#) entails the systematic monitoring of online resources that are associated with cybercrime. This process mitigates the potential risk by continuous surveillance of criminal forums and the black market to identify and address illicit activities. By performing sentiment analysis on the data extracted from dark web platforms [\[26\]](#), AI leverages its threat intelligence on the dark web and identifies the key parameters of the potential along with their area of specialization. This will facilitate the organization, to plan for security alternatives to prevent the attacks.

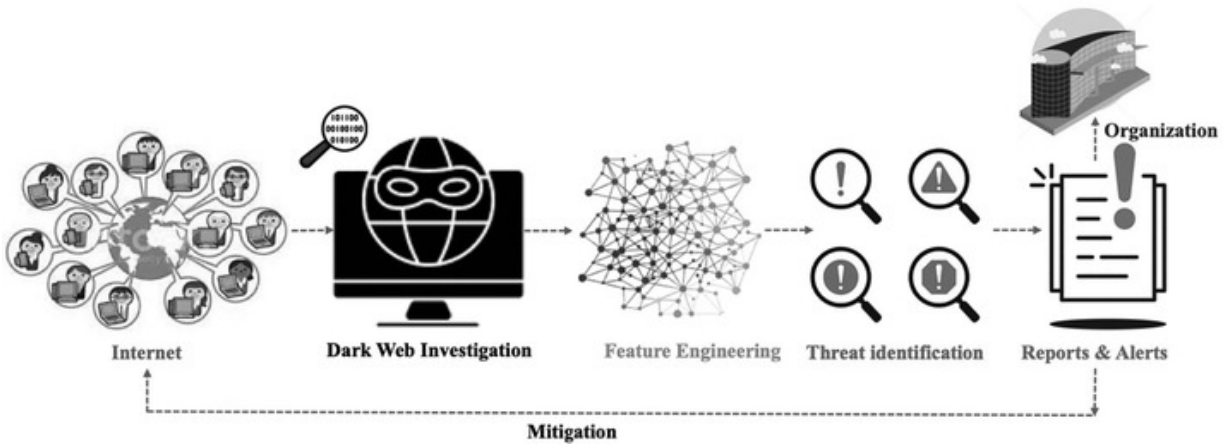


FIGURE 11.4 AI tool for scanning dark web forums. [🔗](#)

In a real-world application, unusual login attempts from foreign IP addresses are easily detected by an AI-powered security system and it automatically blocks the IP address, sends alerts to the security professionals, and triggers a forensic analysis to measure the depth of the potential breach.

AI models can easily analyze data from recent ransomware attacks to identify attacks targeting cloud services and send recommendations to strengthen the cloud security measures, such as multi-factor authentication and backing up data at regular intervals, etc.

11.7 CONCLUSION

In many organizations, software vulnerabilities serve as entry points for cyber-threats, and it is the responsibility of the organizations to be proactive in detecting the passive vulnerabilities and patching them before they exploit the system. With the potential of deep learning, machine learning, and natural language processing algorithms, AI has the capability to dig vast amounts of data to identify potential vulnerabilities, and access them to prioritize the patches based on their level of risk. This protects the organizations from vulnerabilities and enables the organization to stay ahead of cyber attackers and reduce their attack surface. The chapter examines how AI can aid in threat intelligence with various information

gathering techniques. Threat intelligence automated the process of information collection, analyzing, and interpreting data about cyber vulnerabilities to gain insights and strongly defend against them. It can detect emerging threats from the past data, predict attack trends for future, and provide actionable insights to the security teams to respond effectively. Furthermore, the challenges such as algorithm bias, privacy invasion, and explainability, emphasizing the importance of responsible and transparent use of AI in cybersecurity, are discussed. Incorporating AI into cybersecurity strategies offers a promising future for a more secure digital world. By utilizing its capabilities in real-time threat detection, automated response, and proactive threat management, AI can significantly improve our ability to defend against cyber-attacks. By embracing AI as a partner in cybersecurity, we can usher in an era of proactive defense and build a more resilient digital world.

REFERENCES

- 1.Kumar M, Darshan SS, Yarlagaadda V. Introduction to the cyber-security landscape. In *Malware Analysis and Intrusion Detection in Cyber-Physical Systems*. 2023 (pp. 1–21). IGI Global.[↗](#)
- 2.Abacus Technologies. Cybersecurity. www.abacustechnologies.com/cybersecurity/[↗](#)
- 3.Shamiulla AM. Role of artificial intelligence in cyber security. *International Journal of Innovative Technology and Exploring Engineering*. 2019 Nov 1;9(1):4628–4630.[↗](#)
- 4.Dutta A, McCrohan K. Management's role in information security in a cyber economy. *California Management Review*. 2002 Oct;45(1):67–87.[↗](#)
- 5.La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny R. Investor protection and corporate governance. *Journal of Financial Economics*. 2000 Jan 1;58(1–2):3–27.[↗](#)
- 6.Mughal AA. The art of cybersecurity: Defense in depth strategy for robust protection. *International Journal of Intelligent Automation and*

- Computing*. 2018 Mar 11;1(1):1–20. [↗](#)
7. Zhang F, Hines JW, Coble JB. A robust cybersecurity solution platform architecture for digital instrumentation and control systems in nuclear power facilities. *Nuclear Technology*. 2020 Jul 2;206(7):939–950. [↗](#)
 8. Safitra MF, Lubis M, Fakhurroja H. Counterattacking cyber threats: A framework for the future of cybersecurity. *Sustainability*. 2023 Sep 6;15(18):13369. [↗](#)
 9. George AS, George AH, Baskar T. Digitally immune systems: Building robust defences in the age of cyber threats. *Partners Universal International Innovation Journal*. 2023 Aug 25;1(4):155–172. [↗](#)
 10. Zhao L, Zhu D, Shafik W, Matinkhah SM, Ahmad Z, Sharif L, Craig A. Artificial intelligence analysis in cyber domain: A review. *International Journal of Distributed Sensor Networks*. 2022 Apr;18(4):15501329221084882. [↗](#)
 11. Nadella GS, Gonaygunta H. Enhancing cybersecurity with artificial intelligence: Predictive techniques and challenges in the age of IoT. *International Journal of Science and Engineering Applications*. 2024;13(04):30–33. [↗](#)
 12. Khan SU, Eusufzai F, Azharuddin Redwan M, Ahmed M, Sabuj SR. Artificial intelligence for cyber security: Performance analysis of network intrusion detection. In *Explainable Artificial Intelligence for Cyber Security: Next Generation Artificial Intelligence*. 2022 Apr 19 (pp. 113–139). Cham: Springer International Publishing. [↗](#)
 13. Evangelou M, Adams NM. An anomaly detection framework for cybersecurity data. *Computers & Security*. 2020 Oct 1;97:101941. [↗](#)
 14. Uzoma J, Falana O, Obunadike C, Oloyede K, Obunadike E. Using artificial intelligence for automated incidence response in cybersecurity. *International Journal of Information Technology (IJIT)*. 2023 Nov;1(4):12–15. [↗](#)
 15. Alrimawi F, Pasquale L, Mehta D, Yoshioka N, Nuseibeh N. Incidents are meant for learning, not repeating: Sharing knowledge about security

- incidents in cyber-physical systems. *IEEE Transactions on Software Engineering*. 2022;48(1):120–134. doi: [10.1109/TSE.2020.2981310](https://doi.org/10.1109/TSE.2020.2981310).[↗]
- 16.Cascavilla G, Tamburri DA, Van Den Heuvel W-J. Cybercrime threat intelligence: A systematic multi-vocal literature review. *Computers & Security*. 2021;105: 65–69. <https://doi.org/10.1016/j.cose.2021.102258>[↗]
- 17.Al-Rowaily K, Abulaish M, Al-Hasan Haldar N, Al-Rubaian M. BiSAL —A bilingual sentiment analysis lexicon to analyze Dark Web forums for cyber security. *Digital Investigation*. 2015;14:53–62. <https://doi.org/10.1016/j.diin.2015.07.006>[↗]
- 18.Schmitt M. Securing the digital world: Protecting smart infrastructures and digital industries with artificial intelligence (AI)-enabled malware and intrusion detection. *Journal of Industrial Information Integration*. 2023;36: 12–14. <https://doi.org/10.1016/j.jii.2023.100520>[↗]
- 19.Montasari R, Carroll F, Macdonald S, Jahankhani H, Hosseinian-Far A, Daneshkhah A. Application of artificial intelligence and machine learning in producing actionable cyber threat intelligence. In *Digital Forensic Investigation of Internet of Things (IoT) Devices*. 2021 (pp. 47–64). Springer.[↗]
- 20.Parrend P, Navarro J, Guigou F, Deruyver A. Foundations and applications of artificial intelligence for zero-day and multi-step attack detection. *EURASIP Journal on Information Security*. 2018;4: 12–14.[↗]
- 21.Alaeifar P, Pal S, Jadidi Z, Hussain M, Foo E. Current approaches and future directions for Cyber Threat Intelligence sharing: A survey. *Journal of Information Security and Applications*. 2024;83: 61–76. <https://doi.org/10.1016/j.jisa.2024.103786>[↗]
- 22.Raman R, Kumar Nair V, Nedungadi P, Ray I, Achuthan K. Darkweb research: Past, present, and future trends and mapping to sustainable development goals. *Heliyon*. 2023;9(11): 104–107. <https://doi.org/10.1016/j.heliyon.2023.e22269>[↗]
- 23.Ramsdale A, Shiaeles S, Kolokotronis N. A comparative analysis of cyber-threat intelligence sources, formats and languages. *Electronics*. 2020;9(5):824.[↗]

- 24.Saeed S, Suayyid SA, Al-Ghamdi MS, Al-Muhaisen H, Almuhaideb AM. A systematic literature review on cyber threat intelligence for organizational cybersecurity resilience. *Sensors (Basel)*. 2023 Aug 19;23(16):7273. <https://doi.org/10.3390/s23167273>↗
- 25.Santos FCC. Artificial intelligence in automated detection of disinformation: A thematic analysis. *Journal Media*. 2023;4:679–687. <https://doi.org/10.3390/journalmedia4020043>↗
- 26.Schäfer M, Fuchs M, Strohmeier M, Engel M, Liechti M, Lenders V. BlackWidow: Monitoring the dark web for cyber security information. In *2019 11th International Conference on Cyber Conflict (CyCon), Tallinn, Estonia*. 2019 (pp. 1–21). <https://doi.org/10.23919/CYCON.2019.8756845>↗

12 AI and Inclusive Democracy

Transforming Governance for Equitable Societies

*Kushanthi S. Harasgama, Vaishali Arora,
Manvendra Singh, and Tahir Qureshi*

DOI: [10.1201/9781003668657-12](https://doi.org/10.1201/9781003668657-12)

12.1 INTRODUCTION

The loss of personal liberty was as terrible as the invasion of privacy. Moreover, artificial intelligence (AI) also exacerbates the proliferation of AI-powered misinformation, manipulation, and surveillance tools that threaten the integrity of democratic processes, including elections and public discourse. The dissemination of deep fakes, algorithmically generated fake videos, and targeted disinformation campaigns can undermine trust in democratic institutions and erode the foundation of informed civic engagement ([Kaplan, 2019](#)). Automated systems could mean an efficient manner of policy analysis, predictive governance, and the ability to provide more personalized services to constituents with greater transparency and accountability. It can offer empowerment through information, promotion of digital literacy, and introduce participatory platforms in deliberative democracies. These promises and complexities are encumbered by AI, with very serious challenges that should be addressed

with meticulous consideration in democratic regimes. The psychographic profiling of Facebook Users in 2016 in Britain and America is a prominent example. Although it is the tip of the iceberg, there are many other ways that AI endangers democracy outside those voting manipulation cases. The potential of AI can open the gate in the hands of powerful entities, whether governments or corporations and raise concerns about the centralization of power and the erosion of democratic principles. The capability of AI to empower authoritarian surveillance systems or facilitate the suppression of dissent requires the critical need for robust safeguards and regulatory frameworks to protect democratic values, institutions, and human rights.

12.1.1 A BRIEF UNDERSTANDING OF AI

Artificial intelligence, abbreviated as AI, generally refers to the simulation of human intelligence processes by machines, particularly computer systems (Coursera, 2024b). In the mid-20th century, these systems heralded the concept and notion of AI. The researcher Alan Turing in 1950 proposed the Turing Test to determine a machine's ability to demonstrate intelligent behavior (Stanford Encyclopedia of Philosophy, 2021). The Dartmouth Conference in 1956 became a milestone for the birth of AI as a field, with pioneers like John McCarthy, Marvin Minsky (Gregersen, n.d.), Allen Newell, and Herbert Simon ([Frantz, 2003](#)). The 1960s marked the era of early AI development, including expert systems and language translation. In 1990, AI demonstrated progress in fields such as neural networks, machine learning (ML), and natural language processing (NLP), and the 21st century marked a significant improvement in the development of AI, including deep learning, reinforcement learning, and the rise of AI applications in various domains such as image recognition, speech recognition, and autonomous vehicles. Further, AI encompasses a large number of subfields, including ML (Coursera, 2024a), which involves training algorithms to learn from data and make predictions or decisions, and deep learning, which uses neural networks with many layers to analyze complex patterns in large datasets. Other subfields of AI include NLP, which enables computers to

understand and generate human language, and computer vision, which allows computers to interpret and understand visual information from the world (Coursera, 2024b). Moreover, there is intense discussion over when AI can outsmart humans. In an experiment known as the Turing Test, a human interrogator is tested in a blind conversation and asked whether computer-generated or human responses are more natural-sounding (Stanford Encyclopedia of Philosophy, 2021). Until then, we lived in an era of artificial narrow intelligence (ANI), sometimes known as weak AI, in which humans are outperformed by specialized computers programmed to perform specific tasks like text analysis and skill games. ANI comprises cognitive computing, in which machines aid people in doing activities like reading X-rays, making transactions for stockbrokers, and writing contracts for lawyers ([Licklider, 1960](#)). The abilities of it go beyond applying intelligence to any problem and tackling a specific, predetermined set of challenges ([Kaplan, 2019](#)). The artificial superintelligence (ASI) will be achieved when computers can independently exceed even the most intelligent humans ([Bostrom, 2025](#)).

12.2 AI AND DEMOCRACY

AI and democracy meet in the most complex and multiple ways, therefore drawing a skein of interrelations that shape the present and the future, posing severe challenges to governance, societal organization, and individual rights. The very tenets of democracy include liberty, equality, justice, freedom, and equal participation, aiming to ensure all voices are heard and accounted for in the decision-making process. AI, however, has huge implications on the dynamics of power, access to information, and distribution of resources with the kind of new challenges it presents to basic tenets of democracy ([Krasodonski, 2023](#)). Among probably the most important functions of AI, enhancing decision-making processes through the power of automation, predictive analytics, and data-driven insight counts as one of the main promises of this technology in democratic settings. AI systems will be in a position to assist governments in resource

allocation, policymaking, and delivery of public service through the use of huge volumes of data. For example, AI-empowered tools can be utilized to get a feel of public opinion trends, identify current issues, and anticipate societal needs.

In contrast, it can enable policymakers to respond more effectively to the demands and requirements of their constituents. Furthermore, AI has the power to democratize political involvement and information availability. AI algorithms have the potential to promote civic participation, elevate marginalized voices, and facilitate the spread of varied perspectives through platforms such as social media ([Sanders et al., 2024](#)). The more inclusive and representative democracies may result from this, enabling people to hold decision-makers responsible, participate in public discourse, and influence the course of their communities.

However, AI's rapid proliferation is also a severe threat to democracy, its principles and institutions. The most fundamental threat to use of "Weaponize AI" is to disturb and corrupt democratic elections and their processes. It can be achieved physically and physiologically via cyber-attacks that erode people's trust in the democratic process. Secondly, malicious actors can utilize AI to undermine civil society organizations and the free press to erode democratic institutions. Thirdly, AI can affect our fundamental principles of equality, due process, and economic freedom interests. AI alone can resist the most significant features of democracy: transparency, accountability, and fairness ([Polishchuk, 2024](#)).

Moreover, the possibility that AI would worsen already-existing inequality and threaten democratic norms is one of the main worries. AI algorithms can potentially reinforce preexisting biases ([Borton, 2019](#)) in training data, resulting in biased outcomes in domains including public service access, hiring practices, and law enforcement. Moreover, the concentration of AI technologies and resources in the hands of a few strong players, like tech companies or autocratic governments, may exacerbate the divide between the privileged and the underprivileged, undermining democracy's fundamental principles. AI-powered surveillance technologies also threaten civil rights and individual privacy by allowing governments to

watch and control their populace in previously unheard-of ways ([Primer, 2021](#)). The concerns over mass monitoring, government overreach, and the degradation of due process rights are raised by using AI in law enforcement and national security, from facial recognition systems to predictive policing algorithms.

12.3 AI AND ELECTORAL POLITICS

AI is a double-edged sword with the potential to generate such deep fakes to propagate false information during the election campaign. It can sway the democratic process in three main ways: first, on behalf of the voter; next, on behalf of the candidate; and finally, during the vote-counting procedure. With the most advanced AI, it is also possible to simulate anyone, whether a voter or a candidate. Hence, it is straightforward to construct an AI avatar or digital twin of any well-known politician and make any misleading claims or fraudulent pledges, leading people to mistakenly believe that this is the party's manifesto goal, as per Kartik Sharma, author of "Essentials of AI" ([Bhadoria, 2023](#)). The foundation of democracy is laid on free and fair elections, but subsequent current events have demonstrated that elections can be manipulated. AI has the potential to both expand the options for influencing elections and increase the efficacy of current efforts to stifle voting. Further, AI can potentially influence election administration, for example, by disseminating false information to reduce voter turnout ([Noti, 2024](#)). In Argentina's 2023 presidential elections, also known as the first-ever AI-driven election ([Herrera, 2023](#)), the competing parties used AI technologies to develop images and videos for their promotion and launch one another during the election campaign ([Polishchuk, 2024](#)). Another essential instance of elections is the 2023 presidential campaign in Turkey. A video showing President Recep Tayyip Erdoğan's chief opponent, Kemal Kiliçdaroğlu, receiving support from the Kurdistan Workers' Party, a recognized terrorist organization, was circulated by the president's administration in the run-up to the elections. Despite being fake, the video went viral ([Elci, 2023](#); [Polishchuk, 2024](#)). In 2024, in the November general

election of the US, a robocall generated by AI mimicked the voice of President Biden, advising people not to cast a ballot in that particular election. The message was misleading, suggesting voters should “save” their ballots ([Noti, 2024](#)). Moreover, several states are currently on the move, adopting or drafting laws that prohibit political deepfakes or require disclaimers to be displayed to notify voters when AI is used in elections. Legislation, sponsored by members of both parties in Congress, would outlaw disseminating misleading AI-generated content for fundraising or influencing elections at the federal level. Indeed, a group of twenty leading tech companies, including TikTok, Google, X (previously Twitter), Open AI, and Meta (the parent company of Facebook and Instagram), has promised to take proactive measures to identify, monitor, and stop the use of deepfakes and other forms of political meddling ([Noti, 2024](#)). Although there seems to be a lot of interest in stopping AI-based election manipulation, there is still a long way to go before many of the suggested remedies truly protect voters and the political process. Because of this, the Campaign Legal Center keeps pushing lawmakers nationwide to step up efforts and take decisive steps to deal with the particular problems AI poses for democracy.

12.4 CYBER THREATS TO ELECTIONS

Cyber threats to elections also pose a severe risk to the integrity, security, and fairness of the democratic elections. They range from disinformation campaigns to hacking attempts on the electoral system. AI technologies have the potential to boost or malign polls if they are used maliciously to manipulate data or target voters with false information; in 2016, US elections were targeted by Russian intelligence agents ([Liu, 2018](#)). Their effort intended to compromise election equipment and software in twenty-one states. They attacked voting software through cyberattacks, gained access to at least one campaign finance database, and breached the email accounts of one hundred local election officials ([Kaplan, 2019](#)). Moreover, the election hacking is not new, but the emergence of AI technologies has

aggravated it, and AI “is helping hackers attack election systems faster than officials can keep up” ([Kaplan, 2019](#)). Although cyberattacks were early AI users, using ML technology to analyze large amounts of purloined data, they can better identify targets and create plans for triumphing over cyber defenses ([Rossi, 2017](#)). Although monetary offences account for most cybercrimes, military and espionage use of cyber incursions is rising. Malware and cyberattacks can also be used to achieve other strategic objectives, such as political or ideological ones. AI is a vital weapon if the goal is to undermine democratic participation. Fortunately, AI may be employed offensively and defensively ([Rossi, 2017](#)). For instance, the Department of Defense’s DARPA Cyber Grand Challenge winner employed deep learning AI to prevent cyberattacks (DRPA, 2016). Therefore, to protect election integrity against AI-driven misinformation, it must move promptly to develop effective methods and tools.

12.5 PSYCHOGRAPHIC PROFILING

Psychographic profiling frequently entails gathering personal data about someone, which poses serious privacy issues. Customers are growing more hesitant to share their personal information in the current market due to the rising frequency of data breaches and misuse of personal information. Psychographic profiling attempts to interpret human behavior; however, it is not always simple to understand and anticipate complicated human behaviors and preferences with precision. Psychographic segmentation can offer insightful information, although it cannot always fully capture the subtleties of a given person’s behavior. For instance, it is challenging to use psychographic profiling only for long-term consumer segmentation because an individual’s interests and preferences may vary over the course of time (Capital, n.d.). There are ethical issues with the use of psychographic profiling, especially when it comes to targeted advertising and voter behavior manipulation. The biggest controversy surrounding Cambridge Analytica’s “psychographic profiling” of Facebook users was during the 2016 US and UK elections. However, voter manipulation incidents are the

only dangers AI brings to democracy. The extent of constitutional rights is reduced as governmental services are increasingly being outsourced to private companies. Subsequently, assigning these responsibilities to AI permits covert decision-making that is unaffected by public oversight. Predictive policing and AI sentencing in criminal cases are two examples of how unfair societal practices can be subtly reinforced by seemingly impartial methods ([Abrar, 2021](#)). Complex manipulating technologies have advanced to the degree that people now believe their decisions are their own, while in reality, they are “guided” by algorithms. A robust example is “big nudging,” a form of “persuasive computing” “that allows one to govern the masses efficiently, without having to involve citizens in democratic processes” ([Manheim, 2019](#)).

12.6 REGULATION OF AI

The regulation of AI is the advancement of public sector policies and promoting and regulating AI is related to the inclusive regulation of algorithms ([Nemitz, 2018](#)). The regulation of AI is the most emerging issue across the nation and globally. Although, at the beginning of 2016, soft regulation in the form of guidelines was issued to maintain social control over the emerging technologies of AI ([Heder, 2020](#)). The regulation of AI is also significant for promoting AI and managing risks. Hence, the voice of AI regulation is also high in terms of ethics, consumer protection, and cybersecure AI. In 2017, the founder of Space X, Elon Musk, also “*Warns Governors: AI Poses ‘Existential Risk’*”, “*AI is a fundamental risk to the existence of human civilization*” ([Gibbs, 2017](#); [Domonoske, 2017](#)). However, some politicians demonstrate skepticism in relation to the regulation of AI because it is still in development. As per Intel CEO Brian Krzanich, AI is still in its infancy stage, and it is premature to regulate AI technology ([Kharpal, 2017](#)) and oppose harsh regulation of AI.

Moreover, hard and soft law proposals have been proposed, and some legal experts have viewed hard law approaches to AI regulation as having substantial challenges ([Marchant, 2023](#)). Despite the difficulties, the

“pacing problem” is a situation in which established laws and regulations frequently fall behind the rapid advancement of AI technology and the risks and benefits that accompany it ([Marchant, 2023](#)). AI regulation could be navigating from some fundamental principles reviewed by the existing set of principles in 2020 by the Berkman Klein Center for Internet & Society. These eight fundamental principles were recognized by the Beijing and Asilomar Principles, among others: privacy, accountability, safety and security, transparency and explainability, fairness and nondiscrimination, human control of technology, professional responsibility, and respect for human values ([Jessica Fjeld, 2020](#)). The three primary subjects of IA law and regulations are privacy and safety concerns, responsibility and accountability for the systems, and governance of autonomous intelligence systems ([Wirtz et al., 2018](#)). Thus, a large number of national, regional, and international agencies began to take measures for the regulation of AI with different approaches. It has been said that “the United States is following a market-driven approach, China is advancing a state-driven approach, and the EU is pursuing a rights-driven approach” ([Bradford, 2023](#)).

12.7 USA AND AI

In the USA, the regulation of AI is complex and evolving as lawmakers grapple with the mushrooming growth and potential risks associated with this technology. The United States lacks a comprehensive federal law specifically to regulate AI. It promotes a “sectoral approach,” comprising a plethora of federal laws tailored to individual industries, frequently enforced by various authorities and offering a range of standards ([Solove, 2015](#)). Hence, AI in the USA is addressed through a patchwork of laws and regulations that touch on various aspects of AI, such as privacy, consumer protection, employment, and national security. On the other hand, the EU and numerous other advanced countries adopt an omnibus approach, wherein a single legislation governs data collection, utilization, and dissemination uniformly across various businesses. For instance, the “General Data Protection Regulation” (GDPR) (EUR, 2016) 284 of the EU

is a comprehensive law that covers all organizations “established” in the EU that provide goods or services to the EU or monitor individuals within the EU. Its latter characteristics give rise to the prospect of extraterritorial application and enforcement against US companies. However, many US businesses initially believe in a sectoral approach to modify laws to suit their specific requirements.

Furthermore, among the most significant federal laws is the Health Insurance Portability and Accountability Act” (HIPAA), which regulates the use and disclosure of individuals’ health information ([Haydel, 2022](#)). Apart from that, the protection of consumer privacy extends to state regulations such as the “California Consumer Privacy Act” (CCPA), which attempts to protect personal information, including information gathered by AI systems ([Bonta, 2024](#)). Moreover, regulating bias and discrimination in AI systems is another subject. AI systems can inherit biases from the data used to train them, resulting in biased outputs. This is becoming increasingly recognized, especially in recruiting, lending, and law enforcement domains. Anti-discrimination legislation, such as Title VII of the Civil Rights Act of 1964, is a federal statute designed to prevent businesses from engaging in unlawful discrimination against workers or job seekers based on any of the numerous forbidden criteria ([Mayukh, 2023](#)). It may be applicable to AI systems and is enforced by the Equal Employment Opportunity Commission (EEOC) and other agencies. Although there is no explicit federal legislation controlling AI in terms of safety or responsibility, existing product liability laws may be applicable. For instance, under state tort laws, the maker or developer of an AI-powered product may be held accountable if damage or injury results from a flaw or malfunction. However, determining who is liable in AI situations can be difficult, especially if the technology uses an opaque or poorly understood decision-making process. Hence, legislators, business representatives, and advocacy organizations are all involved in the evolving regulatory landscape surrounding AI in the United States. There will probably be more calls for thorough and aggressive regulation to adequately address the

advantages and threats of AI as the technology develops and its influence on society expands.

12.8 EU AND AI

The European Union has been pioneering in establishing a regime regulating AI technology. It has recognized the need for a comprehensive framework to address AI adoption's ethical, legal, and societal implications. In 2019, the European Commission issued ethical standards for reliable AI, defining concepts like transparency, responsibility, equity, and human supervision (European Parliament, 2019). Their most fundamental instrument, the AI Act, was proposed in 2021 and finally passed on 13 March 2024, the world's first comprehensive legislation on AI (European Parliament, 2023). Thus, it aims to regulate AI technology use within the EU. This act seeks to ensure individuals' safety and fundamental rights while promoting innovation and competitiveness in AI technologies. It covers diverse areas of AI, including those used in high-risk sectors such as healthcare, transportation, and law enforcement. It also addresses AI applications that pose risks to essential rights, such as biometric identification systems. The most fundamental feature of the Act is that it is a risk-based approach, which is one of its main characteristics. Based on the amount of risk involved, AI systems are divided into four groups: low risk, high risk, limited risk, and unacceptable risk. Stricter guidelines apply to high-risk AI systems, such as data quality, human monitoring, and conformance tests, apply to high-risk AI systems (European Parliament, 2023). Moreover, EU AI Act could be a game changer in the field of AI. It establishes strict guidelines for the creation and application of ethical AI. Furthermore, with regard to privacy, EU privacy laws are one of the most significant in protecting individuals' personal data and ensuring their fundamental right to privacy. The milestone of EU privacy legislation is the "GDPR," 2016. "It is the strongest privacy and security law in the world" (Council of the European Union, n.d.). All companies must obtain explicit consent from individuals before collecting and processing their personal

data. Thus, companies have to explain how individuals use data and obtain their consent. The GDPR also places stringent obligations on businesses to sustain the security of personal data, mandating that they put in place the necessary organizational and technical safeguards to guard against data breaches and unauthorized access.

In case of violation, companies must report a data breach to the appropriate supervisory body within seventy-two hours and, in some situations, notify the impacted parties as well. Moreover, GDPR also inculcates the notion of data protection by design and by default, requiring companies to consider data protection issues from the outset when designing new products or services and to implement privacy-friendly default settings (ICO, n.d.). to support data protection, EU also EU has enacted additional privacy rules and regulations in addition to the GDPR. For instance, the ePrivacy Directive. It establishes guidelines for using cookies and other related tracking technology and regulates the privacy of electronic communications. The proposed ePrivacy Regulation aims to harmonize with the GDPR, thereby modernizing and fortifying the laws regulating electronic communications. In a nutshell, EU regulation of AI is more comprehensive than the US, which is a fragmented and sector-specific approach in which AI regulation falls under various federal agencies, with some relying on existing laws and non-binding guidelines. This approach prioritizes innovation and self-regulation and is frequently viewed as more industry-friendly. However, the absence of a centralized regulator has raised questions about its efficacy and consistency. Conversely, the EU GDPR and AI Act seek to regulate AI applications across various sectors. Hence, the EU approach has stricter rules on high-risk AI systems and generally takes a more cautious stance, emphasizing risk assessment and transparency.

12.9 CONCLUSION AND SUGGESTIONS

In conclusion, there are advantages and disadvantages to the coexistence of democratic ideals and AI. Although AI can improve democratic processes through greater transparency, citizen interaction, and decision-making, it

poses concerns about privacy, prejudice, and inequality. Ensuring AI maintains democratic principles requires solid regulatory frameworks that prioritize fairness, accountability, and transparency in AI development and application. Furthermore, to cope with emerging challenges and develop AI in a way that benefits society as a whole, it is imperative that policymakers, engineers, ethicists, and civil society collaborate across disciplinary boundaries. We can protect democratic ideals and values while using AI's transformational potential by actively participating in discussion and debate, encouraging diversity and inclusion in AI development, and giving citizens the knowledge and skills to understand and influence AI governance.

REFERENCES

- Abrar, N. A. (2021). *Infamous for manipulating polls, Cambridge Analytica is a non-entity today*. Retrieved April 9, 2024, from www.business-standard.com/article/international/infamous-for-manipulating-polls-cambridge-analytica-is-a-non-entity-today-121012201513_1.html
- Bhadoria, S. (2023, July 12). *Lok Sabha Elections: How AI will be a double-edged sword to boost fake news, curb voting manipulations in upcoming polls*. Retrieved April 5, 2024, from www.livemint.com/news/india/lok-sabha-elections-2024-how-ai-will-be-a-double-edged-sword-to-boost-fake-news-curb-voting-manipulations-11708705430974.html
- Bonta, R. (2024). *California Consumer Privacy Act (CCPA)*. Retrieved April 22, 2024, from <https://oag.ca.gov/privacy/ccpa>
- Borton, L. (2019, May 22). *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*. Retrieved April 2, 2024, from www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/
- Bostrom, N. (2025). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

- Bradford, A. (2023). *The race to regulate artificial intelligence*. Retrieved April 8, 2024, from www.foreignaffairs.com/united-states/race-regulate-artificial-intelligence
- Coursera. (2024a, March 12). Retrieved from: www.coursera.org/articles/natural-language-processing
- Coursera. (2024b, March 14). *What does AI stand for?* Retrieved from www.coursera.org/articles/what-does-ai-stand-for
- Domonoske, C. (2017). *Elon Musk: Artificial intelligence poses 'existential risk'*. Retrieved April 3, 2024, from www.npr.org/sections/thetwo-way/2017/07/17/537686649/elon-musk-warns-governors-artificial-intelligence-poses-existential-risk
- DRPA. (2016). *DARPA celebrates cyber grand challenge winners*. Retrieved April 8, 2024, from www.darpa.mil/news-events/2016-08-05a#:~:text=DARPA%20officials%20this%20morning%20released,by%20team%20ForAllSecure%20of%20Pittsburgh.
- Elci, A. (2023). *AI content is meddling in Turkey's election. Experts warn it's just the beginning*. Retrieved April 3, 2024, from www.euronews.com/next/2023/05/12/ai-content-deepfakes-meddling-in-turkey-elections-experts-warn-its-just-the-beginning
- EUR. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data*. Retrieved April 4, 2024, from <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- European Parliament. (2019). *EU guidelines on ethics in artificial intelligence: Context and implementation*. Retrieved April 6, 2024, from [www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI\(2019\)640163_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf)
- European Parliament. (2023, December 12). Retrieved from www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

- Frantz, R. (2003). *Herbert Simon. Artificial intelligence as a framework for understanding intuition*. Retrieved March 4, 2024, from [https://doi.org/10.1016/S0167-4870\(02\)00207-6](https://doi.org/10.1016/S0167-4870(02)00207-6)
- Gibbs, S. (2017). *Elon Musk: regulate AI to combat 'existential threat' before it's too late*. Retrieved April 2, 2024, from www.theguardian.com/technology/2017/jul/17/elon-musk-regulation-ai-combat-existential-threat-tesla-spacex-ceo
- Gregersen, E. (n.d.). *John McCarthy American mathematician and computer scientist*. Retrieved March 2, 2024, from www.britannica.com/biography/John-McCarthy
- Haydel, E. (2022, February 12). *Health Insurance Portability and Accountability Act*. Retrieved April 5, 2024, from www.ncbi.nlm.nih.gov/books/NBK500019/
- Heder, M. (2020). *A criticism of AI ethics guidelines*. Retrieved April 3, 2024, from 10.22503/inftars.XX.2020.4.5
- Herrera, J. N. (2023). *Is Argentina the first A.I. election?* Retrieved April 2, 2024, from www.nytimes.com/2023/11/15/world/americas/argentina-election-ai-milei-massa.html?auth=login-google1tap&login=google1tap
- Jessica Fjeld, N. A. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Retrieved April 3, 2024, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482
- Kaplan, M. (2019). Artificial intelligence: Risks to privacy and democracy. *Yale J.L. & Tech*, 21, 108–188.
- Kharpal, A. (2017). *A.I. is in its 'infancy' and it's too early to regulate it, Intel CEO Brian Krzanich says*. Retrieved April 3, 2024, from www.cnbc.com/2017/11/07/ai-infancy-and-too-early-to-regulate-intel-ceo-brian-krzanich-says.html
- Krasodonski, A. (2023). *From risk to revolution: How AI can revive democracy*. Retrieved April 2, 2024, from www.chathamhouse.org/publications/the-world-today/2023-10/risk-revolution-how-ai-can-revive-democracy

- Licklider, J. C. (1960). *Man–computer symbiosis*. Retrieved March 4, 2024, from <https://groups.csail.mit.edu/medg/people/psz/Licklider.html>
- Liu, H.-Y. (2018). The power structure of artificial intelligence. *Innovation and Tech*, 10(2), 197–229 . Retrieved April 2024, from [10.1080/17579961.2018.1527480](https://doi.org/10.1080/17579961.2018.1527480)
- Manheim, K. (2019). *Artificial intelligence: Risks to privacy and democracy*. Retrieved April 3, 2024, from <https://law-journals-books.vlex.com/vid/artificial-intelligence-risks-to-790547417>
- Marchant, G. (2023). “Soft Law” governance of artificial intelligence. Retrieved April 3, 2024, from https://escholarship.org/content/qt0jq252ks/qt0jq252ks_noSplash_1ff6445b4d4efd438fd6e06cc2df4775.pdf
- Mayukh, G. B. (2023). *The Buck Stops with You: Artificial Intelligence, Employment, & Title VII of the Civil Rights Act*. Retrieved April 4, 2024, from www.wardandsmith.com/articles/Artificial-Intelligence-Employment-and-Title-VII-of-the-Civil-Rights-Act#:~:text=Title%20VII%2C%20which%20applies%2C%20generally,m,ans%20biological%20sex%2C%20pregnancy%2C%20sexual
- Nemitz, P. (2018). *Constitutional democracy and technology in the age of artificial intelligence*. Retrieved April 2, 2024, from <https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0089>
- Noti, A. (2024). *How artificial intelligence influences elections, and what we can do about it*. Retrieved April 7, 2024, from <https://campaignlegal.org/update/how-artificial-intelligence-influences-elections-and-what-we-can-do-about-it>
- Polishchuk, A. (2024). *AI poses risks to both authoritarian and democratic politics*. Retrieved January 2, 2024, from www.wilsoncenter.org/blog-post/ai-poses-risks-both-authoritarian-and-democratic-politics
- Primer, A. (2021). *Artificial intelligence, human rights, democracy, and the rule of law*. Retrieved March 3, 2024, from <https://rm.coe.int/primer-en-new-cover-pages-coe-english-compressed-2754-7186-0228-v->

[1/1680a2fd4a#:~:text=When%20codified%20into%20law%2C%20these, and%20the%20rule%20of%20law&#x2191](#)

Rossi, N. (2017). *The promises and perils of emerging technologies for cybersecurity*. Retrieved April 3, 2024, from [www.govinfo.gov/content/pkg/CHRG-115shrg28382/pdf/CHRG-115shrg28382.pdf&#x2191](#)

Sanders, N., Schneier, B., and Eisen, N.. (2024). *How public AI can strengthen democracy*. Retrieved April 2, 2024, from [www.brookings.edu/articles/how-public-ai-can-strengthen-democracy/&#x2191](#)

Solove, D. (2015). *The growing problems with the sectoral approach to privacy law*. Retrieved April 4, 2024, from [https://teachprivacy.com/problems-sectoral-approach-privacy-law/&#x2191](#)

Stanford Encyclopedia of Philosophy. (2021, March 12). *The Turing Test*. Retrieved from [https://plato.stanford.edu/entries/turing-test/#:~:text=The%20phrase%20%E2%80%9CThe%20Turing%20Test,to%20d](#)

Wirtz, B. W., Weyerer, J. C., and Geyer, C. (2018). Artificial intelligence and the public sector—applications and challenges. *International Journal of Public Administration*, 42(7), 596–615.[&#x2191](#)

13 AI in Law Enforcement and Counter-Terrorism

Supporting SDG 16 for Human Rights

Bhupinder Singh, Christian Kaunert, Saquib Ahmed, Anjali Raghav, and Saurabh Chandra

DOI: [10.1201/9781003668657-13](https://doi.org/10.1201/9781003668657-13)

13.1 INTRODUCTION

Counter-terrorism refers to the steps taken by governments and global entities to not only prevent but also combat acts of terrorism ([Thapaliya & Sharma, 2023](#)). Counter-terrorism comprises practices, tactics, techniques, and strategies that governments, militaries, police departments, and corporations adopt to combat or prevent terrorism, which includes methods for preventing terrorist acts; detecting potential plans for terrorism occurring, forcing counter terrorists to ensure they can predict if an existing attack is likely ([Dash et al., 2022](#)). Counter-terrorism efforts involve multiple levels of activity, from intelligence collection to law enforcement operations, border control, and legal frameworks to enhance punitive measures or cooperate internationally ([Barocas et al., 2023](#)), also tackling the root causes triggering terrorist activities.

The most important activities in counter-terrorism are intelligence gathering, for information collection and analysis to facilitate not only

identifying possible threats but investigating terrorist networks as well, which also helps prevent attacks ([Narayana & Sreedevi, 2023](#)). It is the duty of law enforcement agencies to capture terrorists, conduct investigations into their activities, and then subject them to due legal processes ([Mhlanga, 2021](#)). All these measures mean they are also essential to border control, which has a direct bearing on the target groups of terrorists and their financing and arms trafficking in connection with terrorism ([Pocher et al., 2023](#)).

Adequate legal frameworks are needed to provide international support for counter-terrorism measures and ensure that terrorists engaged in this kind of activity “to be brought before the court” ([Chen et al., 2018](#)). Such frameworks must be devised in a way that respects human rights and civil liberties, while taking adequate steps on the part of authorities for prevention, investigation, and prosecution of terrorism-related offences. Terrorist activities often go beyond the borders of a country, and hence international cooperation holds much significance in tackling terrorism ([Tyagi et al., 2020](#)). These include intelligence-sharing, border security coordination, and mutual legal harmonization, as well as supporting nations in investigations and prosecutions ([Cao et al., 2021](#)). It involves the cooperation of multilateral organizations such as the United Nations and Interpol.

The other crucial element that effective counter-terrorism entails is responsibility in tackling the root causes of terrorism ([Alazab et al., 2023](#)). This would mean addressing the roots of socio-economic disparities, political grievances, and radicalization from being an attractive concept, while also not inciting ideological extremism. Measures to foster inclusive governance, human rights observance, social justice, and economic development can help in addressing the root causes that lead to terrorism and extremism (Zhang et al., 2022a).

Counter-terrorism is certainly needed to ensure proper protection of individuals and security in society, but counter-terrorism should rather be pursued within the boundaries of human rights norms, respecting rule-of-law principles as well as basic freedoms (Singh, 2023d). To prevent

counter-terrorism measures from being abused and to maintain legitimacy, it is necessary for security efforts to be balanced with civil liberties ([Lee & Shin, 2020](#)). Hence, it implies a lot of things can be done to prevent terrorist attacks which is why counter-terrorism comprises an array of tactics and measures in deterring the intended acts, destroying terror cells and networks so as to act or promote security ([Bhuvaneshwari & Kaythry, 2023](#)). Societies can counter terrorism by integrating intelligence, law enforcement, international cooperation, and tackling root causes while simultaneously maintaining peace, justice, as well human rights ([Li et al., 2020](#)).

These are interlinked and regarded as the new threat to political/social and economic stability globally. There is no accepted definition for this term “terrorism” universally, but it may be taken as a means of intimidation which employs or threatens to employ violence to diffuse fear. The call to counter-terrorism sentient is nothing new and has a long history in itself as there are big challenges. In the present day, terrorism and criminal financing manifest in varied forms of violence targeting a civilian population, military personnel, as well as officials from the State and Central government ([Bhuvaneshwari & Kaythry, 2023](#)). Almost all of those groups used the main brutal type of terror, which was assassination committed to jubilant martyrdom and therefore carried considerable risks for its perpetrators (see [Mühlhoff, 2021](#)).

Some parts of a government or political people ([Saltz et al., 2019](#)) used terrorist attacks as an instrument for politically motivated acts, and to provoke fear among the mass, thus they targeted specific members. This categorization distinguishes between targeted (focused) terrorist attacks and broader, indiscriminate violence. Historically, most terrorist incidents have been primarily targeted in nature, though there has been a notable rise in the scale and frequency of indiscriminate attacks over time. Victims of terrorist attacks are not chosen specifically by personal characteristics, but “are chance as the victims who happen to be at wrong place at wrong time” (Zhang et al., 2022b).

The terrorism and criminal financing utilized different varieties of violence, targeting civilians, military staffs along with officers from state and central governments ([Baker & Robinson, 2020](#)). The thousand-natured challenges of terrorism are not new and have a long history as the most complex entity. At that time, the main mode of dissemination of terror used by nearly all these sets was defection assassination, which also held up some probability for political martyrdom but could mean personal financial danger ([Aziz & Andriansyah, 2023](#)).

13.2 COUNTER-TERRORISM: RELEVANCE

Precisely to measure the attitude of international society about terrorism, the chapter aims at understanding laws against Terrorism worldwide as well as in India. Any attempt to find out what the commonalities and non-commonalities related definition of terrorism are can be an understatement; in other words, this article is nothing else but trying to look beyond/understand what probably went wrong when we think about a universal law ([Rai & Sahu, 2020](#)).

13.3 OBJECTIVES OF THE CHAPTER

Enforcement is a first line of defense to prevent the activities of any terrorist organizations, and this can be through intelligence, surveillance, and investigation ([Sharma et al., 2023](#)). They are aimed at finding out and nullifying the possible threats that will prevent them from occurring in a healthy way so that peace and tranquility may be preserved within society ([Sharma et al., 2023](#)). Furthermore, police forces are on the frontline of bringing terrorists to justice and collecting evidence against them for their prosecution. It is about ensuring that terrorists are brought to justice for their crimes, but it also contributes to strengthening the institutions of law and order ([Varma et al., 2022](#)). Furthermore, including law enforcement in counter-terrorism activities encourages cooperation and facilitates national or international intelligence sharing between police forces, which helps to

create more resilient institutions that are better-established around the world for addressing cross-border security concerns ([Tippins et al., 2021](#)). Furthermore, the effectiveness of counter-terrorism strategies depends on law enforcement agencies acting in a manner that is proactive and thorough to meet Sustainable Development Goals (SDG) 16's broader goals for promoting peaceful societies free from violence ([Lui & Lamb, 2018](#)).

13.4 FACTORS LEADING TO TERRORIST ACTIVITIES

The causes were terrorist activities are interrelated systems of communicating vessels. This means the surplus elements, that is, their joint relations and various existing situations interacting with each other ([Mijwil et al., 2023](#)). Terrorism is a given name and it occurs in many ways, as religious, ethnic, or ideological conflicts fuel terrorism; the consequences of modernization can display in negative condition for human life that actuates terrorist ambitions inside some societies. Alternatively, several other dimensions may be included, similar to political–economical–social–cultural–psychological aspects from the previous history and incidence toward internal power struggle.

13.5 DIVERSION OF REVENUES TO FINANCE TERRORISM

The connection between terrorist activities and criminal financing is more manifest now; the multimodal sources of finance usually vary according to dimensionality and ideology as well as its applicability ([Siwakoti et al., 2023](#)). Due to the need for finances, these terrorist groups and organizations can perform various illegal activities ([Zhang et al., 2020](#)). Most of these illegal operations often have a cross-border dimension and include activities such as drug trafficking, smuggling, human trafficking, wildlife poaching transnational organized crime where the flow over borders which can be harming financial gain or no gains in terms of criminal payload like credit frauds high tolling counterfeits good etc. ([Nassar & Kamal, 2021](#)). There is

a multifarious blend of source of finance to cater to the targets, modus operandi, and scale of terrorism operations as those terrorist organizations who are high-value profit-procuring capabilities get much exposed due to having higher dependants on income revenues ([Alruwaili, 2023](#)).

13.6 INTERNATIONAL LEGAL INSTRUMENTS TO COMBAT TERRORISM

Legal instruments provide a necessary core in the overall response to addressing terrorism at an international level ([Grimm et al., 2021](#)). These devices include a variety of agreements, conventions, and treaties which allow the nations to unite against terrorism ([Dhoni & Kumar, 2023](#)). It contains a complete strategy to counter terrorism, giving guidelines for states on what they should include in their counter terror actions by providing the reference strategic basis according to which security issues compete against freedoms and seize when others present cooperation, with respect to human rights ([Mijwil & Aljanabi, 2023](#)). The progress reflects our shared commitment to prevent and prosecute terrorism, in full compliance with the fundamental principles of justice, cooperation, and respect for human rights by all states as urged in those key global legal non-binding measures ([Sambasivan et al., 2021](#)). These legal frameworks are important steps for a safer and more peaceful global environment in our increasingly integrated world (Singh, 2023d).

13.7 COUNTER-TERRORISM AND UNITED NATIONS: COMMITTEES

The Counter-Terrorism Committee (CTC): This committee is responsible for enhancing the ability of U.N. Member States on a cross-regional basis and was established post 9/11 terrorist attacks on the United States. CTC is a functional body of the UN Security Council, supported by the CTC Executive Directorate for carrying out its policy decisions; providing

counter-terrorism technical assistance to individual countries and conducting expert evaluations on every Member State ([Katyal, 2019](#)).

13.7.1 THE BASIS OF INTERNATIONAL HUMAN RIGHTS LAW

As a declaration, the Universal Declaration of Human Rights has become an inspiration to us all in our endeavor to make amends for injustices, to embrace peace in times of conflict, to remedy societal machinations of repression, and has pledged us to the creation of international human rights law itself. Following this, the declarative text has given rise to a plethora of legally binding international human rights treaties that have collectively been formulated to strive for the realization of human rights and fundamental freedoms for all peoples. It states that it is appropriate to remember that “all human beings are born free and equal in dignity and rights in reiteration of the intrinsic dignity and the equivalent and the inalienableness of all humans rights.” This is the core of our aim – whatever our nationality, place of residence, gender, national or ethnic ancestry, color, religion, language, or any other status, we are all members of the same human family. The commitment of the international community on December 10, 1948 was a shared aspiration for justice and self-esteem. This mutual aspiration for unity has translated into law over the decades, with a variety of treaties, regional agreements, custom laws, general values, and national laws outlining human policies. The Universal Declaration of Human Rights (UDHR) has become a source of inspiration to numerous international human rights treaties and declarations, domestic human rights laws, and regional human rights agreements to promote and protect human rights.

Building on the understandings of the UDHR, where it became operative in 1976 ([Daugherty & Wilson, 2018](#)): They established basic human rights, for example, the right to life, freedom of speech, or equal treatment before law and social welfare, such as the rights to work, housing security, or education ([Świątkowska, 2020](#)). International human rights treaties grow in

scope over time, becoming more specialized and focused both with respect to the major issues that they address as well as particular social groups identified for protection ([Berk et al., 2021](#)). The body of international human rights keeps evolving, expanding, and clarifying the essential freedoms held in the International Bill of Human Rights, dealing with subjects like racial discrimination as well enforced disappearances, the right to avoid torture, including a capacity-related right, disability, just as privileges for ladies kids migrants minorities indigenous people ([Abbas et al., 2024](#)). Governments commit to bringing domestic measures and legislation in line with the treaty obligations and responsibilities through ratification of international human rights treaties ([Sudhakar & Kumar, 2023](#)). The domestic legal system is the primary context for realizing human rights enshrined in international law ([Kochhar et al., 2023](#)). Complaint mechanisms and procedures for individual and group complaints are available at the regional as well as international level to ensure that the local action does not give bread crumbs to let down the International human rights standard. These mechanisms step in for helpless people where domestic legal actions cannot be sought. As a result, the United Nations is increasingly asked to organize in global terms of fighting against terrorism ever more, and the number of universal instruments dealing with special terrorist activities has been elaborated within the framework of the UN system ([Liu et al., 2024](#)). UN Member States have never before agreed to a common strategic and operational framework for combating terrorism, until they adopted the United Nations Global Counter-Terrorism Strategy in September 2006 ([Rajkumar et al., 2023](#)).

13.7.2 ASEAN CONVENTION ON COUNTER-TERRORISM

The Association of Southeast Asian Nations includes, among others, Brunei Darussalam, the Republic of Indonesia, the Kingdom of Cambodia, Lao People's Democratic Republic (Laos), Malaysia, and the Philippines. Everybody is most worried about the extreme danger of terrorism to human lives, infrastructure, and environmental conservation, regional and global

peace, and economic development ([Shukla & Tiwari, 2024](#)). Root cause identification: Root causes behind terrorism and the need of integrating them in counter-terrorism planning ([Paddalwar et al., 2023](#)). A firm commitment should be made to protect human rights, ensuring fair treatment, the rule of law, and due process, including in all appropriate counter-terrorism measures.

13.8 DIVERSION UP TO SDG 16: PEACE, JUSTICE, AND STRONG INSTITUTIONS

SDG 16 Peace, Justice, and Strong Institutions. By the year of commencement, you dive into creating a more sustainable world for all by embarking on your journey at university or further tertiary studies through to graduation. This aim is meant to ensure that equal access to justice for all and tries in the promotion of peaceful, inclusive societies with effective, transparent, accountable institutions ([Odeh & Abu Taleb, 2023](#)). Peace is not just the absence of war, it includes social harmony, respect for human rights, and promotion of non-violence. The cornerstone of a fair and just society is justice, serving to confirm that individual rights are the same as legal process and innovation. Stability, transparency, and rule of law ([Ortiz-Ruiz et al., 2024](#)) need well-functioning institutions to be implemented.

Peace, Justice, and Strong Institutions: SDG 16 recognizes that peace, justice, and strong institutions are mutually reinforcing ([Ali et al., 2024](#)). People should live in a place where they are free to go out and do anything of their choice until or unless it is under the legal minimum threshold. It underlines the need for addressing root causes of conflict and reinforcing social cohesion (as well as) violence reduction in all its shapes. To achieve this aim, legal frameworks should be made stronger and action toward access to justice as well as capacity building for institutions at the local level ([Latif et al., 2024](#)). This involves ensuring more effective legal systems and also greater empowerment of marginalized groups against the system. In addition to all of that, working in the field of fighting corruption

and enhancing transparency are principles for building strong institutions ([Anis et al., 2023](#)).

SDG 16 further identifies the importance of inclusive and participatory governance in which no voice is unheard, thus balancing decisions are taken openly, transparently and with accountability. People, on the other hand, articulate how civil society must be consulted and freedom of expression promoted while human rights defenders are well protected. Likewise, it underscores the importance of establishing effective and accountable global institutions to confront transnational threats like organized crime, terrorism, as well as illicit finance sector.

SDG 16: Promote Justice, Peaceful and Inclusive Societies Ultimately. Even where not explicitly headlined, SDG 16 aims to create a world that works for everyone ... one in which justice is what must prevail; peace just has to become sustainable (#Right?), and institutions truly are resilient! It recognizes that peace and justice are the basis for other human rights as well as a requirement to achieve all the rest of sustainable development goals. The third group is peace, justice, and strong institutions. Investment in it means to create a better future for everyone which rich of diversity, fairness sustainability ([AL-Dosari et al., 2024](#)).

It sets a number of interconnected goals, designed to help achieve sustainable development by fostering peaceful and inclusive societies which will provide access to justice for all and build effective, accountable institutions at all levels. In essence, SDG 16 recognizes that peace, security, and effective institutions are not just enablers but are critical ends in their own right too for the achievement of other SDGs. The well-being and progress of human societies depend fundamentally on peace ([Ambika, 2024](#)). SDG 16: Promote peaceful and inclusive societies for sustainable development, provide access to justice for all, and build effective, accountable, and inclusive institutions at all levels, while highlighting the focus on preventing violence (leading with a culture of peace) in nations. It realizes that development must be sustainable, attainable only when societies are free of violence and insecurity. Necessary for the social

progress, economic growth and protection of human rights require peaceful or stable societies.

This root and branch reform has justice at its core – SDG 16 calling for access to justice for everyone, no matter how poor or socially marginal they may be. It states the need to uphold and promote a rule of law, that all countries have effective, robust, and accountable legal systems (that can accompany fund budgets for you in Syria), and assure human rights. Transparent, accountable and impartial justice systems not only serve to prevent violations of peoples individual rights but also contribute more broadly toward social harmony, trust and the health and well-being of societies ([Lone et al., 2023](#)).

Sustainable development hinges upon institutions. Latvia hopes that by meeting the ideals of SDG 16, it can minimize corruption and build a strong rule-of-law environment. The study highlights the role of anti-corruption, good governance initiatives, and in ensuring that marginalized are included in decision-making. So, if you want to allow for the conditions that can stir all three (economic growth, social progress, and protection of rights) then we need strong institutions. SDG 16 recognizes the interrelation among these goals with respect to the broader development agenda while addressing complex issues related to peace, justice, and strong institutions. It recognizes that there can be no peace and justice without respect for human rights, which are the bedrock of any commitment to open societies where sustainable development can take root. SDG 16 can only be attained jointly by the governments, civil society, business stakeholders, and members working together to help make peace sustainable, rights secure, and systems sound. It is only through these acts of togetherness that humanity can move further and lay the foundation for a world where justice, equality peace lie in every hollow ground of this earth – for those before us and long after our memories are gone ([Messai & Seba, 2023](#)).

13.9 CONCLUDING OBSERVATIONS

As a matter of fact, terrorism and terrorist actions are well documented as the human societies' world enemy to all time. Terrorism in its broadest and most destructive effect is obliteration, radical crimes against humanity. There is compelling evidence supporting various processes of dispossession through contemporary socio-economic and political conditions, which indirectly propel people toward terrorism and that terrorist organizations using terrorist tactics differ in their potential susceptibility to recruitment.

Goal 16 of the United Nations SDGs is to promote peaceful and inclusive societies for sustainable development, provide access to justice for all, and build effective, accountable, and inclusive institutions at all levels. It acknowledges that sustainable development is only possible if the root causes of conflicts are addressed, access to justice is promoted, and accountable governance frameworks are in place. SDG 16 aspires toward societies where people enjoy living free and productive lives, and governments deliver justice without fear or favor. SDG 16 recognizes that sustainable development can only be achieved with transparent, accountable, and inclusive institutions, the rule of law, and human rights for all people. It should be recognized that inclusive decision-making, accountability, and respect for fundamental freedoms are essential to building social cohesion, peace conciliation, and achieving sustainable development. SDG 16's purpose is one of the essential building blocks to a fair and peaceful world for every human being on this planet, probably till generation XXI ([Singh et al., 2023](#)). Therefore, what really needs to be done now is up to such legal and social technologies in the continuing fight against terror-associated crime (Sowmya, 2023).

13.10 FUTURE SCOPE AND SUGGESTIONS

This chapter recommends the use of focused simulations to reinforce better law enforcement in counter-terrorism and this is relevant to SDG 16. These simulations must feature scenario-based training that includes inter-agency collaboration, compliance with laws and legal frameworks, community engagement, cybersecurity measures as well as cross-border cooperation.

Through these practice drills, law enforcement can perfect their game plans and work alongside the government to ensure peace, justice, and strong institutions (SDG 16) in a way that preserves citizen rights as well. This simulates law enforcement's role in bettering counter-terrorism efficiency while also respecting the aims of SDG 16 through interagency-type training and collaboration exercises. They could involve state or local police and intelligence bodies from anywhere on earth – even foreign allies. The experience of these exercises can be used to improve coordination, information sharing, and response times by mirroring real-life scenarios with training institutions capable of addressing terrorism while upholding the rule of law.

Legal Scenario Simulations: To prepare law enforcement and legal professionals for the difficult challenges faced by counter-terrorism efforts, it may be useful to do what is known as “legal scenario simulations.” These simulations can deal with questions of evidence admissibility, fair trial, and ways into counter-terrorism. Maintaining that balance between national security and law enforcement gives policing in these scenarios an improvement through more of this practice.

Counter-Terrorism Community Engagement Simulations: Counter-terrorism strategies frequently aim to engage local communities who are at risk of radicalization, and community engagement is also important when gathering intelligence. When law enforcement agencies run through a variety of community engagement scenarios, they will be able to improve their trust-building and information-gathering skills as well as collaboration with communities. It will promote respect for cultural and religious sensitivities, as well as encourage human rights values – echoing SDG 16.

Cybersecurity and Digital Forensics Simulations: Considering the use of Uber technology in terrorist activities, there should be simulations concerning cybersecurity and digital forensics. Apply for Top & Best Colleges Here. These exercises can strengthen the law enforcement to

monitor and investigate online radicalization, manifestos, cyberattacks as well as how terrorists are using digital tools. This is important for the institutional table stakes required to cope with changing circumstances in security (post Cold War).

Cross-Border Simulations: Terrorism is often a challenge that crosses national borders and requires cross-border cooperation. International law enforcement agency simulations can run joint operations, enact extradition procedures, and share information across boundaries. DMEs can test whether existing norms controlling illicit financial flows (IFF) work, such as extradition treaties or mutual assistance agreements that help to advance SDG 16 and its objective of establishing effective institutions at all levels.

Legal Compliance Reviews: Every simulation should also have a legal compliance review that evaluates the steps taken during natural disasters for conformity to domestic and international law by experts. Through this feedback mechanism, we can determine areas in which law enforcement agencies may need to refine their strategies in accordance with the rule of law and principles against arbitrary detention. *Debriefings and Continuous Improvement:* After each simulation, thorough debriefs allow participants to discuss their process and identify other areas where they can improve. This iterative process continues to evolve, ensuring that law enforcement agencies are continuously improving their tactics and procedures in alignment with the objectives of SDG16, but at the same time knocking terrorism off balance. While still a work in progress, the synthesis of these varied simulation scenarios into law enforcement training programs may go some way to enabling agencies that regularly come together within coalition partnerships respond more effectively and efficiently when it comes to counter-terrorism efforts but at the same time support wider goals enshrined within SDG 16 driving for peace justice and strong institutions as being integral tenets from which global security can be achieved.

REFERENCES

- Abbas, S., Bouazzi, I., Ojo, S., Al Hejaili, A., Sampedro, G. A., Almadhor, A., & Gregus, M. (2024). Evaluating deep learning variants for cyber-attacks detection and multi-class classification in IoT networks. *PeerJ Computer Science*, 10, e1793. [↗](#)
- Alazab, M., Soman, K. P., Srinivasan, S., Venkatraman, S., & Pham, V. Q. (2023). Deep learning for cyber security applications: A comprehensive survey. *TechRxiv*. October 12, 2021. doi: [10.36227/techrxiv.16748161.v1](https://doi.org/10.36227/techrxiv.16748161.v1) [↗](#)
- AL-Dosari, K., Fetais, N., & Kucukvar, M. (2024). Artificial intelligence and cyber defense system for banking industry: A qualitative study of AI applications and challenges. *Cybernetics and Systems*, 55(2), 302–330. [↗](#)
- Ali, H. H., Naif, J. R., & Humood, W. R. (2024, March). Deep learning algorithms for IoT security (survey). In *AIP Conference Proceedings* (Vol. 2885, No. 1). AIP Publishing. [↗](#)
- Alruwaili, F. F. (2023). Optimal hybrid deep learning enabled attack detection and classification in IoT environment. *Computers, Materials and Continua*, 75(1), 99–115. [↗](#)
- Ambika, N. (2024). Early detection of darknet traffic in internet of things applications. *Automated Secure Computing for Next-Generation Systems*, 1, 139–153. [↗](#)
- Anis, G. M., Aboutabl, A. E., & Galal, A. (2023). Machine learning for detecting cybercrime in the banking sector. *Journal of Southwest Jiaotong University*, 58(5), 12–13. [↗](#)
- Aziz, L. A. R., & Andriansyah, Y. (2023). The role artificial intelligence in modern banking: An exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. *Reviews of Contemporary Business Analytics*, 6(1), 110–132. [↗](#)
- Baker, D. J., & Robinson, P. H. (Eds.). (2020). *Artificial Intelligence and the Law: Cybercrime and Criminal Liability*. Routledge. [↗](#)

- Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press. [↵](#)
- Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 50(1), 3–44. [↵](#)
- Bhuvaneshwari, A. J., & Kaythry, P. (2023). A review of deep learning strategies for enhancing cybersecurity in networks: Deep learning strategies for enhancing cybersecurity. *Journal of Scientific & Industrial Research (JSIR)*, 82(12), 1316–1330. [↵](#)
- Cao, L., Yang, Q., & Yu, P. S. (2021). Data science and AI in FinTech: An overview. *International Journal of Data Science and Analytics*, 12, 81–99. [↵](#)
- Chen, Z., Van Khoa, L. D., Teoh, E. N., Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: A review. *Knowledge and Information Systems*, 57, 245–285. [↵](#)
- Cheng, X., Liu, S., Sun, X., Wang, Z., Zhou, H., Shao, Y., & Shen, H. (2021). Combating emerging financial risks in the big data era: A perspective review. *Fundamental Research*, 1(5), 595–606.
- Dash, B., Sharma, P., & Ali, A. (2022). Federated learning for privacy-preserving: A review of PII data analysis in Fintech. *International Journal of Software Engineering & Applications (IJSEA)*, 13(4), 1–13. [↵](#)
- Daugherty, P. R., & Wilson, H. J. (2018). *Human + Machine: Reimagining Work in the Age of AI*. Harvard Business Press. [↵](#)
- Dhoni, P., & Kumar, R. (2023). Synergizing generative AI and cybersecurity: Roles of generative AI entities, companies, agencies, and government in enhancing cybersecurity. *TechRxiv*. doi: [10.36227/techrxiv.23968809.v1](https://doi.org/10.36227/techrxiv.23968809.v1) [↵](#)
- Grimm, P. W., Grossman, M. R., & Cormack, G. V. (2021). Artificial intelligence as evidence. *Northwestern Journal of Technology and Intellectual Property*, 19, 9. [↵](#)

- Hoijsink, M., & Planqué-van Hardeveld, A. (2022). Machine learning and the platformization of the military: A study of Google's machine learning platform TensorFlow. *International Political Sociology*, 16(2), olab036.
- Katyal, S. K. (2019). Private accountability in the age of artificial intelligence. *UCLA Law Review*, 66, 54.[↵](#)
- Kochhar, S. K., Bhatia, A., & Tomer, N. (2023). Using deep learning and big data analytics for managing cyber-attacks. In *New Approaches to Data Analytics and Internet of Things through Digital Twin* (pp. 146–178). IGI Global.[↵](#)
- Latif, S., Boulila, W., Koubaa, A., Zou, Z., & Ahmad, J. (2024). DTL-IDS: An optimized intrusion detection framework using deep transfer learning and genetic algorithm. *Journal of Network and Computer Applications*, 221, 103784.[↵](#)
- Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, 63(2), 157–170.[↵](#)
- Li, L., Zhao, T., Xie, Y., & Feng, Y. (2020). Interpretable machine learning based on integration of nlp and psychology in peer-to-peer lending risk evaluation. In *Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, Proceedings, Part II* 9 (pp. 429–441). Springer International Publishing.[↵](#)
- Liu, C., Chen, B., Shao, W., Zhang, C., Wong, K., & Zhang, Y. (2024). Unraveling attacks in machine learning-based IoT ecosystems: A survey and the open libraries behind them. *arXiv preprint arXiv:2401.11723*.[↵](#)
- Lone, A. N., Mustajab, S., & Alam, M. (2023). A comprehensive study on cybersecurity challenges and opportunities in the IoT world. *Security and Privacy*, 6(6), e318.[↵](#)
- Lui, A., & Lamb, G. W. (2018). Artificial intelligence and augmented intelligence collaboration: Regaining trust and confidence in the financial sector. *Information & Communications Technology Law*, 27(3), 267–283.[↵](#)

- Messai, M. L., & Seba, H. (2023, August). IoT network attack detection: Leveraging graph learning for enhanced security. In *Proceedings of the 18th International Conference on Availability, Reliability and Security* (pp. 1–7).[📄](#)
- Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International Journal of Financial Studies*, 9(3), 39.[📄](#)
- Mijwil, M., & Aljanabi, M. (2023). Towards artificial intelligence-based cybersecurity: The practices and ChatGPT generated ways to combat cybercrime. *Iraqi Journal for Computer Science and Mathematics*, 4(1), 65–70.[📄](#)
- Mijwil, M., Salem, I. E., & Ismaeel, M. M. (2023). The significance of machine learning and deep learning techniques in cybersecurity: A comprehensive review. *Iraqi Journal for Computer Science and Mathematics*, 4(1), 87–101.[📄](#)
- Mühlhoff, R. (2021). Predictive privacy: Towards an applied ethics of data analytics. *Ethics and Information Technology*, 23(4), 675–690.[📄](#)
- Narayana, Y. V., & Sreedevi, M. (2023, October). DDCATF: Deep Learning Approach for Detection of Cybercrime Activities Based on Temporal Features. In *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)* (pp. 462–469). IEEE.[📄](#)
- Nassar, A., & Kamal, M. (2021). Ethical dilemmas in AI-powered decision-making: A deep dive into big data-driven ethical considerations. *International Journal of Responsible Artificial Intelligence*, 11(8), 1–11.[📄](#)
- Odeh, A., & Abu Taleb, A. (2023). Ensemble-based deep learning models for enhancing IoT intrusion detection. *Applied Sciences*, 13(21), 11985.[📄](#)
- Ortiz-Ruiz, E., Bermejo, J. R., Sicilia, J. A., & Bermejo, J. (2024). Machine learning techniques for cyberattack prevention in IoT systems: A comparative perspective of cybersecurity and cyberdefense in Colombia. *Electronics*, 13(5), 824.[📄](#)
- Paddalwar, S., Ragha, L., & Mane, V. (2023). Cyber threat mitigation using machine learning, deep learning, artificial intelligence, and blockchain. In

Intelligent Approaches to Cyber Security (pp. 163–179). Chapman and Hall/CRC. [↵](#)

Pocher, N., Zichichi, M., Merizzi, F., Shafiq, M. Z., & Ferretti, S. (2023). Detecting anomalous cryptocurrency transactions: An AML/CFT application of machine learning-based forensics. *Electronic Markets*, 33(1), 37. [↵](#)

Rai, R., & Sahu, C. K. (2020). Driven by data or derived through physics? A review of hybrid physics guided machine learning techniques with cyber–physical system (CPS) focus. *IEEE Access*, 8, 71050–71073. [↵](#)

Rajkumar, S., Sheeba, S. L., Sivakami, R., Prabu, S., & Selvarani, A. (2023). An IoT-based deep learning approach for online fault detection against cyber-attacks. *SN Computer Science*, 4(4), 393. [↵](#)

Saltz, J., Skirpan, M., Fiesler, C., Gorelick, M., Yeh, T., Heckman, R., & Beard, N. (2019). Integrating ethics within machine learning courses. *ACM Transactions on Computing Education (TOCE)*, 19(4), 1–26. [↵](#)

Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., & Prabhakaran, V. (2021, March). Re-imagining algorithmic fairness in India and beyond. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 315–328). [↵](#)

Sharma, A., Mansotra, V., & Singh, K. (2023). Detection of Mirai Botnet attacks on IoT devices using deep learning. *Journal of Scientific Research and Technology*, 174–187. [↵](#)

Shukla, R. K., & Tiwari, A. K. (2024). Security analysis of the cyber crime. In *The Ethical Frontier of AI and Data Analysis* (pp. 257–271). IGI Global. [↵](#)

Singh, B. (2023a). Federated learning for envision future trajectory smart transport system for climate preservation and smart green planet: Insights into Global Governance and SDG-9 (Industry, Innovation and Infrastructure). *National Journal of Environmental Law*, 6(2), 6–17.

Singh, B. (2023b). Tele-health monitoring lensing deep neural learning structure: Ambient patient wellness via wearable devices for real-time

- alerts and interventions. *Indian Journal of Health and Medical Law*, 6(2), 12–16.
- Singh, B. (2023c). Blockchain technology in renovating healthcare: Legal and future perspectives. In *Revolutionizing Healthcare through Artificial Intelligence and Internet of Things Applications* (pp. 177–186). IGI Global.
- Singh, B. (2023d). Unleashing alternative dispute resolution (ADR) in resolving complex legal-technical issues arising in cyberspace lensing E-commerce and intellectual property: Proliferation of E-commerce digital economy. *Revista Brasileira de Alternative Dispute Resolution-Brazilian Journal of Alternative Dispute Resolution-RBADR*, 5(10), 81–105.
- Singh, N. J., Hoque, N., Singh, K. R., & Bhattacharyya, D. K. (2023). Botnet-based IoT network traffic analysis using deep learning. *Security and Privacy*, e355. [📄](#)
- Siwakoti, Y. R., Bhurtel, M., Rawat, D. B., Oest, A., & Johnson, R. C. (2023). Advances in IOT security: Vulnerabilities, enabled criminal services, attacks and countermeasures. *IEEE Internet of Things Journal*. [📄](#)
- Sowmya, B. J., Hanumantharaju, R., Kumar, D. P., & Srinivasa, K. G. (2023). Identification of authorship and prevention of fraudulent transactions/cybercrime using efficient high performance machine learning techniques. *International Journal of Business Intelligence and Data Mining*, 22(1–2), 144–169.
- Sudhakar, & Kumar, S. (2023, April). ABBDIoT: Anomaly-based botnet detection using machine learning model in the Internet of Things network. In *International Conference on IoT, Intelligent Computing and Security: Select Proceedings of IICS 2021* (pp. 235–245). Singapore: Springer Nature Singapore. [📄](#)
- Swiátkowska, J. (2020). Tackling cybercrime to unleash developing countries' digital potential. *Pathways for Prosperity Commission Background Paper Series*, 33, 2020–01. [📄](#)
- Thapaliya, S., & Sharma, P. K. (2023). Optimized deep neuro-fuzzy network for cyber forensic investigation in big data-based IoT infrastructures.

International Journal of Information Security and Privacy (IJISP), 17(1), 1–22. [↗](#)

- Tippins, N. T., Oswald, F. L., & McPhail, S. M. (2021). Scientific, legal, and ethical concerns about AI-based personnel selection tools: A call to action. *Personnel Assessment and Decisions*, 7(2), 1. [↗](#)
- Tyagi, A. K., Aswathy, S. U., & Abraham, A. (2020). Integrating blockchain technology and artificial intelligence: Synergies perspectives challenges and research directions. *Journal of Information Assurance and Security*, 15(5), 1554. [↗](#)
- Utami, S. H., Purnama, A. A., Hidayanto, A. N., Utami, S. H., Purnama, A. A., & Hidayanto, A. N. (2022). Fintech lending in Indonesia: A sentiment analysis, topic modelling, and social network analysis using Twitter data. *International Journal of Applied Engineering & Technology*, 4(1), 50–56.
- Varma, P., Nijjer, S., Sood, K., Grima, S., & Rupeika-Apoga, R. (2022). Thematic analysis of financial technology (Fintech) influence on the banking industry. *Risks*, 10(10), 186. [↗](#)
- Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022a). Interpretable machine learning models for crime prediction. *Computers, Environment and Urban Systems*, 94, 101789.
- Zhang, Y., Xiong, F., Xie, Y., Fan, X., & Gu, H. (2020). The impact of artificial intelligence and blockchain on the accounting profession. *IEEE Access*, 8, 110461–110477. [↗](#)
- Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022b). Explainable artificial intelligence applications in cyber security: State-of-the-art in research. *IEEE Access*, 10, 93104–93139.

14 AI in Cybersecurity

Safeguarding Cyberspace and Human Rights

Sakthivel M, Suresh Kumar K, Joshua Bapu J, Swathypriyadharsini P, Girimurugan R, and Raghavendran N

DOI: [10.1201/9781003668657-14](https://doi.org/10.1201/9781003668657-14)

14.1 INTRODUCTION

In recent times, it is difficult to find a business, organization, or community that does not use technological devices and the internet for its purposes. The use of artificial intelligence (AI) in cybersecurity offers new possibilities for proactive and adaptive defense strategies. AI techniques have shown significant growth and potential in recent years, with applications ranging from facial recognition to image analysis. Because the use of digital devices and apps on a daily basis by humans has become overwhelming [1]. For many of us, simply controlling how we use technology is an impossible task. While some people may not use the internet enough to keep up with the quick developments in technology, others may be obsessed with it and find it impossible to give it up. Some people may choose to spend hours on their computers or phones rather than socializing [2]. Because the internet is such a strong instrument, we cannot help but want to utilize it for

everything. It is evident that technology has improved our efficiency and productivity in a number of ways. But we also need to think about how it affects our communal and private lives, as well as our psychological and physical wellness [3]. The rise of Internet-connected devices has brought benefits, but cybercrimes have cast a shadow over our lives that cannot be ignored. In this linked and open society, there is always a risk of losing our private information, which has prompted numerous concerns. Our rapid expansion in connectivity has increased our efficiency while also increasing our susceptibility to the risks posed by cybercrime. We need to be aware of how we are connected to the internet [4]. The internet has made it possible for people to connect in previously unheard-of ways. Nevertheless, this has led to weaknesses that cybercriminals might take advantage of. Cybercriminals utilize online platforms to steal information, perpetrate deception, or inflict harm. The phrase “cybercrime” refers to a broad range of cyber-related offenses; essentially, it refers to illicit actions carried out using computers, including but not limited to computer hacking, password mining, Trojan dispersion, online harassment, and theft of personal information [5]. One of the most profitable crimes of the modern day is cyberattacks. Cybercriminals benefit from billions of dollars annually from stolen information, loss of information, and critical infrastructure attacks. Cyberattacks have changed significantly over time and will continue to do so, just like any other type of crime [6]. Cybersecurity has emerged as a crucial component for any infrastructure to avoid being breached by external attacks, including virus infections, malware threats, and attempts to hack into it. However, human error—rather than an outside threat—is the primary cause of the majority of protection lapses and criminal acts [7]. Although viruses and other malicious software have been a problem virtually since the early days of information technology, the significance of data security has only recently become known with the rise in use of the internet [8]. The rise in hacking activity in recent times has made more opportunities available to these hackers on the internet. Having said that, there are numerous risks associated with this exposure. Pirates have the ability to carry out illicit transactions, acquire data from our systems and

servers, and cause disruptions like website crashes [9]. This is a result of the massive worldwide online usage of over five thousand million people, or around 67% of the world's population, and the exponential rise in cyberattacks. Globally, cyberattacks are expected to cause losses of around eight trillion dollars by 2026 [10]. Data corruption or destruction, compromised data being erased or re-established, revenue theft, reduced efficiency, pillaging of proprietary information, stealing of personal and business information, and adverse publicity are just a few of the consequences associated with cyberattacks [11]. Building a more secure system from the ground up, stopping cyberattacks before they occur, and lessening their effects when they do occur are now multidisciplinary issues [12]. As cyberattacks are becoming a bigger issue, we must take precautions to keep ourselves safe. There are numerous methods to accomplish this, but the most crucial thing is to be conscious of the environment around us and the things we perform digitally with our own private information [13]. We may be able to avert the majority of the problems that lurk via the internet simply by being mindful of the risks associated with what we are doing. There are some conventional approaches to defend against cyberattacks, such as a firewall, an antivirus program, and security software for the web, to guarantee the device's integrity and reliability [14]. Using secure credentials and updating them frequently is another method. Other approaches include ensuring that the operating system has the most recent software updates as well as fixes installed [15]. To prevent malware attacks, one can also set up auto-responders and analyze traffic on the network for shortcomings. In addition, we ought to control our online social networking profiles and abstain from accessing unprotected Wi-Fi networks in unsecured locations. We may reduce our risk of becoming the victim of digital identity vandalism, online harassment, and several other hazards by limiting the amount of proprietary and private data that is shared online [16]. Beyond these traditional approaches, which are at minimum solutions, AI is becoming a more popular topic in cybersecurity. AI is now widely used in practically every scientific discipline, including commerce, security forces, defense, and

medical care. The application of AI in cybercrime is expanding so quickly that it is already one of the major global concerns [17]. AI is a powerful technology that is being utilized to tackle a wide range of criminal activities. Finding innovative uses for this technology will be essential for criminal justice agencies throughout the globe to keep up with the rapidly rising tide of cyberattacks. AI is being used in many different ways to combat crime. It is being utilized in the cybercrime space to help identify prospective risks, discover trends that may point to prior illicit conduct, and identify new types of unlawful activities already in place [18]. AI is being utilized, meanwhile, as a component of a larger study area on cybercrime and those who commit it. Cybercrime data is gathered, examined, and applied to create advanced virtual crime scenes that have the ability to foresee crimes before they occur. AI is useful for data extraction, discovering trends, and event anticipation [19]. Additionally, it can be used to recognize and avert cyberattacks. AI systems will eventually be able to analyze network traffic and determine whether different data packets are accessed in an odd way that is not immediately obvious to humans, such as a potential cyberattack. AI is a powerful tool that will only get bigger and more prevalent in our daily lives as it continues to advance.

14.1.1 MOTIVATIONS OF CURRENT RESEARCH

Recent years have seen study on the use of AI in cybersecurity applications, and numerous previous investigations have examined the body of work that has been accomplished in this area. However, there is also a growing trend in research to use AI to create more transparent and understandable services for industries like visual analysis and medical care. Nevertheless, an exhaustive work concentrating on the evaluation of AI-based solutions for a broad range of cybersecurity scenarios is lacking. This study also offers unique, in-depth quantitative recommendations at the end. These results point to a number of gaps that could be addressed by AI techniques and suggest the general path of future study in this area.

14.1.2 CONTRIBUTIONS OF THIS RESEARCH

- To examine various AI approaches and classifiers.
- To investigate the security requirements of systems that AI can solve.
- To ascertain datasets and architectures that are accessible for the AI-based cybersecurity system.
- To examine the most recent, effective AI-based cybersecurity systems and applications.
- To recognize obstacles and areas in need of investigation for AI applications in cybersecurity.

14.1.3 PAPER ORGANIZATION

The remainder of the paper is organized as follows. Section 14.2 describes the existing works in the literature that emphasizes the application of AI techniques to solve the challenges in cybersecurity. Section 14.3 gives a detailed view of the current strategies used in cybersecurity and the machine learning (ML) and deep learning (DL) methods that can be used for attack identification. Section 14.4 presents the results obtained on applying various AI techniques to two different datasets. Section 14.5 discusses the advantages, disadvantages, and future research prospects of the present research. Section 14.6 concludes the current work.

14.2 RELATED WORKS

The works in the literature related to the applications of AI for cybersecurity are discussed in this section. The research elaborated in ref. [\[20\]](#) addresses the prospects and difficulties of applying AI in cybersecurity and offers a thorough overview of this field's usage of AI. The use of AI for cybersecurity, including the investigation of malware, identification of anomalies, and recognition of intrusions, is covered by the investigators along with its potential to address cybersecurity concerns. After that, it goes over the difficulties and potential applications of AI in cybersecurity,

including the requirement for large datasets, the dangers of AI-powered incidents, and the necessity of manual supervision. The potential applications of AI in cybersecurity and how AI can be applied to improve cybersecurity are covered in the final section of the article. The investigators in ref. [\[21\]](#) conducted a thorough analysis of the application of AI in cybersecurity and highlighted the hurdles affecting digital healthcare applications and cybersecurity, such as the growth in quantities of data and intricacy, the scarcity of qualified workers, and the increasing cost of medical treatment. The experts look at how AI might be used for digital safety and intelligent medical care, as well as how it might be able to handle these issues with cybersecurity in the medical field. The authors address the benefits and difficulties of using AI to cybersecurity and healthcare domain, including the requirement for enormous amounts of information, potential hazards of AI being used by perpetrators, and the necessity for manual governance. The work by authors in ref. [\[22\]](#) addresses cybersecurity risks and possible remedies for intelligent energy systems applications. The examiner describes the security breaches that the electricity system confronts, including interruption of service, spyware, trojans, and internal assaults, and suggests ways to combat them. The researcher has addressed multiple prospective remedies, including training staff members about safety risks as well as remedies and putting safety precautions in place. The research in ref. [\[23\]](#) used DL techniques to identify potential security vulnerabilities in interconnected endpoints. The investigator addresses the topic of using DL to detect risks associated with cybersecurity and suggests combining DL techniques to identify corrupted and illicitly obtained files across Internet of Things (IoT) networks. This approach improves classification performance with a 95.67% accuracy rate. Researchers in ref. [\[24\]](#) demonstrate how AI can help with cybersecurity by avoiding, identifying, and minimizing hazards. The application of AI in identifying malicious software, identifying network breaches in infrastructure, examination of vulnerabilities, danger evaluation, and automation of security processes are among the areas in which the researcher recognizes the significance, current prospects, and opportunities of employing AI in

cyber threat minimization. To demonstrate how AI may be used in cybersecurity to accelerate processes, enhance decision-making processes, and discover vulnerabilities more successfully than with standard techniques, a thorough analysis is the basis for this investigation. Using DL techniques, authors in ref. [25] investigated spyware investigation in IoT surveillance. The investigator suggests using an advanced neural network to learn the patterns of both benign and dangerous services on IoT devices as they were not satisfied with existing spyware identification techniques, which have become ineffectual due to the fast rise of malicious software. The suggested strategy, according to the researcher, is a more accurate way to identify spyware in IoT equipment than conventional methods that utilize signatures. However, the consequences of resistance techniques—which make malicious software more difficult to detect by both the suggested deep models and standard signature-driven methods are overlooked by the author. Using unique measurements, the researchers in ref. [26] present an innovative approach for assessing the reliability and integrity of cyber-physical systems. The investigator suggests a method that uses individualized matrices and an ascending danger assessment tool to guarantee more thorough protection and integrity research. Through the combination of standard indicators, the examiner hopes to address the shortcomings of security processes. Hazardous management instances that could cause accidental damages are identified using the suggested approach. Using the personalized matrices, the potential for malicious individuals to exploit these situations is subsequently investigated. Authors in ref. [27] used AI to study cybersecurity solutions for the medical IoT devices using blockchain. The researcher talks about how blockchain technology and AI could enhance privacy and security when applied to healthcare IoT. The authors argue that blockchain can provide a secure and safe platform to store and share health-related information, while AI can be used to spot trends and identify and prevent intrusions. The research scientists in ref. [28] offer a collaborative deep learning framework that recognizes abnormalities in the industrial IoT data by applying the technique of learning historical data relationships. The researcher uses two

real-world datasets for evaluating the hypothesis. A gas pipeline was the primary focus of the first dataset, and water treatment was the objective of the second. With the former dataset and the latter dataset, the model's accuracy rates were 98.42% and 98.82%, respectively. Blockchain technology has been studied by researchers in ref. [29] as a potential security measure for internet-hosted medical databases. A publicly accessible database that is resistant to manipulation is the blockchain. This suggests that without the consent of everyone on the network, electronic health records (EHRs) stored on the blockchain cannot be altered or deleted. The blockchain and the EHR system are connected via a smart contract. EHR access is restricted, and only designated individuals are allowed to access EHRs according to the smart contract. To obtain a high degree of accuracy in the protection and privacy of EHR data in the cloud, the researcher assesses the suggested model using the real-time dataset. The study conducted by authors in ref. [30] aimed to examine the potential financial advantages that businesses could attain through the collection and analysis of information about cyberattacks provided by the American governance. The benefits of receiving this information are directly tied to the discrepancy between the risk level and the recipient organization's presumption regarding the risk level, as the researcher demonstrates using a conceptual framework.

14.3 BACKGROUND

14.3.1 CYBERSECURITY

The continuously evolving research environment is driving constant advances in cybersecurity. The worldwide community for cybersecurity is aware that no threat can be completely eradicated. Therefore, to lessen the negative effects of cyberattacks, studies and advances in technology are crucial. To avoid or mitigate security issues before they cause harm in the digital world, research has looked toward a more preventive approach. Because of the revolutionary principles of technological change found in

the Industry 4.0 environment, cybersecurity threats are becoming one of the biggest concerns confronted by the organizations. Cybersecurity employs a range of techniques, strategies, and tools to guarantee that systems are shielded against attacks and weaknesses. Cyberattacks seek to breach the safety, reliability, and accessibility of linked resources, applications, or infrastructure by gaining access to them. To raise the standards for cybersecurity, sophisticated defense strategies that can handle the wide range and rapidity of attacks must be created. Over time, cybersecurity has changed from being a technical field centered on computer safety to a worldwide concern. It is a subject that company executives are putting more and more emphasis on. An industrial setting with advanced manufacturing facilities, self-driving machines, cyber-physical systems, industrial IoT, remote computing, and big data requires aggressive control of AI-based security concerns. In this way, without requiring human input, AI has the ability to automatically produce significant cybersecurity insights. AI and ML have the potential to be revolutionary technologies for the exchange of data and protection of data in online environments.

14.3.2 ARTIFICIAL INTELLIGENCE

Automated processes and technological advancement across industries have increased as a result of recent advancements in AI technology. Even though AI technologies have many advantages, their use has prompted discussions about their potential negative applications as well. AI is a branch of computer science that creates systems, hypotheses, procedures, and strategies to mimic and extend cognitive abilities in automated machinery. Giving machines human intelligence is the aim of AI. One way to apply AI is through ML, which uses algorithms to examine and learn from data. AI may now be applied in many areas through a technology called DL, which is a component of ML. The foundation of AI is the idea that human intelligence can be precisely defined, allowing computers and software to emulate it. Cognitive reasoning, knowledge, scheduling, digitization, data mining, natural language processing, autonomous systems, emotional

intelligence, and cybersecurity are just a few of the disciplines encompassed by AI. Cybersecurity concerns and AI applications meet on an interdisciplinary level. Computer hacking in cyber–physical systems, on the other hand, is becoming more common as AI technologies become more sophisticated and pervasive and take advantage of the interface that connects physical and cyber aspects. There are several actors in the securitsetting, and attackers look for various kinds of pitfalls to exploit. The intricacy and competence of complicated recurrent risks, malevolent acts in cyberspace, and the commercialization of cybercrime are some examples of these attacks. To put protective measures into place, the cybersecurity community must comprehend how AI can be exploited for cyberattacks and pinpoint its vulnerabilities.

14.3.3 APPLICATION DOMAINS OF AI FOR CYBERSECURITY

The varied applications of AI for cybersecurity are discussed in this section, as shown in [Figure 14.1](#).

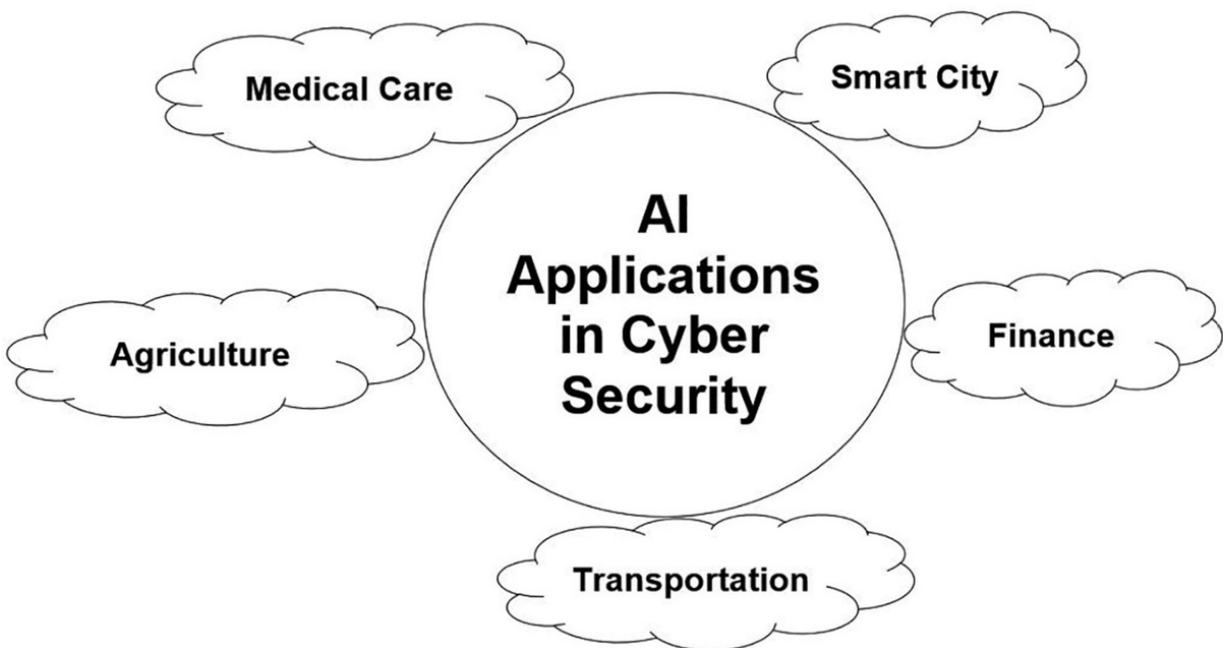


FIGURE 14.1 Applications of AI in cybersecurity. [↗](#)

14.3.3.1 Medical Care

Big data, virtualization, and the IoT combine to create a sophisticated, modern healthcare sector. Improved data connectivity is achieved through the use of the IoT, disruptive production technologies, firmware, hardware, automated machinery, detectors, and other complex information technologies. Innovations in computer technology are helping to improve the standard of healthcare by converting traditional medical organizations into innovative medical facilities. Concerns regarding the weaknesses of the smart healthcare system are emerging as AI plays an increasingly larger part in healthcare. Two key factors that make smart healthcare a good target for cybercrime are an abundance of extremely crucial data and inadequate safety precautions. Cybersecurity breaches include misappropriation of personal health data and infections with ransomware at medical centers, as well as potential threats to embedded surgical instruments. Breach can damage credibility among patients, endanger sophisticated systems of medicine, and jeopardize the lives of people.

14.3.3.2 Smart City

Modern smart cities are able to provide sophisticated amenities for power, transportation, medical services, and amusement to citizens as well as tourists by utilizing real-time ecological surveillance because of the expanding incorporation of data-based AI services like IoT, blockchain technology, and deep learning. But in addition to collecting a range of private-sensitive data from individuals and their social networks, smart city applications also manage government operations, impact the lives of individuals, and raise confidentiality, technological security, and safety concerns. To solve this problem, AI integration with IoT and AI-enabled smart city services can assist in resolving issues and provide elements of openness and comprehension that are necessary for smart city applications to make insightful decisions based on information.

14.3.3.3 Agriculture

Using cutting-edge technology in agriculture, such as the IoT, automated systems, aerial vehicles, cameras, and positioning structures, is known as smart agriculture. Smart farming also relies on AI, virtual and physical reality, virtualization, massive amounts of data, and various other technologies. However, the system is vulnerable to threats to cybersecurity and the infrastructure supporting smart farming due to the installation of several communication modules and AI models. Furthermore, countries whose economies mostly depend on agriculture may suffer from cyberattacks. However, people are unable to understand the relationships between features because most AI models are unknown systems. This is especially important if the system is intended to replicate actual farming activities that have societal consequences, such as vaporization. As a result, a large number of researchers are investigating the possible applications of AI in cybersecurity for smart farming.

14.3.3.4 Finance

AI models, which provide reduced expenses and increased productivity in areas like managing investments, investment recommendations, hazard estimation, financial assistance, and interaction with clients, have drastically changed the financial sector. While the simplicity of integrating AI into these smart financial systems makes them more efficient for all stakeholders, there is also an exponential increase in the risk of cyberattacks. As financial gain has historically been the primary motivation of attackers, smart financial systems are a prime target. The application of AI should be one of the top goals for the smart financial sector to combat money laundering crime that targets smart financial systems.

14.3.3.5 Transportation

While seamlessly integrating intelligent transportation systems, the rise of cutting-edge technologies like automobile networks, the industrial IoT, the distributed ledger, and AI has increased difficulty in managing operations. It may, however, encounter security issues that expose the transportation

networks to hacking. Furthermore, the AI model is impacted by security issues with transportation technologies. Crucial transportation infrastructures, like roadway signal computers, vehicle-to-everything, wireless sensors and networks, and the virtual machine system, have either already been hacked or could be hacked in the future. AI-enabled detection technologies have recently been created to guard against these cyberattacks and stop any cyberthreats on the modern transport system.

14.3.4 SECURITY REQUIREMENTS

14.3.4.1 Network Security

One of the fundamental components of the internet-based and interlinked society we live in today is the network. We can link our gadgets due to the widespread availability of international communications systems, which gives us unparalleled access to a huge range of contemporaneous fashion and worldwide applications and services. A move toward location autonomy and a large-scale outsourcing of data centers to cloud environments are contributing factors to the progressive disappearance of the idea of a controlled network boundary. However, networks, wired or wireless, are the main components that make our contemporary, interlinked society possible. Gateways as well as firewalls are examples of safety measures for network solutions. Using accessibility rules and restrictions that permit communication only with specific other networks or devices, firewalls restrict access to domains or equipment. The greatest network security solution available is the gateway, which creates a secured connection over the internet to thwart both proactive and inactive risks.

14.3.4.2 IoT Security

The ever-growing number of intelligent gadgets has begun to smoothly merge the digital and real worlds. The idea of smoothly connecting web-driven gadgets is captured by the broad concept known as the “Internet of Things,” or “IoT.” Digital twins and cyber–physical systems are all made

possible by this core idea. Security of networks and wireless connections, and mobile system safety are basically combined to form IoT security. When compared to more established devices like desktops and cell phones, IoT devices typically have inadequate security software, which gives hackers a fresh avenue for attack. As all IoT devices are already verified, cybercriminals who manage to take over a mobile phone with a Trojan horse that allows remote access might take control of any device linked to the network. Additionally, IoT devices can be used as a convenient entry point into the network to launch an indirect attack that infects desktops, cell phones, or other devices connected to the network, progressively gaining complete access to the data of the target, data, and behaviors.

14.3.5 CONVENTIONAL CYBERSECURITY STRATEGIES

A set of methods, guidelines, and practices designed to lessen the effects of hacking attempts is called a cybersecurity strategy. It consists of precautions against dangers, including malware, data leaks, and cyberattacks. A cybersecurity plan consists of various elements, including risk assessment, which ascertains the probability of an event occurring and its possible ramifications. Various dangers and weaknesses in an organization are frequently taken into account in risk assessments. For instance, an evaluation might look at if a company’s website is vulnerable to hackers or if its password security is insufficient. Prevention measures can be created to lessen the chance or consequences of cybersecurity incidents after risks have been evaluated. Various conventional strategies used to defend against cyber threats are tabulated in [Table 14.1](#).

TABLE 14.1

List of Conventional Strategies Used to Defend Cyber Threats

S.		
No.	Strategy	Description

S. No.	Strategy	Description
1	Firewall and antivirus	A firewall is a piece of software or physical hardware that keeps a system of computers isolated from the global web to guard against malware infections and virus infections. Organizations can better regulate traffic that comes in and leaves by using firewalls as a barrier between the intranet and the external environment. In a similar vein, antivirus software finds and stops any concerning risks when it is used. This procedure frequently entails finding and eliminating any malware that may be present by scanning the device and/or network.
2	Secure browser extensions	A protected browser plugin keeps you safe online by thwarting dangerous and fraudulent websites that attempt to obtain personal data. It offers defense against both commercial and malicious software. For example, the HTTPS Anywhere is a Firefox addon that compels your internet browser to utilize an SSL-secured rendition of an internet site when one is available to enhance your privacy on the internet.
3	Virtual private networks	The best defense against cyberattacks for data is a VPN. To prevent hackers from accessing data, it encrypts the information and routes it over a secured connection. When a VPN connection is lost, a few providers have a destroy option that blocks all internet activity. This will stop any information from flowing across collaborative links, which might be useful in some circumstances. The fact that the data is encrypted and safe when transferring between servers is the best feature of utilizing a VPN.

S. No.	Strategy	Description
4	Strong and unique passwords	A solid and distinctive password are crucial because they make it more difficult for cybercriminals to figure out what passwords are used on different internet pages and services. Creating a reliable password involves arranging capital and smaller letters, numbers, and characters in a distinctive yet simple-to-memorize fashion. Moreover, a password organizer has the ability to organize passwords, which helps us retain which kind of password is needed for particular internet pages or services.
5	Security paths and updates	Security updates keep computers up to date and it is vital. It is essential to keep software updated for the benefit of the equipment along with everyone on the network. Attackers will attempt to take advantage of users who do not utilize the program and the software itself whenever a security update is made available.

14.3.6 ML TECHNIQUES FOR CYBERSECURITY

The process of applying AI algorithms to analyze data, draw conclusions from the data, and decide on actual occurrences is known as machine learning, or ML. Researchers employ diverse methodologies to manage this copious quantity of information. ML uses many techniques to address data-related issues. The kind of algorithm used in the implementation will depend on the problem that needs to be solved while taking the factors in the training procedure into account. Within the realm of AI-driven cybersecurity investigation, ML is a pertinent topic in this era of digital transformation. It is significant that ML is used to offer strong defense against threats and adaptively reduce the negative effects of cyberattacks. There are three types of ML algorithms, which include reinforcement learning, unsupervised learning, and supervised learning. Applications for unsupervised as well as supervised learning are frequently employed in

cyber–physical incidents, virus identification, surveillance for intrusions, and information confidentiality assurance. A subfield of ML called reinforcement learning requires consecutive acts in an arbitrary manner, regardless of external expertise, bringing it closer to the way people learn. Different ML algorithms that can be applied to cybersecurity applications are presented in [Table 14.2](#).

TABLE 14.2
ML Techniques for Cybersecurity

S. No.	Technique	Description
1	Artificial neural networks (ANNs)	ANNs, which are made of networked neuronal cells that can perform specific mathematical operations on their inputs, are influenced by the structure of the human brain. The initial layer of the network’s neurons is activated by the supplied data, and the result serves as the signal for the neurons in the subsequent layer of the network. In a similar manner, the result from each layer is passed on to an adjacent layer, which then produces the final result.
2	Association rules	Finding previously unidentified pattern connections in the data is the aim of association rule mining. An association rule explains how many traits relate to one another. The relationship that C is present when A and B are present is the essence of this rule. Dimensions for association rules indicate the frequency with which a specific connection appears in the information being analyzed.

S.

No.	Technique	Description
3	Bayesian network	A probabilistic graphical model that depicts the parameters and their connections is called a Bayesian network. To create an acyclic graph with directed behavior, the network is built with nodes serving as the single or multiple independent variables along with directed edges serving as the links between them. The descendants of the nodes rely on their parents. The conditional likelihood form and the arbitrary variable's values are retained by each node. Efficient methods for inference or expert knowledge are used in the construction of Bayesian Networks.
4	Clustering	A collection of methods called clustering is used to identify similarities in complex data without labels. Using a correlation metric, the data is clustered together in this autonomous search for patterns method. Clustering for intrusion detection has the primary benefit of being able to learn from inspection information without having the system administrator to explicitly describe different attack classes.
5	Decision trees	A decision tree is a structure that resembles a tree with leaves that stand for categories and stems that represent the characteristics that combine to form those groups. Testing a data feature values against the nodes of the decision tree results in its labeling.

S. No.	Technique	Description
6	Ensemble learning	The algorithms used for supervised learning often search through the space of assumptions to find the most optimal hypothesis to predict a given situation accurately. Appropriate hypotheses might exist, but they might be difficult to determine. The goal of ensemble methods is to create a better assumption than the most effective one by combining several of them. To develop a strong learner, these methods frequently involve a number of weak learners.

14.3.7 DL TECHNIQUES FOR CYBERSECURITY

A potent ML technique called deep learning, or DL aims to create a network of artificial neurons that mimics the human brain to facilitate analytical learning through comprehension of information. An artificial neural network is a collection of techniques designed to replicate the functions of the human brain by identifying underlying links in a dataset. With the help of numerous layers, DL creates neural networks that can analyze complicated data at huge volumes and make intelligent judgments without the need for human interaction. A significant amount of data relating to cybersecurity that is publicly available online can be processed by DL techniques. Researchers identify harmful activity in information systems brought on by cyberattacks using ML as well as DL techniques. Applications of DL algorithms generate crucial information on the product development process and enable active surveillance in the industrialized setting. DL combined with reinforcement learning together shows exceptional efficacy as well as productivity for cybersecurity technologies that handle increasingly complicated and unpredictable intrusions. DL techniques that are suitable for cybersecurity scenarios are presented in [Table 14.3](#).

TABLE 14.3
DL Techniques for Cybersecurity

S.	No.	Technique	Description
	1	CNN	Neural networks of this kind are made to handle inputs that are kept as arrays of values. The CNN architecture is composed of three different types of layers: convolutional, aggregation, and categorization layers. There are several types of cybersecurity attack detection techniques driven by CNN such as Auditory Model, Single CNN, Layered CNN, and Restricted Load Sharing CNN.
	2	RNN	This is a ML model for mapping consecutive inputs to desired outcomes that was modified from neural networks. Military agencies can apply RNN for sentiment analysis in information dissemination analysis applications. The downside with this type of neural network is rectified by applying a bi-directional approach to RNN applications, which train the network with both current and historical input data.
	3	GANs	This category uses two neural networks with unsupervised ML. One network assumes the function of a generator, while the other one acts as a discriminator. After receiving the input data, the generator network generates the output information that closely resembles the original data. To determine whether the input data is real or fraudulent, the second network receives both genuine and bogus data.

S. No.	Technique	Description
4	Deep reinforcement learning (DRL)	This neural network is an amalgamation of reinforcement learning techniques including Q-learning, Deep Q-Networks, and Policy Gradients with deep neural network algorithms. When perception, reasoning, and behavior are all needed for an intricate process of choice-making, a blend of the two algorithms yields an advantageous response. Although DRL algorithms rely more on memorization for analyzing, they are built on pattern recurrence.
5	Autoencoders	These unsupervised neural networks demonstrate adaptability by learning to encode data in a compressed format. The two main parts of the autoencoder are the encoder and the decoder. Autoencoders are useful for applications involving sensory failure identification, detection of breaches, and connected device protection.
6	DBNs	This is a type of neural network that is multilayered and works on the principle of both supervised as well as semi-supervised learning algorithms. They can be categorized into different variants such as deep Boltzmann machine, restricted Boltzmann machine and a combined technique which is deep restricted Boltzmann machine.

14.4 RESULTS

14.4.1 DATASET DESCRIPTION

Among the 25,192 observations in the dataset selected for network intrusion, 11,743 have been classified as traffic from attackers, and 13,449

are considered routine traffic. Every observation has 41 elements that relate to the traffic input in a variety of groupings, one of which is a reaction column that indicates whether the finding is indicative of a malicious attempt or typical traffic activity. This dataset includes the diverse attacks in four classifications, such as Probe, U2R, DoS, and R2L. In its totality, the dataset consists of 125,973 observations. The truncated version of the dataset—a 30% variation on the original—was utilized for this investigation. There are 157,800 entries in the second dataset, of which 24,301 are classified to be harmless and 133,499 as risky. Every observation has 61 aspects, one of which is a reaction column that specifies if the observation is an assault or typical activity. This dataset is a collection of a large number of produced data of many attack types, including DoS assaults, collecting information, man-in-the-middle attacks, attacks involving injections, and infections with malware. This is an up-to-date, extensive dataset for industrial IoT security investigations. The data are produced by a variety of IoT equipment, including sensors with ultrasonic frequencies, ignition detectors, pH measurement detectors, water quality detectors, digital sensors for humidity and temperature, and water flow indicators.

14.4.2 EXPERIMENTAL EVALUATION

Experiments are conducted on the datasets described above using several machine learning and DL algorithms. ML algorithms such as decision trees, Bayesian networks, random forests, logistic regression, support vector machine, and XGBoost classifiers are used to identify the attacks in Dataset 1 and Dataset 2. The results of applying ML algorithms to Datasets 1 and 2 are presented in [Tables 14.4](#) and [14.5](#). The outcomes are compared with graphical representations in [Figures 14.2](#) and [14.3](#). It can be seen that the decision tree algorithm produced an accuracy of 89.6% with 88.2% precision, 88.5% recall, and 89.1% F1 score. Bayesian networks produced 90.8% accuracy and an F1 score of 90.3% with precision and recall values closer to each other at 89.5% and 89.8%, respectively. Random forests

exhibited an accuracy of 92.6% and an F1 score of 92.2%, whereas logistic regression produced an accuracy of 93.5% and an F1 score of 93.1%. It was observed that support vector machine produced 94.6% accuracy, 92.8% precision, 93.3% recall, and 94.3% F1 score. XGBoost produced the highest accuracy of 95.9% with precision, recall, and F1 score values as 94.2%, 94.6%, and 95.2%, respectively.

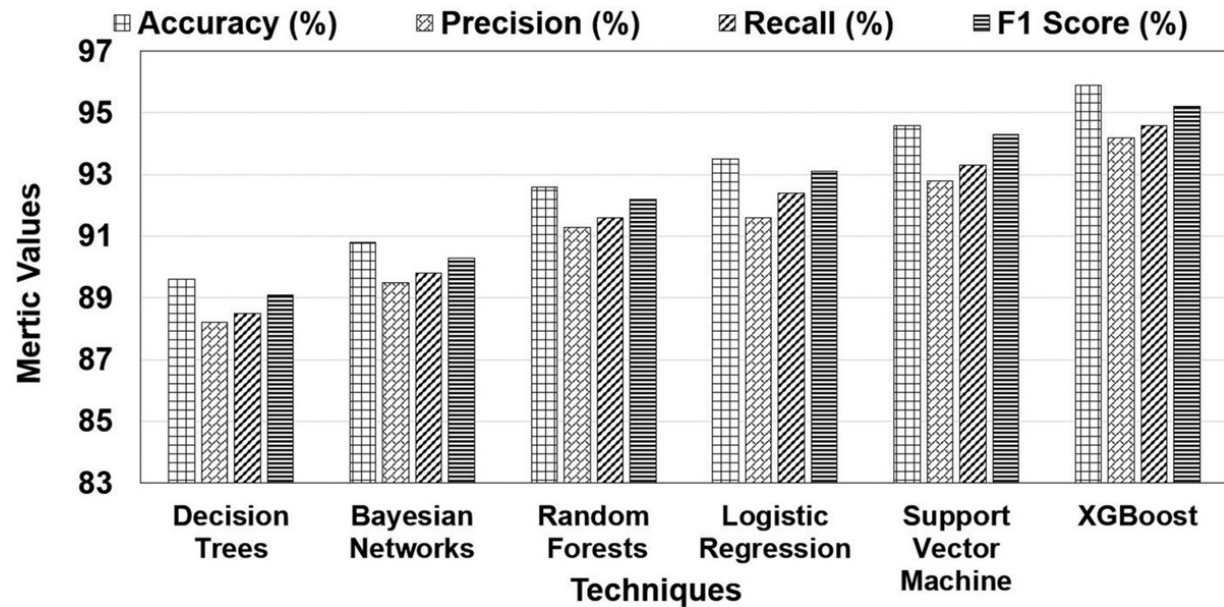


FIGURE 14.2 Performance comparison of ML algorithms on Dataset 1. [↗](#)

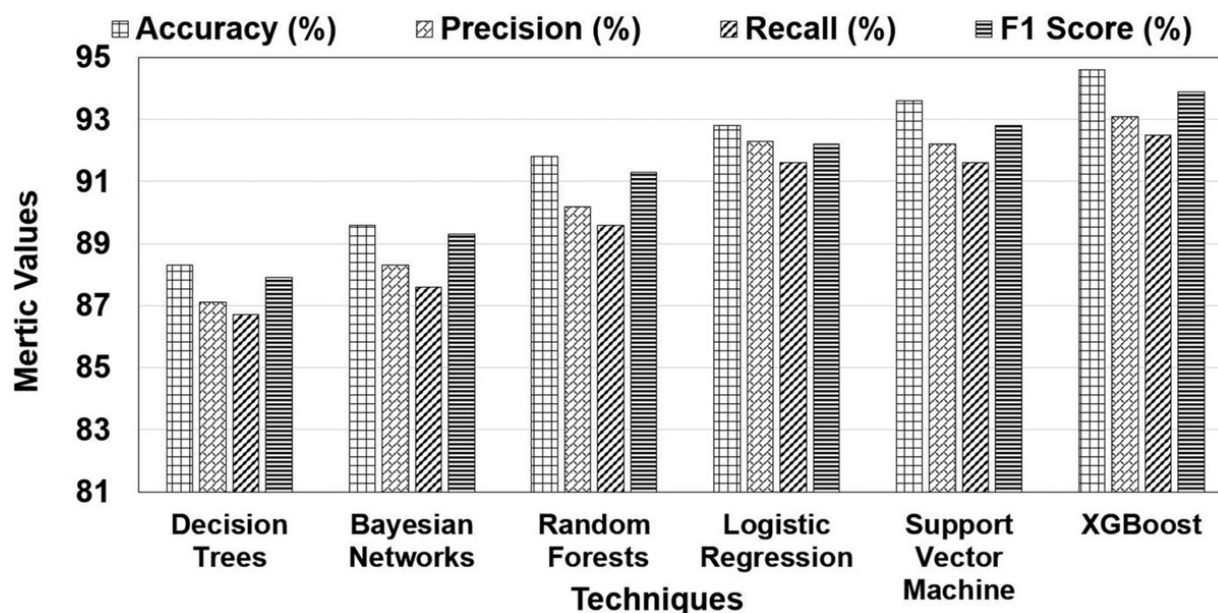


FIGURE 14.3 Performance comparison of ML algorithms on Dataset 2. [↗](#)

TABLE 14.4

Performance Comparison of ML Algorithms on Dataset 1 [↗](#)

Techniques	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Decision trees	89.6	88.2	88.5	89.1
Bayesian networks	90.8	89.5	89.8	90.3
Random forests	92.6	91.3	91.6	92.2
Logistic regression	93.5	91.6	92.4	93.1
Support vector machine	94.6	92.8	93.3	94.3
XGBoost	95.9	94.2	94.6	95.2

Decision tree algorithm produced an accuracy of 88.3%, precision of 87.1%, recall of 86.7%, and an F1 score of 87.9%. Bayesian networks classified the attack and normal scenarios with an accuracy of 89.6% and an F1 score of 89.3%. Random forests produced accuracy, precision, recall, and F1 score of 91.8%, 90.2%, 89.6%, and 91.3%, respectively. Logistic regression algorithm was capable enough to make predictions with 92.8% accuracy and 92.2% F1 score. Support vector machine exhibited 93.3%

accuracy, 92.2% precision, 91.6% recall, and 92.8% F1 score. XGBoost produced higher accuracy and F1 score values of 94.6% and 93.9% correspondingly in identifying the attacks effectively.

TABLE 14.5
Performance Comparison of ML Algorithms on Dataset 2

Techniques	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Decision trees	88.3	87.1	86.7	87.9
Bayesian networks	89.6	88.3	87.6	89.3
Random forests	91.8	90.2	89.6	91.3
Logistic regression	92.8	92.3	91.6	92.2
Support vector machine	93.6	92.2	91.6	92.8
XGBoost	94.6	93.1	92.5	93.9

DL algorithms such as convolutional neural network (CNN), recurrent neural network (RNN), generative adversarial networks (GANs), autoencoders (AEs), and deep belief networks (DBNs) were employed in Datasets 1 and 2 as well, and the performance of the models was analyzed. The obtained results for Datasets 1 and 2 are presented in [Tables 14.6](#) and [14.7](#), respectively, and depicted graphically in [Figures 14.4](#) and [14.5](#). The CNN algorithm was able to make predictions with an accuracy of 91.7%, a precision of 90.3%, a recall of 90.6%, and an F1 score of 91.4%. The RNN algorithm classified samples with 93.8% accuracy and 93.4% F1 score. GAN produced 95.6%, 94.1%, 94.5%, and 95.1% of accuracy, precision, recall, and F1 score, respectively. The AE algorithm exhibited 97.5% accuracy, 96.3% precision, 96.7% recall, and 97.3% F1 score on Dataset 1. DBN performed with the highest accuracy of 98.6%, precision of 97.4%, recall of 97.7%, and F1 score of 98.3% in making predictions.

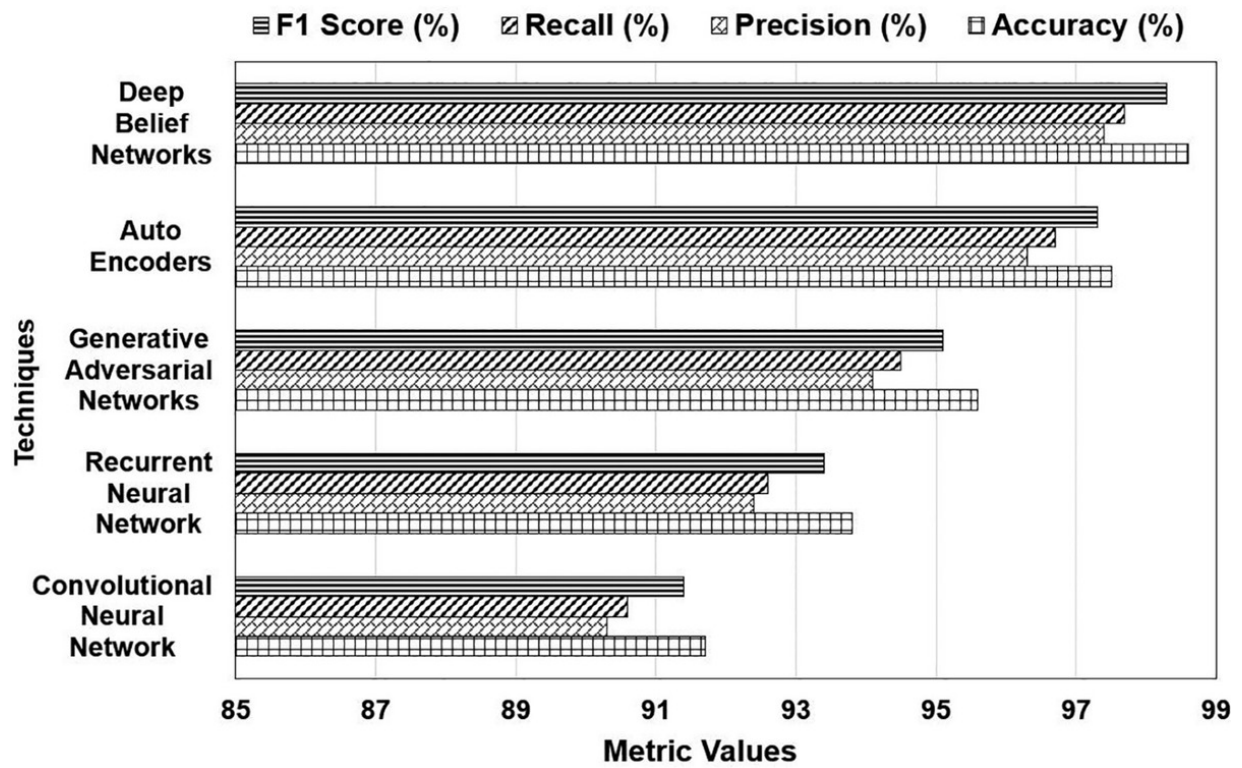


FIGURE 14.4 Performance comparison of DL algorithms on Dataset 1. [↗](#)

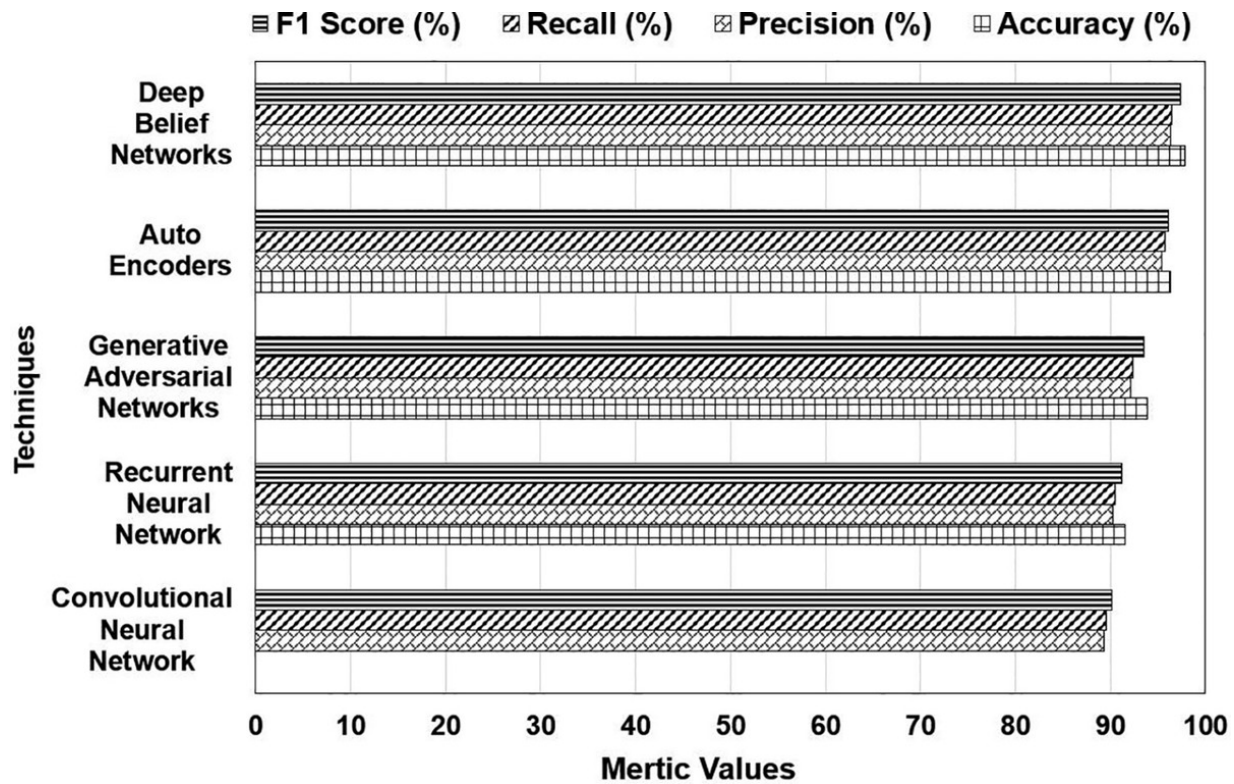


FIGURE 14.5 Performance comparison of DL algorithms on Dataset 2. [↗](#)

TABLE 14.6

Performance Comparison of DL Algorithms on Dataset 1 [↗](#)

Techniques	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
CNN	91.7	90.3	90.6	91.4
RNN	93.8	92.4	92.6	93.4
GANs	95.6	94.1	94.5	95.1
Autoencoders	97.5	96.3	96.7	97.3
DBNs	98.6	97.4	97.7	98.3

Furthermore, the CNN algorithm produced an accuracy of 90.8%, a precision of 89.3%, a recall of 89.6%, and an F1 score of 90.2% on Dataset 2. RNN produced 91.6% accuracy and 91.2% F1 score for classifying the attack as well as regular traffic. GAN produced an accuracy of 93.9% and an F2 score of 93.5% for identifying the attacks. The AE algorithm

exhibited 96.3%, 95.4%, 95.7%, and 96.1% of accuracy, precision, recall, and F1 score, respectively, on Dataset 2. The DBN algorithm produced highest values for accuracy, precision, recall, and F1 score with 97.9%, 96.3%, 96.5%, and 97.4%, respectively.

TABLE 14.7
Performance Comparison of DL Algorithms on Dataset 2 [↗](#)

Techniques	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
CNN	90.8	89.3	89.6	90.2
RNN	91.6	90.3	90.5	91.2
GANs	93.9	92.1	92.4	93.5
Autoencoders	96.3	95.4	95.7	96.1
DBNs	97.9	96.3	96.5	97.4

14.5 DISCUSSIONS

14.5.1 STRENGTHS AND WEAKNESSES OF AI FOR CYBERSECURITY

Researchers in cybersecurity create algorithms to examine and recognize contemporary and developing online risks. As AI systems advance, cyberspace is seeing the emergence of deceptive AI tactics. AI approaches will be used to safeguard the manufacturing setting so that systems can keep learning from attempted attacks. Predictive data analysis will therefore be useful for systems to handle the intricacy of intrusions. AI supports the procedure of surveillance by helping to spot trends of both normal and abnormal activity that may be malevolent. Attacks can be localized and prevented by surveillance. AI technology does not provide complete security for manufacturing environments. The application of these technologies raises a number of ethical questions as well, such as the absence of rules of conduct for automated machinery. The ineffectiveness

of AI to feel and make choices that take ethical concerns into account is a challenge when it comes to decision-making that may have social consequences. Communication between machines, security concerns, accuracy of data, and operative routines are a few of the obstacles that AI techniques in Industry 4.0 have highlighted as necessary to guarantee an edge over others. The rate at which attacks are expanding surpasses the capacity of cyber defenses to safeguard networked facilities. Consequently, the ability of security systems to identify assaults against tangible, portable devices is increased when developments in neural network-based ML and DL algorithms are used for cybersecurity scenarios. By taking into account the potential of robotics and data processing abilities with reliability, precision, and rapidity, AI is employed in the context of cybersecurity to improve defense operations. Promising research has been found using ML and DL algorithms to counter cyberattacks. But malicious individuals can also use AI technology to build new kinds of complex and advanced dangers. For instance, they can use an ML model to create malicious software, personalize scam emails, and launch attacks with greater vigor.

14.5.2 PROSPECTIVE RESEARCH DIRECTIONS

14.5.2.1 Improving Responsiveness, Adaptability, and Durability

To protect our contemporary cyberspace, further research is needed to improve the adaptability, responsiveness, and durability of AI-based cybersecurity systems. Due to cybercriminals' use of deception and camouflage techniques, one major problem with AI in cybersecurity is that there is no assurance the AI algorithm will identify the imminent danger. Furthermore, it could mistakenly label benign activity as malevolent, leading to needless costs. To boost AI's adaptability in cybersecurity, future research should concentrate on enhancing the preciseness of modern threat and recognition of anomalies. Given the substantial rise in intelligent gadgets and facilities, as well as the trend toward cities that are smart, digital replicas, including the metaverse, a particular focus on CPS and IoT devices is a desirable research subject in general. In the upcoming decades,

CPS security will become increasingly important for enhanced human activities.

14.5.2.2 Emphasize on End-Point Security

Future research should emphasize the improvement of the robustness of AI-based security systems by hardening them against malicious assaults. By including more datasets, the subject matter of the work can be expanded, strengthening its significance and improving the work's reliability and versatility. The study's scope might be broadened by selecting new datasets, as network privacy, mobile safety, and stability in a manufacturing connected device setting should be the main areas of focus. These days, with everything stored on the cloud, the old network security perimeter is essentially obsolete. Instead, endpoint security is the new standard for safeguarding our systems and businesses. Thus, it makes sense to concentrate on endpoint security.

14.5.2.3 Incorporating More Enabling Technologies

Adding further ML techniques in addition to existing methods is an extra choice. Additionally, by altering the mechanisms for activation or categories of networks, DL variants can be enhanced. Using various blending techniques is an additional approach to boost the models' forecasting capabilities and reliability. Lastly, for practical implementations in both the private and public sectors, the AI algorithms discussed in this study must be integrated into more enabling technologies with broader security systems. It would be beneficial to conduct more study on how it can be utilized to address the integration problems of AI-powered cyberattack techniques for identification and get over probable roadblocks in the areas of manipulation, information, implementation, and technology.

14.5.2.4 Personal Network Security

It is also feasible to shift your attention from businesses and governments to personal network security. The protection of major organizations and

government agencies receives the majority of financing and investigation, but the ordinary public, particularly personal networks, has received less attention up to this point. Attacks on people and private networks are more likely to occur as long as technical breakthroughs keep accelerating. Furthermore, as remote work becomes more common and the sharing economy grows, consumer security will become more crucial for individuals.

14.6 CONCLUSIONS

AI is a potent framework that adds simplicity and comprehensibility to DL and ML methods for solving complex problems. To protect systems against cybersecurity risks and evaluate security assessments, advanced techniques such as AI are necessary. As a result, this work provided an extensive overview of current research on AI in cybersecurity applications. This work primarily focuses on providing fundamental concepts and ontologies of cutting-edge AI models using crucial resources like a general structure and freely accessible datasets. In addition to this, the latest developments in AI-based cybersecurity systems from various application scenario perspectives and corresponding defensive strategies are also discussed. To follow up on these few difficulties, significant findings, and future possibilities for AI applications in cybersecurity research are analyzed effectively. It is anticipated that investigators, programmers, and security researchers interested in applying AI models to difficult problems in cybersecurity domains would find this work to be a useful resource.

REFERENCES

1. A. Mishra, A.V. Jha, B. Appasani, Emerging technologies and design aspects of next generation cyber physical system with a smart city application perspective, *Int. J. Syst. Assur. Eng. Manag.* 1 (2021), <https://doi.org/10.1007/s13198-021-01523-y>

2. I. Petri, Y. Rezgui, A. Ghoroghi, A. Alzahrani, Digital twins for performance management in the built environment, *J. Ind. Inf. Integr.* 33 (2023), 100445, <https://doi.org/10.1016/j.jii.2023.100445>
3. L. Li, B. Lei, C. Mao, Digital twin in smart manufacturing, *J. Ind. Inf. Integr.* 26 (2022), 32–33, <https://doi.org/10.1016/j.jii.2021.100289>
4. I.H. Sarker, M.H. Furhad, R. Nowrozy, AI-driven cybersecurity: an overview, security intelligence modeling and research directions, *SN Comput. Sci.* 2 (2021), 1–18, <https://doi.org/10.1007/s42979-021-00557-0>
5. Y. Otoum, D. Liu, A. Nayak, DL-IDS: a deep learning–based intrusion detection framework for securing IoT, *Trans. Emerg. Telecommun. Technol.* 33 (2022), 1–14, <https://doi.org/10.1002/ett.3803>
6. E. Tcydenova, T.W. Kim, C. Lee, J.H. Park, Detection of adversarial attacks in AI-based intrusion detection systems using explainable AI, *Hum. Centric Comput. Inf. Sci.* 11 (2021), 11–13, <https://doi.org/10.22967/HCIS.2021.11.035>
7. H. Jmila, M.I. Khedher, Adversarial machine learning for network intrusion detection: a comparative study, *Comput. Netw.* 214 (2022), 109073, <https://doi.org/10.1016/j.comnet.2022.109073>
8. M. Abdullahi, Y. Baashar, H. Alhussian, A. Alwadain, N. Aziz, L.F. Capretz, S.J. Abdulkadir, Detecting cybersecurity attacks in Internet of Things using artificial intelligence methods: a systematic literature review, *Electronics* 11(2022), 1–27, <https://doi.org/10.3390/electronics11020198>
9. C. Zhang, D. Jia, L. Wang, W. Wang, F. Liu, A. Yang, Comparative research on network intrusion detection methods based on machine learning, *Comput. Secur.* 121 (2022), 102861, <https://doi.org/10.1016/j.cose.2022.102861>
10. A. Fatani, M.A. Elaziz, A. Dahou, M.A.A. Al-Qaness, S. Lu, IoT intrusion detection system using deep learning and enhanced transient search optimization, *IEEE Access* 9 (2021), 123448–123464, <https://doi.org/10.1109/ACCESS.2021.3109081>

11. Z. Zhang, H. Ning, F. Shi, F. Farha, Y. Xu, J. Xu, F. Zhang, K.K.R. Choo, Artificial intelligence in cyber security: research advances, challenges, and opportunities, *Artif. Intell. Rev.* 55 (2022), 1029–1053. [↗](#)
12. H.I. Kure, S. Islam, M. Ghazanfar, A. Raza, M. Pasha, Asset criticality and risk prediction for an effective cybersecurity risk management of cyber–physical system, *Neural Comput. App.* 34 (1) (2022), 493–514. [↗](#)
13. S. Jeon, H.K. Kim, AutoVAS: an automated vulnerability analysis system with a deep learning approach, *Comput. Secur.* 106 (2021), 102308. [↗](#)
14. D. Iorga, D.G. Corlatescu, O. Grigorescu, C. Sandescu, M. Dascalu, R. Rughinis, Yggdrasil—early detection of cybernetic vulnerabilities from Twitter, in: *23rd International Conference on Control Systems and Computer Science (CSCS)*, 2021, pp. 463–468. IEEE. [↗](#)
15. T. Saha, N. Aaraj, N. Ajjarapu, N.K. Jha, SHARKS: Smart Hacking Approaches for Risk Scanning in Internet-of-Things and cyber–physical systems based on machine learning, *IEEE Trans. Emerg.* 10 (2) (2021), 870–885. [↗](#)
16. A. Shukla, S. Ahamad, G.N. Rao, A.J. Al-Asadi, A. Gupta, M. Kumbhkar, Artificial intelligence assisted iot data intrusion detection, in: *Proceedings of the 4th International Conference on Computing and Communications Technologies, ICCCT*, 2021, pp. 330–335. IEEE. <https://doi.org/10.1109/ICCCT53315.2021.9711795> [↗](#)
17. S. Zhou, J. Liu, D. Hou, X. Zhong, Y. Zhang, Autonomous penetration testing based on improved deep Q-network, *Appl. Sci.* 11 (2021), 8823. [↗](#)
18. H. Kavak, J.J. Padilla, D. Vernon-Bido, S.Y. Diallo, R. Gore, S. Shetty, Simulation for cybersecurity: state of the art and future directions, *J. Cybersecur.* 7 (1) (2021), 1–13, doi:[10.1093/cybsec/tyab005](https://doi.org/10.1093/cybsec/tyab005) [↗](#)
19. M. Alshehri, Blockchain-assisted cyber security in medical things using artificial intelligence, *Electron. Res. Arch.* 31 (2) (2023), 708–728, doi:[10.3934/era.2023035](https://doi.org/10.3934/era.2023035) [↗](#)
20. S. Tsimenidis, T. Lagkas, K. Rantos, *Deep Learning in IoT Intrusion Detection*, Springer US, 2022, <https://doi.org/10.1007/s10922-021-09621->

21. G. Srivastava, R.H. Jhaveri, S. Bhattacharya, S. Pandya, P.K.R. Rajeswari, Maddikunta, G. Yenduri, J.G. Hall, M. Alazab, T.R Gadekallu, XAI for cybersecurity: state of the art, challenges, open issues and future directions, *arXiv:2206.03585* 1 (2022), 1–33.
<http://arxiv.org/abs/2206.03585>
22. N.N. Abbas, T. Ahmed, S.H.U. Shah, M. Omar, H.W. Park, Investigating the applications of artificial intelligence in cyber security, *Scientometrics* 121 (2) (2019), 1189–1211, doi:[10.1007/s11192-019-03222-9](https://doi.org/10.1007/s11192-019-03222-9)
23. Y. Imrana, Y. Xiang, L. Ali, Z. Abdul-Rauf, A bidirectional LSTM deep learning approach for intrusion detection, *Expert Syst. Appl.* 185 (2021), 115524, <https://doi.org/10.1016/j.eswa.2021.115524>
24. H. HaddadPajouh, A. Dehghantanha, R. Khayami, K.K.R. Choo, A deep recurrent neural network based approach for Internet of Things malware threat hunting, *Futu.Gener. Comput. Syst.* 85 (2018) 88–96, doi:[10.1016/j.future.2018.03.007](https://doi.org/10.1016/j.future.2018.03.007)
25. S. Zhao, S. Li, L. Qi, L. da Xu, Computational intelligence enabled cybersecurity for the internet of things, *IEEE Trans. Emerg. Top. Comput. Intell.* 4 (2020) 666–674, <https://doi.org/10.1109/TETCI.2019.2941757>
26. A. Yazdinejad, M. Kazemi, R.M. Parizi, A. Dehghantanha, H. Karimipour, An ensemble deep learning model for cyber threat hunting in industrial internet of things, *Digit. Commun. Netw.* 9 (1) (2022), 101–110, doi:[10.1016/j.dcan.2022.09.008](https://doi.org/10.1016/j.dcan.2022.09.008)
27. F. Medjek, D. Tandjaoui, N. Djedjig, I. Romdhani, Fault-tolerant AI-driven intrusion detection system for the internet of things, *Int. J. Crit. Infrastruct. Prot.* 34 (2021), 100436, <https://doi.org/10.1016/j.ijcip.2021.100436>
28. B. Bera, A.K. Das, M.S. Obaidat, P. Vijayakumar, K.F. Hsiao, Y. Park, AI-enabled blockchain-based access control for malicious attacks detection and mitigation in IoE, *IEEE Consum. Electron. Maga.* 10 (5) (2021), 82–92, doi:[10.1109/MCE.2020.3040541](https://doi.org/10.1109/MCE.2020.3040541)

29. R. Maeda, M. Mimura, Automating post-exploitation with deep reinforcement learning, *Comput. Secur.* 100 (2021), 102108, doi:[10.1016/j.cose.2020.102108](https://doi.org/10.1016/j.cose.2020.102108)
30. H. Zhang, J.L. Li, X.M. Liu, C. Dong, Multi-dimensional feature fusion and stacking ensemble mechanism for network intrusion detection, *Future Gener. Comput. Syst.* 122 (2021), 130–143, <https://doi.org/10.1016/j.future.2021.03.024>

15 Ethics of AI-Driven Targeted Advertising

Safeguarding Consumer Rights

Gigimon V S and Darsana R S

DOI: [10.1201/9781003668657-15](https://doi.org/10.1201/9781003668657-15)

15.1 INTRODUCTION

While traditional consumer protection laws emphasize physical or financial harm, the digital realm presents new forms of risks that impact consumer autonomy and privacy. This chapter bridges existing literature on data privacy and artificial intelligence (AI) ethics, offering a novel perspective on the inadequacy of conventional laws in addressing the unique challenges of AI-enabled advertising. Drawing from European Union (EU) consumer laws, GDPR, and emerging ethical guidelines, the chapter proposes an integrative framework to regulate AI in advertising. The chapter employs a comparative legal analysis, examining international regulatory efforts, particularly within the EU and the US, with an aim to mitigate privacy risks and promote transparency. Case studies on real-time bidding and personalized ad delivery illustrate practical implications of these issues on consumer rights. An attempt by marketers to influence consumer choices is not a new phenomenon ([Paterson et al., 2021](#); Zard & Sears, 2022). In today's digital age, it has acquired a new form where advertisements are made available to consumers by relying on their information ([Paterson et](#)

[al., 2021](#)). Online advertising has enabled easy access to information and interaction between people. In the digital setup, advertisements can be targeted to people based on the data about individuals which are collected and analyzed using various tools ([Sartor et al., 2021](#)).

Targeted advertising focuses on providing advertisements (ads) to consumers by understanding their interests and needs. Earlier, businesses inferred purchasing attitudes through transactions with customers and then employed strategies to present personalized ads and improve customer relations. The advent of the internet brought new techniques of data collection, storage, transmission, and analysis of data which transformed the targeting landscape. Businesses are able to understand consumer personality and interests by analyzing various types of information which includes demographic information, frequently visited websites, browsing history, location information, health information, etc. In fact, it has enabled businesses to track and make inferences about consumers with whom they had no previous transactions ([Gao et al., 2023](#); [Tzoulia, 2021](#)). This process has been further augmented by employing AI.

AI signifies the ability of machines to collect information and use sophisticated algorithmic functions to learn from it. AI provides models of human intelligence enabling performance of tasks autonomously ([Argan et al., 2023](#); Chaitanya et al., 2023; [Thilagavathy & Kumar, 2021](#)). AI's role in the digital domain is indispensable and the advertising industry is no exception to this ([Gao et al., 2023](#); [Li, 2019](#); [Paterson et al., 2021](#)). Targeting is a crucial application of AI in advertising. It enables ads to effectively reach the right audience with appropriate information at the appropriate moment ([Gao et al., 2023](#); [Thilagavathy & Kumar, 2021](#)).

Consumer profiles become the basis for targeted advertising. Today's online landscape is characterized by search engines, social media platforms, apps, etc. that collect huge amounts of user data which is used to create digital profiles and make predictions about the customer's behavior. Algorithmic processes are employed for the creation of profiles and making predictions. Further, correlations between preferences may be made which helps in the prediction of future behavior. These profiles are used to group

people based on similar characteristics and identifiers (*Filter Bubbles and Targeted Advertising*, 2020; [Paterson et al., 2021](#)).

AI tools have made these processes faster and more efficient, thereby enhancing user experience and customer interactions. AI has revolutionized the marketing sector by delivering personalized advertisements and recommendations to consumers ([Beauvisage et al., 2023](#); Chaitanya et al., 2023). AI's role is not limited to bringing ads catering to the needs of people. It can predict user's needs and wants and suggest user-specific content ([Argan et al., 2023](#)). These tools help them identify the relevant audience and provide them with tailored ads increasing consumer relations and maximizing profit (Chaitanya et al., 2023).

The positive side of targeted advertising is that relevant material is filtered to assist consumer decision-making. It enables consumers to make better choices based on their preferences and interests. As the ads are tailored to consumer preferences, different consumers would be viewing different ads, offers, and even prices. It also saves time and costs as businesses are able to deliver ads to those people who are most likely to positively respond to their ads. Thus, they can save unnecessary expenditure ([Haleem et al., 2022](#); [Paterson et al., 2021](#)).

This, however, presents various risks to consumers ([Paterson et al., 2021](#)). The more data that is available, the better and accurate inferences can be made. Hence, businesses would be interested in any type of information that they may get about consumers. Such data collection, therefore, leads to unwarranted surveillance and privacy breach ([Sartor et al., 2021](#)). Targeted advertising also has the potential to affect consumer autonomy. The stress on personalization could lead to reduction of the scope of available choices being made available to consumers. It could also lead to manipulation of consumer behavior as well as discrimination between individuals and groups.

Collection, monitoring and processing of personal data primarily is a concern of data protection and privacy laws in many countries. They focus on the aspect of giving notice to consumers and on giving the option to consent or withdraw from such practices. However, such processes also

detrimentally affect consumer autonomy and welfare and hence require cognizance of consumer protection law as well ([Paterson et al., 2021](#)). This chapter analyzes the role of consumer protection law in safeguarding consumers from the evils of AI-enabled online targeted advertising. Ensuring fair and ethical practices in advertising is one of the functions of consumer protection law. It not only regulates the way businesses may collect, use, and disclose data for advertising, but also imposes obligations relating to informed consent and privacy.

Traditional consumer protection regimes focus on the physical or financial damage to consumers. However, the harms of targeted advertising are caused due to managing consumer attention which might not result in such damage. Therefore, there is a need for a change in outlook by the legal system which keeps note of the scope of new technologies and their effects on consumer welfare.

The chapter analyzes online targeted advertising, its components, and the various actors involved. It examines how AI is being used in the creation of personalized and targeted ads and the impact these practices have on consumer autonomy and privacy. It analyzes the need for consumer law intervention in addition to data protection law for comprehensive regulatory oversight over such practices.

Considering the dangers raised by AI-driven targeted advertising, a legal and ethical framework must be developed that can adequately respond to technological advancements. This chapter takes note of positive measures adopted in various nations, in pursuit of bringing such a framework. In the absence of an international framework, concerns of targeted advertising that encompasses state boundaries cannot be comprehensively addressed. The study emphasizes on the need for a balance between technological innovation and legal safeguards to ensure a responsible and consumer-friendly landscape.

15.2 TARGETED ADVERTISEMENTS IN THE DIGITAL LANDSCAPE

Targeted advertising is a marketing practice that uses data about individuals to display advertisements that are most likely to appeal to their interests ([Antonio et al., 2022](#); [Paterson et al., 2021](#); [Sartor et al., 2021](#); Zard & Sears, 2022). Unlike traditional marketing techniques that apply uniformly to all viewers, targeted advertising tailors to the expectations of individual recipients or groups of recipients. In the digital domain, ads likely to influence individuals are displayed to them through analysis of information collected from their online behavior ([Galli, 2020](#); [Paterson et al., 2021](#); [Shin & Yu, 2021](#); [Smith, 2015](#); [Tzoulia, 2021](#)). This would include information provided by people for accessing online services as well as information collected from their transactions and online activities (e.g., ads might be shown based on a product that the user has previously searched for). Targeted Advertising therefore involves prior collection and processing of recipients' data ([Tzoulia, 2021](#)). Such data about individuals is used to create customer profiles which are then grouped with other profiles based on common characteristics. These are studied and analyzed to facilitate the matching of ads to users ([Sartor et al., 2021](#)). The data collected may also be sold to other businesses for further use in advertising. This section will delve into the realm of online targeted advertising and the potentials of AI in bringing an enriched customer experience.

There has been a rise in online platforms that encompass search engines, marketplaces, app stores, media streaming services, etc. Many companies run on huge profits with the growth of the internet and online services ([Smith, 2015](#); Zard & Sears, 2022). What makes these services different is that are able to access them for free. Platforms generate revenue by charging those who wish to sell, promote or advertise their product, service or content through them. They therefore act as intermediaries that connect users and businesses. Targeted advertisements are used to increase the availability of ads to the interested users, thereby reducing costs and increasing efficiency of the marketing campaign (Zard & Sears, 2022).

Initially ads were targeted based on the searches of users as well as webpages visited by the users. However, it was later realized that more information could be collected which went beyond the information

expressly given by the user. This would include IP addresses, location information, financial details, patterns in internet usage, health information, etc. ([Beauvisage et al., 2023](#); [Boerman et al., 2017](#); [Paterson et al., 2021](#); [Sartor et al., 2021](#)). These are also analyzed in building user profiles. Various technological tools such as cookies have enabled collection of massive amounts of personal data which the user may not be aware of is being processed. Hence, user activity and preferences can be monitored in order to provide relevant ads that appeal to the customer. This has also led to online advertising flourishing faster than traditional forms of advertising ([Choi & Lim, 2020](#); [Sartor et al., 2021](#)).

Online targeted advertising can be categorized into three types based on how data is collected. It could be based on searches or web pages. It may be done based on characteristics provided by the user while registering a website. Another form is behavioral advertising which is based on observing the behavior of individuals over time, creation of profiles, and then providing ads tailored to user interests. Behavioral targeting has gained prominence with the rise in online platforms. It uses huge amount of personal data of consumers ([Antonio et al., 2022](#); [Boerman et al., 2017](#); [Choi & Lim, 2020](#); [Sartor et al., 2021](#)).

Targeted advertising takes various forms based on how they are delivered to users. This is depicted in [Figure 15.1](#). Display ads are very common while visiting any website. They capture the user's attention in multiple ways. Keyword advertising is based on the search of the user. Users will be displayed ads related to their search and if the user clicks through it, then the advertiser will be charged. Social media platforms provide the best environment for the collection and processing of data ([Gouda et al., 2020](#); [Sartor et al., 2021](#)). This includes browsing history as well as behavioral data such as socio-demographic information. Mobile advertising may be through mobile websites or in-app advertising. Chatbot advertising enables human interactions through text-based dialogue systems (Chaitanya et al., 2023; [Samat et al., 2017](#); [Sartor et al., 2021](#)).

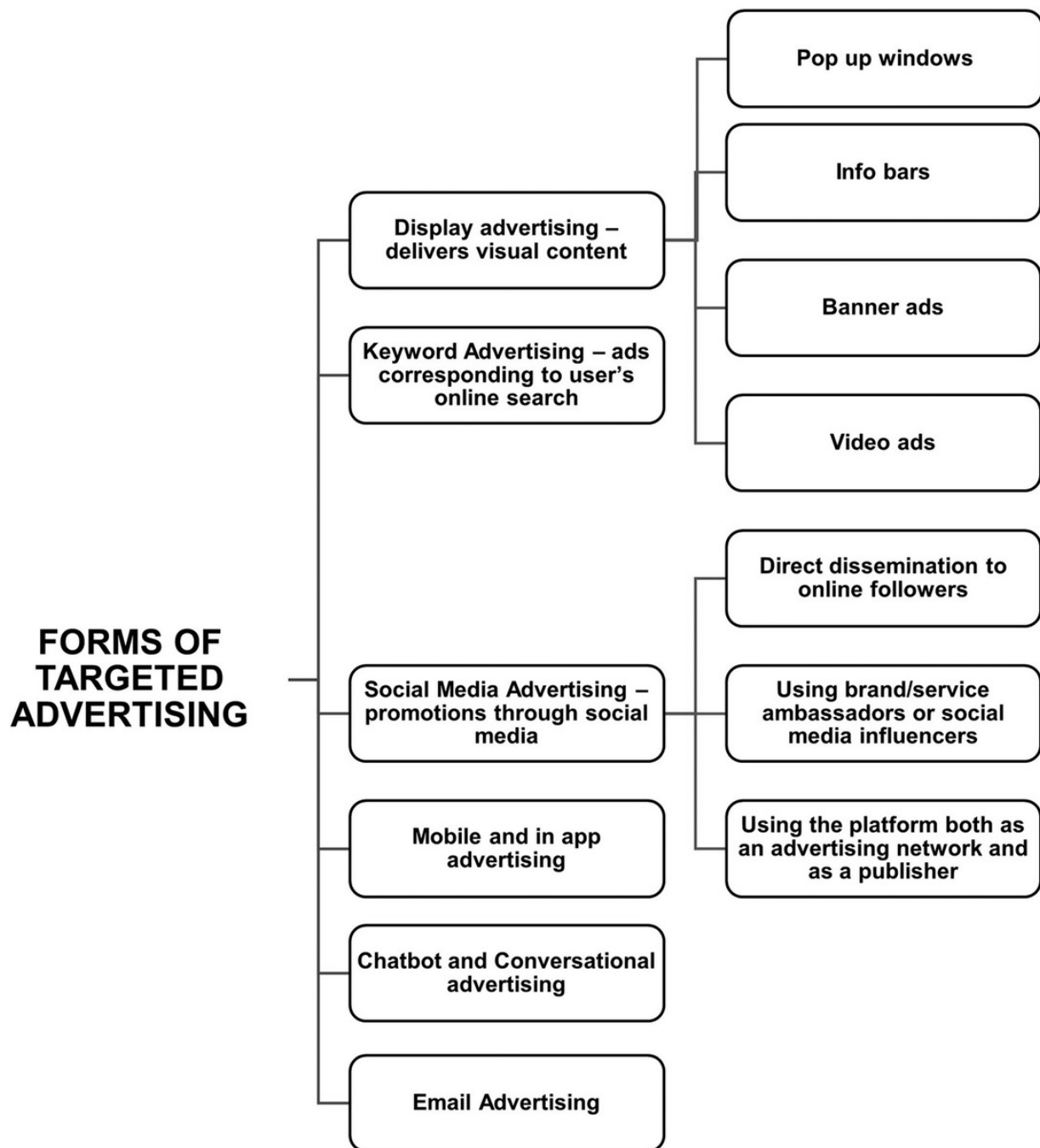


FIGURE 15.1 Forms of targeted advertising. [↗](#)

15.2.1 OVERVIEW OF THE TARGETED ADVERTISING ECOSYSTEM

15.2.1.1 Data

Data is the focal point of the targeted advertising realm. The whole system thrives on user information. By monitoring the online activities of users, data is collected, tracked, and analyzed to create profiles. This is done using powerful analytics and computational technologies to deliver advertisements on the publisher’s website ([Sartor et al., 2021](#)).

Many aspects come within the ambit of the term data such as demographic information, addresses, email, location, IP addresses, browser data, financial and transaction data, data from social media, facial images from CCTV recordings, data from sensor technologies, etc. Data may be classified in multiple ways. It could be volunteered, observed, derived, or inferred. It may also be categorized as first party or third party. It may also be common and sensitive personal data ([Paterson et al., 2021](#); Zard & Sears, 2022). [Table 15.1](#) illustrates the various categorizations of data.

TABLE 15.1
Classification of Personal Data

Classification	Types	Source	Examples
Origin			Data published on social media, data provided during creating online accounts or while registering in websites, credit card information while entering into online transactions
	Volunteered	Through direct actions of individuals	

Classification	Types	Source	Examples
	Observed	Collected by data collectors by tracking individual activity	Data from cookies, sensors, locational information, screen time, facial images
	Derived	Data determined through computations	Time spent on websites; classification based on common attributes
	Inferred	Data generated through the analytic process	Credit score, fraud score, probability of developing diseases
	Common	General data	Name, email address
	Sensitive	Data relating to sensitive personal information	Health, political views, sexual orientation
Special category	First party	Collected by businesses directly from customers/users	
		Acquired from a first party or from third parties through purchase, licensing, etc. or from publicly available information	
Ownership	Third party		

Data is collected using various mechanisms that track and monitor users' online behavior. In some cases, the user is aware of the data collection, whereas in others, this happens without the knowledge of the user ([Sartor et al., 2021](#)). Such data-tracking technologies include cookies, tracking walls, web beacons, device fingerprinting, and dark patterns ([Sartor et al., 2021](#); [Zard & Sears, 2022](#)).

The collected data is compiled, and user profiles are built. This would then be analyzed using various technologies such as big data analytics and cognitive computing. These techniques create groups of people based on similar criteria and new users would also be inserted into these groups or segments depending on similarities. These are then used to predict user behavior as well which assess the probability of how consumers would respond in benefit to the marketers ([Sartor et al., 2021](#)). Based on analytics, customers would be presented with advertisements depending on their group. Hence what is viewed will vary among customers. The data may be used by the business or may be sold or passed on to other businesses or third parties. Data management platforms and data brokers are involved in exchange of data as well as compiling data from multiple sources. The information collected may also be used for retargeting, that is, consumers would receive ads about those products that they have earlier shown interest ([Sartor et al., 2021](#)).

15.2.2 THE TARGETED ADVERTISING NETWORK

The targeted advertising landscape presents a complex structure with multiple actors involved at various levels, having different roles and purposes as indicated in [Figure 15.2](#). Marketers present their offers to consumers through ads with the aim of increasing their sales. Targeted advertisements ensure they are able to effectively communicate the ads to interested consumers and achieve better results for their campaign. This enables monitoring the campaign and also feedback analysis. Moreover, it attracts consumers and builds relationships. Publishers include any website, online platform, service providers, in-app developers, or even smart devices

that provide online content. They not only provide users with ads, but also collect data from users so that they may be utilized in advertising. The publishers generate revenue by delivering marketers' content on their platform and also through the sale or exchange of data collected from users ([Sartor et al., 2021](#); [Smith, 2015](#)).

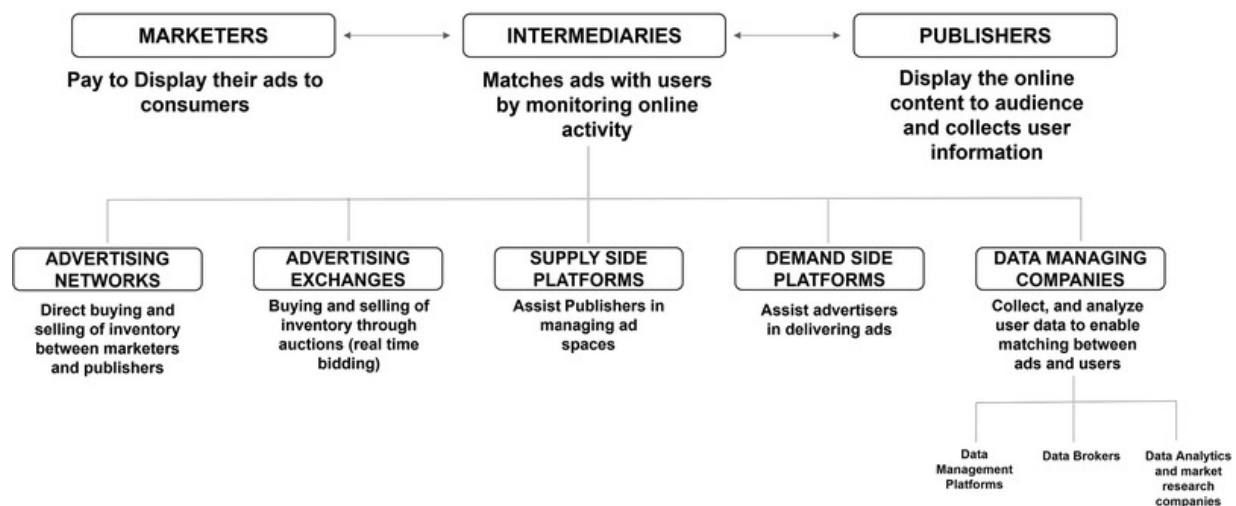


FIGURE 15.2 Targeted advertising system.

An intermediary would include various actors such as advertising networks, demand and supply side platforms, and advertising exchanges ([Zard & Sears, 2022](#)). Intermediaries facilitate the matching of demand and supply. They enable matching of advertising material accurately with user interests. They are involved in tracking and profiling of users based on information received through online activity. Intermediaries facilitate the interaction between marketers, publishers, and consumers ([Sartor et al., 2021](#); [Smith, 2015](#)). The data-managing platforms are involved in collection, compilation, and analysis of data to build user profiles ([Sartor et al., 2021](#)). Intermediaries facilitate trade between advertisers and publishers of the advertising space. Ultimately these ads reach the consumers or the users. Even consumers have a dual role; they are the source of data and also the ones targeted by the advertisements. Their information is processed by the different agencies in the advertising sector ([Sartor et al., 2021](#)).

Real-time bidding has transformed the interaction between advertisers, publishers as well as consumers. It refers to online instantaneous auctions for ad placements. Advertisers are alerted when a user enters a webpage address into their browser. By analyzing the online behavior and preferences of the users, an auction is held to present an ad to the user and the winner will deliver their ads. This happens within fraction of seconds and the ad would be displayed by the time the webpage is loaded on the screen. Everything happens in the backend without the user's knowledge. Ad exchanges and ad networks facilitate real-time bidding ([Mandal et al., 2017](#); [Sartor et al., 2021](#); [Smith, 2015](#)). Advertisers buy the right to present their ads to the optimal audience, selected based on the data available from their online behavior. They are able to choose who they bid for and how much to bid; thereby making it a cost-efficient strategy which allows precise targeting ([Smith, 2015](#); [Tang & Yu, 2022](#)).

The rise of technology enhanced the ability of companies to process large amount of data, thereby enabling nuanced targeting practices. Data is used to infer the interests, lifestyle, values, and preferences of consumers ([Davola, 2021](#); [Zard & Sears, 2022](#)). These data are processed by various agents using various technologies. From the inferences so gained, ads are delivered to users. Moreover, tools are used to understand customer reaction to the ads to understand whether they are interested in engaging with a business. AI tools have boosted the capabilities of businesses to analyze consumer behavior ([Gouda et al., 2020](#)), though it impacts consumer behavior and presents various risks as well.

15.3 AI-ENABLED TARGETED ADVERTISING

AI signifies the development of skills in machines to exhibit intelligence similar to human intelligence to perform tasks (Chaitanya et al., 2023; [Koolen, 2023](#); [Theodoridis & Gkikas, 2019](#)). AI is capable of manifesting cognitive and problem-solving abilities similar to humans ([Haleem et al., 2022](#); [Tzoulia, 2021](#)). In digital marketing, AI plays a role in customer profiling, segmentation, targeting, and prediction ([Beauvisage et al., 2023](#);

[Chaitanya et al., 2023](#)). Data is fed to AI tools to analyze consumer behavior and to predict the likelihood of a positive response from customers ([Campbell et al., 2022](#); [Gouda et al., 2020](#); [Haleem et al., 2022](#)). Through available data, users are scored and ads are matched based on the assessed scores. Use of analytic tools helps platforms monitor online behavior and enables advertisers to tailor their campaigns, generate high sales, and maximize their profits ([Samat et al., 2017](#)). Consumer segmentation allows the display of suitable advertisements to relevant audiences, thereby reducing advertising expenses ([Davola, 2021](#); [Zard & Sears, 2022](#)). AI tools regulate audience scope as well as the content delivered ([Gao et al., 2023](#)). AI tools deepen the pool of information that businesses may use to influence consumer decisions. Statistical models are employed to correlate ads to consumers. Through trial and error, and by applying feedback loops, the algorithm may determine the interests and preferences of the user ([Galli, 2020](#)).

AI leverages data to decide what to display and to whom ([Boyko & Kholodetska, 2022](#); [Chaitanya et al., 2023](#); [Galli, 2020](#); [Haleem et al., 2022](#)). Machine learning techniques enable efficient and fast data analysis. It automatically learns from data by employing algorithms and statistical models ([Chaitanya et al., 2023](#)). AI therefore can automatically process data, identify behavioral patterns, and develop consumer profiles without any kind of human monitoring. Hence, the processing of data, identification of patterns, and making inferences (especially about the future behavior of customers) can be done without explicit programming ([Chaitanya et al., 2023](#); [Galli, 2020](#); [Thilagavathy & Kumar, 2021](#)). They also deliver the targeted content to consumers ([Tzoulia, 2021](#)). By using AI tools, better insights into consumer behavior and preferences can be gained. AI tools help businesses bloom by enhancing the efficiency of the marketing campaigns ([Chaitanya et al., 2023](#); [Haleem et al., 2022](#); [Papić et al., 2023](#); [Theodoridis & Gkikas, 2019](#)). It also enhances real-time bidding systems ([Tang & Yu, 2022](#)).

As the consumer's digital footprint evolves, customer profiles also get constantly updated. With AI, profiles can be made quicker and the wants

and needs of people can be easily understood in real time thereby increasing quality and quantity of marketers' sales ([Argan et al., 2023](#); [Galli, 2020](#); [Gao et al., 2023](#); [Haleem et al., 2022](#)). Algorithms recognize trends and patterns in user behavior through consumer activities and interests. AI may also be used to understand emotions; what consumers are feeling/saying based on available data ([Argan et al., 2023](#); [Gao et al., 2023](#)). The shopping environment will be shaped according to the predicted psychological traits and emotional states ([Galli, 2020](#)). In the actual advertising stage, customers would view information relating to warranties. Incentives, etc., to assess the value of the brand and understand what they want to pay. Post-purchase, feedback analysis helps understand the likelihood of them engaging with the brand again ([Argan et al., 2023](#); [Boyko & Kholodetska, 2022](#)). User interests could be determined based on how they respond to various types of content.

Therefore, AI helps optimize advertising strategies and enhance customer experience ([Argan et al., 2023](#); [Chaitanya et al., 2023](#); De Marcellis-Warin et al., 2022; *Filter Bubbles and Targeted Advertising*, 2020). Through personalization, marketers can achieve better results and high customer satisfaction as they can deliver relevant and engaging content ([Chaitanya et al., 2023](#)). Personalization also increases the engagement of consumers with businesses ([Galli, 2020](#); [Theodoridis & Gkikas, 2019](#)). AI can make accurate predictions about consumer expectations and desires and apply behavior theories to improve the consumer experience ([Chaitanya et al., 2023](#); [Galli, 2020](#)). AI-based advertising relying on psychological and impulsive factors have huge impact on reactions of users ([Argan et al., 2023](#)). As AI helps companies gain a deeper understanding of customer feedback, it helps them address customer concerns easily and efficiently.

The ability of AI-enabled technology to understand patterns and make predictions enables businesses to model consumer behavior to achieve a desired result ([Galli, 2020](#)). The prospects of AI would only keep on evolving, making them more accessible and user-friendly ([Chaitanya et al., 2023](#); [Haleem et al., 2022](#)).

15.3.1 ISSUES TO CONSUMER WELFARE

AI has revolutionized targeted advertising. AI's potential has enabled targeting in a narrowed, precise, and accurate manner ([Argan et al., 2023](#); De Marcellis-Warin et al., 2022; [Theodoridis & Gkikas, 2019](#)). Big data sets are used to automatically track behavior and make predictions of consumer responses. Similar ads would be shown to people with similar responses. Moreover, AI tools identify what to present and when to present it to the consumers (De Marcellis-Warin et al., 2022). The products and offers viewed by one customer could be different from their peers. AI can predict the reactions of customers. Merchants aim to trigger such reactions by providing them with appropriate ads ([Chaitanya et al., 2023](#); [Sartor et al., 2021](#)).

This, however, raises legal questions about consumer rights and autonomy. The growth of digital technology has led to the exploitation of individual biases or those of micro-segments rather than those existing in general among the population ([Davola, 2021](#); [Davola et al., 2023](#); De Marcellis-Warin et al., 2022). As AI enables more precise predictions of consumer behavior, the risks of manipulation are also high. AI can shape consumer preferences and navigate them in specific patterns (De Marcellis-Warin et al., 2022). Not only prices, but also manipulations tend to be personalized (De Marcellis-Warin et al., 2022). The consumer may not be aware of such manipulation and this may also impose unfair contractual obligations ([Chaitanya et al., 2023](#); [Galli, 2020](#)). There is therefore disregard for basic principles of consumer welfare ([Galli, 2020](#)).

More reliance on personalization has created a situation wherein there is a massive collection of personal data and surveillance ([Boerman et al., 2017](#); [Galli, 2020](#); [Hunt, 2018](#); [Sartor et al., 2021](#)). Any element which helps establish correlation; something as small as a click or a message is tracked and processed to extract personality traits. Portable digital devices facilitate the collection of personal and sensitive data. Such data collection influences the content viewed by consumers. This affects consumers' decision-making to the extent that it impacts their economic and political

attitudes. Tracking and AI-based profiling also helps manage user engagement with platforms, thereby leading to the grouping of individuals or the creation of “filter bubbles” (*Filter Bubbles and Targeted Advertising*, 2020; [Sartor et al., 2021](#)). The nature and volume of data collection also raise privacy concerns. Profiling and segmentation are done by combining attributes from multiple sources. Data identifiers and attributes are sold or transferred to third parties. This raises risks of unauthorized data use and also data leakage ([Mandal et al., 2017](#)).

Predictions require large volumes of data. Insufficient data leads to wrong/less accurate predictions. Acquiring sufficient information may be challenging for small/less popular businesses. Depending on the data input, there can be biased recommendations. Further, the manner in which decisions are arrived at is also unknown or could be protected by IPRs ([Chaitanya et al., 2023](#); [Gouda et al., 2020](#); [Tzoulia, 2021](#)).

The nature and scope of targeting and the extent of data collected would be unknown to the consumers. Consumers are not aware or do not understand when, for what reason, and how their data is processed or analyzed by the automated system ([Boerman et al., 2017](#); [Paterson et al., 2021](#); [Samat et al., 2017](#); [Tzoulia, 2021](#)). This opacity takes advantage of their technical illiteracy ([Galli, 2020](#)). Evolutionary machine learning improves the algorithmic performance in a manner that even the creator would not be able to understand. Therefore, the reasoning of the algorithmic process or the manner of its working is out of the hands of the customer. This lack of transparency also hinders the consumer in proving the adversities of targeted advertising ([Tzoulia, 2021](#)).

Consumers are put in a different reality of market conditions as what they receive is filtered based on their profile. Their scope of choice is therefore significantly reduced ([Davola, 2021](#); [Paterson et al., 2021](#); [Tzoulia, 2021](#)). They are basically kept away from full information and the whole range of options, which is the foundation of consumer autonomy ([Paterson et al., 2021](#)). Through targeted recommendations, AI-enabled tools tend to limit consumer choice based on the segment to which consumer is put into by the algorithm. Therefore, consumer choices may be

restricted based on arbitrary criteria such as age, gender, previous purchases, payment capacity, etc. (De Marcellis-Warin et al., 2022; [Hunt, 2018](#)). They are therefore constrained to their own “eco chambers” or filter bubbles” ([Tzoulia, 2021](#)).

AI-enabled targeting advertising techniques also have the potential to influence and manipulate the decision-making capacity of consumers ([Campbell et al., 2022](#); [Davola, 2021](#); [Davola et al., 2023](#); [De Marcellis-Warin et al., 2022](#); [Galli, 2020](#); [Paterson et al., 2021](#); [Sartor et al., 2021](#)). Algorithms predict the behavior of persons and bring them ads which are likely to influence their transactional decisions. This may take advantage of their weaknesses or sensitivities; thus, manipulating them to make decisions which they would not have made otherwise ([Tzoulia, 2021](#)). They also significantly impact customer psychology. AI may profile consumers by exploiting or manipulating their psychological preferences and vulnerabilities ([Campbell et al., 2022](#); [De Marcellis-Warin et al., 2022](#); [Paterson et al., 2021](#); [Sartor et al., 2021](#)). Targeted advertising may also take unfair advantages of the sensitivities and biases of individuals to present them with products which are unlikely to promote their welfare. They therefore place an undue influence on consumer preferences. Predictions may not be beneficial to consumers and in fact might divert the attention of consumers to unwanted things ([Paterson et al., 2021](#)). Smart devices and mobile applications have become a common place for targeted ads ([Smith, 2015](#); [Tzoulia, 2021](#)). Time and space limitations pose problems for users in viewing and erasing unsolicited ads and also in reading and understanding them, thereby forcing them to make unwarranted transactional decisions ([Tzoulia, 2021](#)).

This also may lead to discrimination against individuals based on profiles ([Hunt, 2018](#); [Paterson et al., 2021](#)). Consumers, therefore in addition to being restricted in purchasing choices, may be subjected to unfavorable and differential pricing. Further, certain people or groups of people may be completely excluded from the market ([Paterson et al., 2021](#)). Automated processes segregate people based on identifiers. This leads to the exclusion of people on parameters such as age, gender, political views,

etc. Such discrimination could be due to errors in the algorithm, or due to bias of the creator due to the data fed not being representative of all. Therefore this affects the consumers' ability to access free market and competitive pieces ([Gao et al., 2023](#); [Gouda et al., 2020](#); [Tzoulia, 2021](#)).

Some groups might have not been part of the collected data sets. This data would not be representative of the whole population and therefore would result in marginalization of people. The outcome therefore will be suboptimal and not welfare-oriented especially to those who were not part of the data set ([Paterson et al., 2021](#)). For optimization of services and to provide users with relevant ads, tracking technologies and data brokers are employed. Popular platforms have more customers; the concentration of data gives them a competitive advantage over others ([Smith, 2015](#); [Tzoulia, 2021](#)). This again results in customers receiving partial or insufficient information.

AI also influences consumers' decision-making by giving them false beliefs and molding expectations favorable to them. They employ dark patterns to knowingly confuse users. Dark patterns manipulate users to take specific actions, thereby affecting the consumers' capacity to make free choices ([De Marcellis-Warin et al., 2022](#)).

Data protection and privacy laws form the primary weapons of law to safeguard consumers from the detrimental effects. However, the extent of the impact of AI enables online targeted advertising requires data protection to be supplemented by consumer protection laws as well.

15.4 THINKING BEYOND DATA PROTECTION: ROLE OF CONSUMER LAW

Targeted advertising is dependent on the processing of personal data of users. To regulate the unwarranted collection of data and their misuse, legal responses to targeted advertising, in many jurisdictions are found in their data protection and privacy legislation ([Davola, 2021](#); [Li, 2019](#); [Zard & Sears, 2022](#)). Free and informed consent becomes the central focus of these

laws ([Davola et al., 2023](#); [Tzoulia, 2021](#)). They envisage mandatory disclosure of how user data is processed.

The scope of data protection can be illustrated by the example of the General Data Protection Regulation (Regulation (EU) 679/2016) of the European Union). It regulates automated data processing and profiling. Accordingly, such activities should be lawful, fair, and transparent. Data processing may happen only when sanctioned by the subject or to comply with a legal or contractual obligation. Profiling without prior consent is prohibited. The consent should be free, specific, and informed. Consumers should have an option to refuse or to subsequently withdraw consent. The consumer should be provided with sufficient information for making informed decisions. Such information (details of intermediary proposal of data processing, duration of data storage, consumer's right to opt out) has to be presented in a concise, easily accessible and transparent manner by employing plain and simple language. Moreover, automated decision making should happen in an unbiased and non-discriminatory manner. Discrepancies in data input, wrong modelling, etc. should be prevented, avoided, and rectified ([Tzoulia, 2021](#)).

However, the detrimental effects of targeted advertising make it imperative that regulatory intervention should go beyond the data protection and privacy domain. Even in the presence of Data Protection Laws, consumers are still unaware of how data is drawn, monitored, and used. Factors such as information overload, inaccessibility, time limitations, and the need to access services hinder consumers in making informed decisions. Furthermore, privacy policies are largely oblique and complex, and consumers may not understand the intricacies ([De Marcellis-Warin et al., 2022](#); [Paterson et al., 2021](#)). Reliance on the transparency requirement is therefore not enough. Not only the ads, but the harm is also personalized in targeted advertising. Further, consumer segmentation, manipulation, price discrimination, etc., may be out of the scope of data protection. Hence, the need arises for law to take notice of the strategies that businesses employ to make consumers vulnerable ([Boerman et al., 2017](#); [Davola, 2021](#); [Galli, 2020](#); [Koolen, 2023](#)).

Due to their impact on consumer autonomy and privacy, harms of targeted advertising come within the purview of the consumer protection regime. As publishers monetize their services through ads whereby the consumer is receiving services for free, targeted advertising practices were earlier regarded to be out of the scope of consumer protection. However, they are covered by consumer protection rules firstly as they directly affect consumer welfare and self-determination. Moreover, they leverage consumers' attention and data in exchange for these services ([Zard & Sears, 2022](#)).

Consumer law deals with the relationship between businesses/traders and consumers. It aims to protect consumers from unfair practices and as targeted advertising involves the matching of ads to customers, it is well within its scope ([Galli, 2020](#); [Tzoulia, 2021](#)). Further, consumer manipulation is also a matter of concern of the consumer protection as well as competition regimes ([De Marcellis-Warin et al., 2022](#)). Consumer protection laws take into account the asymmetries of power and potentials of consumer harm and aim to protect consumer autonomy ([Koolen, 2023](#); [Zard & Sears, 2022](#)). Consumer protection can complement privacy laws to ensure maximum consumer welfare ([Paterson et al., 2021](#)).

Consumer protection rules encompass two aspects. Firstly, it aims at providing sufficient information to consumers that enables them to make informed decisions. They will be able to exercise their autonomy in a better manner as they understand the information and will act accordingly ([Davola et al., 2023](#); [Zard & Sears, 2022](#)). This goes beyond disclosure requirements. Hence, issues of transparency can be managed in this paradigm. The protection of unfair contract terms intends to address issues of information asymmetry and in laying down contractual terms which are clear and most favorable to the consumer. For example, the Consumer Rights Directive of the EU requires that consumers be informed about the features of the services, the details of the publishers, and the fact that their data is being tracked or monitored for personalization ([Zard & Sears, 2022](#)). Secondly, it aims to protect consumers from unfair practices. These aim to act in consumers' interests by protecting the consumers from misleading,

manipulative or unfair conduct. These are applicable in the targeted advertising landscape also ([Zard & Sears, 2022](#)).

Targeted advertising using AI technologies has been receiving much criticism as it exploits the vulnerabilities of consumers and may force them to make decisions that may undermine their interests. There is a call for a ban on such advertising practices altogether, or specific forms of targeted advertising due to the harm caused to consumers and how they take advantage of consumers' sensitivities and vulnerabilities. However, an outright ban might not be the ideal solution, considering the realm of technological advancement in which we are part. Measures toward increasing awareness and transparency such as providing necessary warnings about data processing, would keep consumers' attention to harmful targeted advertising practices ([Davola et al., 2023](#); [Paterson et al., 2021](#)).

Consumers should be informed about differential pricing (as under the EU Digital Services Act) and personalized ads. Moreover, consumers also should have an option to "opt-out." For example, under the California Consumer Privacy Act, 2018 websites are required to provide a "Do not Sell my personal information" link. These strategies would reduce transparency concerns. Although this may not be the only solution, it could enhance consumer awareness and rights. However, as in the case of data protection and privacy laws, these fall short of completely addressing all issues ([Paterson et al., 2021](#)).

AI-enabled targeted advertising impacts the decision-making capacity of consumers. Their working is in such a manner that at the choice-making stage itself, the whole picture is concealed from the consumer. Consumer protection regimes must acknowledge this effect and adapt to it. They should incorporate measures for the good governance of AI through the development of an ethical and legal framework. Self-regulation tools and external assessment tools, such as audit frameworks, are mechanisms to ensure the responsible use of AI. This also includes compliance with the rule of law in internal procedures or codes of AI. These may help in the

management of risks induced by automated systems ([De Marcellis-Warin et al., 2022](#); [Farseev, 2023](#)).

The harmful effects of targeted advertising may also be brought under the consumer protection regimes under the categories of misleading, unfair, or unconscionable conduct. These prohibitions ensure enterprises do not take advantage of the lack of bargaining power of consumers. Such an approach would also ensure that businesses do not undermine consumer welfare by exploiting the general consumer ignorance in the digital landscape. The ill effects of targeted advertising can be brought within the ambit of unfair trade practices and misleading conduct typically dealt with under the consumer protection regimes. However, proving this would be cumbersome due to the personalized and variable nature of targeted advertising ([Paterson et al., 2021](#)).

Regulatory scrutiny of privacy policies and notices is also essential. As consumers' role is limited in this respect, a robust regulatory mechanism can ensure greater transparency and monitor unauthorized data use. Strict statutory control over unfair contractual terms can curb unwarranted data tracking and privacy violations ([Paterson et al., 2021](#)). Therefore, a coordinated approach is necessary that involves regulators in the privacy, consumer protection, and even competition regimes. Further, collaboration is also essential between regulators of different jurisdictions ([Paterson et al., 2021](#)). In pursuance of developing an effective framework regulating AI-enabled online targeted advertisements, positive measures adopted in certain countries may be analyzed.

The European Union's consumer protection response to targeted advertising is spread across various laws—the Unfair Commercial Practices Directive, the Consumer Rights Directive, the Unfair Contract Terms Directive, the Digital Content Directive, and the Digital Services Act ([Zard & Sears, 2022](#)). The Unfair Commercial Practices Directive (UCPD) prohibits unfair practices against consumers. It regulates business-to-consumer relationships ([Galli, 2020](#); [Paterson et al., 2021](#); [Tzoulia, 2021](#); [Zard & Sears, 2022](#)). The UCPD has a wider scope covering all commercial practices within its ambit, including targeted advertising practices. It covers

both the information paradigm and the unfair conduct paradigm of consumer protection. A practice would be considered unfair under three circumstances: (1) if it falls within the blacklist of 35 practices explicitly prohibited; (2) if the practice is “misleading” or “aggressive”; (3) when such a practice is contrary to professional diligence ([Galli, 2020](#); [Tzoulia, 2021](#); [Zard & Sears, 2022](#)).

The law determines whether a practice is unfair by analyzing its likelihood of distorting the economic behavior of an average customer ([Galli, 2020](#); [Tzoulia, 2021](#); [Zard & Sears, 2022](#)). Unsolicited commercial communications that influence consumer decision-making come within the ambit of “aggressive” practices. Incomplete, ambiguous, or unclear presentation of information is within the ambit of “misleading” practices. Differential pricing and misrepresentation of market conditions are also unfair practices ([Galli, 2020](#); [Tzoulia, 2021](#)). A publisher’s failure to inform consumers about data processing would be considered a misleading omission. Similarly, the use of dark patterns would amount to ‘aggressive’ conduct.

The ‘omnibus directive’ (Directive 2019/2161) amended consumer protection laws to improve enforcement in the online sector. It mandates providing clear, specific, information to customers about personalisation ([Galli, 2020](#)). Consumer interests are also prohibited under the Digital Services Act which prohibits conduct that deceives, manipulates, or distorts the ability of consumers to make free and informed decisions. It takes note of the risks involved with personalization and places disclosure requirements not only about the publisher/advertiser but also about the parameters of targeting. It also requires that users be given the option to change those parameters. Therefore, users have control over or may refuse profiling. Moreover, the Act bans the segmentation of consumers based on sexual orientation, race, health conditions, etc. The Digital Markets Act prohibits personalization without the user’s consent. It also does not allow combining data from other sources without the user being presented with the choice ([Zard & Sears, 2022](#)).

The US Deceptive Experience to Online User Reduction Act (Detour) Act also provides a positive step in this respect. This is a Bill proposed in the United States Congress which has not yet culminated into legislation. It aims to prohibit deceptive and exploitative online practices that manipulate the user interface, impair autonomy, or segment consumers based on psychological traits without the consent of the users would be deceptive and would come within the ambit of Unfair Practices within which the Federal Trade Commission may impose action ([Galli, 2020](#)). It requires adequate disclosure in case of any experiment or research carried out on the online activity of users. It also envisages the establishment of Professional Standard Bodies to develop guidelines for achieving the objectives of the Act ([Galli, 2020](#)). Such a law covers most of the strategies of targeted advertising such as dark patterns, behavioral targeting, etc., in ensuring responsible and consumer-friendly behavior from businesses.

In terms of remedy, compensation is the remedy envisaged by both Data Protection and Consumer Protection laws envisage compensation as the remedy for such practices. However, with the automated and evolving data processing, placing liability on a natural person might not be possible, and this remains a challenge ([Tzoulia, 2021](#)).

15.5 CONCLUSION

Technological advancements have transformed how businesses reach and engage with consumers. AI is used in various ways to optimize operations and enhance consumer interactions. The use of AI has made a significant impact on online targeted advertising practices. It can clearly identify the potential optimal audience and predict their responses. However, such data processing elicits concerns of transparency, privacy, discrimination, and manipulation. Attributing accountability is also a problem due to challenges in proving unfair conduct by businesses. Further, with machine learning technologies, it would be difficult to determine on whom the liability lies.

A holistic and integrated approach is essential which combines data protection and privacy legislation as well as consumer protection. As

consumer choice is affected, the matter is of importance to the competition regime as well. Law can also make use of AI's potential in helping consumers. Consumers can also employ algorithmic tools to detect manipulations. These may be facilitated through legal requirements and monitoring devices set up by nonprofit organizations.

The issues of targeted online advertisements span to legal, contractual, and ethical dimensions. The internet and digital landscape transcend national boundaries. National laws are yet to completely incorporate the online landscape. Many countries are yet to adopt data protection legislation. They may therefore be behind in adapting to the technological advancements. Due to the global and dynamic nature, there is no international framework laying down the foundations of the same.

Such a framework can ensure uniformity and transparency. It will also place a higher standard of accountability. This framework should also incorporate the broader view aggregating the different systems of law. National legislation based on such a framework can empower consumers and create safe digital spaces. Transparency, security, privacy, autonomy, and self-determination should serve as the foundational pillars of this. A comprehensive framework can ensure a consumer-friendly digital space that balances technological advancements with regulatory measures to mitigate risks.

REFERENCES

- Argan, M., Dinç, H., Kaya, S., & Tokay Argan, M. (2023). Artificial intelligence (AI) in advertising: Understanding and schematizing the behaviors of social media users. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 11(3), 331–348. <https://doi.org/10.14201/adcaij.28331>
- Antonio, B. A., Dela Cruz, K. L., Jimenez, A. I., & Pantoja, E. (2022). Invasion or personalization: An overview on user attitudes towards the privacy issues in targeted advertising in NCR and its effect in consumer

- purchase behavior. *Journal of Business and Management Studies*, 4(2), 38–47. <https://doi.org/10.32996/jbms.2022.4.2.4>
- Beauvisage, T., Beuscart, J.-S., Coavoux, S., & Mellet, K. (2023). How online advertising targets consumers: The uses of categories and algorithmic tools by audience planners. *New Media & Society*, 1, 12–14. <https://doi.org/10.1177/14614448221146174>
- Boerman, S. C., Kruikemeier, S., & Zuiderveen Borgesius, F. J. (2017). Online behavioral advertising: A literature review and research agenda. *Journal of Advertising*, 46(3), 363–376. <https://doi.org/10.1080/00913367.2017.1339368>
- Boyko, N., & Kholodetska, Y. (2022). Using artificial intelligence algorithms in advertising. In *2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT)*, 317–321. IEEE. <https://doi.org/10.1109/CSIT56902.2022.10000819>
- Campbell, C., Plangger, K., Sands, S., & Kietzmann, J. (2022). Preparing for an era of deepfakes and AI-generated Ads: A framework for understanding responses to manipulated advertising. *Journal of Advertising*, 51(1), 22–38. <https://doi.org/10.1080/00913367.2021.1909515>
- Chaitanya, K., Saha, G. C., Saha, H., Acharya, S., & Singla, M. (2023). The impact of artificial intelligence and machine learning in digital marketing strategies. *European Economic Letters*, 13(3). <https://doi.org/10.52783/eel.v13i3.393>
- Choi, J.-A., & Lim, K. (2020). Identifying machine learning techniques for classification of target advertising. *ICT Express*, 6(3), 175–180. <https://doi.org/10.1016/j.icte.2020.04.012>
- Davola, A. (2021). Fostering consumer protection in the granular market. The role of rules on consent, misrepresentation and fraud in regulating personalized practices. *SSRN Electronic Journal*, 1, 25–27. <https://doi.org/10.2139/ssrn.3791265>
- Davola, A., Querci, I., & Romani, S. (2023). No consumer is an island—relational disclosure as a regulatory strategy to advance consumer

- protection against microtargeting. *Journal of Consumer Policy*, 46(1), 1–25. <https://doi.org/10.1007/s10603-022-09530-7>
- De Marcellis-Warin, N., Marty, F., Thelisson, E., & Warin, T. (2022). Artificial intelligence and consumer manipulations: From consumer's counter algorithms to firm's self-regulation tools. *AI and Ethics*, 2(2), 259–268. <https://doi.org/10.1007/s43681-022-00149-5>
- Directive 2019/2161 of the European Parliament and of the Council of 27 November 2019. (2019). *Amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as Regards the Better Enforcement and Modernisation of Union Consumer Protection rules*.
- Farseev, A. (2023). Under the hood of social media advertising: How do we use AI responsibly for advertising targeting and creative evaluation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 1281–1282. ACM. <https://doi.org/10.1145/3539597.3575791>
- Filter Bubbles and Targeted Advertising* (First Edition). (2020). New York Times Educational Publishing.
- Galli, F. (2020). AI and consumers manipulation: What the role of EU fair marketing law? *Católica Law Review*, 4(2), 35–64.
- Gao, B., Wang, Y., Xie, H., Hu, Y., & Hu, Y. (2023). Artificial intelligence in advertising: Advancements, challenges, and ethical considerations in targeting, personalization, content creation, and Ad optimization. *SAGE Open*, 13(4), 21582440231210759. <https://doi.org/10.1177/21582440231210759>
- Gouda, N. K., Biswal, S. K., & Parveen, B. (2020). Application of artificial intelligence in advertising & public relations and emerging ethical issues in the ecosystem. *International Journal of Advanced Science and Technology*, 29(06), 7561–7570.
- Haleem, A., Javaid, M., Asim Qadri, M., Pratap Singh, R., & Suman, R. (2022). Artificial intelligence (AI) applications for marketing: A

- literature-based study. *International Journal of Intelligent Networks*, 3, 119–132. <https://doi.org/10.1016/j.ijin.2022.08.005>
- Hunt, K. M. (2018). *consumeR-IOT: Where everything collides. Promoting consumer Internet of Things protection in Australia*. Bond University. <https://research.bond.edu.au/en/studentTheses/consumer-iot-where-every-thing-collides-promoting-consumer-intern>
- Koolen, C. (2023). Consumer protection in the age of artificial intelligence: Breaking down the silo mentality between consumer, competition, and data. *European Review of Private Law*, 31(2/3).
- Li, H. (2019). Special Section Introduction: Artificial intelligence and advertising. *Journal of Advertising*, 48(4), 333–337. <https://doi.org/10.1080/00913367.2019.1654947>
- Mandal, A., Mitchell, J., Montgomery, H., & Roy, A. (2017). Privacy for targeted advertising. In *2017 IEEE Conference on Communications and Network Security (CNS)*, 438–443.
- Papić, T., Mihajlović, A., & Gajić, J. (2023). Advanced technologies as a framework for sustainable marketing campaigns (AI application in neuromarketing). In *Proceedings of the International Scientific Conference—Sinteza 2023*, 180–184. <https://doi.org/10.15308/Sinteza-2023-180-184>
- Paterson, J. M., Chang, S., Cheong, M., Culnane, C., Dreyfus, S., & McKay, D. (2021). The hidden harms of targeted advertising by algorithm and interventions from the Consumer Protection toolkit. *IJCLP*, 9, 1.
- Regulation (EU) 679/2016 of the European Parliament and of the Council of 27 April 2016 (2016). On the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.
- Samat, S., Acquisti, A., & Babcock, L. (2017). Raise the curtains: The effect of awareness about targeting on consumer attitudes and purchase intentions. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, 299–319. Conference, Santa Clara, CA, USA.

- Sartor, G., Lagioia, F., & Galli, F. (2021). *Regulating Targeted and Behavioural Advertising in Digital Services. How to Ensure Users' Informed Consent*. Think Tank European Parliament. [www.europarl.europa.eu/RegData/etudes/STUD/2021/694680/IPOL_STU\(2021\)694680_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2021/694680/IPOL_STU(2021)694680_EN.pdf)
- Shin, J., & Yu, J. (2021). Targeted advertising and consumer inference. *Marketing Science*, 40(5), 900–922. <https://doi.org/10.1287/mksc.2021.1284>
- Smith, M. (2015). *Targeted: How Technology Is Revolutionizing Advertising and the Way Companies Reach Consumers*. AMACOM/American Management Association.
- Tang, X., & Yu, H. (2022). Towards trustworthy AI-empowered real-time bidding for online advertisement auctioning (Version 1). *arXiv*. <https://doi.org/10.48550/ARXIV.2210.07770>
- Theodoridis, P. K., & Gkikas, D. C. (2019). How artificial intelligence affects digital marketing. In A. Kavoura, E. Kefallonitis, & A. Giovanis (Eds.), *Strategic Innovative Marketing and Tourism* (pp. 1319–1327). Springer International Publishing. https://doi.org/10.1007/978-3-030-12453-3_151
- Thilagavathy, N., & Kumar, E. P. (2021). Artificial intelligence on digital marketing—an overview. *Nveo-Natural Volatiles & Essential Oils Journal (NVEO)*, 9895–9908.
- Tzoulia, E. (2021). Targeted advertising in the digital era: Modern challenges to consumer privacy and economic freedom: The responses of the EU legal order. In *EU Internet Law in the Digital Single Market* (pp. 447–477). Springer.
- Zard, L., & Sears, A. M. (2022). Targeted advertising and consumer protection law in the European Union. *Vanderbilt Journal of Transnational Law*, 56 <https://doi.org/10.31235/osf.io/jbpsm>

16 AI-Generated Content and Copyright

Legal and Ethical Challenges in India

Hammouch Hind, Sushma Singh, and Anushka Singh

DOI: [10.1201/9781003668657-16](https://doi.org/10.1201/9781003668657-16)

16.1 INTRODUCTION

Artificial intelligence (AI) has transitioned from being a simple tool to being a creator in its own capacity, with its influence extending to many domains like natural language processing, picture recognition software, and product recommendation engines. Generative AI systems, such as ChatGPT and DALL-E, showcase the artistic capabilities of AI by generating visuals that correspond to written prompts and providing written responses that are coherent and contextually appropriate. Generative AI techniques give rise to intellectual property problems, such as ownership of works, the violation of third-party intellectual property rights (IPR), concerns about personality rights, and the creation of illegal material. Prior to adopting the benefits of generative AI, businesses must understand the related dangers and develop suitable preventive measures.

The interaction between copyright infringement and AI can impact the progress of AI technology and its prospective uses. Ensuring the security of copyright owners' rights while fostering innovation in AI is crucial for the advancement and growth of the field. This phenomenon has revolutionized

our quality of life, work methods, and creative production; however, it also poses difficulties in our comprehension of progress, equity, and possession. The creative skills of generative AI span several disciplines, such as visual arts and literary works, which disrupt traditional ideas of authorship and copyright. AI algorithms have the ability to examine vast amounts of data, acquiring knowledge and imitating patterns to produce material that demonstrates human ingenuity, including many forms such as visual arts and literary compositions. The favorable influence of AI on copyright concerns stems from its ability to analyze vast quantities of data, detect patterns, and make informed decisions. AI may aid content producers, copyright holders, and enforcement agencies in several copyright-related domains, thus giving them more power and skills.

However, the responsible use of AI in compliance with current legislation can establish a healthy balance between technical progress and copyright safeguarding, necessitating the implementation of ethical and legal frameworks. Collaboration between politicians, stakeholders, and tech experts is essential in exploring the whole potential of AI in addressing copyright issues in the dynamic digital landscape. By embracing AI as a means of promoting positive transformation, we can encourage the development of new ideas, protect the rights of intellectual property, and nurture a vibrant and innovative atmosphere that benefits creators, customers, and society at large. Human involvement may be restricted to supplying data to the AI. Research shows that around 2.5 quintillion bytes of data are produced every day, and this number is continuously rising in the aftermath of the COVID-19 epidemic, which compelled many people to work remotely.

The use of copyrighted resources in AI model training is a legally ambiguous matter, as copyright rules do not provide full protection for any content solely created by AI, irrespective of its genesis from a human-crafted text prompt. Fair use regulations provide a legal framework that permits the use of copyrighted content without the owner's permission in some situations. However, continuing legal conflicts have the potential to disrupt this established norm and introduce uncertainty over the future of

training AI models. Hence, it is crucial to uphold a copyright system that is comprehensive and resilient by carefully considering the concerns of content producers, consumers, and the general public.^{[1](#)}

16.2 UNDERSTANDING GENERATIVE AI, ITS NEXUS WITH COPYRIGHT LAW, AND ITS EFFECTS

Generative AI technologies, such as ChatGPT, Midjourney, Copilot, and Firefly, are machine learning algorithms that utilize user input to produce novel content. These technologies are trained using extensive datasets, which often include hundreds of millions of websites, including textual or graphic information. These training data sets can consist of publicly accessible and unconstrained information (referred to as pure data), encrypted information (such as copyright-protected works), or a combination of both, depending on the technique used by the AI tool developer.

AI-generated text is material that is created by AI. It employs machine learning algorithms to evaluate large amounts of information and generate new outputs by identifying and using noticeable patterns and trends. The importance of generative models rests in their capacity to foster creativity and innovation by generating novel creations in music, writing, or visuals.

Personalization is an essential aspect of generative AI, enabling customized recommendations, customized information, and distinct user experiences. For instance, the incorporation of generative AI into educational programs enables the creation of customized courses that are specifically designed to match the unique learning pace of each individual, hence enhancing accessibility and effectiveness. Artists, designers, and writers might find inspiration in the wide range of results produced by these systems, allowing them to explore different creative possibilities.^{[2](#)}

OpenAI, a leading entity in the industry, has a market valuation above \$90 billion and has established a collaboration with Microsoft, positioning it as the most valuable company worldwide. Presently, Chatbots such as ChatGPT, Claude, and Stable Diffusion serve as illustrations of generative

AI systems. Nevertheless, there are still obstacles to overcome to ensure consistent quality control and the delivery of cohesive results.³

The use of generative AI tools is anticipated to grow as internet connectivity and electricity become more accessible to a larger population, leading to their increased applicability across diverse professions. Nevertheless, closing the disparity between individuals who possess digital technology and those who lack it poses a substantial obstacle.⁴

The intersection between copyright law and generative AI presents substantial obstacles. Traditionally, authors own exclusive privileges over their unique creations, differentiating them based on their imaginative choices and discernment. Nevertheless, AI-generated content is frequently not regarded as authored works because of its lack of aim or originality. However, AI systems are governed by algorithms that analyze and interpret patterns in data. This gives rise to inquiries over ownership and authorship, which can be handled if an individual guides an AI system during its production. The primary objective of the European Union (EU)'s Copyright Directive is to provide clarity about the legal standing of works created by AI, as well as to guarantee appropriate recognition and remuneration.⁵ Copyright infringement concerns encompass the utilization of copyrighted works as data sets for training AI models, the production of outputs that resemble copyrighted content, and the failure to attribute or manipulate rights management information. The issue of whether generative AI models may be deemed as "authors" involves basic inquiries regarding the definition of authorship and machine-generated creations. This is exemplified by the AI-generated painting, "The Next Rembrandt," which demonstrates the challenge of distinguishing between human-made and machine-generated artworks owing to the advanced degree of sophistication in AI-created material. The implementation of AI in generating news items, poetry, music, and designs necessitates a reassessment of copyright regulations to establish ownership and liability in the event of infringement, while also finding a middle ground between authors' rights and technical advancements.⁶

16.3 INDIAN LEGAL RULES AND REGULATIONS FOR GENERATIVE AI

The important contention is to keep pace with the fast-changing AI generative space. We should always refresh and create laws that voyage between innovation and legality or even ethical concerns.

The Copyright Act of India, 1957⁷ (hereinafter referred to as “the Act”) grants authors exclusive rights to their original works for a limited time. These works may include literature, drama, music, and art, among others. Unauthorized use of copyrighted material is regarded as an infringement under this law. However, there is no provision in the Act for recognizing non-human authors, thus leaving copyright protection uncertain on AI-generated content.

The Patents Act 1970⁸ grants the rights for patents in new processes and articles of manufacture, or products that make up an article. It has provisions on when utilization of work protected by copyright is allowed and the circumstances under which use constitutes copyright infringement. Nevertheless, the legal framework of the Act does not contain provisions for inventions or creations that are entirely produced by AI systems.

The 2000 Information Technology Act⁹ controls electronic commerce, as well as cyber security, through various provisions, such as data protection, legal recognition of electronic records, and digital signatures, among others. AI is not directly treated under this law, though its indirect effect is greatly felt through data privacy and cybersecurity issues, as it contains provisions on how to protect them.¹⁰

Concerns about processing and protecting personal data are addressed in the Digital Personal Data Protection (DPDP) Act 2023.¹¹ However, there are some gaps in the DPDP Act, such as allowing scraping of data by AI companies when it is publicly available, thereby raising issues regarding privacy and ethics. Additionally, the “right to be forgotten under the guise of right to correction and erasure of personal data” is established in Section 12 of the DPDP Act, 2023. It requires the deletion of the user’s personal data upon their request. This can function as an effortless method for the

removal of copyright-infringing material. However, practical concerns arise because the AI software is unable to “unlearn” itself once it has been trained on a selected data set. For example, Chat GPT’s “opt-out function” compared to data collection would only disable the chat history. This implies that the training process continues to incorporate outdated data, while only the most recent content is excluded from the training process. It does not offer any method for removing the data that was previously input. This is particularly alarming due to the fact that AI tools are currently being integrated into a variety of platforms, including web browsers, which allows for real-time information to be obtained with each click. It has become important to rectify the gaps and loopholes in the data protection laws, including but not limited to the DPDP Act, especially for the abuse of publicly available data.[12](#)

To overcome the problems caused by generative AI, Indian lawmakers and executives have taken measures to handle its multiple concerns. With this regard, NITI Aayog has been instrumental in launching a “National Strategy for AI” Discussion Paper and organizing the International Conference on Tools with AI (ICTAI) in partnership with private players. Moreover, AIRAWAT—India’s first AI-specific cloud computing infrastructure—will soon be unveiled by NITI Aayog, which also plans to establish an AI research, analytics and knowledge assimilation platform so as to make India a leader among emerging economies in the field of AI. The Draft National Strategy for AI suggested by NITI Aayog lays down a roadmap for AI development and regulation in India that aims at fostering innovation while dealing with ethical and legal issues related thereof; it proposes ethical guidelines such as transparency, accountability and privacy protection along with recommending regulatory sandbox approach toward testing out AI applications within controlled environment which helps us understand their implications better before developing appropriate regulations around them.[13](#)

Healthcare is regulated by the Medical Council of India (MCI) and other bodies, where they ensure conformity with medical standards besides safeguarding patients’ privacy during the usage of various AI tools adopted

therein. Likewise, finance sector sees the Reserve Bank of India (RBI) taking charge over its use, mainly focusing on aspects around cybersecurity, data privacy, along with ethical deployments being made throughout different systems employed within this area. Similarly, Ministry of Road Transport and Highways (MoRTH) regulates self-driving cars powered by computer vision algorithms capable enough to drive autonomously without human intervention but still need adequate rules ensuring safety standards are met while operating them on public roads across nation. It is advantageous to encourage awareness and educate people about AI technologies and their impact, which is the reason to develop an aware society.

India was among 15 nations that signed the charter of the Global Partnership on AI in 2020, executing scientific advancements into positive frameworks to apply appropriate and responsible integration of AI. The Indian government has implemented measures such as the AI Task Force and the “AI for All” policy to leverage AI for economic and social transformations. Nevertheless, the Government of India has implemented measures such as the “AI for All” policy and the formation of the AI Task Force to capitalize on the potential of AI for economic and societal transformation, recognizing its importance in the developmental process.¹⁴

In addition, in its 161st report on “IPR Regime in India,” the Parliamentary Standing Committee supported the suggestions to exclusively categorize AI and its intricate innovations separately under IPRs. The Copyright Act, 1957, was found to be “not well equipped for facilitating authorship and ownership by AI” in the 161st Parliamentary Standing Committee Report. For example, Section 16 of the Act explicitly stipulates that “no person” shall be entitled to copyright protection except in accordance with the terms of this Act. The Report, however, recommended that “patent protection” be extended to AI-generated works to “incentivize innovation, research, and development.” We propose that this rationale be further expanded to encompass protection of copyright for AI-generated content to foster creativity and improve AI-based expressions.

The Ministry of State for Electronics and Information Technology Rajeev Chandrasekhar explained in detail in an interview in June 2023 that the government is prepared to provide “governance” to AI to protect “digital citizens.” Further, the Union Ministry of Commerce and Industry has stated that the existing Indian IPR regime is sufficient in terms of protection of AI generated works in their opinion; there is no need for a new category of rights.¹⁵ Yet, issues remain, and should not be overlooked as demonstrated by the lawsuits and plea for rules for copyright protection of generative AI models by actors, news outlets, and creative writers.¹⁶

16.4 LEGAL ISSUES AND CHALLENGES IN COPYRIGHT LAW OF INDIA WITH GENERATIVE AI

While AI-generated content has become increasingly prevalent in today’s society, there are various legal issues that come into play when discussing the relationship between content and copyright that may require subsequent legal analysis and legislation revision. The major concern in the creative field, as the AI systems grow in the capability of writing/singing/creating/painting without any human input, is that little by little, concepts of authorship and ownership, as well as the whole paradigm of IPR, are being challenged.

The Indian Copyright Act of 1957 safeguards copyright in India. Section 13¹⁷ of the Act provides copyright protection for a range of creative works, such as “literary, musical, theatrical, artistic works, sound recordings, and cinematographic films.” These categories contain a wide array of artistic manifestations. For example, the Act provides legal protection for literary works, including publications, manuscripts, and poetry. It guarantees the protection of creative works in literature, theater, music, art, cinematographic films, and sound recordings against unlawful use and access. Section 14¹⁸ of the Act stipulates that the owner of a work has exclusive rights to perform or allow particular acts such as reproduction, publishing, modification, and translation. These rights are legally referred to as “Copyright.” Furthermore, Section 17¹⁹ of the same Act explicitly

states that the original author of the work is immediately acknowledged as the copyright proprietor. Nevertheless, if the work is produced as a result of a contractual obligation for payment at the employer's behest, in this situation, the employer takes ownership of the work. The present copyright rules face several challenges when it comes to comprehending the implications of generative AI, which will be discussed in detail.[20](#)

16.5 AUTHORSHIP

Pertinently, the Indian Copyright Act, 1957, like the majority of copyright laws across the globe, professes human authorship as its foundation. According to Section 2(d), the author means the person who writes or composes the work. In the context of AI, this raises significant questions: Is it possible to offer consideration of an AI author based on Indian law? If this is not the stance, then who is to be considered as the author of the text—the programmer of the AI software, or the user of AI? This lack of clarity calls for the intervention of legislative or judicial systems to determine who is considered an author of an AI-generated work putting it within the legal framework of India.[21](#)

The attribution of authorship is assigned to the “individual” responsible for creating the work; however, the term “individual” is not clearly defined. The definition of a “person” provided in Section 16 of the General Clauses Act, 1897, which includes “any company or association or body of individuals, whether incorporated or not,” does not offer a definitive solution. Under the rules of the Copyright Act of India, copyright protection is only awarded to human writers. According to Section 171 of the Act, individuals are not eligible to get a copyright unless they meet the requirements specified by the legislation. In 1994, the Act was modified to address scenarios in which computer-generated creative works may be produced. The modification added Section 2(d)(vi) to the statute, which specifically declares that the person responsible for creating computer-generated works is the rightful creator of such works. Section 2(d)(vi)[22](#) of the Act grants authorization to attribute authorship to the individual who is

responsible for the creation of computer-generated works. When filing for copyright registration, Section 14 requires the submission of specific information, including the author's name, nationality, and address. According to Section 2(d) of the Act, it may be deduced that a person who has only provided a single-line input cannot be regarded as the one responsible for the formation of the output. According to Indian law, it is not allowed for anybody to assert authorship, co-authorship, or ownership of an AI-generated work. This is a hurdle to us as the Indian legal system does not acknowledge AI as a legally recognized entity. Consequently, the current legal structure may not effectively handle works created by entities that are neither human nor legally recognized. Granting AI distinct legal status and conferring it with ownership of IPR has the potential to lead to copyright infringements. Furthermore, the current legislation may not adequately deal with these violations as Section 51²³ of the Act specifically specifies that copyright infringement can only be perpetrated by an individual.²⁴

The copyright claim on a computer-generated list in the case of *Navigators Logistics Ltd. v/s Kashif Qureshi*²⁵ was dismissed because there was no human involvement in its creation. The court determined that human participation in the creation process is necessary for the granting of copyright protection in India. The "Significant Input" test sets two crucial criteria for assessing whether an author, who employed AI aid in producing a product, may assert copyright or not. The first criterion is objective and involves determining if there was any human participation in the creation process. Furthermore, the degree of human engagement should be taken into account. The level of human expertise, discernment, and effort dedicated to its formation must be substantial to the point that the result would be fundamentally altered or nonexistent in its absence.²⁶

The lack of specific legislation regarding these situations might result in legal ambiguities and disputes between authors, operators of AI systems, and businesses that employ their results. Conventional copyright rules revolve around human authorship, prompting inquiries over the recognition of AI systems as creators of artistic works. The US Copyright Office in the

United States has explicitly declared that works created only by AI systems do not qualify for copyright protection due to their absence of human creativity.²⁷ Conversely, the Copyright Directive of the EU recognizes that copyright can be given to works produced with the aid of AI or other computer-generated methods. EU has drafted an AI Act,²⁸ which can be adopted by India. The contrasting methods exemplify the current discourse about the involvement of AI in the creative procedure and the assignment of authorship.²⁹

16.6 ORIGINALITY

Indian copyright law protects works that are original and exhibit a modicum of creativity. An author is awarded protections under copyright law if their artistic work or product satisfies the necessary criteria to be considered an “original” creation. According to the Copyright Act in India, for a creation to be eligible for copyright protection, it must meet the requirement of having a minimum level of inventiveness, as established by the Supreme Court’s decision in the case of *Eastern Book Company & Ors v/s D.B. Modak & Anr.*³⁰ The court determined that a “minimum degree of inventiveness” is necessary, characterized by a significant difference rather than a minor one.

To be considered original, a work must represent an independent and creative expression of ideas. The originality test, as laid down in cases like *Eastern Book Company* case that requires that the expression should be new or different from what has already been said before about it. However, this is where things get interesting with AI-generated content—such texts often come from massive data sets which mean they might not meet this criterion for novelty because they draw upon large quantities of information while generating them. So how can we determine if something created by an AI passes muster under Indian copyright law’s originality requirements, especially given that many times these kinds of works heavily reference other existing ones?

The legislation safeguards literary and creative creations that are deemed to be “original.” Nevertheless, as per the dominant idea, AI is presently incapable of generating information that is genuinely “original.” AI-generated works are commonly seen as adaptations or alterations of preexisting knowledge that the AI has acquired, evaluated, and been taught on. This viewpoint originates from the recognition that AI systems depend on datasets that are shaped by the prejudices and constraints of their human developers. Put simply, the results of AI are regarded as being taken from already existing data rather than being truly new or creative. Nevertheless, a result derived from a basic prompt, regardless of its uniqueness compared to other works, cannot be protected by copyright as it lacks the substantial effort often exerted by human writers when crafting a single-line prompt.³¹

The Supreme Court deliberated on many criteria of “originality” to establish the criterion for India. The “sweat of the brow” theory is the minimum need for awarding copyright safeguards to the author; however, it does not apply in India. According to this doctrine, copyright rights are obtained by the author via their diligent and concerted work. A significant case for comprehending this matter is *University of London Press Ltd v/s Tutorial Press Ltd*,³² in which a publisher created a compilation of question papers that were published by the University of London. Copyright is granted based on the level of care and effort put into creating something, rather than requiring further inventiveness. The “Sweat of the Brow” criterion was deemed insufficient for granting copyright protection and was seen as excessively favoring the rights of the owner, therefore creating an imbalance in copyright protection. It also examined the American criterion of “modicum of creativity,” which asserts that a certain degree of inventiveness is necessary for a product to be eligible for copyright protection. The court determined that the criterion was excessively stringent. The Canadian standard, which has been ultimately adopted, stipulates that the work in question must be a manifestation of the author’s expertise and discernment. The exercise of ability and judgment should not be overly simplistic to the point that it can be only performed as a mechanical task. This further reinforces the case for why a result produced

by a one-line “prompt” cannot be afforded protection. Ultimately, it was determined that the human expertise and discernment required in producing the supplementary components encompassed legal knowledge, skill, and the author’s judgment, so rendering the SCC version of verdicts eligible for copyright protection. The criterion of novelty is seen excessively stringent for copyright, and there is no justification for establishing a distinct criterion for AI-generated works. Hence, AI-generated creations in India can fulfil the necessary criteria of “originality” regardless of whether or not they were drawn from pre-existing information.

16.7 OWNERSHIP

In most cases, the author is the copyright holder in India unless the work is done under employment or a contract indicating otherwise. Determining ownership for AI-generated works presents problems. If AI cannot be an author, should the programmer or user default to ownership? Also, who owns collaborative works between humans and AI?

Establishing ownership of computer-generated outputs is a multifaceted matter within the realm of copyright law that extends across several legal jurisdictions. In the Indian context, the Act does not provide protection specifically for “computer-generated work.” Instead, it grants protection to the author of such works generated by computers, who is described as “the person responsible for creating the work.” India’s stance on copyright law regarding AI-generated outcomes closely aligns with that of the United States, as both countries do not recognize AI as the sole proprietor of the output.[33](#)

16.8 ECONOMIC AND MORAL RIGHTS

Under Indian law, authors have economic rights (e.g., reproduction and distribution) as well as moral rights (e.g., right to be identified as having an authorship). Applying these rights to AI-generated works can be difficult because it is unclear which entity should possess them. For example, what

about moral rights for an AI-generated work that incorporates elements from multiple human authors' works?

Furthermore, AI currently lacks consciousness and therefore not be granted moral entitlements. Moreover, acknowledging AI as the creator of copyrightable works might give rise to several complexities. For example, an AI lacks the capability to independently transfer or license the rights to the work it creates, and it is unable to determine if the work has been criticized, so making moral rights ineffective. In the absence of any *mens rea*, enforcing human culpability in AI-created products will be exceedingly challenging. In the case of AI-generated work in India, it can be seen that if the work is defamatory, obscene, or goes below public morals, the requirement of "intention to harm" for defamation cannot be fulfilled by the AI. The inquiries that arise from contemplating a person as the originator of AI works are virtually boundless. Nevertheless, current laws, both at the international and national levels, have not yet established any recognition of authorship or imposed legal responsibility for any violations that may arise from works made by AI.

16.9 INFRINGEMENT AND LIABILITY

The Act also contains provisions against infringing upon these protected works in an unauthorized way. Section 51 of the act defines infringement when an unauthorized person reproduces, distributes, performs, adapts, or translates the work. Infringement can also entail unauthorized copying, selling, renting, importing, public display, and digital storage. There are, however, some like private use; criticism and review; reporting of current events; educational purposes; and public library use, which are exceptions under "fair use" in Section 52.³⁴ Provisions for remedies, on the other hand, include civil remedies (Section 55³⁵) such as injunctions, damages, accounts of profits and delivery up of infringing copies as well as criminal sanctions (Section 63³⁶) involving imprisonment and fines; seizure of infringing copies. Enforcement can be pursued through civil suits in district or High Courts and criminal prosecution via police complaints.

In India, as well as in other places, AI systems have the potential to unintentionally generate content that violates current copyrights. Establishing responsibility for such a violation is a complex task. Who should bear liability: the AI programmer, the user, or the entity responsible for training the AI? The current Indian legal framework does not have explicit provisions for such situations, which requires courts to interpret existing statutes in an innovative manner.³⁷

Currently, according to the existing legislation, no legal authority holds an AI system responsible for copyright infringement. Instead, the responsibility for copyright infringement will always be placed on a human being. When a work violates the rights of a copyrighted work, the creator of the violating work is often held responsible for the violation. If the inventor of an AI is seen to be the author of any works produced by the AI, they will be held accountable for any copyright violations committed by the AI.

The interaction between AI training procedures and copyright infringement demonstrates a connection that merges technical progress with issues over intellectual property. The creative process of AI is refined by being exposed to extensive datasets, which may include copyrighted materials. This training process, although essential for the advancement of AI, poses important concerns regarding its compatibility with copyright standards.

16.10 FAIR USE AND EXCEPTIONS

The Indian Copyright Act includes fair dealing provisions which permit limited use of copyrighted materials without permission under certain conditions such as research or criticism. Applying this to AI-generated works involves determining if using copyrighted material by an AI during its training amounts to fair use, especially given that vast amounts of data processed by them, making it hard to ensure compliance with standards on fair usage.

Applying the “fair use” and “transformative use” exceptions to computer-generated and computer-aided results poses distinct issues. Fair

use permits the utilization of copyrighted content without obtaining permission, if specific conditions are met, including criticism/critiques, commentary, parody, or educational objectives. Transformative usage pertains to the alteration or adjustment of a piece of work to generate something novel and distinct. Nevertheless, when it comes to AI-generated creations, the application of these principles brings up inquiries regarding the level and impact of human involvement, as well as the transformative aspect of their utilization. To fully utilize the capabilities of AI, courts must also analyze fair use rules within the framework of computer-generated and computer-aided results, taking into account the intention, characteristics, and impact of the usage. For example, the United States has witnessed several copyright lawsuits concerning the use of transformative usage in relation to works created by AI. In the *Naruto v/s Slater*³⁸ case, a monkey used a wildlife photographer's camera to take a selfie. The court determined that the monkey did not have the legal right to claim copyright ownership. In the *Warner Bros. Entertainment Inc. v/s RDR Books*³⁹ case, a fan-created Harry Potter lexicon was determined to violate copyright laws. The court examined the extent to which the lexicon transformed the original work and its effect on the market.⁴⁰

Furthermore, AI systems may face a possible concern regarding the originality or distinctiveness of the material they produce. The terms of service of AI platforms explicitly admit that the result generated for similar inquiries may not be unique among different users. Imagine a scenario in which one user asserts copyright ownership of a certain result, and subsequently, another user independently arrives to the identical result and likewise asserts copyright ownership of it. If a significant number of such claims emerge, it has the potential to generate complications. In contrast to trademark law, copyright law does not acknowledge the notion of honest and contemporaneous usage, as it is improbable for two persons to create an identical play or book. In contrast to conventional copyright infringement cases, when one party is obviously responsible, it would be challenging to anticipate the result in these particular circumstances.⁴¹

According to Section 52 of the Act, fair dealing is allowed for private or personal use, including study. This legal approach was taken from the United States. An essential question that Courts must confront when considering the extension of legal rights to AI Technologies is whether the use of AI systems for training and learning purposes would be considered private or personal use, without involving any human other than the AI system itself. Additionally, it will be necessary to evaluate whether it is appropriate for a business body such as OpenAI to engage in private usage for its own learning and development, even if this learning results in a competing product.

Does the defense of private or personal use, as stated in “Section 52(1)(a)(i)” of the Act, apply solely to individuals or does it also include corporations or other entities? Regarding research use, it might be argued that collecting information from expressions without exposing any individual to the expression could be considered protected research use under “Section 51(1)(a)(i)” of the Act. Crucially, the clarification to Section 52(1)(a) explicitly states that storing anything for fair use in a private or personal capacity, such as for research purposes, does not constitute infringement. However, these problems will only come up if the Courts initially consider the preservation and use of data for training purposes to be eligible for safeguarding under Section 14 of the Act.⁴²

In previous copyright infringement cases, the defendants’ websites or OTT platforms enabled the unauthorized streaming of copyrighted sound recordings. They argued for exemptions under “Section 52(1)(a)(i)”²² of the Act, which specifically relates to private or personal use. Nevertheless, both judges determined that if the online streaming and distribution of sound recordings includes a charge or income derived from sponsors or other parties, it qualifies as commercial exploitation and is not covered by the exemption for personal use. When AI systems produce output for commercial or public use, it may not be considered fair usage unless it is significantly transformative according to certain criteria. For instance, the results acquired from searching for particular portions from a chapter of a book in ChatGPT might fall short of the criteria defined for fair use.

Despite taking into account the educational use exception outlined in “Section 52(1)(i)” of the Copyright Act, OpenAI may nevertheless face legal responsibility. The reason for this is that there is a differentiation between producing and disseminating course packets to a restricted group of students who do not include a prospective market for the works in issue, and permitting hundreds or even millions of individuals to access substantial portions of literary works.[43](#)

16.11 REGULATORY AND POLICY CHALLENGES

India needs a regulatory framework that can account for peculiarities associated with AI-generated work. Current legislation may require amendments so as provide for authorship, ownership, and liability clearly. Also, enacting specific rules and policies governing AI-generated content creation would help pre-emptively tackle these issues. The government’s Digital India initiative, coupled with increased adoption of AI across various sectors, calls upon relevant authorities to revise their laws accordingly.[44](#)

16.11.1 INTERNATIONAL STANDARDIZATION

The worldwide reach of AI technology requires India’s copyright laws to be made uniform with those in other countries. To better operate at the complicated crossroads of AI and copyright law, India should take into account its peculiarities as it works with agreements such as the Berne Convention and TRIPS. To come up with workable regulations for AI, this calls for collaborating with global organizations and following what is considered as international standards. It is important to update or create rules that foster invention without infringing on rights while dealing with generative AI which keeps changing.

16.11.2 ETHICAL ASPECTS

The ethical considerations should be what matter most when developing policy and regulation. To make sure that AI contributes positively toward India's cultural and creative landscape, it needs to be transparent, accountable, and fair. There are also other concerns related to generative AI, such as ethical, technical societal matters. These include misuse of synthesized media like deep fakes or breach of intellectual property amongst others. For instance in November 2023 a huge number of social platforms including Twitter and Facebook shared a viral deep fake video featuring Rashmika Mandanna. Additionally, if biased or limited training data is used by generative models, then they will produce outputs reflecting those biases, thereby potentially discriminating against certain groups while reinforcing existing social imbalances. To align with societal values and norms, ethical guidelines need to be strengthened for the development of AI.^{[45](#)}

16.11.3 DEFICIENCY OF ACCOUNTABILITY FOR AI

The challenge of attributing responsibility to AI for its outcomes is a crucial one. Assessing the intelligence of AI directly becomes challenging when the stuff it generates is provocative, obscene, or goes against societal morals. Although repairs might involve the deletion of information or the termination of the AI, attributing authorship to the AI can be hard in the absence of explicit responsibility. The intricacies of AI-based tasks require the creation of a complete framework to effectively tackle the specific challenges connected with them. As the area of AI advances, it is crucial to evaluate and modify copyright laws to strike a balance between fostering innovation and ensuring responsibility in the digital age.

16.11.4 DECIPHERING ROYALTY

The author possesses an indisputable entitlement to receive money. When the creator of a work is AI, it becomes challenging to ascertain the cost and shift to AI. There will be inquiries on the authority responsible for

determining compensation and if the amount of compensation should be fair and proportional.

16.12 PRESENCE OF BIASES AND THE UNAUTHORIZED USE OF COPYRIGHTED CONTENT BY GENERATIVE AI SYSTEMS

Generative AI systems, which depend on algorithms and training data, have the potential to introduce biases and unintentionally include copyrighted information without permission, which raises concerns about legal responsibility and accountability. These systems have the ability to generate new works or modified versions of protected information, which makes it difficult to distinguish between originality and derivative production. According to experts, it is anticipated that by 2026, about 90% of online content might be created by AI. This emphasizes the need to address biases and copyright problems. An investigation employing Stable Diffusion, an AI tool that generates images based on prompts, uncovered a recurring pattern of biased outcomes. Specifically, the tool tended to depict high-paying occupations as being predominantly occupied by individuals with light skin and men, while low-paying jobs were associated with individuals with dark skin and women in certain roles. This example highlights the significance of integrating a variety of data sources to prevent the reinforcement of preconceptions. To tackle these difficulties, it is necessary to have international collaboration, technology advancements, regulatory changes, education, and the participation of many stakeholders. The task of striking a balance between copyright protection and promoting “innovation and creativity” is a substantial regulatory problem in the ever-changing AI environment.[46](#)

In India, copyright law meets AI-generated content at an ever-changing crossroads where legal ambiguity reigns supreme. Addressing authorship, originality ownership infringement liability requires legislative clarity as well as judicial precedent. With the advancement of AI systems within our country’s borders, it becomes necessary for us as Indians to create laws that

protect creators' rights but also foster creativity, such acts may not have been thought about before. Ongoing conversations between lawmakers, judges, technologists, artists, creators, policymakers, etc., will go a long way toward ensuring robustness and relevance in this area, considering how fast things change during times like these when everything seems connected together through some sort of network or another.

16.13 JUDICIAL POSITION OF AI IN INDIA

The intersection between copyright laws relating to AI-generated contents has not yet reached its peak in Indian courts due lack thereof many reference points from relevant cases. However, there exist certain principles which can aid understanding of these matters. The Indian courts have taken strong measures to prevent the exploitation of AI techniques for copyright infringement. Below are some examples in which courts attempt to interpret the discrepancies between rules regarding AI and copyright laws.

So, the cases explained below also analyze the potential challenges with the same respect.

In the case of *Anil Kapoor v/s Simply Life India & Ors*,⁴⁷ the Court has granted an order prohibiting the use of AI to produce counterfeit or altered information, particularly for commercial purposes. Its objective was to safeguard the individual's rights pertaining to their individuality. Actor Anil Kapoor filed a lawsuit against Simply Life India for unauthorized use of generative AI techniques to create fabricated situations featuring him. Simply Life India received a court order from the Delhi High Court, which declared that using AI to create material infringed against Kapoor's personality rights. This case highlights the significance of "privacy and publicity" rights in relation to AI-generated material.

The case of *Ankit Sahni v/s US Copyright Office* serves as a prime example of the ongoing controversy about the ownership of AI creations and the rights to intellectual property. In India, there was a temporary time in which copyright protection was expanded to encompass works created by AI. Nevertheless, this decision was subsequently reversed when a

withdrawal notice was sent. In 2019, the Delhi High Court dismissed a copyright suit about a computer-generated list, citing, among other reasons, the absence of human involvement. In 2020,⁴⁸ the Copyright Office acknowledged Raghav, an AI tool, as a co-author of an artwork it created, with the creator of the AI tool. In 2021, Mr. Ankit Sahni, the owner of an AI-powered painting program called “RAGHAV,” made an effort to obtain copyright protection for an artwork named “*Suryast*.” At first, the Indian Copyright Office rejected an application that sought to identify the AI (“RAGHAV”) as the sole creator of the artwork. Following that, a fresh application was filed, which included both the owner of the AI and the AI itself as co-authors for a different artwork, and this application was approved. However, after a year, the Copyright Office sent a notice of withdrawal, requesting clarity on the legal status of the AI entity known as Raghav. This action signifies a reconsideration of the recognition of AI as a legitimate copyright author in India. This marked the inaugural instance in India where an AI tool was acknowledged as the creator of a copyrighted work. However, later on, the Copyright Office released a notice of withdrawal, indicating that it was the responsibility of the applicant to inform the Copyright Office about the legal status of the AI tool. This ruling perhaps disregarded the original purpose of the Indian legislation, which explicitly acknowledges and safeguards computer-generated works.

In the case of *Mattel Inc. and Ors v/s Jayant Agarwalla and Ors*⁴⁹ Mattel, the proprietor of the Barbie brand, filed a lawsuit against Jayant Agarwalla for producing and distributing computer-generated images of Barbie dolls using AI. The Delhi High Court issued an order in favor of Agarwalla, determining that his utilization of AI to produce and market Barbie pictures violated Mattel’s trademark rights. This case underscores the capacity of AI to produce unlawful information, as well as the imperative for trademark owners to protect their legal interests in the digital domain.

In *Rupendra Kashyap v/s Jiwan Publishing House Pvt. Ltd.*,⁵⁰ the High Court of Delhi used a conventional approach in this case, addressing the copyright dispute between the Central Board of Secondary Education and

the question papers. The Court concluded that the CBSE has the authority to claim copyright unless it can provide proof of personal contribution in the creation of the question papers, considering its position as an artificial organization. According to the Indian copyright legislation, authorship may only be ascribed to an individual. This stance was also reinforced in the matter of *Tech Plus Media Private Ltd. v/s Jyoti Janda*.⁵¹

In this instance of *Tech Plus Media Private Ltd. v/s Jyoti Janda*,⁵² the Court upheld that a juristic person cannot be credited as the author, but can still retain the copyright ownership. The Delhi High Court reaffirmed this interpretation in 2019 in the matter of *Navigators Logistics Ltd. v. Kashif Qureshi & Ors.* The case of Navigators Logistics revolved around a copyright dispute about a computer-generated list. The Court rejected the claim because it lacked evidence of human involvement. This is consistent with the stance in the United States, where authorship cannot be exclusively ascribed to AI. The court dismissed a copyright claim over a computer-generated list, emphasizing the absence of human involvement. Thus, the situation in India corresponds to the approach in the United States, where AI is not eligible for exclusive authorship.

This case of *Najma Heptulla v/s Orient Longman Ltd. & Ors.*,⁵³ dealt with a biographical work's ability to be copyrighted, where effort and novelty were seen important components for establishing protection. The court stated that ideas alone were not enough but rather they needed expression too thus making them copyrightable. AI-generated materials usually involve processing and reorganizing data which already exists, hence may find it hard to meet the requirements set by this decision.

In *Manager ICICI Bank Ltd. v/s Prakash Kaur & Ors.*,⁵⁴ the Supreme Court had an opportunity to show how it approaches new technologies and ethical considerations even though there was no direct relationship between ethicality technology use or copyright protection in this case; however, the apex court emphasized on adopting new techs ethically as well as responsibly. The implication being that when dealing with content created by machines, one needs to look at issues like data and the potential for bias and infringement, among others.

In *R.G. Anand v/s Deluxe Films*⁵⁵ it was elucidated that when a film based on a play infringed copyright law was brought before judges, they argued about similarities showing substantial similarity and idea–expression dichotomy while acknowledging general concepts do not attract legal safeguards. Hence, such works could not undergo scrutiny under idea expression, as these AI only repeat existing data without any unique way presenting them.

In *Yahoo India Pvt. Ltd. v/s Akash Arora & Anr.*,⁵⁶ the dispute involved internet trademark infringement thus setting stage for courts determine how deal with digital age advancements concerning various aspects including but not limited adapting traditional laws into modern times so that they can be applicable even where necessary. The court’s recognition of the need to adapt existing principles to recognize new contexts suggests that judicial minds are open toward developing appropriate rules against creations produced through AI systems.

In the 2016 case of *Mareta v/s Google Inc.*,⁵⁷ the US District Court determined that measures taken to prevent copyright infringement and safeguard privacy should be interpreted in a comprehensive manner to encompass “new technologies.” The Indian Courts have also embraced this fundamental premise of fast change.⁵⁸

16.14 SUGGESTIONS FOR IMPROVING COPYRIGHT REGULATION OF GENERATIVE AI

With the ongoing evolution of AI technology, it is imperative that existing copyright rules be modified to incorporate these breakthroughs. The existing legislation fails to provide explicit guidelines for AI-generated material or innovations, resulting in a state of legal ambiguity. It is imperative to make revisions that specifically address the inclusion of laws regarding AI authorship and patent rights for innovations created by AI. India necessitates all-encompassing AI law that specifically tackles the distinct concerns presented by AI technology, encompassing matters of responsibility, transparency, and accountability.

16.14.1 ADAPTING COPYRIGHT LAWS WITH EVOLVING ADVANCEMENTS

As AI technologies continue to evolve, existing copyright laws need to adapt to accommodate such advancements. The existing laws do not unambiguously cover AI-based content or innovations giving rise to a lack of legal certainty. The laws should be changed so that they can accommodate AI elements in writing as well as the patentability of their inventions. To address various issues of liability, transparency, and accountability among others, India must formulate comprehensive AI legislation that deals with peculiar concerns presented by AI. This may need reevaluating and revising copyright laws to accurately represent the changing environment and tackle emerging forms of creativity and ownership. Recently, Japan made a ruling stating that utilizing creative works to train AI models does not infringe upon the country's copyright law. Thus, this enables AI model trainers to gather data without obtaining the explicit consent of the data owner. Japan has established a precedent by being the first country to take action in this manner. However, the utilization of preexisting artworks to “educate” AI models to imitate the style and produce output similar to that of artists is a matter of worry, as it has the potential to devalue their work.⁵⁹

16.14.2 ALL STAKEHOLDER APPROACH

To effectively utilize generative AI material, it is necessary to adopt a method that involves all stakeholders. This approach should include “consultation and negotiation” among different parties to overcome copyright concerns. It is essential to establish data usage and governance rules for AI projects and require AI businesses to designate compliance officers who will be liable for copyright protection, audits, and evaluations. This is crucial for ensuring proper oversight and adherence to regulations.

16.14.3 COPYRIGHT REGISTRATION AND MANAGEMENT

AI has the ability to streamline the “copyright registration” process by automation of the administrative chores and helping authors effectively manage their copyright portfolios. AI-driven solutions can streamline the process of registering creative works, automatically produce copyright notifications, and assist authors in managing their copyrighted “content, licenses, and permissions.”⁶⁰

16.14.4 EXPANDING FAIR USE APPROACH

It is crucial to broaden the fair use approach, as the utilization of copyrighted content for teaching purposes is often deemed fair use. It is necessary to provide specific exemptions for text and data mining to prevent individuals from claiming copyright protection for AI-generated work when they use copyrighted information for training purposes. The fair use doctrine can be determined by applying the four-factor test established in the 1996 case of *Civic Chandran v/s C. Ammini Amma*,⁶¹ a decision by the Kerala High Court. This test takes into account the intended use of utilization, the nature of the work under copyright, amount and substantiality of the piece used, and influence on the value of the copyrighted work. It is imperative to provide explicit instructions regarding the utilization of AI in copyright applications to guarantee transparency. Additionally, it is vital for AI training to be acknowledged by the judiciary as “research” under the fair dealing framework. Ensuring a harmonious equilibrium between safeguarding the rights of copyright owners and promoting the progress of AI innovation is vital for the development and progression of AI technology. To ensure the continuous development and future uses of AI, it is crucial to tackle the increasing problem of copyright infringement while also supporting innovation.⁶²

16.14.5 COPYRIGHT FILTERING AND MONITORING

Copyright filtering utilizes AI technology to automatically analyze and filter internet information to detect potential copyright violations. By comparing user-uploaded or shared information with a database of copyrighted works or reference materials, these systems utilize a mix of content identification techniques, such as picture recognition, audio analysis, and text matching. Copyright filters identify and highlight content that may possibly infringe copyright laws, allowing those who hold copyright or platform administrators to conduct additional research or take appropriate action. Although copyright “filtering and monitoring” technologies are generally efficient for enforcing copyright, they can occasionally generate inaccurate findings or overlook fair use exceptions. AI algorithms may erroneously classify non-infringing information as infringing owing to variables like as resemblances, context, or technological constraints. Implementing measures to mitigate false positives and ensuring that fair use exceptions are fully evaluated, are essential to prevent the suppression of legal uses of copyrighted content.[63](#)

16.15 SUGGESTIONS FROM THE 161ST PARLIAMENTARY REPORT

The “161st Parliamentary Report” proposes the establishment of a distinct set of rights for AI and AI-related works. Moreover, it is necessary to begin a conversation regarding the practicality of the following recommendations: Firstly, the proposal is to create an autonomous AI Council responsible for evaluating and deciding whether to approve or reject AI systems. A “Certification Board” for AI systems would be established to assess them based on predetermined criteria. Secondly, it was suggested to establish and execute systems of responsibility and oversight. AI platforms like GPT-4 are commercialized, and any unauthorized use of copyrighted material must result in financial recompense for the breach of IPR, while also following our ethical guidelines. Furthermore, it is imperative to establish efficient mechanisms for transparency and cyber security, specifically with the disclosure of data and the guarantee of the right to erasure, as outlined in

Section 12 of the DPDP Act of 2023. Additionally, it is proposed to establish an AI Fund to promptly address any future violations, implement punitive measures to deter the exploitation of children’s data, and safeguard the Right to Privacy as guaranteed by Article 21.[6465](#)

16.15.1 CONTENT IDENTIFICATION, DETECTION AND AUTOMATED TAKEDOWN

AI-powered content identification systems are essential for enforcing copyright laws as they can automatically detect copyrighted material in many formats, including photos, videos, and audio recordings. AI utilizes machine learning algorithms to examine extensive data and cross-reference it with a database of copyrighted works, facilitating the effective identification of unlawful use or infringement. These tools, such as “content recognition” technology employed by sites like YouTube, assist in the proactive identification and administration of copyrighted content.[66](#)

16.15.2 COPYRIGHT EDUCATION AND AWARENESS

AI-powered chatbots, virtual assistants, and interactive platforms have the capability to offer consumers accurate information and assistance regarding copyright laws, fair use, and the process of getting appropriate permits or licenses for the use of copyrighted content. AI systems can facilitate user education on rights and duties, so minimizing inadvertent infringement and fostering adherence to copyright restrictions.

16.15.3 TRACKING LICENSING AND ROYALTIES

AI may assist in monitoring the utilization and dissemination of intellectual material, therefore assisting copyright owners in guaranteeing appropriate licensing and fee remittances. AI algorithms have the capability to oversee digital platforms, detect occurrences of unauthorized usage, and furnish data for the computation and collection of royalties. It is crucial to

acknowledge that although AI may be a significant tool in tackling copyright and infringement problems, it is not a comprehensive answer. Human discernment, legal acumen, and continuous policy formulation are essential to guarantee a well-rounded and equitable approach to copyright enforcement on the internet. Although AI has the ability to effectively tackle copyright and infringement problems, there are several obstacles and factors that must be taken into account. The advancement of AI technology poses a difficulty in dealing with copyright and infringement concerns, as novel kinds of infringement like deep fakes and other emergent tactics arise. Collaborative endeavors such as the “Coalition for Content Provenance and Authenticity” (C2PA), “Project Origin, and the Content Authenticity Initiative” (CAI) that involve technology developers, copyright experts, and legislators are crucial. These initiatives aim to proactively guarantee that AI-driven solutions stay successful in combating emerging infringement tactics. The focus on innovation is seen in solutions such as the “Authentication of Media via Provenance” (AMP), which aims to enhance journalism, combat misinformation, and preserve the integrity of digital content.⁶⁷

The convergence of AI-generated material and copyright rules in India poses intricate legal challenges that need meticulous examination and proactive restructuring. India can establish a legal framework that safeguards intellectual property and encourages the progress of AI technology by examining issues related to authorship, originality, data usage, liability, moral rights, and economic rights, as well as advocating for regulatory change. Striking a balance is crucial to promote innovation and protect the rights and interests of artists in the digital era.

16.16 CONCLUSION

The convergence of copyright infringement with AI in India presents several legal obstacles. As AI technology progresses, it is crucial to find a middle ground that respects the rights of copyright owners while also promoting innovation in AI. This involves modifying copyright laws,

acknowledging the concept of fair use in the age of AI, and establishing strong governance mechanisms. At present, it is not feasible to provide copyright registration to AI without comprehensive legislative reforms that can address all potential gaps and uphold the fundamental principles of copyright law.

However, the Copyright Act of 1957 in India does not specifically include works generated by AI. Nevertheless, it does acknowledge the contributions and instructions provided by humans who developed them. However, human creators bear exclusive responsibility for the main authorship and legal culpability for any violations of copyright. The use of AI-generated material for commercial purposes highlights the need for an update in copyright laws to guarantee the safeguarding of copyright holders' rights. Lawmakers must navigate a complex labyrinth of AI technology while ensuring the protection of IPR, data privacy, and fostering innovation.

To ensure future advancements in technology, it is imperative to establish a balance between copyright protection and AI. The emergence of AI-generated artwork poses difficulties to traditional notions of authorship, originality, and copyright, while simultaneously presenting opportunities for novel partnerships and advancements. The recent copyright disputes in India have not only brought attention to these concerns but have also established significant legal precedents for future cases and policy-making processes. With the rapid progress of AI technology, there will arise substantial number of issues in reconciling human rights with the rights of AI producers. This will pose a daunting task for politicians and legal experts worldwide. Striking a balance between regulation and incentivizing is crucial when addressing the use of AI-driven innovation. The existing intellectual property framework has to be modified to synchronize with the capabilities of AI. Laws based on conventional frameworks have challenges in effectively addressing the matter of determining authorship for information produced by AI. Therefore, it is crucial to reassess current legislation to promote innovation while also safeguarding the rights of artists. An innovative and inclusive innovation ecosystem may be created in

India by implementing a forward-thinking legislative framework that harnesses the advantages of AI.

The convergence of technology and copyright law is not a new phenomenon. Google justified the usage term “transformative” in relation to its search engine. To assure legal compliance in data collection for training AI models, AI developers must incorporate licensing and reimbursement procedures for intellectual property owners into their framework. Therefore, it is essential to properly get data, provide transparency, and preserve the origin of AI-generated content. Therefore, it is important to provide licensing information and appropriately remunerate artists.

According to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules of 2021, AI businesses have the option to appoint compliance officers who are responsible for supervising and implementing copyright protection measures. By doing this, it achieves a harmonious equilibrium between safeguarding the rights of copyright holders and promoting the advancement of AI research. Given India’s expanding user base of AI technologies, it is crucial to implement proactive policies and laws to effectively manage the integration of AI into the creative area. Hence, it is necessary to achieve a careful equilibrium between protecting the interests of human artists and embracing the promise of AI. Therefore, it is necessary to establish efficient legal, ethical, and regulatory structures that can utilize generative algorithms to maximize the potential of generative AI, while still maintaining intellectual property principles. This will promote innovation and creativity.

Protecting IPR for AI-generated material poses many obstacles that need meticulous examination and adjustment within legal and regulatory structures. The current copyright legislation provides safeguards for the intellectual property of artists, but it faces new challenges due to the emergence of AI. The expansion of Generative AI technology requires a reevaluation of copyright legislation to strike a balance between the rapid generation of content and the protection of the IPR of original artists. By using proactive procedures, it guarantees a more ethical and legally

compliant approach to utilizing this transformative technology. Copyright rules must adapt to changing technologies and processes, particularly with the growing use of AI, which presents potential infringement risks. Policy-makers, legal experts, and stakeholders should modify copyright rules to tackle the distinctive difficulties posed by AI. This should be done while considering the interests of authors, consumers, and developers, and while supporting innovation that respects copyright values.

Ultimately, to address the unique challenges posed by generative AI in content production across several fields, it is crucial for policymakers, legal professionals, and individuals with vested interests to contemplate modifying the Copyright Acts. In the future, as we navigate the copyright environment, it will be crucial to find a middle ground between promoting innovation and respecting copyright norms. This will include managing the conflicting interests of AI tool owners, customers, and inventors.

NOTES

1. Lian, A. The Legal Implications of AI-Generated Content in Copyright Law. *INDIAai*. Accessed June 21, 2024. <https://indiaai.gov.in/article/the-legal-implications-of-ai-generated-content-in-copyright-law>↵
2. Mannary, N.R., S. Shahriar, and T. Wang. 2023. Artificial Intelligence in Developing Countries: The Impact of Generative Artificial Intelligence (AI) Technologies for Development. *Information Development* (September). www.researchgate.net/publication/373934747↵
3. IP and Frontier Technologies. Generative AI Navigating Intellectual Property. *WIPO*. Accessed June 21, 2024. www.wipo.int/about-ip/en/frontier_technologies/pdf/generative-ai-factsheet.pdf↵
4. Jain, S., and A. Agrawal. 2024. Indian Copyright Law and Generative AI. *Lexology*. Saikrishna & Associates. Accessed June 21, 2024. www.lexology.com/library/detail.aspx?g=d32f6864-e555-4acd-ae76-cfd9d558f44e↵
5. Agrawal, A. 2024. Indian Copyright Law and Generative AI—Part 1. *IPRMentLaw*. Accessed June 21, 2024.

<https://iprmentlaw.com/2024/05/04/indian-copyright-law-and-generative-ai-part-1/>

6. Moitra, K., and K. Vallabh. 2021. Copyright in Works Created by Artificial Intelligence: Issues and Perspectives. *Lexology*. Bharucha & Partners. Accessed June 21, 2024. www.lexology.com/library/detail.aspx?g=4513277a-6571-40f1-923d-c09ec5366fdd
7. The Copyright Act, 1957 (14 of 1957).
8. The Patents Act, 1970 (39 of 1970).
9. The Information Technology Act, 2000 (Act No. 21 of 2000).
10. Sanyal, N., S. Mishra, and N. Sharma. 2024. Intersection of Intellectual Property Rights and AI-Generated Works—Part I. *Bar and Bench*. www.barandbench.com/law-firms/view-point/intersection-intellectual-property-rights-ai-generated-works-part-i
11. The Digital Personal Data Protection Act, 2023 (No. 22 of 2023).
12. News. 2023. CHATGPT, AI Content: All about Legal Challenges Pertaining to Copyright under Indian Law. *Business Today*. www.businesstoday.in/technology/news/story/chatgpt-ai-content-all-about-legal-challenges-pertaining-to-copyright-under-indian-law-399393-2023-09-22
13. NITI Ayog. (2018). National Strategy for Artificial Intelligence. *NITI*. www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf
14. Roy, A. 2024. New AI Law Will Guard Rights of Content Creators: Experts. *The Economic Times*. https://economictimes.indiatimes.com/tech/technology/new-ai-law-will-guard-rights-of-content-creators-experts/articleshow/109142779.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst
15. Anon. 2024. Existing IPR Regime Well-Equipped to Protect AI Generated Works, No Need to Create Separate Category of Rights. *Press Information Bureau*. Ministry of Commerce & Industry. <https://pib.gov.in/PressReleasePage.aspx?PRID=2004715>

16. Ojha, S. 2023. Who Owns AI-Generated Works? Here's What the Laws Say on Copyright Issue. *India Today*. www.indiatoday.in/law/story/chatgpt-ai-generated-content-copyright-ownership-complexities-india-2439165-2023-09-22
17. Works in which copyright subsists, Copyright Act of 1957.
18. Meaning of Copyright, Copyright Act of 1957.
19. First owner of copyright, Copyright Act of 1957.
20. Verma, A. 2023. The Copyright Problem with Emerging Generative AI. *SSRN* (August 10): 1–17. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4537389
21. Rana, L. 2019. Artificial Intelligence and Copyright—The Authorship. *Mondaq*. S.S. Rana & Co. Advocates. www.mondaq.com/india/copyright/876800/artificial-intelligence-and-copyright--the-authorship
22. “in relation to any literary, dramatic, musical or artistic work which is computer-generated, the person who causes the work to be created.” Copyright Act, 1957.
23. When copyright infringed, Copyright Act, 1957.
24. Chavi.2023. The Nexus Between Artificial Intelligence and Copyright Act. *IIPRD*. Accessed June 15, 2024. www.iiprd.com/exploring-the-nexus-between-artificial-intelligence-and-copyright-act-1956/
25. AIR 2018 DEL 1483.
26. Chabra, H., and G. Pandey. 2024. Balancing Indian Copyright Law with AI-Generated Content: The “Significant Human Input” Approach. *IJLT*. Accessed June 21, 2024. www.ijlt.in/post/balancing-indian-copyright-law-with-ai-generated-content-the-significant-human-input-approach
27. Wadhwani, S. 2023. AI-Generated Content Cannot Be Copyrighted. *Spiceworks Inc.* www.spiceworks.com/tech/artificial-intelligence/news/us-copyright-law-ai-generated-content/
28. Corrigendum EU. AI Act Amendment 2024 of 2021. www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf

29. Raj, N., and M. Banta. 2023. Legal Implications of AI-Created Works in India. *Mondaq*. Khurana and Khurana. Accessed June 21, 2024. www.mondaq.com/india/copyright/1348418/legal-implications-of-ai-created-works-in-india
30. AIR 2008 SC 809.
31. Chesterman, S. 2024. Good Models Borrow, Great Models Steal: Intellectual Property Rights and Generative AI. *Policy and Society* (February 12). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4590006 or [10.1093/polsoc/puae006](https://doi.org/10.1093/polsoc/puae006)
32. [1916] 2 Ch 601.
33. Pathak, P. 2024. Challenges in Protecting Copyright for AI-Generated Content. *IJCRT* 12, no. 2 (February). <https://ijcrt.org/papers/IJCRT2402092.pdf>
34. Certain acts not to be infringement of copyright, Copyright Act, 1957.
35. Civil remedies for infringement of copyright, Copyright Act, 1957.
36. Offence of infringement of copyright or other rights conferred by this Act, Copyright Act, 1957.
37. Aggrawal, A. 2024. Indian Copyright Law and Generative AI: Part 2— Transformative and Extractive Use. *IPRMentLaw*. Accessed June 21, 2024. <https://iprmentlaw.com/2024/05/29/indian-copyright-law-and-generative-ai-part-2-transformative-and-extractive-use/>
38. No. 16-15469 (9th Cir. 2018).
39. 575 F.Supp.2d 513 (SDNY 2008).
40. Agarwal, S., and C. Bhardwaj. 2021. The Dilemma of Copyright Law and Artificial Intelligence in India. *SSRN Electronic Journal*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3818280 or <https://dx.doi.org/10.2139/ssrn.3818280>
41. Admin. 2024. Conundrum of Copyright over the Content Created by AI Tool. *IIPRD*. www.iiprd.com/conundrum-of-copyright-over-the-content-created-by-ai/ or <https://doi.org/10.56042/jipr.v29i2.1205>
42. Vishnu, S. 2024. Navigating the Grey Area: Copyright Implications of AI Generated Content. *Journal of Intellectual Property Rights* 29, no. 2. <https://or.niscpr.res.in/index.php/JIPR/article/view/1205>

43. Prakash, S. 2023. Copyright Ownership of AI Generated Content in India. *Mondaq*. Sujata Chaudhri IP Attorneys. www.mondaq.com/india/copyright/1295978/copyright-ownership-of-ai-generated-content-in-india
44. Divyaprashiktha, T.K., and G.S. Vikashani. 2024. Development of AI and Issues Pertaining to Copyright—A Critical Study. *International Journal of Scientific Research in Engineering and Management* 08, no. 01 (January 15): 1–13. <http://dx.doi.org/10.55041/IJSREM27618>
45. Jain, M., and A. Sharma. 2023. Untangling the Mysteries of AI-Generated Content and Copyright. *Live Law*. www.livelaw.in/articles/untangling-the-mysteries-of-ai-generated-content-and-copyright-243749
46. Banerjee, K. 2024. AI and Copyright Infringement: Can AI Be a Creator? *Mondaq*. www.mondaq.com/india/copyright/1428552/ai-and-copyright-infringement-can-ai-be-a-creator
47. CS(COMM) 652/2023 and I.A. 18237/2023-18243/2023.
48. Dairy no. 13646/2020-CO/A; RoC no. A-135120/2020.
49. 2008 (153) DLT 548.
50. 1996(38)DRJ81.
51. Anon. 2023c. Artificial Intelligence and Copyright in India. *Bytescare Blog*. <https://bytescare.com/blog/artificial-intelligence-and-copyright-in-india>
52. 2014 (60) PTC 121 (Del).
53. AIR1989DELHI63.
54. AIR 2007 SUPREME COURT 1349.
55. 1978 AIR 1613.
56. 78(1999)DLT285.
57. 15-CV-04062-LHK.
58. Global Patent Filing. 2024. AI-Generated Creative Works: Challenges & Opportunities in Copyright Domain. *GPF*. Global Patent Filing. Accessed June 18. www.globalpatentfiling.com/blog/AI-Generated-Creative-Works-Challenges-Opportunities-in-Copyright-Domain

59. Sharma, S., and N. Dadhich. 2023. Guardians of Genius: Securing Tomorrow's Generative AI via Copyright Protection. *Cyril Amarchand Blogs*. Accessed June 21, 2024. <https://corporate.cyrilamarchandblogs.com/2023/10/guardians-of-genius-securing-tomorrows-generative-ai-via-copyright-protection/>
50. Wukhari, S., S.U. Hassan, and Y. Aleem. 2024. Impact Of Artificial Intelligence on Copyright Law: Challenges and Prospects. *JLSS* 5, no. 4: 647–656. <http://dx.doi.org/10.52279/jlss.05.04.647656>
51. 16 PTC 329 (Kerala).
52. Jain, A., D. Gupta, and K. Mahan. 2024. Harnessing the Power of AI Revolutionizing Copyright Issues. *Chase India*. www.chase-india.com/media/kakhaqja/harnessing-the-power-of-ai.pdf
53. Anon. 2023. A Brief Look at the Copyright Issues Raised by Generative AI. *IKIGAI Law*. Accessed June 23, 2024. www.ikigailaw.com/article/11/a-brief-look-at-the-copyright-issues-raised-by-generative-ai#_ftn9
54. Personal Life and Liberty, Constitution of India, 1950.
55. Department Related Parliamentary Standing Committee on Commerce. July 23, 2021. 161st Report on Review of the Intellectual Property Rights Regime in India. *Parliament of India Rajya Sabha*. https://files.lbr.cloud/public/2021-07/161_2021_7_15.pdf
56. ASCI and Khaitan & Co. Generative AI and Advertising. *ASCI Online*. Accessed June 20, 2024. www.ascionline.in/wp-content/uploads/2023/08/Kco-WhitePaper-GenAI-D11-DIGITAL_LowRes.pdf
57. Srivastava, A., and R. Jalan. 2023. The Copyright Conundrum: Exploring the Impact of Generative AI on Intellectual Property Rights. *GSLC*. Accessed June 17, 2024. <https://gslc.in/the-copyright-conundrum-exploring-the-impact-of-generative-ai-on-intellectual-property-rights/>

17 AI's Role in Cybersecurity

Balancing Innovation with Privacy and Rights

*Rahul Joshi, Suman Kumari, Krishna Pandey,
Laith Farhan, and Amr Badr*

DOI: [10.1201/9781003668657-17](https://doi.org/10.1201/9781003668657-17)

17.1 INTRODUCTION

Information and communication technologies are developing at a rapid pace, ushering in the digital age. The internet has been crucial in people's and businesses' shift to digital solutions and technologies. The digital revolution is exciting because of the advancement of cutting-edge technologies like artificial intelligence (AI). AI is intelligence manifested through computers; these machines can sense their environment and react accordingly to accomplish tasks. The best term to use when machines imitate human behavior to learn or solve problems is AI [1, 2]. Neural networks, fuzzy systems, machine learning (ML), deep learning (DL), computer vision, natural language processing (NLP), cognitive computing, evolutionary computation, and algorithms inspired by nature are some of the subdomains of AI [3]. AI is seen as a critical component of society's digital revolution. It is expected that future applications will bring about significant modifications. Autonomous vehicles, Alexa, robotic advisers,

traditional bots, OTT platform suggestions, email spam filters, and more are a few real-world AI applications. The three primary subcategories of AI are artificial narrow intelligence, artificial general intelligence, and artificial super intelligence. It can also be categorized according to its value to the organization.

The five categories of AI are as follows: analytic (supplier risk assessment, sensitivity analysis), functional (robots, Internet of Things (IoT) solutions), text (text recognition, speech recognition), interactive (chatbots, chatbots), and visual (computer vision, augmented reality) [4]. Not only is AI a smart revolution in cybersecurity, but it also offers many possibilities. Systems with AI capabilities may be taught for safe human-machine communication [5]. AI-based systems may be taught to identify new threats, provide threat warnings, and protect data critical to an enterprise. To ensure that technology is a friend rather than an adversary, it is necessary to advocate implementing specific ethical practices in AI [6]. Cybersecurity solutions powered by AI provide various systems, data, and application capabilities. Security for networks, clouds, malware, IoT, etc., is a subset of system security. Data security is linked to danger detection, new attack identification, and cloud information safety [7]. AI in cybersecurity aids organizations in cost reduction, threat response speed, and breach reaction.

AI has three fundamental characteristics within the realm of cybersecurity:

1. *Detection*: In modern times, corporations use artificial intelligence to identify potential threats accurately. Over half of firms have integrated AI-powered cybersecurity technologies. AI-based detection is a distinct capacity of behavior analysis to monitor and identify abnormal traffic consistently. The predominant use involves ML and DL methods.
2. *Prediction*: The prediction function has the second-highest utilization rate. According to the research findings, 35% of firms substantially utilize AI to forecast and anticipate cyber risks. AI facilitates accurate predictions by analyzing vast data through systematic training. AI

enables the process of automation. Assists enterprises in identifying vulnerabilities and fortifying their network to defend against cyber-attacks.

3. *Response*: AI is currently evolving in its ability to respond to threats. Approximately 18% of firms use AI to combat cyber-attacks. A novel protection system has been built using automation to identify threats in real-time scenarios.

The purpose of incorporating cybersecurity is not just to protect the boundaries but to determine the worth of the assets at risk, pinpoint the threat's location, and prioritize the measures to establish a comprehensive defensive system that guarantees uninterrupted service. Critical infrastructures like the supply of water, petrochemical sites, power plants, and transportation infrastructure systems have been the focus of cyberattacks over the past few years. The goal of these attacks is to take down power and obtain sensitive data without authorization [8]. The integration of physical equipment and networked software and sensors has increased the complexity of the technological system, in addition to the growing risks. The International Electrotechnical Commission (IEC) suggests that the best way to address the complexity and susceptibility of cyber systems is to adopt comprehensive strategies that include processes, people, and technology, such as AI. Nevertheless, integrating AI-driven technology with cybersecurity has advantages and disadvantages. While one aspect of AI supports the ability of a system to withstand and recover from cyber-attacks, another aspect strengthens the prevention of unauthorized access. The research on AI-based applications is quite varied, yet the security worry outweighs the preventive measures. The study on cyber defense revealed an expectation to use AI to automate the system for analyzing large amounts of data and conducting necessary inspections to detect cyber threats [9].

17.2 EVOLUTION OF CYBERSECURITY AND THREATS

Cybersecurity encompasses a range of definitions and motivations that differ across various organizations and global standards. Cybersecurity, as defined by the ISO/IEC 27032:2023(en) standard, pertains to protecting internet-connected systems, including hardware, software, and data, from cyber assaults, harm, or unauthorized entry. Cybersecurity refers to various methods, technologies, and measures to safeguard systems, networks, and data against potential online dangers. To protect systems, data, devices, programs, and networks from cyberattacks, cybersecurity uses technology, protocols, and techniques [10]. Reducing the likelihood of damaging attacks and offering sufficient defense against them are the objectives. Another often-used term is information security, which encompasses protecting all digital and physical data resources. Over the last several years, there have been many prominent instances of cyber-attacks [11]. In 2017, the Equifax hack exposed the personal data of 143 million users, including information such as date of birth, address, and passwords. In 2018, cybercriminals successfully breached Marriott International's server and illicitly duplicated data from around 500 million people. Cybersecurity encompasses several sectors, from commerce to mobile computing [12]. [Figure 17.1](#) shows the different types of cybersecurity available in the industry.

DIFFERENT TYPES OF CYBER SECURITY

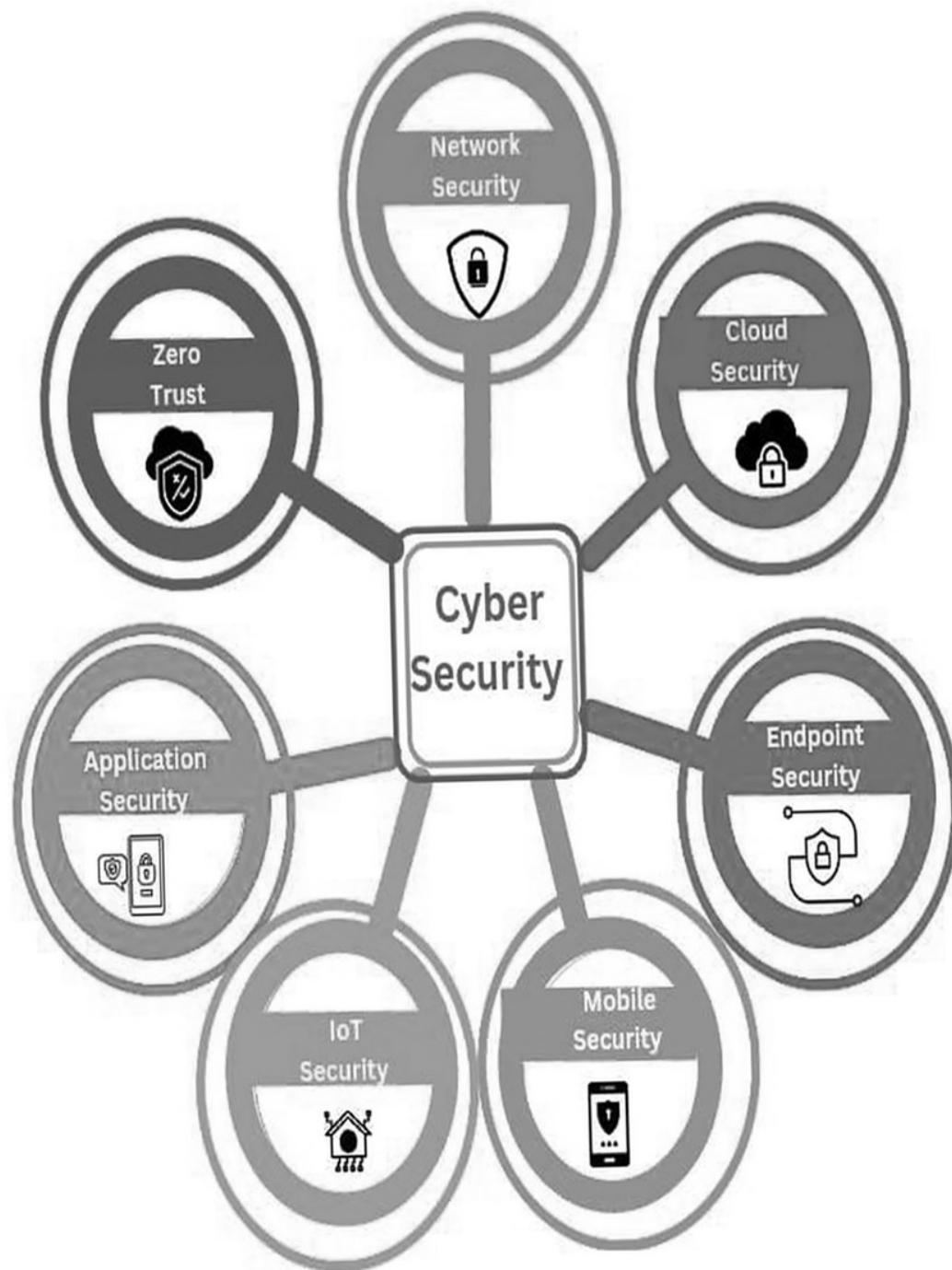


FIGURE 17.1 Different types of cybersecurity available in the industry. [🔗](#)

Application Safety: The objective is to safeguard the device or secure the program. Adequate security measures are implemented during the design phase before the deployment of the device, for some instances of application security, robust passwords. Occasionally, the user must undergo two-step authentication, answer security questions, and take other similar measures to ensure proper user authentication [[13](#)].

Network Security: Diverse methodologies are used to safeguard against malevolent software, unauthorized access to data, and similar security breaches. A firewall is a defensive barrier between a client's network and the external environment. It can obstruct or permit network traffic based on security configurations. Email security is crucial for safeguarding the network due to the prevalence of phishing attacks [[14](#)]. Consequently, it is necessary to develop tools to analyze both incoming and outgoing emails to monitor and prevent phishing attempts.

Operational Security: This pertains to the management of risks associated with data resources. Authorized users within an organization need proper instruction on the most effective methods for storing and safeguarding personal and professional data inside a network. Typically, organizations hire risk management specialists to ensure the availability of a contingency plan.

End-User Security: Providing end users in an organization with education could help protect the system from unintentional infections. Among other things, they can connect unknown USB devices or delete dubious email attachments.

Information Security: This aims to guarantee the preservation of information integrity and secrecy while it is stored and throughout its transfer [[15](#)].

Calamity Recovery: Every organization must establish disaster recovery rules to restore its operations and data effectively. Maintaining the organization's operational capability at the same level as before is

necessary. Organizations often develop a business continuity strategy to continue functioning even if specific resources are unavailable [[16](#)].

Cloud Security: Today, it is common for individuals to keep the bulk of their data in the cloud. Instead of using conventional hard drives for storage, a variety of online platforms are used, including Microsoft's OneDrive, Google Drive, as well as Apple iCloud, and others. This includes safeguarding the end user interface, mitigating the risk of human error that might compromise the network, implementing a robust backup plan, and ensuring the safety of information storage, among other factors [[17](#)].

17.3 CYBER THREATS

This malevolent behavior often involves digital data theft, destruction, or disruption. Attackers may use harmful code to compromise other computers. Thus, it is essential to search for and address cyber threats. Every organization must use cyber-attack prevention methods [[18](#), [19](#)].

The field of cybersecurity primarily faces the following threats:

1. *Cybercrime:* It encompasses actions by a group or a person to cause disruption or financial benefit.
2. *Cyberterrorism:* It is the deliberate use of electronic means to cause harm or instill fear to disrupt or generate panic inside a system.
3. *Cyber-Attack:* Occasionally, it is associated with acquiring information, often motivated by political factors.

The following are a few critical cybersecurity hazards that every cybersecurity practitioner must know:

1. *Malware:* It is a term for harmful software that is a pervasive cyber threat. Historically, hackers would damage users' systems by luring them into clicking on harmful email attachments, links, and similar elements. Once activated, malware can disrupt network connectivity, install

counterfeit software, and significantly impair system operations. Malicious software belonging to this category comprises Trojan horses, ransomware, adware, spyware, and botnets [20].

2. *Phishing*: It refers to the deceptive practice of using email to fraudulently obtain sensitive information from users, including passwords, login credentials, and credit card data.
3. *SQL*: Known as Structured Query Language, is used by hackers to infiltrate a server by inserting malicious code. The goal is to gain control over the database, potentially exposing confidential data stored within [21].
4. *Emotet*: It is a highly destructive form of malware, a significant financial threat. Certified Information Systems Auditor (CISA) categorizes it as an advanced modular banking Trojan that primarily functions as a downloader.
5. *Denial of Service*: The computer system experiences excessive traffic and cannot meet the legal requirements [22]. The network becomes unfeasible in such circumstances, and any organization cannot perform vital activities.
6. *Man in the Middle Attack*: This cyber danger involves a hacker intercepting the communication between two parties. As a result, traffic disruptions might occur, allowing cyber thieves to pilfer and isolate information. Primarily, this assault happens while using an unprotected Wi-Fi network, whereby the hacker may interrupt data or use malware.
7. *Recent Cybersecurity Menaces*: In 2019, there was an observed occurrence of the Dridex Malware assault, which had a global impact on infrastructure, public entities, government systems, and several industries. The UK Cyber Security Center recommends patching devices and ensuring that antivirus software is updated and activated. In 2020, Romance Scams occurred, in which hackers used chat rooms, dating sites, and apps. It impacts a significant portion of the population and results in substantial financial losses.

17.4 ADVANCEMENT OF AI IN CYBERSECURITY

The advancement of technology enables the fast development of AI in the 21st century. Various decades of study have led to the rapid development of AI. Alan Turing provided the foundation in the mid-20th century by introducing the Turing test to assess machine intelligence [23]. John McCarthy further developed this idea by introducing the notion of generality in AI [24]. Although it was not until later in the century that conceptual ideas were transformed into actual implementations, these advancements have consistently propelled progress. In the 1990s, the technology field saw a significant increase in computing power and hopeful advancements in data collection and processing technologies. ML has increasingly gained prominence and prevalence across several fields of application. The rapid advancement of internet technology, along with the exponential growth in computer power, has shown the immense potential of neural network architecture for conducting complicated analysis [25]. In the late 1990s, the U.S. military recognized the importance of AI in safeguarding the National Information Infrastructure. The paper addressed the border security aspect of various threats, including internal and transnational attacks. It also highlighted the interconnectedness of military, government, and civilian information systems. The paper stressed the importance of implementing enhanced measures to adapt to the ever-changing technology, as this is vital for ensuring the continued relevance and effectiveness of the National Information Infrastructure. Several applications for AI used in cybersecurity are shown in [Figure 17.2](#).

10 APPLICATIONS OF AI IN CYBER SECURITY

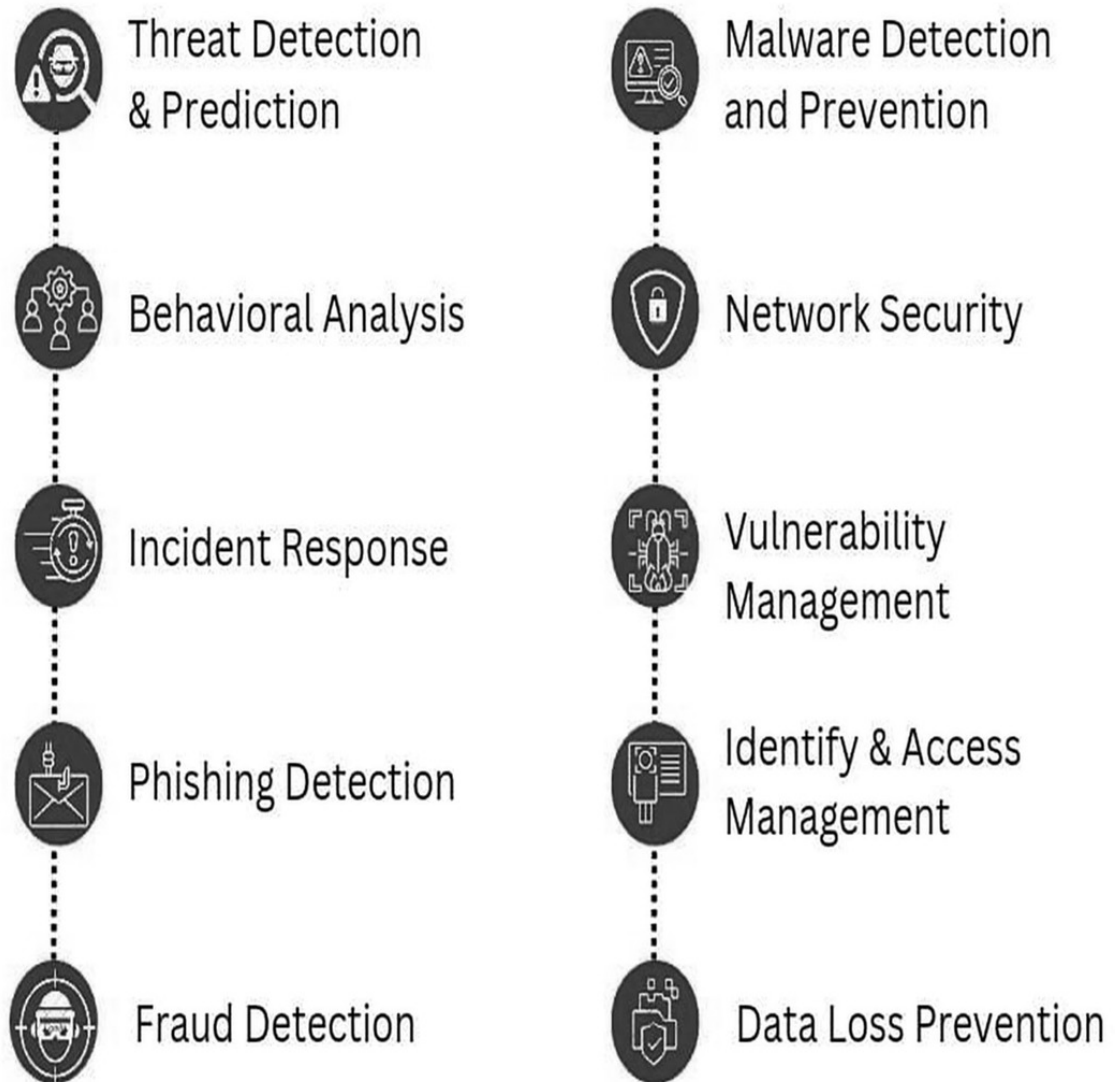


FIGURE 17.2 Applications of AI in cybersecurity. [↗](#)

In the early stages of AI development, the National Infrastructure Protection Center planned to integrate Sandia National Laboratories' AI-based software into the Arms Control Treaty Monitoring System. In this system, an embedded system was designed using Intelligent Agent Sensors to identify and respond to attacks on vital infrastructures [26]. In the 21st century, development significantly transitioned from a stable state to a fast

acceleration due to the rise of Big Data and advanced AI models. Advancements in algorithms and the progress in hardware capabilities such as Graphics Processing Units [27] have empowered AI to analyze and learn from this overwhelming amount of data efficiently. Statistical approaches mainly influenced the first ideas of AI during that period. These algorithms are performed satisfactorily in security event management systems for network monitoring and anomaly avoidance.

The development of intrusion and prevention techniques included using packet filtering, pattern matching, and flagging records of occurrences to build an anomalous network intrusion system. Statistical models were used to observe substantial deviations from normal or anticipated behaviors. “To employ statistical preprocessing and neural network classifications to identify network threats and flaws, the researchers created the Hierarchical Intrusion Detection System along with the Generalized Anomaly and Fault Threshold System” [28]. Activity profiles generated and classified by multiple probability density functions are used in the HIDE model. These profiles are analyzed to identify typical activity patterns in the measured parameters. Despite the considerable potential, statistical models still had several limits. Several dynamic network traffic components are included in the system monitoring processes when working with cybersecurity data. Furthermore, the fluctuation of network traffic patterns over time might be challenging for statistical models that rely on previous data for predictive analysis or anomaly identification. Using established patterns and historical data in the detection process may also lead to inefficiencies in identifying previously unseen attacks, popularly known as zero-day attacks.

Furthermore, advanced models may require significant computing capacity, leading to resource-intensive tasks in extended and intricate network settings. The limitations of statistical models may be overcome by using more sophisticated network models driven by AI technology. These advanced models are better equipped to handle ambiguous connections, inadequate data in harmful activities, and unexpected patterns seen in cyber assaults. Bayesian networks were initially developed to handle complicated situations, and Bayesian ontologies were introduced to improve

cybersecurity measures [29]. Bayesian ontologies and probabilistic reasoning derived from Bayesian networks may effectively tackle cybersecurity concerns by facilitating semantic interoperability across diverse systems and promoting knowledge exchange. The uncertainty inherent in changing cybersecurity data, which is often noisy and irregular, may be effectively described using ontologies. Probabilistic ontologies enhance conventional ontologies by including terms that connect statistical patterns and probabilistic relationships within a specific domain. Probabilistic components with Web Ontology Language emphasize the equilibrium between expressiveness and traceability in conveying uncertainty. When it comes to cybersecurity, the process of discovering and assessing risks sometimes entails the interpretation of patterns and abnormalities within data. Probabilistic ontologies enhance this process by offering a structure for comprehending the probability of certain occurrences or actions, essential for identifying possible security breaches or assaults. Cybersecurity must capture and comprehend uncertainty to conduct security analysis. Traditional graph models cannot provide network defense mechanisms due to their inability to handle unforeseen vulnerabilities in real time. To solve these issues, the Bayesian network was used [30].

To protect against these ever-changing variations, a suggested approach called Adversary Courses of Action utilizes classical AI planning to create a series of exploits that map from the starting state to the end action, which includes the aims of an attacker [31]. The study examines the Behavioral Adversary Modeling System for simulating potential behaviors of insider threats. Hoffmann's METRIC-FF planner and the Planning Domain Definition Language are used to generate Courses of Action and a simplified blueprint of a document management system. This technology can replicate the rapid creation of complex threat models, which may help operators foresee vulnerabilities and harmful actions by insider threats.

In the early 2000s, rule-based systems were the most common method for identifying abnormal patterns associated with cyber risks. The architecture of the Sensing Engine and Intelligent Threat Prevention was

unveiled. It made the use of multiple rule bases from real-time traffic monitoring and expert knowledge systems possible. The Intelligent Threat Sensing Engine (ITSE) and the Intelligent Prevention Engine (IPE) were two of the engines included in the architecture [32]. AI agents enhanced interactions across different computing environments in an advanced World Wide Web interface computing [33].

Utilizing cyber signatures in mobile computing enables improved authentication and security. The authors also explored advanced authentication approaches, including intelligent trees and agent language processing, to provide online solid interaction mechanisms. AI rules based on genetic algorithms were created to detect network intrusions with an AI-rule-integrated intrusion detection system (IDS). These rules focus on capturing the temporal and geographical information of network traffic. This IDS incorporates procedures such as translating network connections and traffic patterns into rules using genetic algorithms. These methods include phases of chromosomal evolution, including selection, recombination, and mutation. A framework was developed to enhance the analysis of malware behavior threats by grouping known and undiscovered malware, including viruses, Trojan horses, and worms, as the effectiveness of the anti-virus scanner diminishes in detecting a wide range and a large amount of malware [34]. The proposed method efficiently computes and assigns vectors as behavioral patterns by using a vector space model based on ML. The framework uses a Malware Instruction Set to encrypt system calls and the arguments they provide using numerical identifiers while keeping an eye on system calls and activities. Although the method does not eliminate malware threats, it greatly enhances the ability to recognize and respond to new malware developments. Since the 2010s, AI has evolved significantly in cybersecurity, moving from rule-based approaches to more dynamic and adaptable learning tactics. Nevertheless, with the advancement of AI technology, adversarial AI has grown more adept at avoiding detection and executing complex assaults. A detection method based on Transformer models is developed to categorize traffic data and identify abnormalities in system logs. This mechanism's performance is superior to

that of current recurrent neural network (RNN) and long short-term memory (LSTM) models [35]. Transformer models are more successful in realizing and analyzing the complicated connection among extended patterns of sequential data than LSTM or RNN-based models. The self-attention approach decreases the computational expense and improves the detection of intricate anomaly patterns by capturing the relationships within the patterns.

17.5 INNOVATIVE AI TECHNIQUES IN CYBERSECURITY

AI, a vast field including advanced technology, consists of several subcategories designed to enhance cybersecurity applications. Utilizing AI-based algorithms may more effectively tackle the substantial issues posed by unprecedented and polymorphic threats in the realm of cyberspace compared to traditional ways.

17.5.1 ML MODELS

ML models allow computers to make judgments or predictions based on selected data without explicit programming. Performing data preprocessing before inputting it into the ML system is the first and crucial stage in enhancing AI's capabilities in cybersecurity. The process, which ranks the most important characteristics or determines feature significance, is frequently understandable and yields insights into the data. Understanding the decisions made by the model in these approaches in detail is crucial to being able to recognize abnormal behavior. Feature engineering is performed on the processed data and trained data to identify abnormalities or outliers in harmful patterns [36]. The basic version of ML is more computationally efficient than DL models. However, it needs the ability to effectively evaluate intricate patterns and lengthy sequences of data to detect abnormalities. Nevertheless, DL models are well recognized for their complexity and lack of interpretability, posing difficulties in explaining their conclusions. ML models may perform very well in areas where

transparency is essential [37]. To counter Distributed Denial-of-Service attacks on future IoT devices, a strong IDS named IntruDTree was suggested. This system utilizes ML algorithms [38]. The research highlighted ML, specifically tree-based models, which can acquire knowledge from preprocessed security datasets derived from IoT devices. Incorporating IntruDTree into the IDS decision-making process improves efficiency by using lower feature dimensions and prioritizing relevant characteristics according to their significance, reducing computing complexity. Tree-based models are more successful than other classic ML models than generalized intrusion detection methods. However, ML models are less successful than DL models in detecting changes in real-time as network traffic and data communication increase as a result of the incorporation of computing devices with communication capabilities.

17.5.2 DL AND NEURAL NETWORKS

DL has an input, hidden layer(s), and output layer. A critical understanding of brain neural networks underpins DL models. Modeling deep networks requires understanding how neurons interact to make proactive judgments about known and unknown events. Many hidden layers might be used to train artificial neural networks to adapt to complicated patterns and analyze relative data efficiently. The most popular DL architecture, convolutional neural network (CNN), was widely praised for its efficiency and exceptional success. The RNN uses loops to recover information lost in hidden layers owing to gradient descent [39], overcoming CNN's limitations. "Long Short-Term Memory (LSTM) networks and Autoencoders, which have cells to control information flow and decoders to reconstruct the output from the compressed input, are the most sophisticated RNN technology" [40]. Transformer models use the self-awareness technique to extract long-range relationships from data [41]. This method is especially effective for analyzing complex, sequential network traffic data patterns to spot advanced cyber threats. These DL models construct adept anomaly and IDS to discover complex patterns and abnormal activities. The

models can learn from big datasets to discover known and new abnormalities, categorize risks, and understand possible security breaches. Using Deep Neural Network (DNN) in a Software-Defined Networking (SDN), a flow-based IDS was created to manage network packet data [42]. SDN offers dynamic forwarding rule updates and global network awareness to detect unexpected threats. In-vehicle Controller Area Networks are a vital security risk as safety-critical cars integrate more computer equipment like Electronic Control Units. An unsupervised deep belief network-based IDS was developed to improve the Controller Area Network (CAN) detection ratio and solve the ECU-ECU message broadcast vulnerability. Classifying statistical patterns and mapping complicated non-linear linkages made the suggested DBN more successful than typical Artificial Neural Network-based (ANN) IDS, which uses predetermined attack patterns or specification-based techniques to discover anomalies. Traditional Machine Learning (ML) suffers from a mutation in massive data analysis. However, Deep Learning (DL) models can extract tiny perturbations from complex characteristics. Slight variations in attack patterns may make emerging IoT devices susceptible to the associated system, but DL can capture hidden dangerous patterns for self-learning and compression.

17.5.3 EXPERT SYSTEMS AND THEIR RELEVANCE

Expert systems mimic problem-solving and automated decision-making processes by combining expert human knowledge-based systems with rule-based inference engines. They are a subdomain of AI. Expert System's advising skills enhance the system's resilience and strength in cybersecurity. Rule-based IDS exemplify this notion by using pre-established rules to identify and monitor aberrant actions inside a network or system. Operators are promptly notified of any anomalies when traffic conforms to a rule or set of rules, enabling them to make timely judgments. The Rules and decision-tree-based IDS was introduced to categorize network traffic as malicious or benign. It makes use of decision trees and an embedded rule-driven expert system [43]. This fog computing framework is

based on a three-tier system. It makes use of a rule-based classifier with Java RIPPER (JRIP) and decision tree modeling with Reduced Error Pruning (REP) Tree. The outputs of these two concurrent processes are fed into the final classifier, which is constructed using Forest PA. The results are further refined by this final classifier, which produces a more accurate depiction of intricate data patterns. The Pattern-based Intrusion Detection Engine is recommended to be used with a rule-based system called ExS Shell, a Java-based IDS. Its objective is to assess user behavior within the system [44]. ExS, in the context of industrial IoT applications, may be advantageous for cybersecurity management due to its ability to monitor the extensive volume of network traffic effectively. The suggested hybrid architecture uses rule-based feature selection to effectively decrease the number of training features and enhance the performance of the DL-based classifier [45]. Expert systems are essential in cybersecurity as they utilize expert knowledge and rule-based reasoning to improve security operations, identify threats, respond to incidents, and manage compliance. They function as vital instruments for firms aiming to maintain strong and aggressive cybersecurity measures.

17.5.4 NATURAL LANGUAGE PROCESSING

Using textual analysis, NLP models strive to identify and separate threat alerts and malware code insertion. The process of parsing and extracting pertinent information from security logs [46] assists in identifying events. Current ML models in this field must be more accurate because they depend on manually identifying essential characteristics. DL models also encounter similar difficulties as they need embedded words to properly represent the content of an email exchange. NLP may be used with ML and DL models to effectively classify email content and accurately detect phishing assaults. The cognitive study conducted by ML models informed the use of NLP to construct domain ontologies via a two-step process: the symmetry stage and the modified machine stage [47]. A cybersecurity analyzer prototype was developed using ontologies to provide a

comprehensive architecture with critical components and an implementation initializer. The formation of the structure occurs in four distinct phases. Initially, the input is provided as a document, which then undergoes cognitive analysis based on ML. Subsequently, the data is stored and visualized, culminating in the creation of REST APIs that enable the integration of all these functionalities. The cyber analyzer can effectively examine cyber documents for forensic purposes and identify hostile entities using NLP techniques. Security analysts often have challenges finding internet material relevant to cyber threats promptly due to the vast amount of web content and open-source texts. To resolve this problem, a suggested solution is to develop an automated system that utilizes a naturally embedded approach called Doc2Vec [48] to extract hazardous cyber material from publicly accessible web data. The integrated technique uses a natural language processor to extract cybersecurity-specific text from feed data selectively. NLP minimizes the need for manual involvement during routine operations by examining unorganized data such as security content, operation logs, and threat information. It then offers significant insights to cybersecurity experts, enabling them to take proactive measures.

17.6 AI-POWERED TOOLS FOR ENHANCING CYBERSECURITY

Cyber thieves used AI in deep locker malware, resulting in severe consequences for several enterprises. This virus can circumvent robust cybersecurity solutions using AI models to infiltrate networks via facial and voice recognition techniques [49]. This event has sparked interest in the integration of AI into the field of cybersecurity. Hence, to safeguard networks from such types of assaults, AI for cybersecurity must be considered a must now. Nowadays, all organizations, including startups, invest significant resources in developing AI-powered systems. These systems can evaluate a vast amount of data to assist cybersecurity experts in identifying possible threats. As a result, this trigger implements preventive

actions to address them. Various methods now use AI to enhance cybersecurity. These tools are very beneficial in achieving optimal security.

Some of the widely used tools include:

Vectra Cognito: This platform can identify hackers in real-time situations by using AI. This program is designed to automate the process of identifying and categorizing threats. Cognito uses a behavioral detection algorithm, logs, network data, and cloud events to find hidden attackers within IoT devices and workloads.

The Cognito platform consists of two components:

1. *Cognito Detect*: It is used to locate hidden tracking equipment. Using DL, ML, and data science, this function is performed in real time.
2. *Cognito Recall*: This tool assists in identifying vulnerabilities present in past data. The presence of actionable context accelerates the investigation of threats.

Symantec's Targeted Attack Analytics (TAA): Symantec Security undertook the development of this tool. The goal is to uncover covert and focused assault. AI and ML enhanced the efficiency and effectiveness of the processes, knowledge base, and capacities. The purpose of this assault was to get access to the operational network of energy organizations.

Sophos's Intercept X Tool: Sophos is a British company that specializes in hardware and software safety. Intercept X utilizes a DL neural network, which operates similarly to the human brain. Before completing a folder, X attempts to get many attributes from the folder and then conducts a thorough investigation. Ultimately, a determination is reached on whether a folder is authentic or malicious. The execution of this operation is completed within a time frame of 20 milliseconds.

Darktrace Antigena: Darktrace is an effective self-defense method. Antigena can enhance the essential capabilities of Darktrace by

identifying and replicating the operations of digital antibodies. Consequently, they can disable threats and infections. The operation is executed autonomously, without human intervention, using data on previous assaults, legislation, and other relevant information.

IBM QRadar Advisor: It utilizes IBM Watson technology. AI is used to detect any vulnerabilities that cyberattacks may exploit automatically. QRadar Advisor utilizes cognitive thinking to provide valuable insights and expedite the response process. This tool enables security professionals to assess dangers and minimize the likelihood of overlooking them. Notable characteristics of this tool include automated event analysis, Identification of high-priority risks, Provision of sophisticated justifications, User insights, and Insights for important assets.

17.7 PROSPECTS AND ADVANCEMENT

The ability to process vast amounts of data and generate practical insights has brought about a revolutionary period for cybersecurity. This technical innovation improves the ability to adapt and effectively protect against ever-evolving and complex cyber threats. In today's interconnected digital world, AI's role in safeguarding digital assets and maintaining privacy will become more and more important as it continues to progress. The subsequent sections systematically delineate the contributions of different AI strategies in bolstering cybersecurity measures, elucidating their development and methodologies.

17.7.1 ANOMALY DETECTION

Aberration from typical conduct, as determined by the establishments and resilience prerequisites, is recognized and alleviated by using various statistical models. Utilizing these models enables security staff to promptly make effective decisions by identifying patterns in the assault vectors. Nonetheless, the model is often used in concert with other approaches to

precisely determine the actual nature of the threat and offer preemptive analysis. An anomaly detection approach using a Gaussian Mixture and a decision tree model [50] was presented to incorporate network misuse detection. The model takes into account any discrepancies in the set of data or logs that may lead to inaccurate or, worse, a high false rate in the detection process. The decision tree model utilizes its most favorable qualities to determine the optimal choice between the typical pattern and the predefined attack patterns. This approach assumes that observations are derived from several Gaussian distributions with parameters that are not preset. The expectation maximization technique is used to estimate these parameters. This approach effectively detects attacks that exhibit similarities to standard distribution patterns. The hybrid model categorizes the regular data pattern into various subsets to gain a deeper understanding and analyze the regular pattern more effectively, aiming to identify patterns associated with attacks. The Gaussian Mixture Model-based (GMM) detection mechanism can enhance the system by addressing the limitations of existing techniques in adequately addressing the vulnerability of the AMI. It eliminates the need for external information and a predefined threshold [51].

17.7.2 SIGNATURE-BASED DETECTION

In the literature, signatures are explored in the detection mechanism to identify significant aspects of data patterns, including time dependency and time urgency [52]. Signature-based models are divided into two approaches. They either use a ruleset to protect against unknown or undefined behaviors in the network, or they use patterns to identify anomalous behavior instances from typical traffic [53]. Rule-based models use a predetermined set of criteria to identify the attack and standard patterns in data. Commonly used tools for this type of detection include Snort, Suricata, Zeek, and others. In addition to these open-source IDS tools, several more rule-based methods have been created. A framework to detect risks in cross-country product pipeline networks based on fuzzy rules is one such method. To

handle imprecise, uncertain, or subjective data and provide advanced and dynamic risk investigation in an intricate and changeable setting, such as a pipeline system with limited data, the model combines the Fuzzy Rule Base methodology with Grey Relations Theory. Ransomware, phishing, and business email exploitation have all shown increases in frequency during pandemics. To lessen the impact of potential enemies and threats, data mining techniques and rule-based fuzzy logic are used [54]. In pattern-matching techniques, the deviation from historical and standard data is measured. It is difficult to identify even the smallest departure from the right pattern type. However, it has been accomplished with a variety of models, including absolute median deviation, which improves both the detection and the prevention procedure.

17.7.3 CLOUD SECURITY AND ENCRYPTION

A framework called EEIBDM is presented for energy-efficient industry-scale data management in an IoT environment [55]. This framework facilitates the creation of a digital twin system that incorporates a virtual representation of the Industrial IoT and utilizes advanced techniques such as reinforcement learning and federated learning to enhance the security of cloud-based data management for big data in the industrial sector. The encryption process was enhanced by including DNA-based Huffman coding, which introduced a more intricate and less foreseeable approach to encoding data, drawing inspiration from the resilience shown in biological systems [56].

17.7.4 INCIDENT RESPONSE AND MITIGATION

Techniques for security monitoring and detection help identify unwanted occurrences. Assessing the discovered threat and addressing the vulnerability are the next steps. Modern technological infrastructures' detection techniques, preventive measures, and attack response mechanisms have benefited greatly from AI. The growing interest and advancement of

automated infrastructures give rise to security problems, which can be efficiently tackled by AI models such as statistical decision trees for making optimal decisions or Neural Network (NN) models for automatically resolving anomalous occurrences once detected. By continuously monitoring and analyzing events from devices on the endpoint to identify potentially detrimental behaviors or identified infiltration, AI can enhance endpoint security [57]. By quickly identifying and removing flaws at the device level, this protection method lessens the threat's influence on the network as a whole. DL models are utilized in wireless sensor networks to detect potential threats. This is followed by a prevention mechanism based on CNNs that effectively prevents sensor attacks, achieving high accuracy [58].

17.7.5 THREAT INTELLIGENCE AND DECEPTION TECHNOLOGY

Modern security management systems often go beyond active protection measures and use proactive methods. These systems might identify susceptible areas in specific parts of the infrastructure or monitoring devices within the system to monitor cybercriminals' activity or data flow. "Honey Pot" is the term for the strategy of luring enemies, which helps system analysts gather important data about attack signatures and attack methods used by attackers. One possible use for Honey Pot's research is to model the IDS or even divert the attack from important targets [59]. By examining this data, security analysts and forensic specialists can identify and separate new attack patterns, tactics for injecting malware, and trends in exploitation [60].

17.8 OBSTACLES AND CONSTRAINTS

AI offers a multitude of advantages to cybersecurity systems. However, certain obstacles still prevent it from being considered a crucial security tool [61].

Here are some of the factors:

1. *Cybercriminals and AI*: This technology can be used for both defensive and offensive purposes. Regulations govern cybersecurity, whereas hackers enjoy autonomy. The decreasing cost of developing programs has made it easier for hackers to manipulate technology for malicious intentions. Cybercriminals are leveraging AI technology to carry out rapid, precise, effective, and destructive hacking activities, introducing new and harmful attacks. Hackers meticulously analyze and enhance their malicious software to overcome AI-powered security measures, enabling them to carry out increasingly sophisticated attacks.
2. *Sources of Information*: Organizations must allocate significant financial, temporal, and human resources to meet security needs. The development and sustainability of AI systems depend on factors such as memory requirements, training programs, computational capabilities, and access to information.
3. *Collection of Data*: These are essential for training AI models. Cybersecurity experts must be proficient in handling diverse datasets, including malware, abnormalities, and harmful code. Organizations may need more resources and time when obtaining accurate datasets.
4. *High Adoption Barriers*: Organizations must provide enough financial resources to build and sustain AI systems. As technology advances, the costs of building excellent servers reduce, making them more affordable. At that point, the effectiveness of automating security increases and becomes a need for enterprises to operate on the cloud. Key obstacles to the adoption and implementation of this technology include the challenges in acquiring skilled personnel and the complexities associated with data management and the use of AI tools.
5. *Neural Fuzzing*: Vulnerability identification involves the assessment of a large volume of random input data. This procedure takes place inside the software. While fuzzing may provide significant advantages, hackers can exploit the target system's vulnerabilities using neural networks. Microsoft devised a methodology to enhance software security by developing a procedure that makes it challenging to exploit additional programs.

17.9 SOCIO-ECONOMIC EFFECT

It is crucial to acknowledge that technological advancements profoundly affect both the quality of human existence and the nature of professional obligations. The complex influence of AI in cybersecurity affects the technical revolution and brings about changes in all parts of society. This section will examine the impact of changes in societal interaction in the digital realm and how technological advancements will shape future social life. The transformation also has implications for the duties of individuals in labor, their drive to pursue employment that needs specific skills, and the potential for whole sectors of the workforce to become obsolete. Having a complete knowledge of the consequences of AI in our current digital environment requires insights into social and workforce trends.

17.9.1 SOCIETAL IMPACT

AI-powered technologies are constantly enhancing their mechanisms to provide several advantages via the exchange of personal data. The advancement fosters confidence and dependability on the digital platform, where utilizing personal information is safe and advantageous in many aspects [62]. Nevertheless, efforts to protect this sensitive data are sometimes the most efficient. These protocols frequently fall short of protecting the reputation of the individual who believes that AI-based technological platforms are trustworthy [63]. Due to the possibility of data breaches and service outages, smart home assistants pose a serious security risk. This is primarily due to the open nature of the voice channel, reliance on AI assistance, and the complicated software system architecture. Similarly, increased use of ChatGPT may also indirectly improve the efficacy of automated hacking, social engineering attacks, network payload creation, advanced malware creation, email phishing, and other associated operations. One additional factor in the expansion of cybersecurity technology via the assistance of AI is the limited accessibility to cutting-edge resources for everyone. Insufficient resources, particularly financial

constraints, might lead to digital segregation across various businesses. The restriction may attract the focus of attackers to carry out advanced cyber assaults on these vulnerable networks and services. Ethical concerns have accompanied every invention, and AI is no exception. Ensuring the ethical use of AI requires agreement across states, experts, scholars, and even people.

17.9.2 CYBERSECURITY WORKFORCE

Incorporating AI into cybersecurity significantly impacts the labor market in the present technological environment. The dependence of AI on high-caliber data requires a workforce with expertise in data-centric abilities to evaluate and comprehend model behaviors thoroughly. This proficiency is essential for creating resilient and fortified cybersecurity applications. The emergence of AI has redirected attention from monotonous, hands-on assignments to positions that need analytical reasoning and proficiency in managing intricate situations. As a result, cybersecurity specialists will have more time to devote to strategic decision-making and thorough framework administration as opposed to laborious tasks that require less critical thought [64]. This development draws attention to a discernible trend: the move toward a skill set that emphasizes strategic and critical thinking. The workforce in cybersecurity needs to adapt to this shift by acquiring skills that AI cannot replicate. This encompasses not just the technological prowess but also a profound comprehension of AI's ethical, legal, and societal ramifications in cybersecurity. Although it may decrease certain employment positions, it also creates chances for professionals to enhance their expertise and participate in more significant and influential tasks. The significance of possessing interdisciplinary abilities is growing, as cybersecurity experts must strike a balance between technological expertise and comprehension of ethical, legal, and social factors.

17.10 FUTURE DIRECTIONS

The future of cybersecurity hinges on the capacity to foresee, adjust to, and address these difficulties with efficacy. An approach to consider is the correlation between developing technologies and AI-driven models, such as Quantum Computing. This computing approach can potentially revolutionize present encryption technologies, greatly changing cybersecurity. Additionally, it allows the creation of quantum-resistant algorithms, enhancing data security to a higher degree. Therefore, the potential future of AI in cybersecurity involves enhancing quantum encryption methods such as Quantum Key Distribution, which might provide robust security measures against contemporary dynamic assaults [65]. Blockchain is a rapidly developing technology that guarantees IoT networks' secure and transparent administration. It is instrumental in ensuring data integrity and facilitating informed decision-making processes. Blockchain technology combined with quantum computing may improve AI security models. By incorporating Quantum-inspired techniques into the blockchain architecture, data transfer may be made more trustworthy [66]. Explainable artificial intelligence (XAI) is of utmost importance in shaping the future trajectory of AI in cybersecurity. The objective is to tackle the issue of the opaque nature of DL models and enhance the transparency and comprehensibility of AI's decision-making process. In cybersecurity, it is crucial to comprehend the underlying reasoning behind AI-powered security selections to establish confidence and ensure dependability. Ensuring the explainability of AI models in cybersecurity promotes trust and justice and improves the effectiveness of AI systems in critical security applications [67].

17.11 CONCLUSION

In many fields and applications, the role of AI has been both acknowledged and questioned. Nevertheless, cybersecurity is of utmost importance and cannot afford any shortcomings in AI approaches. Security is the only barrier between malicious individuals and confidential data and services. Thus, for future scientists and programmers to make the required

adjustments and modifications to make well-informed decisions, they must understand the limitations and strengths of AI. This study aims to explore the vast potential of AI in cybersecurity while acknowledging the potential risks if the necessary limits are not considered during its implementation. The evolution of AI serves as a guide for researchers to comprehend the progression of changes that have led to the development of modern AI technology. The resilience of AI may be preserved by ensuring that the advancement of AI keeps pace with the progression of cyber threats. By incorporating additional technologies into AI-integrated security systems, it is possible to successfully protect against threat actors or, at the very least, reduce their impact. Furthermore, the influence of AI on cybersecurity is causing changes in other sectors. It leads to a workforce transformation, pushing it towards more intellectually challenging positions requiring a wide range of skills. At the same time, it creates opportunities for development and adaptability in a rapidly changing technological environment.

REFERENCES

- 1.Liu, Jiaying, et al. “Artificial intelligence in the 21st century.” *IEEE Access* 6 (2018): 34403–34421.[↗](#)
- 2.Joshi, Rahul, Krishna Pandey, and Suman Kumari. “A metaverse: significance and dimensions in various sectors.” In *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 2024.[↗](#)
- 3.Katiyar, Sapna. “Cyber security using artificial intelligence.” In *Cyber Security Using Modern Technologies: Artificial Intelligence, Blockchain and Quantum Cryptography*, 111, 2023. Routledge.[↗](#)
- 4.Chen, Jiayang, and Kay Choong See. “Artificial intelligence for COVID-19: rapid review.” *Journal of Medical Internet Research* 22.10 (2020): e21476.[↗](#)

- 5.Okutan, Ayşe, and Can Eyüpoğlu. “A review on artificial intelligence and cyber security.” In *2021 6th International Conference on Computer Science and Engineering (UBMK)*. IEEE, 2021. [↗](#)
- 6.Truong, Thanh Cong, Quoc Bao Diep, and Ivan Zelinka. “Artificial intelligence in the cyber domain: offense and defense.” *Symmetry* 12.3 (2020): 410. [↗](#)
- 7.Lu, Yang. “Artificial intelligence: a survey on evolution, models, applications and future trends.” *Journal of Management Analytics* 6.1 (2019): 1–29. [↗](#)
- 8.Plèta, Tomas, et al. “Cyber-attacks to critical energy infrastructure and management issues: overview of selected cases.” *Insights into Regional Development* 2.3 (2020): 1–4. [↗](#)
- 9.Tyugu, Enn. “Artificial intelligence in cyber defense.” In *2011 3rd International conference on cyber conflict*. IEEE, 2011. [↗](#)
- 10.Li, Yuchong, and Qinghui Liu. “A comprehensive review study of cyber-attacks and cyber security; emerging trends and recent developments.” *Energy Reports* 7 (2021): 8176–8186. [↗](#)
- 11.Rahim, Noor Hayani Abd, et al. “A systematic review of approaches to assessing cybersecurity awareness.” *Kybernetes* 44.4 (2015): 606–622. [↗](#)
- 12.Al-Mhiqani, Mohammed Nasser, et al. “Cyber-security incidents: a review cases in cyber–physical systems.” *International Journal of Advanced Computer Science and Applications* 9.1 (2018): 1–9. [↗](#)
- 13.Wu, Dazhong, et al. “Cybersecurity for digital manufacturing.” *Journal of Manufacturing Systems* 48 (2018): 3–12. [↗](#)
- 14.Bhamare, Deval, et al. “Cybersecurity for industrial control systems: a survey.” *Computers & Security* 89 (2020): 101677. [↗](#)
- 15.Von Solms, Rossouw, and Johan Van Niekerk. “From information security to cyber security.” *Computers & Security* 38 (2013): 97–102. [↗](#)
- 16.Nishant, Rohit, Mike Kennedy, and Jacqueline Corbett. “Artificial intelligence for sustainability: challenges, opportunities, and a research agenda.” *International Journal of Information Management* 53 (2020): 102104. [↗](#)

17. Thakur, Kutub, et al. "An investigation on cyber security threats and security models." In *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing*. IEEE, 2015. [↵](#)
18. Humayun, Mamoon, et al. "Cyber security threats and vulnerabilities: a systematic mapping study." *Arabian Journal for Science and Engineering* 45 (2020): 3171–3189. [↵](#)
19. Razzaq, Abdul, et al. "Cyber security: threats, reasons, challenges, methodologies and state of the art solutions for industrial applications." In *2013 IEEE Eleventh International Symposium on Autonomous Decentralized Systems (ISADS)*. IEEE, 2013. [↵](#)
20. Kianpour, Mazaher, and Shahid Raza. "More than malware: unmasking the hidden risk of cybersecurity regulations." *International Cybersecurity Law Review* 5.1 (2024): 169–212. [↵](#)
21. Prasetyo, Stefanus Eko, Haeruddin Haeruddin, and Kelvin Ariesryo. "Website security system from denial of service attacks, SQL injection, cross site scripting using web application firewall." *Antivirus: Jurnal Ilmiah Teknik Informatika* 18.1 (2024): 27–36. [↵](#)
22. Uddin, Ryhan, Sathish A. P. Kumar, and Vinay Chamola. "Denial of service attacks in edge computing layers: taxonomy, vulnerabilities, threats and solutions." *Ad Hoc Networks* 152 (2024): 103322. [↵](#)
23. Turing, Alan M. *Computing Machinery and Intelligence*. Springer Netherlands, 2009. [↵](#)
24. McCarthy, John. "Generality in artificial intelligence." *Communications of the ACM* 30.12 (1987): 1030–1035. [↵](#)
25. Seuba, Xavier. "Big data, AI and border enforcement of intellectual property rights." *Intelligence* 4 (1982): 707–748. [↵](#)
26. Thomas, James H. *Managing Risk to the National Information Infrastructure*. 1998. U.S. Army War College. [↵](#)
27. Stuart, Jeff A., and John D. Owens. "Multi-GPU MapReduce on GPU clusters." In *2011 IEEE International Parallel & Distributed Processing Symposium*. IEEE, 2011. [↵](#)

28. Manikopoulos, Constantine, and Symeon Papavassiliou. "Network intrusion and fault detection: a statistical anomaly approach." *IEEE Communications Magazine* 40.10 (2002): 76–82. [↵](#)
29. Costa, Paulo CG, Kathryn B. Laskey, and Ghazi AlGhamdi. "Bayesian ontologies in AI systems." In *Proceedings of the Fourth Bayesian Modelling Applications Workshop. Twenty-Second Conference on Uncertainty in Artificial Intelligence*, Cambridge, MA. AUAI Press, 2006. [↵](#)
30. Xie, Peng, et al. "Using Bayesian networks for cyber security analysis." In *2010 IEEE/IFIP International Conference on Dependable Systems & Networks (DSN)*. IEEE, 2010. [↵](#)
31. Boddy, Mark S., et al. "Course of action generation for cyber security using classical planning." In *International Conference on Automated Planning and Scheduling*, Monterey, CA. AAAI, 2005. [↵](#)
32. Parati, Namita, Latesh Malik, and A. G. Joshi. "Artificial intelligence based threat prevention and sensing engine: architecture and design issues." In *2008 First International Conference on Emerging Trends in Engineering and Technology*. IEEE, 2008. [↵](#)
33. Nourani, Cyrus F., and R. M. Moudi. "Intelligent cyberspace interfaces and wap generic computing." *IKS, St. Thomas, Virgin Islands* (2002). [↵](#)
34. Rieck, Konrad, et al. "Automatic analysis of malware behavior using machine learning." *Journal of Computer Security* 19.4 (2011): 639–668. [↵](#)
35. Wu, Zihan, et al. "RTIDS: a robust transformer-based approach for intrusion detection system." *IEEE Access* 10 (2022): 64375–64387. [↵](#)
36. Wang, H., et al. "Machine learning basics." *Deep Learning* (2016): 98–164. [↵](#)
37. Srinidhi, Chetan L., Ozan Ciga, and Anne L. Martel. "Deep neural network models for computational histopathology: a survey." *Medical Image Analysis* 67 (2021): 101813. [↵](#)
38. Sarker, Iqbal H., et al. "Intrudtree: a machine learning based cyber security intrusion detection model." *Symmetry* 12.5 (2020): 754. [↵](#)

39. Du, Simon, et al. "Gradient descent learns one-hidden-layer CNN: don't be afraid of spurious local minima." *International Conference on Machine Learning*. PMLR, 2018. [↗](#)
40. Mahmoud, Said Elsayed, et al. "Network anomaly detection using LSTM based autoencoder." In *Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*. ACM, 2020. [↗](#)
41. Han, Kai, et al. "Transformer in transformer." *Advances in Neural Information Processing Systems* 34 (2021): 15908–15919. [↗](#)
42. Tang, Tuan Anh, et al. "DeepIDS: deep learning approach for intrusion detection in software defined networking." *Electronics* 9.9 (2020): 1533. [↗](#)
43. Ferrag, Mohamed Amine, et al. "RDTIDS: rules and decision tree-based intrusion detection system for Internet-of-Things networks." *Future Internet* 12.3 (2020): 44. [↗](#)
44. Malek, Zakiyabanu S., Bhushan Trivedi, and Axita Shah. "User behavior pattern-signature based intrusion detection." In *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*. IEEE, 2020. [↗](#)
45. Awotunde, Joseph Bamidele, Chinmay Chakraborty, and Abidemi Emmanuel Adeniyi. "Intrusion detection in industrial Internet of Things network-based on deep learning model with rule-based feature selection." *Wireless Communications and Mobile Computing* 2021.1 (2021): 7154587. [↗](#)
46. Sarker, Iqbal H., Md Hasan Furhad, and Raza Nowrozy. "AI-driven cybersecurity: an overview, security intelligence modeling and research directions." *SN Computer Science* 2.3 (2021): 173. [↗](#)
47. Georgescu, Tiberiu-Marian. "Natural language processing model for automatic analysis of cybersecurity-related documents." *Symmetry* 12.3 (2020): 354. [↗](#)
48. Mendsaikhan, Otgonpurev, et al. "Identification of cybersecurity specific content using the Doc2Vec language model." In *2019 IEEE 43rd Annual*

- Computer Software and Applications Conference (COMPSAC)*. Vol. 1. IEEE, 2019. [↗](#)
- 49.Soni, Neha, et al. “Artificial intelligence in business: from research and innovation to market deployment.” *Procedia Computer Science* 167 (2020): 2200–2210. [↗](#)
- 50.Das, Rishabh, and Thomas H. Morris. “Machine learning and cyber security.” In *2017 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*. IEEE, 2017. [↗](#)
- 51.Shafique, Kinza, et al. “Internet of Things (IoT) for next-generation smart systems: a review of current challenges, future trends and prospects for emerging 5G-IoT scenarios.” *IEEE Access* 8 (2020): 23022–23040. [↗](#)
- 52.Prabhakar, Priti, et al. “Cyber security of smart metering infrastructure using median absolute deviation methodology.” *Security and Communication Networks* 2022.1 (2022): 6200121. [↗](#)
- 53.Giakoumakis, E., G. Papaconstantinou, and Emmanuel Skordalakis. “Rule-based systems and pattern recognition.” *Pattern Recognition Letters* 5.4 (1987): 267–272. [↗](#)
- 54.Zahra, Syed Rameem, et al. “Detecting Covid-19 chaos driven phishing/malicious URL attacks by a fuzzy logic and data mining based intelligence system.” *Egyptian Informatics Journal* 23.2 (2022): 197–214. [↗](#)
- 55.Stergiou, Christos L., and Kostas E. Psannis. “Digital twin intelligent system for industrial IoT-based big data management and analysis in cloud.” *Virtual Reality & Intelligent Hardware* 4.4 (2022): 279–291. [↗](#)
- 56.Mukherjee, Pratyusa, et al. “KryptosChain—a blockchain-inspired, AI-combined, DNA-encrypted secure information exchange scheme.” *Electronics* 12.3 (2023): 493. [↗](#)
- 57.Elhadad, Ahmed, et al. “Fog computing service in the healthcare monitoring system for managing the real-time notification.” *Journal of Healthcare Engineering* 2022.1 (2022): 5337733. [↗](#)
- 58.Chandre, Pankaj Ramchandra, Parikshit Mahalle, and Gitanjali Shinde. “Intrusion prevention system using convolutional neural network for

- wireless sensor network.” *IAES International Journal of Artificial Intelligence* 11.2 (2022): 504. [↗](#)
59. Priya, V. S. Devi, and S. Sibi Chakkaravarthy. “Containerized cloud-based honeypot deception for tracking attackers.” *Scientific Reports* 13.1 (2023): 1437. [↗](#)
60. Li, Zhenyuan, et al. “AttacKG: constructing technique knowledge graph from cyber threat intelligence reports.” In *European Symposium on Research in Computer Security*. Cham: Springer International Publishing, 2022. [↗](#)
61. Sinaeepourfard, Amir, et al. “Cybersecurity in large-scale smart cities: novel proposals for anomaly detection from edge to cloud.” In *2019 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC)*. IEEE, 2019. [↗](#)
62. Anica-Popa, Ionuț, et al. “The integration of artificial intelligence in retail: benefits, challenges and a dedicated conceptual framework.” *Amfiteatru Economic* 23.56 (2021): 120–136. [↗](#)
63. Gupta, Maanak, et al. “From ChatGPT to ThreatGPT: impact of generative AI in cybersecurity and privacy.” *IEEE Access* 1 (2023): 30–31. [↗](#)
64. Dawson, Jessica, and Robert Thomson. “The future cybersecurity workforce: going beyond technical skills for successful cyber performance.” *Frontiers in Psychology* 9 (2018): 744. [↗](#)
65. Xu, Feihu, et al. “Secure quantum key distribution with realistic devices.” *Reviews of Modern Physics* 92.2 (2020): 025002. [↗](#)
66. Abd El-Latif, Ahmed A., et al. “Quantum-inspired blockchain-based cybersecurity: securing smart edge utilities in IoT-based smart cities.” *Information Processing & Management* 58.4 (2021): 102549. [↗](#)
67. Capuano, Nicola, et al. “Explainable artificial intelligence in cybersecurity: a survey.” *IEEE Access* 10 (2022): 93575–93600. [↗](#)

18 Transforming Education

The Role of AI-Driven Chatbots in Advancing Equal Access, Accountability, and Empowerment in Educational Systems

*Santu Karmakar, Ankur Nandi, Tapash Das,
and Tarini Halder*

DOI: [10.1201/9781003668657-18](https://doi.org/10.1201/9781003668657-18)

18.1 INTRODUCTION

If there is one thing that we all agree on, it has to be the fact that artificial intelligence (AI) has brought a significant change in every sector, and education is no exception. Creative solutions to age-old problems are being provided by AI technology, such as the popular chatbots that have now been adapted in various fields. Chatbots – using AI are increasingly being used as powerful tools for personalized learning and timely academic support (Labadze et al. 2023). By advocating for responsible AI implementation, the chapter advances discussions on protecting human rights within AI-powered educational systems and raises new questions about maintaining fairness and transparency in digital learning. This contribution highlights

practical frameworks for incorporating AI ethically into education, inspiring future research on human-centered digital learning solutions.

Education has never really been limited to a certain time, geographical borders, or normally closed educational environments. These challenges often contribute to the complexity of serving diverse student populations with high-quality instruction. However, these barriers are being broken down as chatbots increasingly infiltrate educational environments. Interactive chatbots can assist students in better learning and offer answers to their questions, deliver bespoke educational information tailored according to the individual needs of every student with rapid feedback using advanced machine learning (ML) and natural language processing (NLP) ([Kuhail et al. 2023](#)). Chatbots are being applied in education not just for study-related purposes (Okonkwo and Ade-Ibijola 2021), they aid in administrative tasks also, thereby supporting quality teacher–student contact and providing with data-backed insights to inform teaching practices ([Kooli 2023](#)). With the help of technology and having so many uses for it, chatbots are seen as mandatory parts in advanced education frameworks to grow student engagement, efficacy and accessibility. Students are using chatbots even in education and not just for academic support ([Labadze et al. 2023](#)). They also assist in administration, encourage good teacher-to-student communication, and provide valuable analytics to shape teaching practices ([Labadze et al. 2023](#)). Given their versatility, chatbots are in a sweet spot to take precedence as integral pieces of modern learning systems that might quicken student engagement, efficiency, and accessibility ([Annamalai et al. 2023](#)).

Also, with the continual development of AI technology, chatbots in education are able to do more. In addition to essentially answering questions and delivering tailored feedback, some more advanced chatbots are starting out with personalized learning experiences that have adapted their ways for each unique user ([Labadze et al. 2023](#)). This allows them to adjust question difficulty, provide extra materials and even recommend learning approaches depending on how well the student is doing or prefers ([Pashler et al. 2008](#)). Furthermore, the integration of chatbots with various

other educational technologies – including Learning Management Systems (LMS) and virtual classrooms – is improving their ability to fulfil a significant purpose. For instance, chatbots have now become capable of monitoring a student's progress and schedule study in real time, while also allowing the integration with online resources to provide an all-encompassing learning experience ([Javaid et al. 2023](#)). Thus, the built-in connectivity provides real-time updates that ensure prompt support is available to students as they make their way through an educational journey. Chatbots also contribute to creating a more comprehensive learning experience, allowing multilingual support systems and accessibility options (Ralston et al. 2019). This is especially useful for students that language has difficulty or those with disabilities who need some form of accommodation. Chatbots also provide support in various languages and formats, which can help ensure all students have equal access to educational resources ([Wollny et al. 2021](#)). But the implementation of chatbots in education is not devoid of challenges. Matters involving data privacy, ethical concerns around the use of AI and an incessant need to improve all require significant attention. Much like we teach our kids not to cheat, it is necessary that chatbots function within ethical guidelines and ensure data security at a high standard for them to be successful and accepted as tools in educational environments. The future of chatbots in education is exciting. The interaction and learning experience become even more advanced as the AI technology progresses. Then, chatbots can get even better at interpreting complex queries and delivering nuanced responses to help deliver a more meaningful educational experience. Their role in education is going to be more cool hats for a while yet providing awesome stuff for next-generation student learning, engagement, and success in the age of all things digital.

18.1.1 HISTORY OF CHATBOTS

The fascinating journey of chatbots dates back to several decades during which major technological advancements and human–computer interaction

paradigms have evolved. Some pivotal points in the emergence of chatbots are provided below:

- *1960s: The Birth of Chatbots:*
 - *ELIZA (1966):* With Joseph Weizenbaum at MIT, ELIZA is the earliest chatbot. It simulated a Rogerian psychotherapist by rephrasing users' statements as questions. Simple though it was, ELIZA showed how a machine could make human conversation ([Adamopoulou and Moussiades 2020](#)).

1. *The 1970s and the Age of Early AI & NLP:*

- *PARRY (1972):* Created by psychiatrist Kenneth Colby, it was made to imitate an imagined individual with paranoid schizophrenia. It was better than ELIZA in that it employed an early model of human emotion ([Adamopoulou and Moussiades 2020](#); [ZEMČÍK 2019](#)).

2. *AI and NLP Developments: Early 1990s:*

- *A.L.I.C.E. (1995):* This version of A.L.I.C.E. (1995) (Artificial Linguistic Internet Computer Entity), developed by Richard Wallace, utilized AIML (Artificial Intelligence Markup Language), was one of the most visible utilizations in a period where chatbots were much more commonly found on webpages such as this than they are today. Person Identification: A.L.I.C.E. achieved success, winning multiple Loebner Prizes in the early 2000s ([Adamopoulou and Moussiades 2020](#)).
- *Jabberwacky (1997):* Jabberwacky was established by Rollo Carpenter as an entertaining effort to mimic natural human chat. One of the earliest examples as a learning chatbot, it adapted its responses based on interactions from users ([Adamopoulou and Moussiades 2020](#)).

3. *In the 2000s: Virtual Assistants Emerge:*

- *SmarterChild (2001)*: Built by ActiveBuddy SmarterChild was an intelligent, always-on chat robot designed to support continuous conversations through AOL Instant Messenger or MSN Messenger with the ability to learn from its users and provide answers faster than humans ([Adamopoulou and Moussiades 2020](#)).
- *Siri (2011)*: Acquired by Apple (Giffin 2024), this was the first virtual assistant that shipped with iPhones. It used state-of-the-art speech recognition and natural language understanding to do things for you, from finding the right music video on YouTube ([Adamopoulou and Moussiades 2020](#)).

4. 2010 – The Decade of AI Chatbots:

- *Watson (2011)*: IBM’s Watson became famous for defeating human contestants at the game show Jeopardy! against human champions. Later, new industries also benefit from Watson’s powerful NLP, like Healthcare, customer service, etc. ([Adamopoulou and Moussiades 2020](#)).

5. 2021 Revisted: Smart AI and Conversation Centeres Everywhere:

- *OpenAI’s ChatGPT (2020)*: This release brought chatbots to a new level of maturity, with the ability of commercial chatbot platforms to reply using a transformer. This could produce real human text given a prompt, allowing for the creation of better chatbots that can understand context and more complex language.

18.2 REVIEW OF RELATED LITERATURE

TABLE 18.1

Review of Related Literature on Chatbots in Education↵

Title(Citation)	Summary of Main Findings

Title(Citation)	Summary of Main Findings
Role of AI chatbots in education: systematic literature review (Labadze et al. 2023)	Using chatbots in learning offers many advantages, as implemented with AI referenced above. They provide for doing homework to students, personalized learning experiences and support in skill development. Chatbots can be useful for educators; they save time, provide help with pedagogical ways etc. But there are also significant hurdles such as the faith in AI chatbots and its capabilities, ethical aspects when it comes to using these solutions.
Educational chatbots (ECs) for project-based learning: investigating learning outcomes for a team-based design course (Kumar 2021)	For instance, by encouraging the learner community to work together and compete with one another attending teams through productive ECs, they may be stimulated socially in practice and complicate their collaboration while providing superior learning results. Their use, however, does not interfere with affective motivational components of learning outcomes (i.e., the perception that they are or should learn more; arousal level and motivation toward various behaviors needed to learn) process validity during Si-Sa- analyzing moves nor creative self-efficacy determining efficacy..

Title(Citation)	Summary of Main Findings
Chatbots Applications in Education: A Systematic Review (Okonkwo and Ade- Ibijola 2021)	This analysis revealed benefits such as quick and personalized services for both students and educators. Additionally, the review identified existing research on the topic, challenges to overcome when implementing chatbots, and promising areas for future research in this field.
Artificial Intelligence in Higher Education: The State of the Field (Crompton and Burke 2023)	We see a troubling trend where researchers increasingly find work in education departments. Most of the focus is on undergraduate students, but particularly in writing and reading areas. Applications of AIED encompass assessment and evaluation, prediction in educational contexts which can support many lines of work including AI assistance, managing student learning. The review identifies these gaps in the literature and therefore emphasizes that there is a requirement for future studies investigating new AI tools such as ChatGPT to be employed within educational settings.

Title(Citation)	Summary of Main Findings
Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions (Kooli 2023)	While systems implement artificial intelligence and chatbots can provide powerful service to human expertise in the case of education or research, they may also present some pitfalls for ethical use. But the pathway is clear and, while there may be room for development, current practices show that AI developments will require novel assessment practices locked with imagination. Key is continuous adaptation to AI enhancing collaboration instead of competition – living with #AI, sustainability and perpetual change. Enabling solutions should focus on sensitizing all and sundry, making people aware, enacting the needed laws and cementing ethical values to shield research systems. In conclusion, AI and chatbots are simple to complex learning tools depending on the context they wish to be used in existing or new systems where making integration strategies innovative becomes key components of effective use.
AI in Education: A Systematic Literature Review (Tahiru 2021)	In the education sector, AI improves collaborative learning and eases administration burdens. But intelligent others remain grounded in research, and there is little evidence to suggest that they scale beyond Intelligent Tutoring Systems. The research on AI in education has been growing since 2016, particularly in developed countries. Some challenges include ethical concerns, data ownership, and AI–human labor substitution. The formulation of policy as well as empirical research on AI effects (economic and societal) are necessary. To support programmers indefinitely, the curricula are to be updated and AI is to be trained continuously.

This related literature review on chatbots in education shows the hopeful returns and painful pitfalls of utilizing such tools to achieve educational

goals ([Table 18.1](#)). Chatbots serve a range of benefits like personalized learning experiences, efficient administrative and project-based teamwork support. They are readily available to help students and teachers, which in turn saves time as well as delivers personalized assistance that can lead to better learning outcomes. But of course, their deployment is not that easy. These include questions of trustworthiness, actual and potential ethical implications around their use, as well as deployment strategies for AI systems in practice. Studies suggest that despite being beneficial for facilitating learning and administration, chatbots fall short in terms of motivating students and maintaining ethical practices. In education, AI is an emerging field of growing research interest targeting at developing novel tools or polishing the existing applications. The authors call for much more research about the effects of AI, as well as extensive ethical rules and novel integration approaches. Chatbots and AI tools, if well-tailored to make optimal use in education, can be designed for productive results through continuous adaptation and sound policy implementation only. Overcoming these bottlenecks will be important to making more of AI in education and ensuring that it is developed responsibly.

18.3 OBJECTIVES

The objectives of this study are as follows:

1. Impact of Chatbots on Education
2. Exploring the Ethical Considerations of Implementing Chatbots in Education
3. Challenges in Chatbot Integration
4. Opportunities for Academic Enhancement

18.4 METHODOLOGY

The extensive search was performed in academic resources such as JSTOR, ScienceDirect, Google Scholar, and other academic databases. The

keywords used for the search were artificial intelligence, education, chatbots, and e-learning technology. The search was limited to the English language and article type within the search period, ranging from 2013 to 2023. This search returned a total of 91 eligible articles and abstracts, which were subsequently assessed for inclusion with respect to their topic relevance, the quality of evidence in their support or methodology-source type as well as time of publication. Any article not peer-reviewed or published from a major archive of scholarly articles/journals was removed. Finally, 31 relevant articles and links were found following the screening process for manual extraction. The selected articles were reviewed in-depth, and the main results, research methods used for conducting the study, and the limitations of these study findings were examined. This is a thematic analysis with the main purpose of finding patterns and trends related to artificial intelligence in education. This meant investigating possible uses and abuses of AI in education, along with their larger-scale consequences for pedagogical practice. For a qualitative synthesis, data gathered from the available articles reviewed were categorized and grouped into thematic clusters. The emergent themes were used to make inferences and draw conclusions about both the real and potential impact of artificial intelligence in education.

18.5 FINDINGS

18.5.1 APPLICATIONS OF CHATBOTS IN EDUCATION

18.5.1.1 24/7 Accessibility with Chatbot Support

Chatbots in education are like having a helpful assistant that is always available (Khidir and bin Sa'ari 2022). This allows them to support students and teachers with question answering or a bit of extra help at any time, 24 hours per day. This continual availability enables learning and teaching to proceed regularly without any disruptions ([Karandish 2021](#)). Chatbots are able to handle a variety of questions, from fairly simple knowledge-based

queries through more complex academic forms. This ensures that help can be available any day or night to the students. This 24/7 assistance can be especially helpful to a distance learner, or someone in another time zone.

18.5.1.2 Quick Responses

This sucks, and it is super annoying to get questions answered 3 days later... Teachers and staff often get asked the same questions every day. AI could benefit students to quickly have the answer of frequently asked questions by bringing in automated support and smart conversations ([Javaid et al. 2023](#)). This will save precious time of educators and help students receive target information quickly, rather than working around long delays ([Javaid et al. 2023](#)). On top of this, chatbot responses are usually very quick, which may strengthen student engagement and reduce frustration, resulting in a more positive learning experience.

18.5.1.3 Personalize Learning

One of the big trends these days is personalization in education. AI promotes customizable learning programs, which help individualize the experiences and preferences of each student ([Mageira et al. 2022](#)). It evaluates the past learning histories and learns from those weaknesses to allow for better personalization of course materials. It supports different learning styles and needs, which makes education more inclusive for everyone.

18.5.1.4 Student Feedback

Student feedback is invaluable in the context of helping to maintain course materials, facilities and an overall quality learning experience for students. Institutions of learning rely on a good reputation, and there is little doubt that they are assisted by both high performance in league tables along with news stories about the contentment of British students. Furthermore, by collecting the feedback in real time also enables higher institutions for

quick identification and fixing of concern areas which ultimately results in continuous improvement in delivery and better learning outcomes.

18.5.2 EXPLORE ETHICAL AND PRIVACY CONCERNS

18.5.2.1 Privacy and Data Security

Issue: Chatbots collect and process data including the names, contact details, academic or learning behaviors even sometimes emotional aspects of students ([Hasal et al. 2021](#)).

Ethical Consideration:

- *Data Protection:* Schools have to institute robust encryption mechanisms and data storage practices. There is a need to conduct regular security audits and vulnerability assessments to secure the privacy of data against any unauthorized access and breaches ([Hasal et al. 2021](#)).
- *Informed Consent:* Parents and students should be made well aware of data used, collected and things like what data they collect, why it is collected, how it will be used, and who might have access to it. Consent should be obtained in a transparent manner ([Drachslers and Greller 2016](#)).
- *Regulatory Compliance:* One important aspect is that new institutes must also comply with the relevant data protection laws in different parts of world. For example, General Data Protection Regulation (GDPR), which is true in Europe and Family Educational Rights Privacy Act (FERPA). These laws include, for instance, the right of access or to rectification and erasure of personal data ([Pittman et al. 2023](#)).
- *Anonymization and Minimization:* When students are being named in a paper, data should be anonymized where possible. Only collect data essential to the chatbot role ([Hasal et al. 2021](#)).

18.5.2.2 Bias and Fairness

Issue: Bias of chatbots may mirror and even sustain inherent prejudices among its training data, resulting in unfair results ([Kooli 2023](#)).

Ethical Consideration:

- *Diverse and Representative Training Data:* Developers will be using diverse and inclusive datasets that can take all possible combinations of student attributes into account to minimize bias ([Paullada et al. 2021](#)).
- *Bias Detection and Mitigation:* You need to run periodic audits and tests to identify such biases so that you can minimize them. This means examining the chatbot conversations for evidence of bias or discrimination ([Xue et al. 2023](#)).
- *Algorithmic Transparency:* Institutions need to guarantee that the decisions made by chatbots become clearer as well. This involves not only generating technical models but also rendering the black-box underlying algorithms interpretable for non-specialist users and articulating how decisions are reached ([Felzmann et al. 2020](#)).
- *Fairness Metrics:* Introduce fair metrics to measure if the interaction between the bot and user is impartial. For instance, monitoring performance within a specific demographic (Garcia Valencia et al. 2023).

18.5.2.3 Autonomy and Dependency

Issue: An over-reliance on chatbots prevents students from strengthening their individual problem-solving skills and critical thinking ([Kooli 2023](#)).

Ethical Consideration:

- *Complementary Role:* Instead, chatbots are intended to facilitate human teaching, and not replace it entirely. Instead, they ought to promote students' critical thinking abilities and encourage the gathering of multiple information ([Wollny et al. 2021](#)).
- *Educational Design:* When implementing chatbots, encourage active learning. Encourage students to think critically about solutions or

walk through problem-solving techniques ([Chang et al. 2023](#)).

- *Balanced Integration:* Balance the use of chatbots with traditional training methodologies and human intervention. This reduces dependence on automated systems ([Mitan 2024](#)).
- *Monitoring Usage:* Record usage of chatbots by students to understand when overuse may be happening. It is also possible to design interventions that promote more autonomous learning behaviours ([Hidayat-ur-Rehman 2024](#)).

18.5.2.4 Transparency and Accountability

Issue: The mechanism behind chatbots can be hard to discern for users, as they do not show direct input-based responses or recommendations ([Khurana et al. 2021](#)).

Ethical Consideration:

- *Clear Disclosure:* Inform users explicitly that they are talking to a chatbot. Describe the purpose, functionality, and restriction of chatbot explicitly ([Kuhail et al. 2023](#)).
- *Feedback and Recourse:* Allow users to provide feedback, report problems, and obtain human help. Those procedures should be in place to handle complaints ([Kooli 2023](#)).
- *Audit Trails:* Keep a history of chatbot interactions, so the record can provide help to follow up on an issue and know who was involved. This provides accountability for issues and considerations ([Mitan 2024](#)).

18.5.2.5 Consent and Choice

Issue: Students may not always have a clear choice about whether or not to interact with chatbots, particularly if these tools are integrated into mandatory educational platforms ([Merele et al. 2023](#)).

Ethical Consideration:

- *Voluntary Participation:* Students should be given a choice to decide whether they interact with the chatbots and not vice versa. In addition, participation should be voluntary with alternative local custodian support mechanisms in place ([Mendoza et al. 2022](#)).
- *Informed Consent:* You should indicate clearly what the chatbot does, how it works, and how all data is collected before students. Obtention of informed consent from students or their guardians ([Mendoza et al. 2022](#)) is paramount.
- *Opt-Out Mechanism:* Make sure there is a clear and easy way for students to opt out of using your chatbot. Ensure that this option will not penalize them with their access to a variety of educational resources ([Adamopoulou and Moussiades 2020](#)).
- *Alternative Support:* Make certain that students who opt not to use chatbots enjoy access to the same level of human support and resources (Gupta and Chen 2022).

18.5.2.6 Emotional and Psychological Impact

Issue: Chatbots could negatively affect the mental health of students, particularly when they are used for providing quasi-counsellor-type support (Gupta and Chen 2022).

Ethical Consideration:

- *Boundaries and Limitations:* Create parameters for what a chatbot can/cannot do, especially for categories such as Mental health support ([Khurana et al. 2021](#)).
- *Human Oversight:* Make sure chatbot interactions at least on sensitive topics have some sort of human oversight. This is useful to give proper health care and assistance when it is necessary ([Mendoza et al. 2022](#)).
- *Emotional Awareness:* This includes capturing negative emotions and reacting to them in a similar way as any human agent would. That

means we are supportive and human in our replies while helping students to the actual humans as needed.

18.5.2.7 Equity and Access

Issue: Not all students have the same access to technology that is required for chatbots adding to social inequalities in education ([Kooli 2023](#)).

Ethical Consideration:

- *Provision of Resources:* Ensure that each and every student has the basic requirements for accessing hardware like a laptop or internet connectivity. One such way may be to provide either device or set up flexible computer labs ([Chang et al. 2023](#)).
- *Inclusive Design:* Develop chatbots that are accessible for students with disabilities, such as visual, auditory, or cognitive impairments. Lately, those functionalities have been making cameo appearances in some Android phone models via the use of screen readers and voice commands ([Chang et al. 2023](#)).
- *Support Programs:* Design programs to assist the needs of individuals from low-income backgrounds. Equally importantly, this support will need to include digital literacy training; grants and subsidies, as well as that chatbots are both culturally appropriate ([Chisom et al. 2024](#)).
- *Monitoring and Assessment:* Monitor the effects of chatbots on various student groups so that differences in access to, and outcomes from using them can be identified and rectified ([Kooli 2023](#)).

18.5.2.8 Quality and Accuracy of Information

Issue: Chatbots can offer erroneous, obsolete, or misleading data which could compromise the students' learning ([Tlili et al. 2023](#)).

Ethical Consideration:

- *Regular Updates:* The information that a chatbot refers to should be updated in a frequent interval and the data source must be reliable

and authentic. This implies that collaboration with educators and professionals has been a theme throughout (Okonkwo and Ade-Ibijola 2021).

- *Fact-Checking Mechanisms*: Develop a system for fact-checking the information shared by chatbots. It also involves cross-checks with reliable educational resources (Okonkwo and Ade-Ibijola 2021).
- *Feedback Loops*: Implement feedback loops to allow students and educators to report errors, or suggest improvements in the chatbot's knowledge base ([Khurana et al. 2021](#)).
- *Transparency of Sources*: Make the sources, chatbot use information to execute a task, transparent to customers. This is not only necessary, but it also increases your visibility and helps users to verify the information on their own (Gupta and Chen 2022).

18.6 CHALLENGES IN CHATBOT INTEGRATION

1. *Quality and Accuracy*:

- *Information Reliability*: It is crucial your chatbots can deliver correct and up-to-date information. Outdated and inaccurate content will not only mislead the students, but also hurt trust in the system ([Sebastian 2023](#)). Finally, you need to continually update and review the chatbot's knowledge base as missing out on your customers' queries can make it highly unrealizable.
- *Handling Complex Queries*: Complex or subtle questions that involve a more nuanced level of knowledge which cannot be handled by chatbots will intervene in humans and therefore lack quality explanation ([Rane 2023](#)). This requires a smooth hand-off process between chatbots and human educators so that students get complete answers.

2. *Privacy and Security*:

- *Data Privacy:* Given that chatbots are responsible for handling such delicate student data, it is really important to safeguard all this information from breaches and any kind of misuse. This is necessary for compliance with data protection regulations such as GDPR and FERPA ([Hamilton et al. 2021](#)). Institutions are required to put in place strong privacy policies and highly secure data encryption methods.
- *Cybersecurity Risks:* Chatbot systems should be protected from various potential cyber threats, such as hacking and data theft to reduce the risk of information leakage in educational institutions ([Yang et al. 2023](#); [Hasal et al. 2021](#)). The basic principles are: be aware, know the problems so you can detect them, and always have a plan for regular security audits and updates to avoid such vulnerabilities.

3. *Technical Limitations:*

- *Natural Language Understanding:* However, the modern chatbots are not perfect yet and they can misunderstand human communications sometimes while processing natural languages so the result becomes frustrating as well ([Izadi and Forouzanfar 2024](#)). Efforts should be made to improve NLP algorithms for only useful and meaningful information will reach the end users.
- *Integration with Existing Systems:* It can be seen that, while chatbots could help learners prepare for mobile devices more than students with computers, the process of combining existing educational and database platforms seamlessly is technically challenging and requires considerable resources ([Bratić et al. 2024](#)). Ensure uninterrupted user experience and data management.

4. *Cost and Resource Allocation:*

- *Initial Investment:* The cost of designing, installing, and training chatbots is high – a large nest egg ([Zhang et al. 2023](#)). Hence, they

cannot convince these institutions to invest in buying a copy of the book without conducting cost–benefit analyses.

- *Maintenance and Updates:* Continuous maintenance, updates, and improvements are necessary to keep the chatbot functional and relevant, which can strain institutional resources ([Zhang et al. 2023](#)). Allocating a dedicated team for ongoing support is crucial for long-term success.

5. User Engagement and Acceptance:

- *Resistance to Change:* Students, educators, and administrative staff may resist adopting new technologies, preferring traditional methods of learning and communication ([Mumtaz 2000](#)). Effective change management strategies and training can help ease the transition.
- *Engagement Levels:* Not all students may find interacting with a chatbot engaging or effective, which could impact their learning experience and outcomes ([Essel et al. 2022](#)). Ensuring the chatbot is user-friendly and interactive can help improve engagement.

6. Over-Reliance and Reduced Human Interaction:

- *Dependence on Technology:* There is a risk that students may become overly reliant on chatbots, potentially diminishing critical thinking skills and face-to-face interactions with educators and peers ([Rathore 2022](#)). Educators should encourage a balanced use of technology and foster critical engagement.
- *Balancing Human and AI Interaction:* Striking the right balance between human interaction and AI assistance is crucial to maintain the quality of education and personal connection ([Feijóo et al. 2020](#)). Institutions should design hybrid models that leverage the strengths of both.

7. Cultural and Language Barriers:

- *Language Limitations:* While chatbots can offer multi-language support, nuances and dialects in language can still pose challenges, potentially leading to misinterpretation or miscommunication ([Younis et al. 2024](#)). Continuous improvements in linguistic capabilities and contextual understanding are necessary.
- *Cultural Sensitivity:* Chatbots must be programmed to understand and respect cultural differences, which can be complex and multifaceted ([Chaves et al. 2021](#)). Regular updates and input from diverse cultural perspectives can enhance the chatbot's effectiveness in different cultural contexts.

18.6.1 OPPORTUNITIES FOR ACADEMIC ENHANCEMENT

1. *Personalized Learning Experiences:*

- *Adaptive Learning:* Chatbots offer individual student performance analysis, they analyze the learning pace and style of a student on the basis of which they modify the complexity factor in content accordingly ([Kooli 2023](#)). The differentiation is personalized in a way that ensures every child has access to what is suggested in each topic area, whatever happens or may seem like at their level of ultimate excellence and enthusiasm. In the meantime, this adaptive system could be used to identify weak spots for individual students and provide additional support or resources as are necessary based on their input.
- *Customized Feedback:* They offer guidance on the homework so that students can be aware of where they are good and what aspects need to improve ([Chang et al. 2023](#)). The feedback is much quicker, individually appropriate, and focuses on promoting students' work with specific effective guidance for development which enables a deeper layer of potential recurrence back into teaching. Because the feedback is immediate mistakes are

corrected right away – this results in solid learning pathways being built – you get better and faster at a high rate, with comprehension.

2. Accessibility:

- *24/7 Availability:* Chatbots are available around the clock, allowing students to access learning resources and assistance anytime, anywhere ([Mendoza et al. 2022](#)). Students who may have less traditional schedules or live in a different time zone can still access the support they need because of this 24/7 availability. This flexibility allows tailoring of instruction to various learning styles and life situations, thus increasing access and the versatility needed by modern students.
- *Support for Diverse Learners:* They can offer resources in multiple languages and cater to students with special needs, ensuring that education is inclusive and accessible to all ([Zobel et al. 2023](#)). With this kind of flexibility, educational materials become more available to the different population of learners and inclusivity is promoted and gaps bridged. Chatbots allow for different students at varying aptitudes and backgrounds to have a tailored learning experience which paves the way to an equitable school environment where every student has equal access in reaching their maximum potential.

3. Efficient Administrative Support:

- *Streamlined Processes:* Chatbots can handle routine administrative tasks such as scheduling, reminders, and answering common queries, freeing up time for educators to focus on teaching ([Majeed et al. 2023](#)). This efficiency enables educators to spend extra time on instruction and students, which in turn boosts overall productivity. Reduce administrative burdens, but also the net experience for students as a whole.

- *Instant Information:* They can provide instant responses to frequently asked questions about course details, deadlines, and campus facilities ([Neupane et al. 2024](#)). This quick availability of information aids students in keeping themselves organized and up to date on important matters too, which means less confusion for them to deal with; they can have a more seamless academic life this way as well. Chatbot will give immediate response to the students, therefore as they have their planning and decision-making is improved because all the necessary information is at the supply house.

4. *Interactive Learning Tools:*

- *Engaging Content:* Interactive quizzes, games, and simulations are delivered via a learning management system ([Kuhail et al. 2023](#)). As these interactive elements can enhance the learning experience and help learners understand difficult concepts better, they have a much more important role than just to make studying fun. The incorporation of game-based elements in learning can inspire students to learn more, turning education into a fun and efficient process.
- *Virtual Study Assistance:* Chatbots can host study groups, provide practice questions, and walk students through difficult topics with step-by-step explanations; this virtual assistant promotes collaborative learning where it makes sure that students comprehend the subject not just on those terms of spoon-fed ([Pérez et al. 2020](#)). Chatbots engage learners appropriately and effectively by targeting support, fostering peer collaboration, which cultivates their ability to think critically and act upon complex problems.

5. *Enhanced Student Support:*

- *Academic Advising:* Chatbots can assist with course selection, degree planning, and career advice, helping students make

informed decisions about their academic journey ([Wang et al. 2023](#)). This student-centered approach to academic advising ensures that every single step is guided for the students and made known with all possible options open. Essential guidance provides situational advice, which guides students on their educational direction

- *Mental Health Support:* They can provide resources necessary for mental health and wellness, and direct the students to appropriate support services when needed ([Sweeney et al. 2021](#)). Taking the initiative to address mental health in this way has the potential to vastly improve students' quality of life and consequently make a positive impact upon their grades. Offering timely resources and support for mental health is beneficial because it helps access a supportive learning environment which has long-term advantages.

6. *Data-Driven Insights:*

- *Performance Analytics:* Chatbots, equipped to monitor student performance data analyze natural language interaction, and identify trends and opportunities for intervention ([Wang et al. 2023](#)). They allow educators to analyze the data and tailor their individual instructional strategy to serve students better, thereby uncovering trends, as well as informing decisions, so you can use data to ensure a greater learning experience.
- *Predictive Analytics:* They can predict student outcomes based on engagement and performance data, allowing for proactive support and intervention ([Chang et al. 2023](#)). To stay ahead of the curve, educators can identify areas for concern early and intervene strategically to support struggling students. The predictive insights allow to create proactive means of intervention and support so students would not fall behind academically.

18.7 CONCLUSION

The advent of chatbots in education represents a paradigm shift with the potential to significantly enhance the learning experience for students and streamline the teaching process for educators. Providing 24/7 availability, immediate responses, and tailored learning experiences are different types of applications that currently exist as virtual assistants in modern educational settings. Chatbots deliver such essential assistance as they work 24/7, meaning that you are no longer restricted to the office hours or time zones. This constant availability means that students can get help when they need it, which is a huge benefit, especially for those studying remotely or in different time zones. Instant answers to frequently asked questions eliminate further delays in sharing information and, as a result, generate more effective and attractive learning. On a personal learning level, chatbots will leverage education content according to the student knowledge background and favorite way of studying in addition time. Utilizing data from past engagements or performance, chatbots can deliver a customized educational intervention to individual students, thus enabling an effective and inclusive learning experience. This adaptability not only improves the entire learning experience but also notifies us of where students may need more help. On the flip side, chatbots streamline administrative experiences to ensure they are simple and fast through automating mundane requests such as booking meetings, giving reminders, or answering questions. The automating capability fosters an atmosphere where educators can spend more time on the teaching part and less on administrative burdens, resulting in increased productivity of a streamlined educational environment. There are also interactive study tools and virtual instructors that clients can interact with to customize their learning. However, even a bowlful of good does come with its own set of obstacles when it comes to integrating chatbots into education. Security, privacy, and bias ethical considerations such as protections on privacy (data/trade/personal identity), data security and bias etc. must be addressed cautiously. As chatbots handle and process a lot of sensitive information, it is important to take the necessary data protection precautions. If the institutions can check for their compliances with data protection, and are

good in providing information as to how their data will get utilized, then all is well set. Particularly when it comes to chatbot interactions, addressing biases and retaining algorithmic transparency are necessary in order not only to avoid perpetuating existing prejudices but also to ensure fair treatment of all users. That, along with technical constraints (the quality of NLP and the compatibility between chatbots to existing systems), increase these difficulties. Chatbot responses need to be continuously updated, and improvements must always be made for response reliability as well as integration with educational platforms. It is important to carefully analyze and then incrementally budget the costs of setting up, running or customizing a chatbot system. Another worry is that we might develop an over-reliance on chatbots, and therefore human interaction could start to fade. Chatbots need to be balanced with traditional teaching resources so that students are not just spoon-fed and only given major answers, but also have the chance to develop critical thinking skills themselves. Keeping chatbot use integrated with human interaction is important to prevent the degradation of education quality. Additionally, and for all students to leverage chatbots as a resource the cultural and language-specific specifics of each region should be approached. To bridge the gap and promote cultural inclusiveness in education allows development of chatbots that offer multilingual support to understand and honor cultural differences.

18.8 DISCUSSION

One of the coolest things about chatbots in education is that they are available 24/7. This means that students and teachers can get help and information whenever they need it, without any delays. It is super convenient and makes the learning process much more efficient ([Khidir and bin Sa'ari 2022](#); [Karandish 2021](#)). Chatbots are great for dealing with repetitive questions, which frees up educators to focus on the more complex stuff ([Javaid et al. 2023](#)). But that's not all! Chatbots also help with personalized learning. They analyze each student's data and then customize the educational content to match their unique needs, learning pace, and

preferences ([Mageira et al. 2022](#)). This means that students get targeted support and can improve in the areas where they need it most. It is like having a personal tutor right at your fingertips! And here is another way chatbots make a difference: they gather and analyze student feedback. This feedback is super valuable because it helps improve the course materials, facilities, and overall learning experience ([Kuhail et al. 2023](#)). The use of chatbots in schools comes with important ethical concerns. First, let's talk about data protection and privacy. Chatbots collect a lot of sensitive information, like students' personal details, academic behaviors, and even emotional states. According to [Hasal et al. \(2021\)](#), schools need to use strong encryption and conduct regular security checks to keep this data safe from hackers. They also need to follow laws like GDPR in Europe and FERPA in the USA to protect students' rights regarding their data ([Pittman et al. 2023](#)). Next, informed consent is crucial. Parents and students must know what data is being collected, why it is collected, and who can see it. [Drachsler and Greller \(2016\)](#) say that getting consent in a clear and transparent way helps build trust. Bias and fairness are also big issues. Chatbots can sometimes reflect the biases in their training data, leading to unfair outcomes ([Kooli 2023](#)). To prevent this, developers should use diverse datasets and regularly check the chatbots for biases ([Paullada et al. 2021](#); [Xue et al. 2023](#)). Making chatbot decisions clear and understandable is also important ([Felzmann et al. 2020](#)). Another concern is that relying too much on chatbots might stop students from developing their problem-solving skills and critical thinking. Chatbots should support, not replace, human teachers ([Kooli 2023](#)). They should be designed to encourage students to think critically ([Chang et al. 2023](#)). Schools should also monitor how much students use chatbots to prevent over-dependence ([Hidayat-ur-Rehman 2024](#)). Transparency and accountability are key. Users should know they are talking to a chatbot and understand its purpose and limits ([Khurana et al. 2021](#)). There should be ways for users to give feedback and get help from a human if needed ([Mitan 2024](#)). Consent and choice are also important. Students should have the option to choose whether or not to use chatbots. There should be easy ways to opt out without facing any penalties,

and there should be other forms of support available ([Mendoza et al. 2022](#); [Adamopoulou and Moussiades 2020](#)). It's also important to ensure that all students have access to the technology they need to use chatbots, including those from low-income backgrounds ([Chang et al. 2023](#); [Chisom et al. 2024](#)). Chatbots need to provide accurate and reliable information. This means regularly updating the data they use, having systems to check facts, and allowing users to report errors ([Tlili et al. 2023](#); Okonkwo and Ade-Ibijola 2021). With chatbots, collecting and processing feedback is a breeze, so educational institutions can maintain high standards and keep their students happy. Integrating chatbots into educational systems sure has its benefits, but let's not overlook the challenges that come with it. One of the most important things to consider is making sure that the information provided by these chatbots is top-notch in terms of quality and accuracy. I mean, we definitely do not want students getting misled or losing trust because of outdated or incorrect info, right? ([Sebastian 2023](#)). Another challenge is that chatbots sometimes struggle with handling complex or nuanced questions. That is when human intervention becomes necessary to give students comprehensive answers that hit the mark. It is like the chatbots need a little backup from us humans ([Rane 2023](#)). Because privacy is super important, especially when chatbots are dealing with sensitive student information. Educational institutions have to make sure they are robust data protection regulations like GDPR and FERPA. And of course, they need to be on guard against any potential cybersecurity threats ([Hamilton et al. 2021](#); [Yang et al. 2023](#); [Hasal et al. 2021](#)). Chatbots are revolutionizing education by enhancing personalized learning, accessibility, and administrative efficiency. They adapt content to match individual student needs, offering tailored feedback that fosters deeper understanding and engagement ([Kooli 2023](#); [Chang et al. 2023](#)). Their 24/7 availability ensures that students can access resources anytime, promoting flexible learning ([Mendoza et al. 2022](#)). Furthermore, chatbots support diverse learners by providing materials in multiple languages and catering to special needs, thereby promoting inclusivity ([Zobel et al. 2023](#)). On the administrative side, chatbots streamline processes by handling routine tasks

like scheduling and reminders, freeing educators to focus on teaching ([Majeed et al. 2023](#)). They provide instant responses to common queries, improving the efficiency of information dissemination ([Neupane et al. 2024](#)). Chatbots enhance student engagement through interactive tools like quizzes and simulations and offer academic and mental health support, guiding students through their educational journey([Wang et al. 2023](#); [Sweeney et al. 2021](#)). Finally, by tracking and analyzing performance data, chatbots offer valuable insights for educators to identify trends and predict student outcomes, thereby enabling proactive interventions ([Chang et al. 2023](#)).

REFERENCES

- Adamopoulou, Eleni, and Lefteris Moussiades. 2020. "Chatbots: History, Technology, and Applications." *Machine Learning with Applications* 2 (December):100006. <https://doi.org/10.1016/j.mlwa.2020.100006>.↵
- Annamalai, Nagaletchimee, Radzuwan Ab Rashid, Umair Munir Hashmi, Misrah Mohamed, Marwan Harb Alqaryouti, and Ala Eddin Sadeq. 2023. "Using Chatbots for English Language Learning in Higher Education." *Computers and Education: Artificial Intelligence* 5:100153. <https://doi.org/10.1016/j.caeai.2023.100153>↵
- Bratić, Diana, Marko Šapina, Denis Jurečić, and Jana Žiljak Gršić. 2024. "Centralized Database Access: Transformer Framework and LLM/Chatbot Integration-Based Hybrid Model." *Applied System Innovation* 7 (1): 17. <https://doi.org/10.3390/asi7010017>↵
- Chang, Daniel H., Michael Pin-Chuan Lin, Shiva Hajian, and Quincy Q. Wang. 2023. "Educational Design Principles of Using AI Chatbot That Supports Self-Regulated Learning in Education: Goal Setting, Feedback, and Personalization." *Sustainability* 15 (17): 12921. <https://doi.org/10.3390/su151712921>↵
- Chaves, Ana Paula, Jesse Egbert, Toby Hocking, Eck Doerry, and Marco Aurelio Gerosa. 2021. "Chatbots Language Design: The Influence of

- Language Variation on User Experience,” *ACM Transactions on Computer-Human Interaction* 29 (2): 1–38.
<https://doi.org/10.1145/3349537.3351901>
- Crompton, Helen, and Diane Burke. 2023. “Artificial Intelligence in Higher Education: The State of the Field.” *International Journal of Educational Technology in Higher Education* 20 (1): 22.
<https://doi.org/10.1186/s41239-023-00392-8>
- Drachsler, Hendrik, and Wolfgang Greller. 2016. “Privacy and Analytics.” In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge – LAK ’16*, 89–98. New York: ACM Press.
<https://doi.org/10.1145/2883851.2883893>
- Essel, Harry Barton, Dimitrios Vlachopoulos, Akosua Tachie-Menson, Esi Eduafua Johnson, and Papa Kwame Baah. 2022. “The Impact of a Virtual Teaching Assistant (Chatbot) on Students’ Learning in Ghanaian Higher Education.” *International Journal of Educational Technology in Higher Education* 19 (1): 57. <https://doi.org/10.1186/s41239-022-00362-6>
- Feijóo, Claudio, Youngsun Kwon, Johannes M. Bauer, Erik Bohlin, Bronwyn Howell, Rekha Jain, Petrus Potgieter, Khuong Vu, Jason Whalley, and Jun Xia. 2020. “Harnessing Artificial Intelligence (AI) to Increase Wellbeing for All: The Case for a New Technology Diplomacy.” *Telecommunications Policy* 44 (6): 101988.
<https://doi.org/10.1016/j.telpol.2020.101988>
- Felzmann, Heike, Eduard Fosch-Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2020. “Towards Transparency by Design for Artificial Intelligence.” *Science and Engineering Ethics* 26 (6): 3333–61.
<https://doi.org/10.1007/s11948-020-00276-4>
- Garcia Valencia, Oscar A., Supawadee Suppadungsuk, Charat Thongprayoon, Jing Miao, Supawit Tangpanithandee, Iasmina M. Craici, and Wisit Cheungpasitporn. 2023. “Ethical Implications of Chatbot Utilization in Nephrology.” *Journal of Personalized Medicine* 13 (9): 1363. <https://doi.org/10.3390/jpm13091363>

- Giffin, Kyle. 2024. "The Creator of Siri: "Make Your Prototype Magic"" <https://scet.berkeley.edu/the-creator-of-siri/> Gupta, Sambhav, and Yu Chen. 2022. "Supporting Inclusive Learning Using Chatbots? A Chatbot-Led Interview Study." *Journal of Information Systems Education* 33 (1): 98–108. <https://jise.org/Volume33/n1/JISE2022v33n1pp98-108.html>.
- Hamilton, Clovia, William Swart, and Gerald M. Stokes. 2021. "Developing a Measure of Social, Ethical, and Legal Content for Intelligent Cognitive Assistants." *Journal of Strategic Innovation and Sustainability* 16 (3) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3906226.[↵]
- Hasal, Martin, Jana Nowaková, Khalifa Ahmed Saghair, Hussam Abdulla, Václav Snášel, and Lidia Ogiela. 2021. "Chatbots: Security, Privacy, Data Protection, and Social Aspects." *Concurrency and Computation: Practice and Experience* 33 (19): 19–21. <https://doi.org/10.1002/cpe.6426>.[↵]
- Hidayat-ur-Rehman, Imdadullah. 2024. "Examining AI Competence, Chatbot Use and Perceived Autonomy as Drivers of Students' Engagement in Informal Digital Learning." *Journal of Research in Innovative Teaching & Learning* 1: 33–34. <https://doi.org/10.1108/JRIT-05-2024-0136>.[↵]
- Izadi, Saadat, and Mohamad Forouzanfar. 2024. "Error Correction and Adaptation in Conversational AI: A Review of Techniques and Applications in Chatbots." *AI* 5 (2): 803–41. <https://doi.org/10.3390/ai5020041>[↵]
- Javaid, Mohd, Abid Haleem, Ravi Pratap Singh, Shahbaz Khan, and Ibrahim Haleem Khan. 2023. "Unlocking the Opportunities through ChatGPT Tool towards Ameliorating the Education System." *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 3 (2): 100115. <https://doi.org/10.1016/j.tbench.2023.100115>[↵]
- Karandish, David. 2021. "7 Benefits of AI in Education." *The Journal*. <https://thejournal.com/articles/2021/06/23/7-benefits-of-ai-in-education.aspx>[↵]
- Khidir, Mohd Lutfi Bin Mohd, and Saiful Nizam bin Sa'ari. 2022. "Chatbot as an Educational Support System." *EPRA International Journal of*

Multidisciplinary Research (IJMR), May, 182–5.
<https://doi.org/10.36713/epra10328>

- Khurana, Anjali, Parsa Alamzadeh, and Parmit K. Chilana. 2021. “ChatrEx: Designing Explainable Chatbot Interfaces for Enhancing Usefulness, Transparency, and Trust.” In *2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 1–11. IEEE.
<https://doi.org/10.1109/VL/HCC51201.2021.9576440>
- Kooli, Chokri. 2023. “Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions.” *Sustainability* 15 (7): 5614. <https://doi.org/10.3390/su15075614>
- Kuhail, Mohammad Amin, Nazik Alturki, Salwa Alramlawi, and Kholood Alhejori. 2023. “Interacting with Educational Chatbots: A Systematic Review.” *Education and Information Technologies* 28 (1): 973–1018.
<https://doi.org/10.1007/s10639-022-11177-3>
- Kumar, Jeya Amantha. 2021. “Educational Chatbots for Project-Based Learning: Investigating Learning Outcomes for a Team-Based Design Course.” *International Journal of Educational Technology in Higher Education* 18 (1): 65. <https://doi.org/10.1186/s41239-021-00302-w>
- Labadze, Lasha, Maya Grigolia, and Lela Machaidze. 2023. “Role of AI Chatbots in Education: Systematic Literature Review.” *International Journal of Educational Technology in Higher Education* 20 (1): 56.
<https://doi.org/10.1186/s41239-023-00426-1>
- Mageira, Kleopatra, Dimitra Pittou, Andreas Papasalouros, Konstantinos Kotis, Paraskevi Zangogianni, and Athanasios Daradoumis. 2022. “Educational AI Chatbots for Content and Language Integrated Learning.” *Applied Sciences* 12 (7): 3239.
<https://doi.org/10.3390/app12073239>
- Majeed, Asim, Ayesha Asim, and Paul Bocij. 2023. “Reframing the Impact of Innovative Learning Technologies on University Students and Lecturers to Save Time and Improve Learning Challenges & Opportunities.” 7546–53. <https://doi.org/10.21125/edulearn.2023.1964>

- Mendoza, Sonia, Luis Martín Sánchez-Adame, José Fidel Urquiza-Yllescas, Beatriz A. González-Beltrán, and Dominique Decouchant. 2022. "A Model to Develop Chatbots for Assisting the Teaching and Learning Process." *Sensors* 22 (15): 5532. <https://doi.org/10.3390/s22155532>
- Merelo, Juan J., Pedro A. Castillo, Antonio M. Mora, Francisco Barranco, Noorhan Abbas, Alberto Guillén, and Olia Tsivitanidou. 2023. "Chatbots and Messaging Platforms in the Classroom: An Analysis from the Teacher's Perspective." *Education and Information Technologies*, May. <https://doi.org/10.1007/s10639-023-11703-x>
- Mitan, Electra. 2024. "Artificial Intelligence Role and Impact on Education." 8035–45. <https://doi.org/10.21125/edulearn.2024.1898>
- Mumtaz, Shazia. 2000. "Factors Affecting Teachers' Use of Information and Communications Technology: A Review of the Literature." *Journal of Information Technology for Teacher Education* 9 (3): 319–42. <https://doi.org/10.1080/14759390000200096>
- Neupane, Subash, Elias Hossain, Jason Keith, Himanshu Tripathi, Farbod Ghiasi, Noorbakhsh Amiri Golilarz, Amin Amirlatifi, Sudip Mittal, and Shahram Rahimi. 2024. "From Questions to Insightful Answers: Building an Informed Chatbot for University Resources," May. <http://arxiv.org/abs/2405.08120>
- Okonkwo, Chinedu Wilfred, and Abejide Ade-Ibijola. 2021. "Chatbots Applications in Education: A Systematic Review." *Computers and Education: Artificial Intelligence* 2: 100033. <https://doi.org/10.1016/j.caeai.2021.100033>.
- Chisom, Onyebuchi Nneamaka, Chika Chioma Unachukwu, and Blessing Osawaru. 2024. "Review of AI in Education: Transforming Learning Environments in Africa." *International Journal of Applied Research in Social Sciences* 5 (10): 637–54. <https://doi.org/10.51594/ijarss.v5i10.725>
- Pashler, Harold, Mark McDaniel, Doug Rohrer, and Robert Bjork. 2008. "Learning Styles." *Psychological Science in the Public Interest* 9 (3): 105–19. <https://doi.org/10.1111/j.1539-6053.2009.01038.x>

- Paullada, Amandalynne, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. 2021. "Data and Its (Dis)Contents: A Survey of Dataset Development and Use in Machine Learning Research." *Patterns* 2 (11): 100336. <https://doi.org/10.1016/j.patter.2021.100336>
- Pérez, José Quiroga, Thanasis Daradoumis, and Joan Manuel Marquès Puig. 2020. "Rediscovering the Use of Chatbots in Education: A Systematic Literature Review." *Computer Applications in Engineering Education* 28 (6): 1549–65. <https://doi.org/10.1002/cae.22326>
- Pittman, F. Paul, Abdul Hafiz, and Andrew Hamm. 2023. "Data Protection Laws and Regulations USA 2023–2024." ICLG. <https://iclg.com/practice-areas/data-protection-laws-and-regulations/usa>
- Ralston, Kennedy, Yuhao Chen, Haruna Isah, and Farhana Zulkernine. 2019. "A Voice Interactive Multilingual Student Support System Using IBM Watson." In *2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1924–9. IEEE. <https://doi.org/10.1109/ICMLA.2019.00309>
- Rane, Nitin. 2023. "Chatbot-Enhanced Teaching and Learning: Implementation Strategies, Challenges, and the Role of ChatGPT in Education." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4603204>
- Rathore, Bharati. 2022. "Exploring the Potential Impacts of Chatbot Software/Apps (ChatGPT) on Education: Benefits, Drawbacks, and Future Prospects." <https://ijmirm.com/index.php/ijmirm/article/view/18>
- Sebastian, Glorin. 2023. "Exploring Ethical Implications of ChatGPT and Other AI Chatbots and Regulation of Disinformation Propagation." *SSRN Electronic Journal* 1: 55–57. <https://doi.org/10.2139/ssrn.4461801>
- Sweeney, Colm, Courtney Potts, Edel Ennis, Raymond Bond, Maurice D. Mulvenna, Siobhan O’neill, Martin Malcolm, et al. 2021. "Can Chatbots Help Support a Person’s Mental Health? Perceptions and Views from Mental Healthcare Professionals and Experts." *ACM Transactions on Computing for Healthcare* 2 (3): 1–15. <https://doi.org/10.1145/3453175>

- Tahiru, Fati. 2021. "AI in Education." *Journal of Cases on Information Technology* 23 (1): 1–20. <https://doi.org/10.4018/JCIT.2021010101>
- Tlili, Ahmed, Boulus Shehata, Michael Agyemang Adarkwah, Aras Bozkurt, Daniel T. Hickey, Ronghuai Huang, and Brighter Agyemang. 2023. "What If the Devil Is My Guardian Angel: ChatGPT as a Case Study of Using Chatbots in Education." *Smart Learning Environments* 10 (1): 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Wang, Ting, Brady D. Lund, Agostino Marengo, Alessandro Pagano, Nishith Reddy Mannuru, Zoë A. Teel, and Jenny Pange. 2023. "Exploring the Potential Impact of Artificial Intelligence (AI) on International Students in Higher Education: Generative AI, Chatbots, Analytics, and International Student Success." *Applied Sciences* 13 (11): 6716. <https://doi.org/10.3390/app13116716>
- Wollny, Sebastian, Jan Schneider, Daniele Di Mitri, Joshua Weidlich, Marc Rittberger, and Hendrik Drachsler. 2021. "Are We There Yet? – A Systematic Literature Review on Chatbots in Education." *Frontiers in Artificial Intelligence* 4 (July). <https://doi.org/10.3389/frai.2021.654924>
- Xue, Jintang, Yun-Cheng Wang, Chengwei Wei, Xiaofeng Liu, Jonghye Woo, and C. -C. Jay Kuo. 2023. "Bias and Fairness in Chatbots: An Overview," September. <https://doi.org/10.48550/arXiv.2309.08836>
- Yang, Jing, Yen-Lin Chen, Lip Yee Por, and Chin Soon Ku. 2023. "A Systematic Literature Review of Information Security in Chatbots." *Applied Sciences* 13 (11): 6355. <https://doi.org/10.3390/app13116355>
- Younis, Hussain A., Taiseer Abdalla Elfadil Eisa, Maged Nasser, Thaeer Mueen Sahib, Ameen A. Noor, Osamah Mohammed Alyasiri, Sani Salisu, Israa M. Hayder, and Hameed AbdulKareem Younis. 2024. "A Systematic Review and Meta-Analysis of Artificial Intelligence Tools in Medicine and Healthcare: Applications, Considerations, Limitations, Motivation and Challenges." *Diagnostics* 14 (1): 109. <https://doi.org/10.3390/diagnostics14010109>
- Zemčík, Mgr. Tomáš. 2019. "A Brief History of Chatbots." *DEStech Transactions on Computer Science and Engineering*, no. aicae (October).

<https://doi.org/10.12783/dtcse/aicae2019/31439>.↵

Zhang, Juliana J. Y., Asbjørn Følstad, and Cato A. Bjørkli. 2023. “Organizational Factors Affecting Successful Implementation of Chatbots for Customer Service.” *Journal of Internet Commerce* 22 (1): 122–56.

<https://doi.org/10.1080/15332861.2021.1966723>.↵

Zobel, Theresa, Hendrik Steinbeck, and Christoph Meinel. 2023. “Towards Inclusive Education: A Review of and Guide to Accessible Features in Chatbots for MOOCs.” In *2023 IEEE Learning with MOOCS (LWMOOCS)*, 1–5. IEEE.

<https://doi.org/10.1109/LWMOOCS58322.2023.10306062>↵

19 AI, Consent, and GDPR

Ensuring Privacy and Rights in the Digital Age

*Kittisak Jermsittiparsert, Saurabh Chandra,
and Bhumika Muchan*

DOI: [10.1201/9781003668657-19](https://doi.org/10.1201/9781003668657-19)

19.1 INTRODUCTION

The fundamental building block of European Union (EU)’s digital future is trust, most importantly with respect to personal data of individuals. Understanding the need to modernize data laws, the European Commission set off on a journey in January 2012 in revitalizing data protection regulations throughout the EU. The General Data Protection Regulation (GDPR) contains many detailed rules and requirements for obtaining consent. It is crucial to determine when consent is the right “lawful basis.” In the 21st century, our society demands the use of notice and choice. We want to know that entities who hold our personal information are first communicating it to us in some way. Then we want the luxury of making an informed decision: Do we consent? Or do we withdraw? (But those are just two options out of many. A consent request could carry any number of outcomes in this way.)

This journey led to GDPR which came into force on May 25, 2018. This tremendous regulation marked a huge separation from the obsolete data protection directives of the 1990s, and a shift in stance toward a stronger and more comprehensive way of data privacy in the digital age.

The GDPR is groundbreaking not just because it introduces substantial reforms to the EU data protection regime, but because of its worldwide extra-territorial impact. Organizations worldwide that were conducting business with the EU were forced to make drastic changes to their operational and financial planning to comply with GDPR requirements. Even though GDPR was an EU regulation, through its extra-territorial reach, organizations outside of the EU were compelled to adopt the technological requirements of GDPR. Through its application beyond the territory and nationally adopted law, GDPR can be regarded as the benchmark legislation that has created a globally recognized standard of privacy conditions (Voigt & Von dem Bussche, 2017).

Not simply an accountability obligation, instead, the GDPR instates a principle transformation in the data safety landscape. The Regulation pertains to determining the sustainable change introduced by means of rapid knowledge expansion and globalization of economy by means of the reinforcement of the privilege of data subjects and the accessibility of data subjects over their personal information. Central to the GDPR's ethos is the principle of empowering individuals to assert control over their data, thereby fostering trust in digital transactions and bolstering confidence in the digital economy.

This paper aims at undertaking a critical evaluation of consent in the context of the GDPR discussing its complexity, issues, and consequences on organizational entity in the context of the current complex data protection landscape. Thus, analyzing the legal provisions of GDPR, the current theoretical and practical discourses, it aims at describing the complex and significant meaning of consent in the GDPR context and its enabling role in the processes of compliance, transparency, and individuals' protection in the digital environment.

19.2 LITERATURE REVIEW

The GDPR is a complex of rules on data protection in relation to unauthorized processing of personal data of citizens of the EU. It relates to virtually all activities concerning a person's digital persona and access to personal and organizational resources, especially in cyberspace, to safeguard the rights of data subjects (Cormack, 2020). The GDPR regulates processing operations and brings radical changes to IT processes; it does affect all the member states of the EU (Murphy, 2018). It affects the research undertakings in aspects such as consent, contractual provisions, data protection, and notification of the breach (Murphy, 2018). The GDPR has helped in sensitization on data protection and hastened efforts toward compliance; however, it has also elicited worry on constitutional rights ([Tsekoura & Panagopoulou, 2020](#)).

The GDPR has thus altered consumer consent and a firm's actions in many ways as an outcome. [Matos and Adjerid \(2021\)](#) showed that GDPR-consent made consumers opt-in for the firms' data types and uses, thereby benefiting the firms. However, consumers did not allow permissions for their personal information to be used for sensitive or related purposes. This underlines the requirement for the clear and effective expression of data allowances currently stressed by [Rodríguez Ayuso \(2022\)](#), especially when dealing with the data related to minors. Even though consent is highlighted as one of the key legitimating grounds under the GDPR, [Manganello \(2020\)](#) questions the very quality of consent's existence within the digital space and calls for further scholarly investigation of the topic and possible changes in the EU approach.

GDPR is one of the major advances in data protection law with a duty to protect the privacy of citizens and to frame the information society ([Magalhaes, 2021](#)). It brings into operation a series of rights for data subjects, such as "the right to information, right of access, right to rectification, right to erasure, right to restriction, right to data portability, and right to object." The GDPR also responds to the problems of the so-called information society and attempts to reconcile technology and rights

(Mitrou, 2017). Though data protection has been practiced alongside the GDPR, the new provision of Article 8 in GDPR still remains a subject for researchers (Pudas, 2018).

The recent enforcement of GDPR has proven rather difficult for organizations mainly concerning erasure (Prethus, 2018), interdepartmental cooperation ([Mendoza, 2019](#)) and more so on resource limitations, especially for SMEs ([Freitas & Silva, 2018](#)). To overcome these challenges, a structured approach has been suggested to GDPR compliance that implies the modeling of the business processes and major compliance aspects (Capodieci, 2020). The overall effectiveness of the GDPR with the different organizations is depicted by these studies as they emphasize on the importance of the establishment of a systematic and synergistic approach to the implementation of GDPR on the management of data, communication across departments, and resource utilization.

It is therefore correct to say that the idea of consent as provided in GDPR is grey and would continue to develop over time. According to [Carolan \(2016\)](#), there are three models of the user consent in EU jurisprudence with the active consent model being favored. However, this model may not explain all the situational factors that may elicit consent. [Matos and Adjerd \(2021\)](#) explains that effective consent can enhance the opt-in action of the consumers although the utilization of permissions for massive and simple data may remain limited. This implies that even though there is a probability of enhanced consent benefiting the protection of the privacy, it cannot completely solve the problem. Continuing such analysis, [Manganello \(2020\)](#) also expresses doubts about the relevance of consent in the digital context, emphasizing the difference between the theoretical concept of consent and one's actions associated with autonomy. This highlights the link between the theoretical idea of consent and the real-world actions that express personal autonomy. In the context of European data protection, consent is generally understood as a key mechanism through which individuals exercise control over their own data and choices.

Concerning consent, GDPR gives importance especially to the users: "Freely given, specific, informed, and unambiguous." [Trakman et al. \(2020\)](#)

and Vegas et al. (2020) have noted various hurdles and factors affecting the process of getting this consent and its management. [Trakman et al. \(2020\)](#) recognizes a need for legal convergence in the context of consent, while Vegas et al. (2020) has classified the issue and offered a consent solution: a decentralized yet collaborative platform. Pudas (2018) gives a historical view and origin of data protection as well as the integration of Article 8 to the GDPR, asserting the problem of consent. Altogether, these works demonstrate the relevance of user consent in data protection and the necessity of having proper legal means by which compliance with GDPR could be reached.

Some of the challenges include the comprehensiveness of GDPR, which covers most processes; the fact that most recommendations given are qualitative; and the issue of data networks due to increased complexity (Sirur, 2018). Comparing the GDPR with data protection laws in third world countries, such as Japan, intensifies these challenges ([Kato, 2020](#)). Yet another complication that GDPR brings is data anonymization, which is also important in the case of text data ([Francopoulo & Schaub, 2020](#)). Altogether, these studies provide a rather complex and challenging image of gaining an authentic and permissible agreement under the GDPR framework. India's legal safeguarding of data is relatively advanced, with the Information Technology Act 2002 and its modifications providing minimal protection (Verma, 2014). Drafted based on the GDPR, the Personal Data Protection Bill, 2018, further strengthens these shields ([DevaPrasad & SuchithraMenon, 2020](#)). However, some questions relate to the efficiency of this framework, primarily in terms of the database/data protection demarcation (Raju, 2017). A comparative analysis of data protection legislation in India and other countries explains the need to have these two categories of data and the inadequacies of the IT Act to provide adequate protection to personal information (Garg, 2011). Therefore, cultural, economic, and political factors inform the protection of data laws in India (Basu, 2010). However, the country has not adequately tackled equality, and it still lacks a unified set of laws to govern data protection (Greenleaf, 2011). According to [Greenleaf \(2020\)](#), the Personal Data

Protection Bill, 2019 drafted by the committee led by a former Supreme Court Justice Srikrishna, has a large percentage of the rights and duties of the other leading international data protection legal instruments but lacks enforceability. Given this, it appears that the GDPR consent is multi-layered, necessitating multi-level compliance. [Daoudagh et al. \(2021\)](#) has proposed a privacy-by-design approach that takes the GDPR text and makes it machine-readable, followed by generating access control policies from it. [Kostić and Vargas Penagos \(2017\)](#) emphasizes the concept of consent in the paper, thereby highlighting the criteria for identifying consent that is “freely given.” In the article “Management of Consents and Revocations: A Practical Workflow Approach,” Besik (2020) describes the practical orientation of the workflows for managing consent and revocation in an organization. [Pandit et al. \(2019\)](#) introduces a consent ontology called GConsent, where Pandit models the information concerning consent under the GDPR; this way, Pandit instills the applicability of semantic web technologies. Overall, the presented research demonstrates the necessity of developing a systematic and flexible approach toward consent management according to the GDPR. It is essential to describe consent and consider GDPR as one of the seven legal bases for data processing ([Pandit et al., 2019](#)). It may be difficult to validate patients’ consent, which is often vague and uninformed ([Drozd & Kirrane, 2020](#)). Organizations can employ a semantic model of consent to overcome this issue and comply with GDPR standards (Fatema, 2017). In summary, the GDPR has placed a greater emphasis on consent, transparency, ensuring that individuals are informed and aware of all processes related to their consent ([Breen et al., 2020](#)). Hence, it is impossible not to note the importance of understanding consent when considering GDPR for data controllers and data subjects. Researchers have extensively studied the efficiency of consent mechanisms under GDPR. [Oh et al. \(2021\)](#) developed a method to assess a website’s compliance with the GDPR consent requirements, while [Machuletz and Böhme \(2019\)](#) reported that if consent prompts have a default button, users may approve more purposes than they intended. Consistent with this observation, [Matos and Adjerid \(2021\)](#) has established that GDPR consent

enhances consumers' opt-in rate and firm performance, especially in the marketing context. Besik (2020) proposed the use of workflows as a solution, where the organization assumes the role of monitoring and compliance with consent and revocation. Collectively, these studies demonstrate the issues that may arise and the potential consequences for consent mechanisms under GDPR. Understanding the challenges of working with GDPR consent is crucial, as various models and solutions enrich its definition. In the process of developing an opt-out model for service-level consent in public health research, [Howarth et al. \(2022\)](#), provides insight into the perspective of both the protection of individuals's rights and the mandate for data analysis. The concepts of fairness, transparency, and accountability recognized under the GDPR particularly buoy this approach. In line with this, [Abdulah and Aseri \(2020\)](#), affirms that GDPR has international implications and has made it necessary to look for technical solutions in seeking consent for data, thus acknowledging the policy as having an improving effect on data protection and security. Therefore, these studies underscore the importance of considering practical and ethical issues when implementing GDPR consent procedures. Thus, the GDPR is one of the most complex laws present today, as it regulates nearly all the processes associated with using, handling, or encountering any data. This is especially the case when it comes to consent within the country's culture. The GDPR uses the term "consent" in various situations with actual or potential impact. Additionally, it outlines the concept of an individual providing their personal consent for data use.

This project's research questions delve into the intricacies of what consent means in the context of the GDPR.

This research aims to address the following inquiries:

1. How vital is the understanding of consent in understanding the concept of GDPR?
2. What are the specific elements that make up consent under the GDPR, and why are they important for upholding the law's principles and requirements?

3. How much do the subjects of individual consent really understand about the way in which their personal data is being used?
4. What effects do technological advancements like AI data processing and automatic consent tools have on the way people give and get consent in the GDPR era?

The concept of GDPR is wide and has many subtopics that need to be understood. In the present paper, the prime focus is put into the study of the concept of “Consent” under GDPR. It studies the need, elements, and importance of GDPR. The paper studies the same and all other topics are excluded from the study.

19.3 CONSENT AND GDPR

Article 4 of GDPR defines “consent” as “any freely given, specific, informed, and unambiguous indication of the data subject’s intentions by which he or she, through a statement or clear affirmative action, indicates agreement to the processing of personal data relating to him or her.” The following conditions must be met for consent to be provided: “it must be freely given, precise, informed, unmistakable, and unequivocal affirmative action.” In the absence of coercion or undue influence from the data controller, consent is not considered freely given. The permission would not be valid if the data controllers refused to provide the data user with access to their information or offered a downgraded version of their service in exchange for their refusal or later withdrawal of consent ([Clarke et al., 2019](#)).

The second form of permission required by the GDPR is specific, informed, and unambiguous. It prevents data controllers from using unclear legalese that can cause users of the data to mistakenly accept the terms of consent. No longer are unclear statements permitted from data processors. For instance, we might process your personal information. The statement should instead be more explicit and clarify what it desires. The GDPR specifies explicit affirmative action as the third level of consent. The third

type of consent demands that the user opt-in according to their conditions; no secret boxes are permitted. For instance, the person in charge of the data must tell the user about cookies and get their permission before using them. Additionally, the GDPR allows data users the flexibility to withdraw their consent at any moment.

19.3.1 NEED FOR GDPR

There is no escaping the fact that personal information is valuable. Data enables the development of business models, customer knowledge, successful marketing efforts, and the creation of new products and services. However, there is a requirement for safe use based on accepted guidelines, much like with many other materials. In the past few years, Facebook, eBay, Equifax, and Uber have all made news for scandals involving personal data breaches. Personal data belonging to hundreds of millions of people, including social security numbers, addresses, credit ratings, etc., was exposed. The GDPR not only makes it obvious that a person owns their personal data, but it also threatens to inflict heavy fines on businesses that break the law. In Europe, data protection and privacy are seen as essential elements of a long-lasting democracy. These requirements will be protected by the GDPR, which is an improvement over the old EU data privacy law.

19.3.2 NEED FOR CONSENT OF GDPR

Many individuals wrongly believe that organizations must obtain consent before processing personal data; however, consent is simply one of six legal justifications, and you should only use it if all other options have been exhausted.

The other lawful grounds are:

- *A Contract with the Individual:* To meet an obligation under an agreement with an employee, or to supply work-related items or services at the request of an employee.

- *Compliance with a Legal Obligation:* When a specific purpose for processing data is mandated by law.
- *Vital Interests:* For instance, when handling data is necessary to preserve someone's physical safety or life.
- *A Public Task:* to finish off work or official duties that are in the public interest, for instance. This often applies to public institutions including government agencies, educational institutions, hospitals, and the police.
- *Legitimate Interests:* When a private company has a compelling reason to process an individual's private information without getting their permission but doing so would not harm their rights or freedoms.

However, there may be circumstances in which consent is the most suitable legal basis, so one must be conscious of one's responsibilities ([Tikkinen-Piri et al., 2018](#)).

19.3.3 VALID CONSENT

When it comes to determining what constitutes legitimate permission, the GDPR is very clear:

The term “consent” refers to “any freely given, precise, informed, and unequivocal expression of the data subject's intentions through which he or she, through a statement or through a blatant affirmative action, expresses assent to the processing of personal data related to him or her.”

For consent to be deemed legitimate, the following four conditions must be satisfied:

- *Freely-Given:* This indicates that the user has an option as to whether or not to provide personal information. An internet service cannot be used or navigated without explicit consent.

- *Specific*: Visitors must expressly consent before their information is used for any reason, including sharing for analytics or marketing.
- *Informed and Unambiguous*: Each sort of information you gather and how it is utilized should be disclosed to users. Once users have been informed, consent must be clear and understandable.
- *Clear Affirmative Action*: Any action that establishes consent, such as checking a box or clicking a button, is a clear, affirmative action ([Custers et al., 2019](#)).

Key Elements of Valid Consent under GDPR		
Element	Description	Example
Freely Given	The data subject has a genuine choice and can refuse or withdraw consent without detriment.	A user can choose not to subscribe to a newsletter without losing access to the main service.
Specific	Consent must be given for one or more specific purposes.	A user agrees to data processing for receiving newsletters but not for targeted advertising.
Informed	The data subject must be informed about the data processing activities clearly and understandably.	A pop-up informs users that their data will be used for improving service features, with an option to read the full privacy policy.
Unambiguous	Consent must be given through a clear affirmative action.	A user ticks a checkbox to agree to data processing terms, rather than through pre-ticked boxes or inactivity.

All users will be fully informed of how their information is collected and used, as well as the options available to them regarding their own privacy, because of this precise and comprehensive definition of consent, which is

intended to do away with earlier ambiguous or implied methods of obtaining consent. It is reasonable that consumers and their elected officials want to have more control over personal information now that advertising cookies and user tracking are ubiquitous.

19.4 TYPES OF CONSENT

19.4.1 ORAL CONSENT

The GDPR articles do not expressly forbid oral consent. Instead, the GDPR just mandates that there be enough proof of permission in the form of documentation. The answer to the question is provided by Recital 32, which states that an oral declaration may be acceptable as a clear affirmative act that is adequate for consent.

19.4.2 IMPLIED CONSENT

When a user does not expressly accept the processing of their personal data, but consent is impliedly granted by their behavior, this is known as implied consent. This kind of consent is frequently utilized in marketing, as businesses gather data about individuals through their online browsing patterns and use it for individualized advertising.

Without the individual's express consent, this is done. In this case, consumers must opt out to stop the use of their personal information.

19.4.3 EXPLICIT CONSENT

A person must be given the chance to approve the collection, use, or sale of their data to receive explicit consent. This indicates that the person is aware of the request to use or share their personal information and that they consent to it. Companies must make sure they are transparent about the data they gather and why it is being collected. Typically, explicit consent is given orally or in writing.

19.4.4 ACTIVE CONSENT

Explicit consent can take the form of active consent. By using active behaviors like clicking a button or checking a box, it is possible to obtain an individual's express consent to utilize cookies on their browser. The aim of the cookies that the data subjects agree to is clearly disclosed to them (Klimas, 2008).

19.4.5 PASSIVE CONSENT

When users are not given the option to actively decide whether to accept cookies, passive consent results. Users of data consent without using a button or checkbox. Some websites capture user information through passive means, such as banner ads. Your permission to the collection and use of your data is demonstrated by your continued use of the website. Passive consent is typically regarded as a form of implied permission.

19.4.6 OPT-OUT CONSENT

By choosing not to decline the opt-out consent banner when it is displayed to users, they have granted their approval. Consent is given if it cannot be clearly withheld. Opting out is accomplished in writing. In this strategy, a company will give users the option to opt out via a pop-up window, banner web page, or other appropriate interface.

A well-liked method of seeking authorization to use customers' personal data for other reasons, such as targeted advertising, is the opt-out consent model. This is so that people can prevent their data from being gathered, used, sold, or revealed ([van Erkel et al., 2024](#)).

19.4.7 OPT-IN CONSENT

To do this, data controllers must opt in and receive the explicit consent of the data subjects. Unless they voluntarily consent, the user is by default opt-out of tracking and data processing.

Users are given the option to accept or reject cookies when they see an opt-in banner informing them that they are being used.

19.4.8 INFORMED CONSENT

Businesses make sure to fully disclose all potential outcomes to users and secure their approval before moving forward. Giving someone all the knowledge they need to make an informed decision about a particular issue is the aim of informed consent.

19.4.9 ELEMENTS OF VALID CONSENT

The term “consent” of the data subject is defined in Article 4(11) of the GDPR as “any freely given, explicit, informed and unambiguous expression of the data subject’s desires by which he or she, by a statement or by a clear affirmative action, expresses assent to the processing of personal data relating to him or her.”

If we break down that definition, the four components of a valid consent under the GDPR are as follows: (1) freely offered; (2) particular; (3) knowledgeable; and (4) unambiguous. The consent would be deemed void if any one of these components were lacking.

The consent of the data subject must first be freely granted. The WP29 interprets this to mean that the data subject must be given “genuine choice and control” over a “granular” range of processing purposes and have the ability to refuse or later revoke their consent in a way that is just as simple as giving it in the first place. Additionally, when there is a “clear imbalance between the data subject and the controller” (such as in the context of employment) or if fulfillment of a contract is contingent on assent to processing that is not necessary for that performance, consent is often not regarded as freely given ([Santos et al., 2021](#)).

The second need is that the consent be particular, meaning that it must be connected to “one or more specified goals” and that the data subject must

have a choice regarding each. The WP29 states that to meet this requirement, data controllers must make sure of three things:

1. A purpose specification to prevent function creep,
2. Specificity in consent requests, and
3. A clear difference between consent-related information and unrelated information.

The third requirement is informed consent. To ensure that the data subject knows the options available to them and what they would be agreeing to, sufficient information about the processing must be disclosed to them “in an understandable and easily accessible manner, using clear and plain language” prior to getting their consent. The WP29 makes a strong case for the significance of this component, stating that “if the controller does not disclose accessible information, user control becomes illusory and consent will be an invalid foundation for processing.”

Finally, the consent has to be clear. The individual whose personal data is being processed must express their intentions “by a statement or by a clear affirmative action” indicating their consent. This implies that it must always be “provided through an active motion or declaration” and that it must be “clear that the data subject has agreed to the specific processing,” according to WP29. Therefore, “silence, pre-ticked boxes, or inaction should not constitute agreement,” according to Reason 11 (Romansky, 2022).

19.4.10 EXPLICIT CONSENT

On occasion, express consent may be necessary. Specific consent is one of the exceptions to the GDPR’s prohibition on the processing of particular categories of data, and it is one of the derogations for the transmission of personal data to third countries. It can also be employed to substantiate judgments about data subjects that are predominantly based on automated processing, such as profiling, and that have legal repercussions for or have a

significant impact on the data subject (another term for “automated individual decision-making”).

19.4.11 DEMONSTRATING CONSENT

Data controllers must also be able to “show that the data subject has consented to the processing of his or her personal data,” as required by Art 7(1) of the GDPR. “Controllers are permitted to establish techniques to comply with this requirement in a way that is fitting in their regular activities,” the WP29 states. To comply with this duty, data controllers should not, however, gather more information than is essential or keep it for longer than is necessary. 16 — i.e., while considering how to show compliance with the GDPR, privacy professionals should have the concepts of data reduction and storage limits in the back of their minds.

19.4.12 FREE CONSENT

The law in the United States has long battled with the issue of whether customer permission has been freely supplied in a take-it-or-leave-it service contract. The GDPR adopts a complex strategy that very well may do away with this kind of consent.

It will take some time to determine how strictly the GDPR will be enforced in this regard. However, many companies will have to err on the side of caution if they do not want to run the danger of being subject to a future enforcement action that makes them destroy the data they wrongfully collected.

According to a draft paper published by the Article 29 Working Party, consent must involve “genuine choice and control.” In other words, consent is not freely granted if it cannot be withdrawn. Additionally, it is doubtful that non-negotiable consent contained in a contract’s terms and conditions will be freely granted. A smartphone app that needs GPS location services for behavioral advertising is the Working Party example. In other words, consent is not freely provided if it is required as payment for a service. Only

when there is a “direct and objective link” between processing and performance may non-negotiable consents be considered legitimate. This will only be the case when it is absolutely necessary for the performance.

It is even more challenging to establish consent as a legal basis for processing when the controller and data subject are in unequal positions of authority. The majority of employee data processing at work cannot be justified by consent, according to the Article 29 Working Party, as employees would not feel free to object without running the danger of repercussions. The ability of public authorities to lawfully seek consent must also be taken into account.

19.5 THE PRACTICAL APPLICATION OF GDPR CONSENT

The best effective privacy programs have long understood that consumers’ consent should not be an all-or-nothing choice. Customers desire alternatives that are more specific to their preferences, such as the type of communication (email or SMS) or frequency (once a day or once a month). When a firm does not offer them adequate options for interaction, they are much more likely to say “nothing.” Additionally, GDPR supports proper consent in Article 8 by requiring explicit, verifiable, specific consent that is freely provided and simple to withdraw. Last but not least, firms that lose current consumers due to improperly aggressive marketing activities will be at a competitive disadvantage due to the expense of obtaining a customer.

19.5.1 BROAD CONSENT

A brief description of the wide consent practice is required to get things started. A pertinent overview can be provided from the three perspectives of broad consent’s method and application, advantages and justifications, and acceptance and support. Researchers may be required to obtain consent to perform genetic research using biological samples from competent people and associated data. One form of consent that researchers could request is

broad consent. A single interaction between the research subject and the genetic researcher, or the genomic research infrastructure, is sufficient for broad consent. According to the circumstances, the study participant may be asked permission for three different forms of activities throughout this engagement:

1. Their biological sample will be removed, and pertinent information about their health, lifestyle, and genomics will be gathered.
2. The collection, preservation, and, if necessary, processing of their biological specimen and related data for use in scientific inquiry.
3. At the time of collection, it was not clear how their collected biological sample and related data would be used in the future.

19.5.2 CASE LAWS

The concept of consent, as defined by the GDPR, holds significant importance within the realm of privacy and data protection. This signifies the autonomy of the individual with regard to their personal data, guaranteeing that the processing of their data is solely upon explicit consent. The intricacies of consent application and interpretation under the GDPR, nevertheless, are manifold and have been illuminated by several case laws.

19.5.2.1 Planet 49 GmbH

The case of Planet49 GmbH was significant because it clearly laid out the European Court's teeth clamping on "consent" in the GDPR. The CJEU found the pre-checked boxes a highly problematic form of consent for something as significant as giving a website the Okay to install cookies on a user's device. The CJEU ruling clarified that a user's action must signify their consent to be valid, and consent given under an unclear scenario cannot be U-turned later. This was seen in many circles as the case that guaranteed that "consent" would be the vaguest "Grounds A" for a legal battle (Wiedemann, 2020).

19.5.2.2 Schrems II

The Schrems II ruling mainly dealt with data flows from the EU to the United States. Nevertheless, the judgment also had massive implications for consent as a legal basis for data transfers. The situation with consent was laid bare: without proper safeguards against abusive governmental surveillance that is at all levels indistinguishable from privacy invasion, consent is worthless. This is true for consent under standard contractual clauses and for consent under the Privacy Shield decision of the European Commission. And it is also true for consent in Covered Data Transfers under GDPR Article 49(1)(a) (Tzanou, 2020).

19.5.2.3 Fashion ID

The case of Fashion ID revolved around the thorny problem of Germany's Süddeutsche Zeitung. When its news website updated, it hosted an embedded 'Like' button from Facebook. The result was that Süddeutsche collected as well as Facebook the user's IP address and browser signature, which was transmitted to the latter, information that could be personal without the user necessarily knowing it. The CJEU's decision called into question the use of such buttons as well as similar social media tools whose real purpose might also be to collect user data—as platforms have long done. This ruling places the onus on platforms (and on other joint controllers) to ensure that users' consent holds up under legal scrutiny.

19.5.2.4 Google Spain SL, Google Inc. v Agencia Española de Protección de Datos (AEPD), Mario Costeja González

This instance, despite coming before the GDPR, is noteworthy for apprehending how consent has developed under EU data protection law. The "right to be forgotten" is a requirement that search engines, such as Google, must typically comply with when people want certain kinds of information about themselves to be removed from the search engine's results. It is not so much that the DMA instance originally established what people often call the "right to be forgotten"; it is more an instance that

established the significance of consent when individuals cede their personal data to search engines and other entities.

19.5.2.5 Facebook Belgium

Belgium's Data Protection Authority contested Facebook's use of cookies and social plugins for not gaining the required valid consent from users. The Belgian court later sought a ruling from the CJEU. The CJEU's decision is in keeping with earlier cases it had on the topic, and it seems to be doubling down on how and when clear, informed, and freely given consent is required, especially when it comes to tracking and profiling by online service providers.

19.5.2.6 Orange Romania SA

The CJEU considered the case of Orange Romania SA, which raised the question of whether a company could require customers to allow the company to store copies of their identity documents as a condition of concluding a contract with that company. The CJEU decided that if the customer did not have a real choice to say no and could not conclude the contract without first consenting, then any such consent was not freely given.

19.5.2.7 Bundesverband der Verbraucherzentralen und Verbraucherverbände—Verbraucherzentrale Bundesverband e.V. v. Planet49 GmbH

The CJEU has echoed its position on consent obtained via pre-ticked boxes, a situation that is very similar to the earlier Planet49 GmbH case decided by the same court. The CJEU has now said, in a case concerning a German football club, that the use of pre-ticked checkboxes does not meet the requirement of Article 7(2) of the GDPR for consent to be “freely given, specific, informed, and unambiguous.” For consent to be valid under the GDPR, the CJEU added, what matters is whether the user's action “clearly

and unambiguously expresses the user's agreement to the processing of personal data.

19.5.2.8 Düsseldorf Kreis (DPA Decision)

German data protection authorities—working together on a team called Düsseldorf Kreis—issue decisions and guidelines on all kinds of privacy issues. Some appeals have gone before them, and they have issued decisions in several cases. Generally, the German authorities have been paying a lot of attention to what forms of consent are and are not legally a third rail in privacy practices. They have not said a whole lot about what specifically in the employment context they are good with and what in particular makes them nervous. That has been the role of numerous stories in the security press in the past few years and warnings from the public speaking side of DPAs.

19.5.2.9 Belgian Data Protection Authority v. Facebook Ireland Ltd.

The Belgian Data Protection Authority (DPA) has ruled that Facebook did not get proper permission from people to collect their data—a decision that means the company could pay a fine of up to €100 million (\$124 million). In a case that has been ongoing for months (and which picked up attention around the same time as the Cambridge Analytica uproar), the DPA ruled that Facebook didn't do enough to inform people in Belgium about the data it was collecting. It also ruled that the company's user-tracking mechanisms (like cookies and social plugins) were “time bombs” that could potentially turn Facebook into a full-blown Big Brother.

19.5.2.10 IAB Europe and the Transparency and Consent Framework

IAB Europe's Transparency and Consent Framework (TCF), which is widely applied in the online advertising field to deal with GDPR compliance, has lost another strong supporter. The TCF's lack of a valid legal basis for the online processing of internet users' personal data potentially puts ad industry actors in breach of the principles of the EU's

GDPR, the Belgian DPA said in its 15 July decision, finding against the TCF. In response to adtech industry players, a great many of the internet companies, must promptly initiate immediate cessation of their OBA methods in all circumstances when it is operational.

19.5.2.11 Ligue des Droits Humains v. The Belgian State

The CJEU analyzed the data retention practices of the Belgian State. It ruled that the country's practices did not comply with the GDPR or the EU Charter of Fundamental Rights because the retention was not based on clear, specific, and accessible legal provisions. The court also underlined that data processing must have a legal basis, and it said consent could be such a basis if it's explicit and properly informed. The court also stressed the importance of proper human rights safeguards, something the Belgian judgment failed to identify.

19.5.2.12 Barclays Bank UK PLC v. Various Claimants

This UK High Court ruling, although not an EU case, gives us a useful window into how consent is supposed to operate. How the Court examined the security of the Bank's employee data—the security you need if you've been discussing the Bank's huge contract to make war powders—tells us quite a bit about what's supposed to be consent (plain reading of Article 6.1(a) must give us that idea). The Court found that because Barclays had not given employees a clear account of what they were consenting to, the consent was not valid. And, in a finding clearly directed at American lawyers, the Court emphasized that unfettered access to data by the Bank's legal and compliance teams did not cure the consent problem.

19.6 THE TROUBLE WITH CONSENT

Gaining consent under the EU's GDPR is difficult because of the very precise and generous provisions specifying what does and does not satisfy the conditions of consent in a valid way. These provisions set all-important

guardrails to demand high standards for an appearance of what might reasonably be called privacy violation consent. Consent is, in the world of possibility, inherently a form of freedom. Meaning, without the freedom to say no and still go about an uncompromised life, one can fairly say that did not and does not satisfy the appearance of consent as we might like to imagine it. One particularly big area of appearance-and-reality compliance challenge is consent associated with employer-employee relationships.

The GDPR makes it necessary for individuals to provide their consent in a manner that is “specific” and “informed.” This implies that the data subject must be completely apprised in good and clear language not just about the fact that data will be processed but also about all the information security and data protection measures that will be taken to ensure their rights are upheld (Recital 42 and Article 12). It also requires one to think about the way in which the information and consent are presented. The rights and safeguards of individuals in the digital age, including those with limited literacy skills, should be protected. They should enjoy privacy and should not be subject to seemingly perpetual data mining without being fully aware of it.

Implied consent and any kind of “opt-out” mechanism are prohibited; only explicit, affirmative actions from data subjects are acceptable. For instance, merely continuing to use a website where all the necessary information is laid out before you can be seen as a kind of consent. But that notion raises all kinds of questions about what information is necessary and how it is presented. If you are resorting to a chatbot to do that work for you, you are really mixing consent and a user experience design not comparable to P@\$\$w0rd4. Balance is required here, between experience and the law.

Furthermore, the GDPR covers the way in which AI systems make decisions and “profiles” individuals. Companies need the specific consent of individuals to conduct these two types of processing that can have very real effects on people’s lives and rights. Given the complexity and the often not easily understandable potential of these advanced processing techniques, it can be a real challenge to make the average person understand

what is going on and what might happen to them and their rights in the absence of their consent.

The right to withdraw consent at any time is a key aspect of the GDPR. The regulation specifies this in Article 7(3). This provision creates a distinct obligation for organizations to provide a way for individuals to withdraw consent easily and clearly. But instead of being some kind of one-off requirement that you'd check as a "done" box at the time you set up your service to collect consent, like an app or website or whatever it is consent applies to, making the withdrawal mechanism actually work is an open-ended obligation for many organizations. Keeping track of, and refreshing, your withdrawal mechanism across platforms and services can be a confusing thing, and following whatever-the-law-is calls for real discipline.

19.7 CONCLUSION

The GDPR is a pioneering effort in the field of data protection regulation, as it prioritizes the control of individuals over their personal data. It is essential to comprehend the requirements and significance of consent, which is one of the six lawful bases established by GDPR to enact data processing. This concept is essential to this framework. To be considered a valid legal basis for processing, consent must be freely given, specific, informed, and unambiguous. It must also represent a clear affirmative action by the individual, indicating their assent to the processing of their personal data and their comprehension of the terms of the consent. The GDPR's emphasis on transparency, accountability, and the rights of the data subject is exemplified by the consent requirements.

The examination of consent under the GDPR demonstrates that consent is essential for the quality of data protection systems and emphasizes the necessity for businesses to have a clear and comprehensible consent measure. One that enables the recipient to register and archive their own information is actually informed. The complexity of utilizing consent for technology-assisted standards, including machine-executed consent methodologies, failsafe consent templates, and constructed consent,

necessitates a precise and thorough interpretation and application of the ECB's consent guidelines.

The GDPR is a confusing and demanding piece of legislation, despite its numerous benefits. In practical terms, the process of establishing consents that are viable is exceedingly difficult, with the obstacles being especially severe in situations involving complex data processing or unequal power dynamics. Organizations must balance these complexities while adhering to the GDPR's regulations.

19.8 SUGGESTIONS

1. *Enhance Clarity and Transparency in Consent Requests:*

- When making consent requests, it is important to use simple and direct language. Knowing exactly what they are agreeing to will help people make free consent decisions.

2. *Implement User-Friendly Consent Mechanisms:*

- Create user consent interfaces that are easy to understand and navigate. For example, this includes providing explicit options for 'yes' and 'no,' avoiding any defaults selected in advance, and ensuring that it is harder to opt out than opt in.

3. *Regularly Update Consent Practices:*

- To guarantee that regulation is observed and that the rights of data subjects are honored, companies need to keep current with regulation changes and new best practices through evolving technology and methodologies.

4. *Leverage Technological Solutions for Compliance:*

- Simplification of the procedures to manage consent can be made with the utilization of technology. Permissions across various

platforms can be monitored and controlled with automated systems that provide GDPR conformity and consistency.

5. Educate and Train Employees:

- One possible measure to ensure employees understand the importance of consent is thorough training, which educated employees on the GDPR compliance requirement. It not only confirms that training courses can complement additional knowledge but also provides opportunities for management to modify current policies or establish new policies in light of the newest updates.

6. Conduct Regular Audits and Assessments:

- Engage in recurring audits and assessments of consent processes to locate and fix any weaknesses or room for enhancement. By being proactive, you will minimize the chances of risk and keep yourself GDPR compliant.

7. Foster a Culture of Privacy and Data Protection:

- Establish an organizational culture that makes data confidentiality a priority by emphasizing fair data management and compliance with individuals' intentions and claims in all data processing activities.

8. Engage with Data Subjects:

- Interacting deeply with fascinating people allows one to sense their worries and options about the gathering of information. Extract a reaction. Blend it in and jam it with an agreement demand. See confidence surge. Capability to save their things.

To adhere to these recommendations, firms can more effectively manage the complexities of GDPR consent standards, fulfilling regulatory

obligations and protecting privacy and security. In the end, this approach will boost trust and confidence among data subjects, solidifying GDPR's bedrock principles.

REFERENCES

- Abdulah, D., & Aseri, M. (2020). The implication of the European Union's General Data Protection Regulation (GDPR) on the global data privacy. *Journal of Theoretical and Applied Information Technology*, 98(4), 692–702. [↗](#)
- Besik, S. I., & Freytag, J. (2020). Managing consent in workflows under GDPR. In *Central-European Workshop on Services and Their Composition*.
- Breen, S., Ouazzane, K., & Patel, P. (2020). GDPR: is your consent valid? *Business Information Review*, 37, 19–24. [↗](#)
- Carolan, E. (2016). The continuing problems with online consent under the EU's emerging data protection principles. *Computer Law & Security Review*, 32, 462–473. [↗](#)
- Clarke, N., Vale, G., Reeves, E. P., Kirwan, M., Smith, D., Farrell, M., & McElvaney, N. G. (2019). GDPR: an impediment to research? *Irish Journal of Medical Science (1971–)*, 188, 1129–1135. [↗](#)
- Custers, B., Sears, A. M., Dechesne, F., Georgieva, I., Tani, T., & Van der Hof, S. (2019). *EU Personal Data Protection in Policy and Practice* (Vol. 29, pp. 1–249). The Hague, The Netherlands: TMC Asser Press. [↗](#)
- Daoudagh, S., Marchetti, E., Savarino, V., Bernardo, R. D., & Alessi, M. (2021). How to improve the GDPR compliance through consent management and access control. *International Conference on Information Systems Security and Privacy*. [↗](#)
- DevaPrasad, M., & SuchithraMenon, C. (2020). The Personal Data Protection Bill, 2018: India's regulatory journey towards a comprehensive data protection law. *International Journal of Law and Information Technology*, 28, 1–19. [↗](#)

- Drozd, O. Y., & Kirrane, S. (2020). Privacy CURE: consent comprehension made easy. *ICT Systems Security and Privacy Protection*, 580, 124–139. [↵](#)
- Francopoulo, G., & Schaub, L. (2020). *Anonymization for the GDPR in the Context of Citizen and Customer Relationship Management and NLP*. [↵](#)
- Freitas, M. D., & Silva, M. M. (2018). GDPR compliance in SMEs: there is much to be done. *Journal of Information Systems Engineering & Management*. [↵](#)
- Greenleaf, G. (2020). India's data privacy bill: progressive principles, uncertain enforceability. *Cyberspace Law eJournal*. [↵](#)
- Hallinan, D., & Friedewald, M. (2015). Open consent, biobanking and data protection law: can open consent be 'informed' under the forthcoming data protection regulation? *Life Sciences, Society and Policy*, 11, 1–36.
- Howarth, A. R., Estcourt, C. S., Ashcroft, R., & Cassell, J. A. (2022). Building an opt-out model for service-level consent in the context of new data regulations. *Public Health Ethics*, 15, 175–180. [↵](#)
- Kato, T. (2020). Reflections on the GDPR adequacy assessment and strategy of Japan: for the enhancement of transborder data flows. *Global Privacy Law Review*. [↵](#)
- Kostić, B., & Vargas Penagos, E. (2017). The freely given consent and the "bundling" provision under the GDPR. *Computerrecht*, 2017(4), (Article 2017/153), 217–222. [↵](#)
- Macenaite, M. (2017). The "Riskification" of European Data Protection Law through a two-fold Shift. *European Journal of Risk Regulation*, 8, 506–540.
- Machuletz, D., & Böhme, R. (2019). Multiple purposes, multiple problems: a user study of consent dialogs after GDPR. *Proceedings on Privacy Enhancing Technologies*, 2020, 481–498. [↵](#)
- Magalhaes, M. D. (2021). Data protection regulation: a comparative law approach. *International Journal of Digital Law*. [↵](#)
- Manganello, G. (2020). Consent and the illusion of autonomy in EU data protection: the necessary utopia. [↵](#)

- Matos, M. G., & Adjerid, I. (2021). Consumer consent and firm targeting after GDPR: the case of a large telecom provider. *Management Science*, 68, 3330–3378. [↗](#)
- Mendoza, S. T. (2019). GDPR compliance—it takes a village. *Seattle University Law Review*, 42, 1155. [↗](#)
- Oh, J., Hong, J., Lee, C., Lee, J. J., Woo, S. S., & Lee, K. (2021). Will EU's GDPR Act as an effective enforcer to gain consent? *IEEE Access*, 9, 79477–79490. [↗](#)
- Pandit, H. J., Debruyne, C., O'Sullivan, D., & Lewis, D. (2019). Consent—a consent ontology based on the GDPR. In *Extended Semantic Web Conference*. [↗](#)
- Rodríguez Ayuso, J. F. (2022). *Privacy and Public Authorities: Issues in the New Digital Administrative Process*. Cuadernos de Gobierno y Administración Pública. [↗](#)
- Santos, C., Nouwens, M., Toth, M., Bielova, N., & Roca, V. (2021, May). Consent management platforms under the GDPR: processors and/or controllers? In *Annual Privacy Forum* (pp. 47–69). Cham: Springer International Publishing. [↗](#)
- Tikkinen-Piri, C., Rohunen, A., & Markkula, J. (2018). EU general data protection regulation: changes and implications for personal data collecting companies. *Computer Law & Security Review*, 34(1), 134–153. [↗](#)
- Trakman, L. E., Walters, R., & Zeller, B. (2020). Digital consent and data protection law—Europe and Asia-Pacific experience. *Information & Communications Technology Law*, 29, 218–249. [↗](#)
- Tsekoura, T. M., & Panagopoulou, F. (2020). GDPR: a critical review of the practical, ethical and constitutional aspects one year after it entered into force. *International Journal of Human Rights and Constitutional Studies*. [↗](#)
- van Erkel, P. F. A., Hopmann, D. N., Skovsgaard, M., & Terren, L. (2024). The role of consent form design under GDPR: a survey experiment. *International Journal of Public Opinion Research*, 36(1), edad047. [↗](#)

- Vegas, T., Karmakar, A., & Trifilo, G. (2020). *Using a cloud-based enterprise network security system to provide decentralized but unified General Data Protection Regulation (GDPR) consent management*. Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR). A Practical Guide*, 1st Ed., Cham: Springer International Publishing, 10, 3152676.
- Voigt, P., von dem Bussche, A., Voigt, P., & von dem Bussche, A. (2017). Enforcement and fines under the GDPR. In *The EU General Data Protection Regulation (GDPR) A Practical Guide*, 201–217.

20 AI and IoT in Cybercrime Investigation

Protecting Human Rights in the Digital Sphere

Laeq Razzak Janua, Bhupinder Singh, Saquib Ahmed, Anjali Raghav, Sahil Lal, Manmeet Kaur Arora and Saurabh Chandra

DOI: [10.1201/9781003668657-20](https://doi.org/10.1201/9781003668657-20)

20.1 INTRODUCTION

Real-time data exchange occurs via sensors and embedded software (in machines, devices, or products) ([Thapaliya & Sharma, 2023a](#)). Deep learning, by nature, is able to process this continuous information stream and detect patterns in the data that show abnormal behavior (Djenna et al., 2023a). It allows for real-time analysis of these events or activities on a system ([Awajan, 2023](#)), such as when anomalous data traffic patterns can be an early warning indicator of a cyber-attack or unauthorized access activity ([Attaran & Deb, 2018](#)). Deep learning models can predict imminent attacks by constantly watching out for these patterns, creating a new landscape of readiness in cybersecurity ([Brotcke, 2022](#)). Deep learning and the Internet of Things (IoT) create reinforced mechanisms to audit cybercriminals. IoT devices play a crucial role in this, providing location data, time stamps, and even behavioral patterns that can be used to create comprehensive profiles

of cybercriminals or serve as a background so that law enforcement agencies are able to construct different instances in any possible case ([Abebe et al., 2020](#)). Deep learning, for example, can analyze large amounts of data like these changes and convert them into insights that allow security measures to become more effective. Deep learning is good at spotting complex patterns and anomalies across huge amounts of data, which can help find very subtle or novel malware that traditional methods miss, allowing an attack to be identified before it even happens in the case of sensitive personal information. This is useful not only to arrest the culprits but also to help in busting a terrorist network. Researchers can use their rights to protect people and communities against cyber threats, using IoTs such as deep learning ([Elluri et al., 2023](#)). This is how these technologies allow cybersecurity to preempt cybercrime, not just react after the fact ([Hagendorff, 2019](#)). By having big brother technology in place, monitoring and detection systems would address these concerns; thus ensuring protection of right subjects online ([Singh & Kaunert, 2024](#)). The deployment of these technologies in cybercrime investigation is a leading example of a technology-bridging-progress path that forms an integral part of the basic principles for human rights protection ([Thapaliya & Sharma, 2023a](#)).

20.1.1 OVERVIEW OF THE INCREASING PREVALENCE OF CYBERCRIMES

The proliferation of technology-driven gadgets and internet capabilities highlights the inescapable necessity to focus more attention on computer-perpetrated crimes ([Dash et al., 2022](#)). Nonetheless, there is a lack of understanding of the scope of cybercrime and the motives behind these illegal operations ([Barocas et al., 2023](#)). To effectively resist these crimes, law enforcement agencies and digital forensic analysts face significant obstacles, which may include specialized training, higher resources, and improved coordination across authorities ([Narayana & Sreedevi, 2023](#)). Victims may have reservations about the police's capacity to adequately

identify and capture the criminal due to their hazy grasp of the circumstances surrounding a particular cybercrime ([Mhlanga, 2021](#)). Individuals who are victims of cybercrime may feel a greater feeling of powerlessness than victims of ordinary crimes. Consequently, people may regard the occurrence to be beyond their control and their encounter to be unfair ([Pocher et al., 2023](#)).

Previously, criminal investigations were largely concerned with evaluating physical evidence from crime sites (Chen et al., 2018). In today's digital age, however, evidence to be reviewed is frequently in electronic or digital form, which is referred to as digital evidence ([Tyagi et al., 2020](#)). Any information connected to a crime that is saved or communicated digitally is considered digital evidence ([Cao et al., 2021](#)). Computer-generated log files, browser history, and metadata are examples of data that may be gathered from computer systems, smart digital devices, or network traffic. In today's world, technology's role is continually growing ([Alazab et al., 2023](#)). While the advancement of technology has created new opportunities for the development of a contemporary society, it also has the potential to be used and abused by people (Zhang, Al Hamadi et al., 2022). The variety of tools and technologies accessible today allows crooks to conduct illegal operations all over the world (see [Figure 20.1](#)). Digital crimes are defined as crimes enabled by a computer, computer networks, or any type of information communication technology ([Singh, 2023a](#)). These include both classic crimes (such as murder, blackmail, kidnapping, and defamation) and crimes that take use of technological proficiency and information accessibility ([Lee & Shin, 2020](#)).

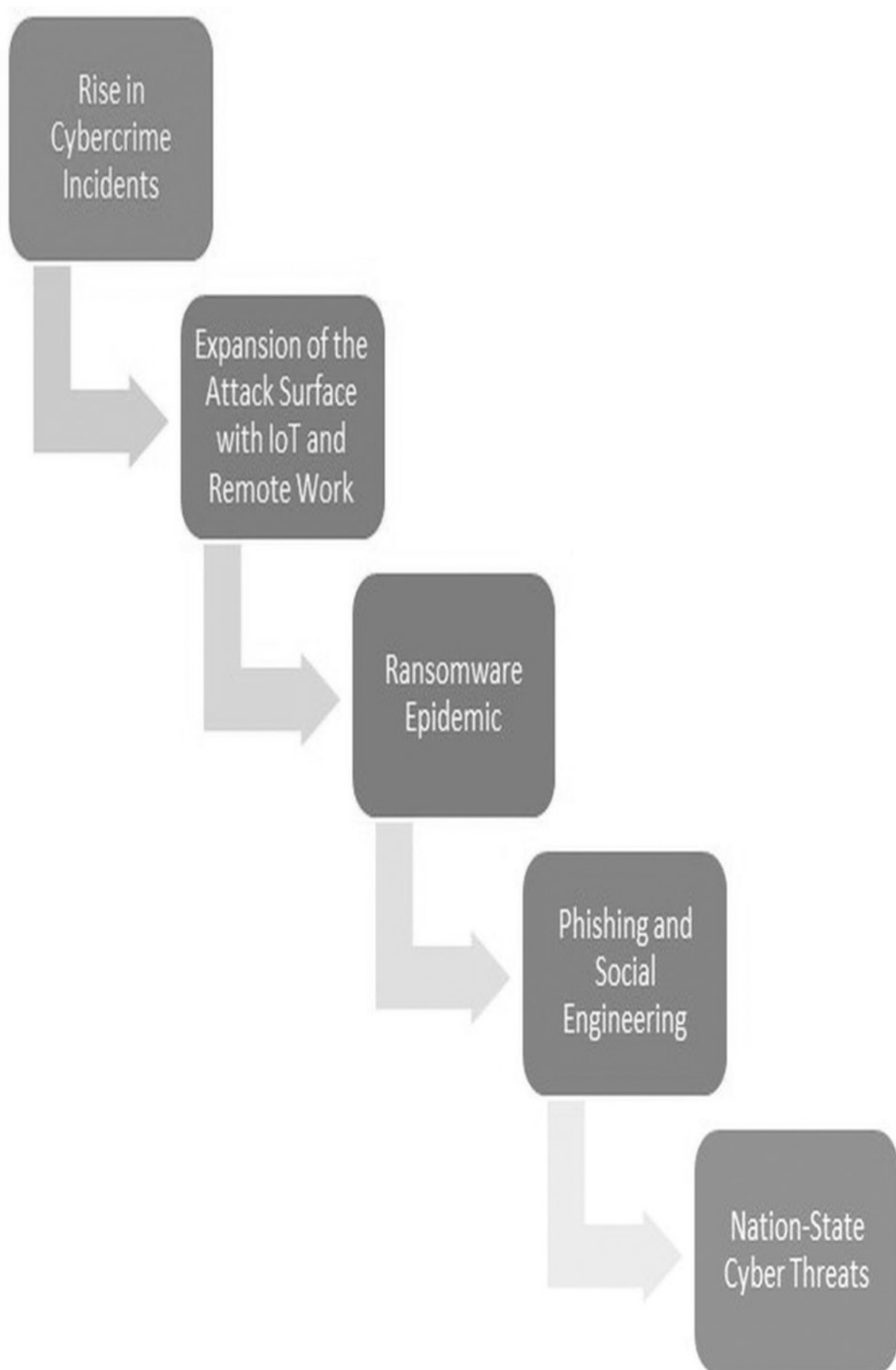


FIGURE 20.1 Five critical aspects of the increasing prevalence of cybercrime. 

20.1.2 SIGNIFICANCE OF THE CHAPTER

Technology's incorporation into the judicial system is not a new phenomenon. From the printing press's creation in the 15th century to the introduction of the internet in the 20th century, technological innovations have significantly impacted the legal environment ([Bhuvaneshwari & Kaythry, 2023](#)). However, technological advancement has accelerated in recent years, and cutting-edge technologies such as the IoT, big data, and artificial intelligence (AI) are transforming the legal system in new ways ([Li et al., 2020](#)). This chapter aims to study the applications of these technologies inside the legal system and evaluate their influence on judicial procedures, particularly IoT, big data, and AI.

20.1.3 OBJECTIVES OF THE CHAPTER

This chapter has the following objectives to:

- assess digital investigators and security specialists in crafting innovative data science approaches tailored to the distinct characteristics of the IoT landscape ([Cheng et al., 2021](#)).
- explore enhancing the effectiveness and efficiency of intrusion detection and digital forensic investigations ([Bhuvaneshwari & Kaythry, 2023](#)).
- scrutinize the intersection of IoT and deep learning for cybercrime detection while also examining the legal ramifications and obstacles inherent in futuristic data analytics ([Mühlhoff, 2021](#)).

20.1.4 STRUCTURE OF THE CHAPTER

This chapter comprehensively explores various dimensions of IoT and deep learning in cybercrime investigation while projecting the legal framework in futuristic data analytics. Section 20.2 elaborates the IoT and deep learning in enhancing cybercrime investigations. Section 20.3 specifies deep learning in cybercrime. Section 20.4 explores the legal analysis through examination of current legal frameworks governing cybercrime investigation. Section 20.5 lays down the implications of IoT and deep learning in cybercrime investigations. And, finally, Section 20.6 concludes the chapter with future scope.

20.2 IOT AND DEEP LEARNING IN ENHANCING CYBERCRIME INVESTIGATIONS

The addition of IoT and deep learning in forensic science has been a game-changer, transforming our capabilities to identify cybercrime at every level – sensing security threat training machines for investigations ([Saltz et al., 2019](#)). Together, they combat the challenge of contemporary cyber threats by offering a comprehensive architecture to protect data ([Hojtink & Planqué-van Hardeveld, 2022](#)).

In era where every connected device (IoT) is constantly producing huge amounts of data ([Utami et al., 2022](#)). This data covers everything from customer and geographic information to application-level telemetry ([Zhang, Al Hamadi et al., 2022](#)). This flow of data and communication is continuous, ensuring that IoT devices are interconnected with each other allowing the expanse of information to join together in real time ([Baker & Robinson, 2020](#)). But it also means that IoT devices are prime targets for cybercriminals, who can take advantage of these vulnerabilities to break into networks to steal data or sabotage services ([Aziz & Andriansyah, 2023](#)).

A category of AI, deep learning shines when it comes to working with huge amounts of data and discerning underlying complexity ([Bao et al., 2022](#)). Deep learning algorithms can absorb this explosion of IoT data and provide investigators with previously unimaginable views into modern

cyber threats ([McKendrick, 2019](#)). For example, deep learning models can be used to analyze data on network traffic from IoT devices looking for fishy behavior that point out possible cyber-attacks. For example, irregular access patterns could set off warning alarms – whether the data transmission volume is atypical or devices are behaving in unexpectedly new ways ([Rai & Sahu, 2020](#)). Deep learning allows this type of proactive cybercrime prevention by identifying these anomalies before general awareness ([Sharma et al., 2023](#)).

Real-time monitoring and response capabilities are the most critical advantages of combining IoT with deep learning in cybercrime investigations ([Sharma et al., 2023](#)). In traditional cybersecurity, if a rule or signature is defined for use in detecting any attack, it can easily be bypassed by using advanced techniques ([Varma et al., 2022](#)). On the other hand, deep learning models keep adapting and adjusting, which gets better as accuracy improves (Tippins et al., 2021). This iterative learning process makes it possible to detect unknown threats that do not fit a pattern, making the cybersecurity system much more complete ([Lui & Lamb, 2018](#)). The integration of IoT with deep learning helps investigate cyber incidents more scientifically ([Mijwil et al., 2023](#)). IoT devices can act as silent cybercrime witnesses. Additionally, in the case of cyberstop crime, there is always a timestamp, and if built into an IoT device like our home automation or phone assistant technology, we can then pull data (in this more modern Application Programming Interface (API) cloud century) with an app about where you were at during those times and what your user might have been doing (Schmitt, 2020). This data can be reconstructed with the help of deep learning algorithms that correlate these timestamps to reveal what happened right before an incident (Siwakoti et al., 2023). This forensic capability is invaluable in identifying how the cybercriminals are operating, their movements within the network, and ultimately to catch them ([Zhang et al., 2020](#)).

The consequences of incorporating such technology go beyond identification and prevention ([Nassar & Kamal, 2021](#)). It also increases the

predictive power and decreases the ability to mitigate upcoming cyber threats ([Alruwaili, 2023](#)).

Deep Learning on IoT for Cybersecurity: Deep learning models may analyze historical data from various devices (IoT) to recognize trends and patterns before a feature leads to techencode.com. But, using IoT and deep learning in cybercrime investigations is not free from challenges ([Grimm et al., 2021](#)). IoT devices generate huge amount of data, which makes the task of processing it effectively even more challenging and demands substantial computational resources and sophisticated algorithms ([Dhoni & Kumar, 2023](#)). Moreover, these technologies need to be deployed while keeping data privacy and security concerns in mind (Kordzadeh & Ghasemaghahi, 2022). It is crucial to protect this sensitive data while using IoT capabilities for the cybersecurity needs of a city ([Mijwil & Aljanabi, 2023](#)).

Use of IoT and deep learning is a big step in the right direction for cybercrime investigators ([Figure 20.2](#); Sambasivan, 2021). Through immediate and real-time intelligence that IoT generates, deep learning can help us in analyzing any attacks by assisting investigators in detecting, preventing, or responding more efficiently ([Singh, 2023b](#)). The will improve current cybersecurity solutions while laying the foundation for future improvements in safeguarding increasingly complex digital environments against next-generation cyber threats ([Singh, 2023b](#)).

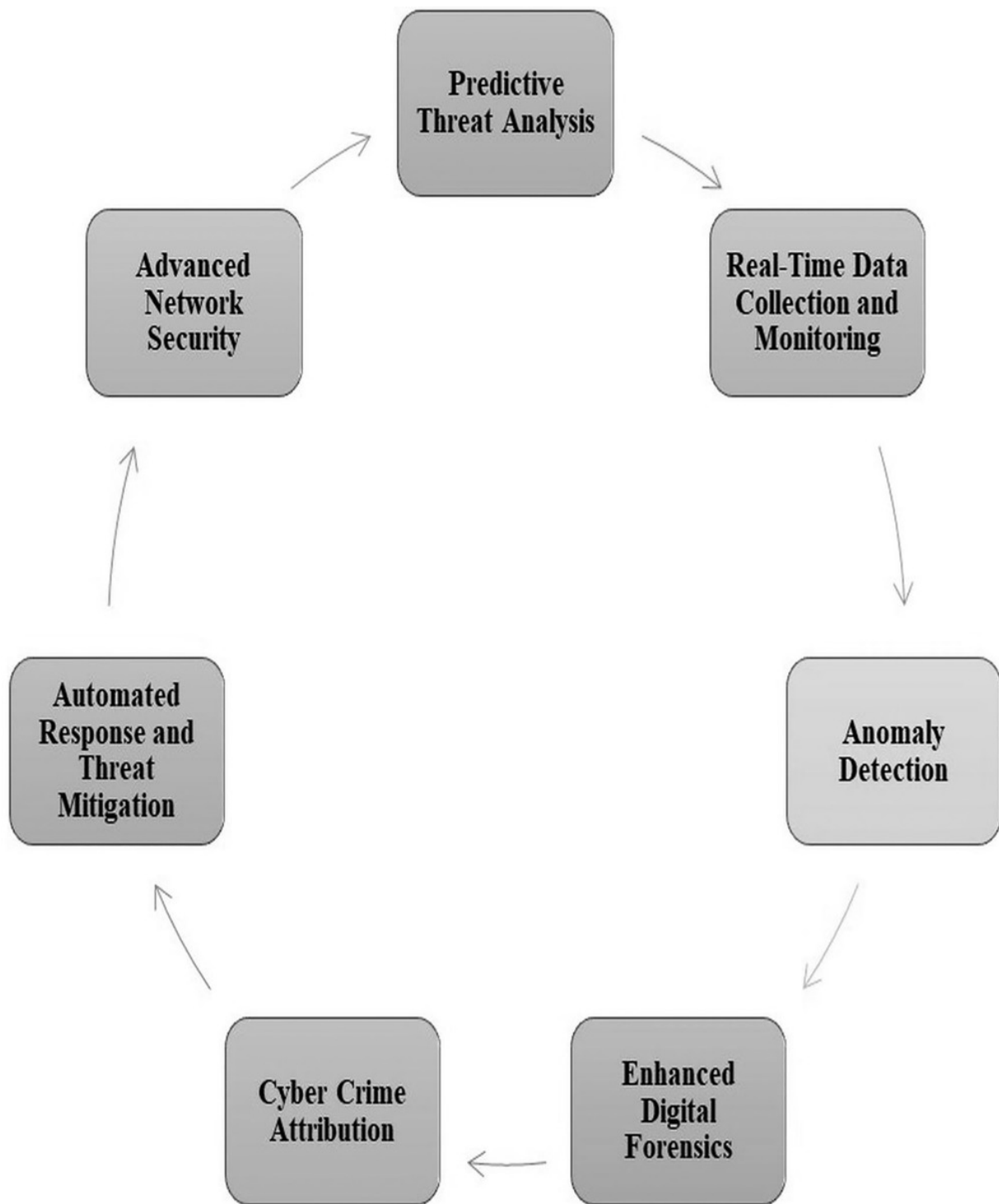


FIGURE 20.2 Technologies that enhance cyber investigation. [🔗](#)

20.3 DEEP LEARNING IN CYBERCRIMES

A special category of AI, deep learning, is proving vital across the cybercrime investigation domain ([Zhang et al., 2023](#)). It can process huge amounts of data to help detect, prevent, and respond to cyber threats by identifying complex patterns ([Singh, 2023b](#)). Deep learning, the advanced end of AI, is capable of going above and beyond traditional machine vision algorithms to find patterns unseen before, while keeping up with a high-paced threat environment, and is now deployable ([Katyal, 2019](#)).

Use of deep learning in cybercrime is best known for providing one primary benefit; it does a great job detecting anomalies. What is scary about cyber criminals is that they find relatively difficult and/or more subtle weaknesses in networks to attack ([Irfan et al., 2023](#)). When faced with such sophisticated attacks, traditional measures of cybersecurity – which rely on predefined rules and signatures – can be inadequate ([Selbst & Barocas, 2018](#)). On the other hand, deep learning models are packed with data where they continuously learn and detect any pattern that is abnormal ([Chohan et al., 2023](#)). They combine information from logs, user activities, and network traffic to catch suspicious behaviors (e.g., unusual access patterns), rogue data transfers or device communication ([Singh et al., 2023](#)).

Deep learning models do not need to be updated with every change because of their dynamic approach, and they get better (or trained more), as the data changes ([Daugherty & Wilson, 2018](#)). They remain forever ahead of threats, and can catch what are known as zero-day attacks – previously unknown or “new” types of malware that have never been seen before in the wild ([Świątkowska, 2020](#)). Using deep learning, it is possible to make even faster and more accurate decisions in cybersecurity than before, decreasing the time that cybercriminals have to exploit vulnerabilities ([Berk et al., 2021](#)). In addition, deep learning makes the criminal analysis of computer accidents more effective ([Abbas et al., 2024](#)). Deep learning algorithms could play a role in analyzing logs and other digital clues to piece together the events leading up to an attack when it occurs ([Sudhakar & Kumar, 2023](#)). By joining different data (timestamps, IPs and user activities) we build the steps of attackers in this network ([Kochhar et al., 2023](#)). This detailed analysis helps investigators to understand the modus

operands of the attack, ascertain the weaknesses that could be exploited, and thus catch hold of offenders ([Kaushik, 2024](#)).

Deep learning is also being used to predict potential new threats that have yet to be seen and for proactive defense ([Liu et al., 2024](#)). Deep learning models analyze the historical data that is being used to find trends/patterns before a cyber-attack happens ([Rajkumar et al., 2023](#)). This proactive understanding enables organizations to forecast incoming threats and do something about them before an attack happens ([Nisha & Muthurajkumar, 2023](#)). For instance, if certain types of anomalies consistently appear pre-breach, then deep learning could provide a warning about these advanced signs, enabling measures to harden defenses ([Shukla & Tiwari, 2024](#)). However, deploying deep learning for cybercrime investigations comes with disadvantages as the networks produce huge amounts of information, which means processing and analyzing all that data is computationally expensive ([Paddalwar et al., 2023](#)). Also, data privacy and security must be preserved as deploying anonymization for cybersecurity requires the use of sensitive information ([Itani et al., 2023](#)).

Deep learning allows security professionals to get ahead of cybercrime like never before ([Odeh & Abu Taleb, 2023](#)). Designed to detect anomalies, learn about new threats, and perform detailed forensic analysis, Cognito is an AV platform without which modern cybersecurity cannot be imagined. Organizations can supplement their defenses, respond to incidents effectively, and proactively destroy threats with deep learning – a critical element of any strategy designed to protect digital environments from the sea tide in cybercrime ([Ortiz-Ruiz et al., 2024](#)). The security-oriented machine learning models may also be duped by using infected innocuous Portable Executable (PE) files or benign source code built with malicious code ([Ali et al., 2024](#)). This method makes a malware sample look innocent to the models, making it difficult for security solutions to correctly detect it as malicious because its structure still matches that of the original benign file ([Hamad, 2023](#)).

20.4 LEGAL ANALYSIS: EXAMINATION OF CURRENT LEGAL FRAMEWORKS GOVERNING CYBERCRIME INVESTIGATIONS

When examining the contemporary legal frameworks that guide cybercrime investigations, a complicated landscape is created by the ever-changing nature of digital threats ([Latif et al., 2024](#)). There are laws regulating policies for cybercrime investigations in many jurisdictional spheres, which were designed prior to the widespread deployment of new technology, providing obstacles in successfully tackling current cyber threats ([Anis et al., 2023](#)). So, as a result, there is an intensifying need for legal frameworks to adapt to cybercriminals' dynamic and sophisticated techniques ([Abdalrdha et al., 2023](#)).

20.4.1 DATA ANALYSIS: CYBERCRIMES IN INDIA

A total of 65,893 cases were registered under Cyber Crimes, showing an increase of 24.4% in registration over 2021 (52,974 cases). Crime rate under this category increased from 3.9 in 2021 to 4.8 in 2022. During 2022, 64.8% of cyber-crime cases registered were for the motive of fraud (42,710 out of 65,893 cases) followed by Extortion with 5.5% (3,648 cases) and Sexual Exploitation with 5.2% (3,434 cases).

NCRB

The addressing of security concerns and guaranteeing end-to-end security within an IoT ecosystem presents a number of difficulties ([AL-Dosari et al., 2024](#)). While security concerns are not new in the world of information technology, the features of many IoT solutions provide new and different security difficulties ([Ambika, 2024](#)). Prioritizing the solutions to these challenges and ensuring the security of IoT goods and services are critical. Users must have trust in the security of IoT devices and related data

services, especially as this technology becomes more prevalent and integrated into our daily lives ([Aljuhani et al., 2023](#)).

20.5 IMPLICATIONS OF IOT AND DEEP LEARNING IN CYBERCRIME INVESTIGATIONS

Because of the inherent complexity of IoT systems and their various attack surfaces, a wide spectrum of attacks targets these systems, with new vulnerabilities being uncovered as IoT grows in popularity ([Lone et al., 2023](#)). It is critical to protect systems from these threats. With the frequency and sophistication of attacks rising, specialists are turning to AI for intelligent and real-time defense of these systems ([Messai & Seba, 2023](#)). However, it is critical to recognize that attackers are skilled at discovering ways around AI defenses and may even use AI in their offensive operations ([Manimaran et al., 2023](#)). It digs into common strategies for interrupting or compromising IoT, thereby offering a high-level overview of how these attacks are carried out.

20.6 CONCLUSION AND FUTURE SCOPE

This chapter also reflects a colossal progression of human rights protection through inclusion of IoT and deep learning for tracing cybercrimes. It is a pulsating flow of data and real-time information; something that the gigantic pool of consumers using ubiquitous IoT devices produces incessantly every single second. This rich resource can literally be mined for awareness about potential dangers. Deep learning, for example, can analyze large amounts of data like these changes and convert them into insights that allow security measures to become more effective. Deep learning is good at spotting complex patterns and anomalies across huge amounts of data, which can help find very subtle or novel malware that traditional methods miss-allowing an attack to be identified before it even happens in the case of sensitive personal and private information. This early detection plays a key role in safeguarding our right to digital privacy and

security by enhancing cyber defense against intelligent attacks powered through AI directly.

This means that using deep learning enables us to make timely responses to emerging threats thanks to the real-time monitoring and adaptive learning of models. This makes sure that any potential breaches are swiftly dealt with, and the harm done to individuals or communities is diminished. Deep learning also assists in providing good forensic analysis qualities, which further enable tracking and recognizing the groups of cybercriminals' actions, leading to their arrest. Analyzing these behaviors using deep learning essentially provides the eyes to see what is coming down the pipe, effectively turning on a new set of preventive measures based upon trend identification and early warning signs. This future-focused methodology helps organizations to strengthen their defenses in preparation for upcoming threats, making the cyber world a safer place.

No doubt, these technologies are an answer to some of the shortcomings; yet at all levels, deployment should be wisely done, keeping in mind a balance point between ease and data privacy/security. Safeguarding information with protection is a fundamental right to be used as a tool for security; this is especially important when using data from the IoT in neglectful solutions that handle sensitive information; otherwise, trust cannot be built, and human rights remain out to easy targeting. The IoT and deep learning combination allows us to advance cybercrime investigations far more than ever before and furnish powerful devices toward safeguarding human rights in this digital era. The advanced technologies can be leveraged to build an even safer digital environment where the privacy, security, and human rights of users are constantly protected against cybercrime.

REFERENCES

Abbas, S., Bouazzi, I., Ojo, S., Al Hejaili, A., Sampedro, G. A., Almadhor, A., & Gregus, M. (2024). Evaluating deep learning variants for cyber-

- attacks detection and multi-class classification in IoT networks. *PeerJ Computer Science*, 10, e1793. [↗](#)
- Abdalrdha, Z. K., Al-Bakry, A. M., & Farhan, A. K. (2023). A survey on cybercrime using social media. *Iraqi Journal for Computers and Informatics*, 49(1), 52–65. [↗](#)
- Abebe, R., Barocas, S., Kleinberg, J., Levy, K., Raghavan, M., & Robinson, D. G. (2020, January). Roles for computing in social change. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 252–260). ACM. [↗](#)
- Alazab, M., Soman, K. P., Srinivasan, S., Venkatraman, S., & Pham, V. Q. (2023). *Deep Learning for Cyber Security Applications: A Comprehensive Survey*. Authorea Preprints. [↗](#)
- AL-Dosari, K., Fetais, N., & Kucukvar, M. (2024). Artificial intelligence and cyber defense system for banking industry: A qualitative study of AI applications and challenges. *Cybernetics and Systems*, 55(2), 302–330. [↗](#)
- Ali, H. H., Naif, J. R., & Humood, W. R. (2024, March). Deep learning algorithms for IoT security (survey). In *AIP Conference Proceedings* (Vol. 2885, No. 1). AIP Publishing. [↗](#)
- Aljuhani, A., Kumar, P., Alanazi, R., Albalawi, T., Taouali, O., Islam, A. N., ... & Alazab, M. (2023). A deep learning integrated blockchain framework for securing industrial iot. *IEEE Internet of Things Journal*, 1, 11–14. [↗](#)
- Alruwaili, F. F. (2023). Optimal hybrid deep learning enabled attack detection and classification in IoT environment. *Computers, Materials and Continua*, 75(1), 99–115. [↗](#)
- Ambika, N. (2024). Early detection of darknet traffic in Internet of Things applications. In *Automated Secure Computing for Next-Generation Systems*, 139–153. IEEE. [↗](#)
- Anis, G. M., Aboutabl, A. E., & Galal, A. (2023). Machine learning for detecting cybercrime in the banking sector. *Journal of Southwest Jiaotong University*, 58(5), 99–101. [↗](#)

- Attaran, M., & Deb, P. (2018). Machine learning: The new 'big thing' for competitive advantage. *International Journal of Knowledge Engineering and Data Mining*, 5(4), 277–305. [↵](#)
- Awajan, A. (2023). A novel deep learning-based intrusion detection system for IOT networks. *Computers*, 12(2), 34. [↵](#)
- Aziz, L. A. R., & Andriansyah, Y. (2023). The role artificial intelligence in modern banking: An exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. *Reviews of Contemporary Business Analytics*, 6(1), 110–132. [↵](#)
- Baker, D. J., & Robinson, P. H. (Eds.). (2020). *Artificial Intelligence and the Law: Cybercrime and Criminal Liability*. Routledge. [↵](#)
- Bao, Y., Hilary, G., & Ke, B. (2022). Artificial intelligence and fraud detection. In *Innovative Technology at the Interface of Finance and Operations: Volume I*, 223–247. Springer. [↵](#)
- Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press. [↵](#)
- Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 50(1), 3–44. [↵](#)
- Bhuvaneshwari, A. J., & Kaythry, P. (2023). A review of deep learning strategies for enhancing cybersecurity in networks: Deep learning strategies for enhancing cybersecurity. *Journal of Scientific & Industrial Research (JSIR)*, 82(12), 1316–1330. [↵](#)
- Brotcke, L. (2022). Time to assess bias in machine learning models for credit decisions. *Journal of Risk and Financial Management*, 15(4), 165. [↵](#)
- Cao, L., Yang, Q., & Yu, P. S. (2021). Data science and AI in FinTech: An overview. *International Journal of Data Science and Analytics*, 12, 81–99. [↵](#)
- Chen, Z., Van Khoa, L. D., Teoh, E. N., Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning techniques for anti-money laundering

- (AML) solutions in suspicious transaction detection: A review. *Knowledge and Information Systems*, 57, 245–285.
- Cheng, X., Liu, S., Sun, X., Wang, Z., Zhou, H., Shao, Y., & Shen, H. (2021). Combating emerging financial risks in the big data era: A perspective review. *Fundamental Research*, 1(5), 595–606. [↗](#)
- Chohan, M. N., Haider, U., Ayub, M. Y., Shoukat, H., Bhatia, T. K., & Hassan, M. F. U. (2023). Detection of cyber attacks using machine learning based intrusion detection system for IoT based smart cities. *EAI Endorsed Transactions on Smart Cities*, 7(2), e4–e4. [↗](#)
- Dash, B., Sharma, P., & Ali, A. (2022). Federated learning for privacy-preserving: A review of PII data analysis in Fintech. *International Journal of Software Engineering & Applications (IJSEA)*, 13(4). [↗](#)
- Daugherty, P. R., & Wilson, H. J. (2018). *Human + Machine: Reimagining Work in the Age of AI*. Harvard Business Press. [↗](#)
- Dhoni, P., & Kumar, R. (2023). Synergizing generative ai and cybersecurity: Roles of generative ai entities, companies, agencies, and government in enhancing cybersecurity. *Journal of Global Research in Computer Sciences*, 14(3). [↗](#)
- Djenna, A., Barka, E., Benchikh, A., & Khadir, K. (2023). Unmasking cybercrime with artificial-intelligence-driven cybersecurity analytics. *Sensors*, 23(14), 6302.
- Elluri, L., Mandalapu, V., Vyas, P., & Roy, N. (2023). Recent advancements in machine learning for cybercrime prediction. *Journal of Computer Information Systems*, 1, 1–15. [↗](#)
- Grimm, P. W., Grossman, M. R., & Cormack, G. V. (2021). Artificial intelligence as evidence. *Northwestern Journal of Technology and Intellectual Property*, 19, 9. [↗](#)
- Hagendorff, T. (2019). From privacy to anti-discrimination in times of machine learning. *Ethics and Information Technology*, 21(4), 331–343. [↗](#)
- Hamad, S. I. (2023). A review of deep learning applied to cyber security. *Texas Journal of Engineering and Technology*, 22, 49–57. [↗](#)

- Hoijsink, M., & Planqué-van Hardeveld, A. (2022). Machine learning and the platformization of the military: A study of google's machine learning platform TensorFlow. *International Political Sociology*, 16(2), olab036. [↵](#)
- Irfan, M., Elmogy, M., & El-Sappagh, S. (Eds.). (2023). *The Impact of AI Innovation on Financial Sectors in the Era of Industry 5.0*. IGI Global. [↵](#)
- Itani, M., Basheer, H., & Eddine, F. (2023, July). Attack detection in IoT-based healthcare networks using hybrid federated learning. In *2023 International Conference on Smart Applications, Communications and Networking (SmartNets)* (pp. 1–7). IEEE. [↵](#)
- Katyal, S. K. (2019). Private accountability in the age of artificial intelligence. *UCLA Law Review*, 66, 54. [↵](#)
- Kaushik, K. (2024). Leveraging deep learning techniques for securing the internet of things in the age of big data. In *Applying Artificial Intelligence in Cybersecurity Analytics and Cyber Threat Detection*, 309. IGI USA. [↵](#)
- Kochhar, S. K., Bhatia, A., & Tomer, N. (2023). Using deep learning and big data analytics for managing cyber-attacks. In *New Approaches to Data Analytics and Internet of Things through Digital Twin* (pp. 146–178). IGI Global. [↵](#)
- Kordzadeh, N., & Ghasemaghahi, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388–409.
- Latif, S., Boulila, W., Koubaa, A., Zou, Z., & Ahmad, J. (2024). DTL-IDS: An optimized intrusion detection framework using deep transfer learning and genetic algorithm. *Journal of Network and Computer Applications*, 221, 103784. [↵](#)
- Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, 63(2), 157–170. [↵](#)
- Li, L., Zhao, T., Xie, Y., & Feng, Y. (2020). Interpretable machine learning based on integration of NLP and psychology in peer-to-peer lending risk evaluation. In *Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October*

- 14–18, 2020, *Proceedings, Part II* 9 (pp. 429–441). Springer International Publishing. [↗](#)
- Liu, C., Chen, B., Shao, W., Zhang, C., Wong, K., & Zhang, Y. (2024). Unraveling attacks in machine learning-based iot ecosystems: A survey and the open libraries behind them. *arXiv preprint arXiv:2401.11723*. [↗](#)
- Lone, A. N., Mustajab, S., & Alam, M. (2023). A comprehensive study on cybersecurity challenges and opportunities in the IoT world. *Security and Privacy*, 6(6), e318. [↗](#)
- Lui, A., & Lamb, G. W. (2018). Artificial intelligence and augmented intelligence collaboration: Regaining trust and confidence in the financial sector. *Information & Communications Technology Law*, 27(3), 267–283. [↗](#)
- Manimaran, M., Dhar, M., Norabuena-Figueroa, R., Mahaveerakannan, R., Saraswathi, S., & Selvakumarasamy, K. (2023). Implementing machine learning-based autonomic cyber defense for IoT-enabled healthcare devices. *Journal of Artificial Intelligence and Technology*, 3(4), 162–172. [↗](#)
- McKendrick, K. (2019). *Artificial Intelligence Prediction and Counterterrorism*. London: The Royal Institute of International Affairs-Chatham House, 9. [↗](#)
- Messai, M. L., & Seba, H. (2023, August). IoT network attack detection: Leveraging graph learning for enhanced security. In *Proceedings of the 18th International Conference on Availability, Reliability and Security* (pp. 1–7). ACM. [↗](#)
- Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International Journal of Financial Studies*, 9(3), 39. [↗](#)
- Mijwil, M., & Aljanabi, M. (2023). Towards artificial intelligence-based cybersecurity: The practices and ChatGPT generated ways to combat cybercrime. *Iraqi Journal for Computer Science and Mathematics*, 4(1), 65–70. [↗](#)

- Mijwil, M., Salem, I. E., & Ismaeel, M. M. (2023). The significance of machine learning and deep learning techniques in cybersecurity: A comprehensive review. *Iraqi Journal For Computer Science and Mathematics*, 4(1), 87–101. [↗](#)
- Mühlhoff, R. (2021). Predictive privacy: Towards an applied ethics of data analytics. *Ethics and Information Technology*, 23(4), 675–690. [↗](#)
- Narayana, Y. V., & Sreedevi, M. (2023, October). DDCATF: Deep Learning Approach for Detection of Cybercrime Activities Based on Temporal Features. In *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)* (pp. 462–469). IEEE. [↗](#)
- Nassar, A., & Kamal, M. (2021). Ethical dilemmas in AI-powered decision-making: A deep dive into big data-driven ethical considerations. *International Journal of Responsible Artificial Intelligence*, 11(8), 1–11. [↗](#)
- Nisha, S. R., & Muthurajkumar, S. (2023). Semantic graph based convolutional neural network for spam e-mail classification in cybercrime applications. *International Journal of Computers Communications & Control*, 18(1), 55–56. [↗](#)
- Odeh, A., & Abu Taleb, A. (2023). Ensemble-based deep learning models for enhancing IoT intrusion detection. *Applied Sciences*, 13(21), 11985. [↗](#)
- Ortiz-Ruiz, E., Bermejo, J. R., Sicilia, J. A., & Bermejo, J. (2024). Machine learning techniques for cyberattack prevention in IoT systems: A comparative perspective of cybersecurity and cyberdefense in Colombia. *Electronics*, 13(5), 824. [↗](#)
- Paddalwar, S., Ragha, L., & Mane, V. (2023). Cyber threat mitigation using machine learning, deep learning, artificial intelligence, and blockchain. In *Intelligent Approaches to Cyber Security* (pp. 163–179). Chapman and Hall/CRC. [↗](#)
- Pocher, N., Zichichi, M., Merizzi, F., Shafiq, M. Z., & Ferretti, S. (2023). Detecting anomalous cryptocurrency transactions: An AML/CFT application of machine learning-based forensics. *Electronic Markets*, 33(1), 37. [↗](#)

- Rai, R., & Sahu, C. K. (2020). Driven by data or derived through physics? A review of hybrid physics guided machine learning techniques with cyber–physical system (CPS) focus. *IEEE Access*, 8, 71050–71073. [↵](#)
- Rajkumar, S., Sheeba, S. L., Sivakami, R., Prabu, S., & Selvarani, A. (2023). An IoT-based deep learning approach for online fault detection against cyber-attacks. *SN Computer Science*, 4(4), 393. [↵](#)
- Saltz, J., Skirpan, M., Fiesler, C., Gorelick, M., Yeh, T., Heckman, R., & Beard, N. (2019). Integrating ethics within machine learning courses. *ACM Transactions on Computing Education (TOCE)*, 19(4), 1–26. [↵](#)
- Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., & Prabhakaran, V. (2021, March). Re-imagining algorithmic fairness in india and beyond. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 315–328).
- Schmitt, M. (2020). Artificial intelligence in business analytics, capturing value with machine learning applications in financial services.
- Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham L. Rev.*, 87, 1085. [↵](#)
- Sharma, A., Mansotra, V., & Singh, K. (2023). Detection of mirai botnet attacks on IoT devices using deep learning. *Journal of Scientific Research and Technology*, 1, 174–187. [↵](#)
- Shukla, R. K., & Tiwari, A. K. (2024). Security analysis of the cyber crime. In *The Ethical Frontier of AI and Data Analysis* (pp. 257–271). IGI Global. [↵](#)
- Singh, B. (2023a). Blockchain technology in renovating healthcare: Legal and future perspectives. In *Revolutionizing Healthcare Through Artificial Intelligence and Internet of Things Applications* (pp. 177–186). IGI Global. [↵](#)
- Singh, B. (2023b). Federated learning for envision future trajectory smart transport system for climate preservation and smart green planet: Insights into global governance and SDG-9 (Industry, Innovation and Infrastructure). *National Journal of Environmental Law*, 6(2), 6–17. [↵](#)

- Singh, B. (2023c). Tele-health monitoring lensing deep neural learning structure: Ambient patient wellness via wearable devices for real-time alerts and interventions. *Indian Journal of Health and Medical Law*, 6(2), 12–16.
- Singh, B. (2023d). Unleashing alternative dispute resolution (ADR) in resolving complex legal-technical issues arising in cyberspace lensing e-commerce and intellectual property: Proliferation of e-commerce digital economy. *Revista Brasileira de Alternative Dispute Resolution-Brazilian Journal of Alternative Dispute Resolution-RBADR*, 5(10), 81–105.
- Singh, B., & Kaunert, C. (2024). Integration of cutting-edge technologies such as Internet of Things (IoT) and 5G in health monitoring systems: A comprehensive legal analysis and futuristic outcomes. *GLS Law Journal*, 6(1), 13–20. [↵](#)
- Singh, N. J., Hoque, N., Singh, K. R., & Bhattacharyya, D. K. (2023). Botnet-based IoT network traffic analysis using deep learning. *Security and Privacy*, e355. [↵](#)
- Siwakoti, Y. R., Bhurtel, M., Rawat, D. B., Oest, A., & Johnson, R. C. (2023). Advances in IOT security: Vulnerabilities, enabled Criminal Services, attacks and countermeasures. *IEEE Internet of Things Journal*, 1, 55–57.
- Sudhakar, & Kumar, S. (2023, April). ABBDIoT: Anomaly-Based Botnet detection using machine learning model in the Internet of Things network. In *International Conference on IoT, Intelligent Computing and Security: Select Proceedings of IICS 2021* (pp. 235–245). Singapore: Springer Nature Singapore. [↵](#)
- Swiątkowska, J. (2020). Tackling cybercrime to unleash developing countries' digital potential. *Pathways for Prosperity Commission Background Paper Series*, 33, 2020–01. [↵](#)
- Thapaliya, S., & Sharma, P. K. (2023a). Cyber forensic investigation in IoT using deep learning based feature fusion in big data. *International Journal of Wireless Information Networks*, 30(1), 16–29. [↵](#)

- Thapaliya, S., & Sharma, P. K. (2023b). Optimized deep neuro fuzzy network for cyber forensic investigation in big data-based IoT infrastructures. *International Journal of Information Security and Privacy (IJISP)*, 17(1), 1–22.
- Tippins, N. T., Oswald, F. L., & McPhail, S. M. (2021). Scientific, legal, and ethical concerns about AI-based personnel selection tools: A call to action. *Personnel Assessment and Decisions*, 7(2), 1.
- Tyagi, A. K., Aswathy, S. U., & Abraham, A. (2020). Integrating blockchain technology and artificial intelligence: Synergies perspectives challenges and research directions. *Journal of Information Assurance and Security*, 15(5), 1554. [↵](#)
- Utami, S. H., Purnama, A. A., Hidayanto, A. N., Utami, S. H., Purnama, A. A., & Hidayanto, A. N. (2022). Fintech lending in Indonesia: A sentiment analysis, topic modelling, and social network analysis using Twitter data. *International Journal of Applied Engineering & Technology*, 4(1), 50–56. [↵](#)
- Varma, P., Nijjer, S., Sood, K., Grima, S., & Rupeika-Apoga, R. (2022). Thematic analysis of Financial Technology (Fintech) influence on the banking industry. *Risks*, 10(10), 186. [↵](#)
- Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2023). *Dive into Deep Learning*. Cambridge University Press. [↵](#)
- Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022). Interpretable machine learning models for crime prediction. *Computers, Environment and Urban Systems*, 94, 101789.
- Zhang, Y., Xiong, F., Xie, Y., Fan, X., & Gu, H. (2020). The impact of artificial intelligence and blockchain on the accounting profession. *IEEE Access*, 8, 110461–110477. [↵](#)
- Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable artificial intelligence applications in cyber security: State-of-the-art in research. *IEEE Access*, 1, 12–14. [↵](#)

21 Role of Artificial Intelligence in Protection of Human Rights with Special Reference to International Law

Tarun Kumar Kaushik, Saquib Ahmed, Ravish, Anurag Singh, Anjali Raghav, and Saurabh Chandra

DOI: [10.1201/9781003668657-21](https://doi.org/10.1201/9781003668657-21)

21.1 INTRODUCTION

If this definition is correct, then artificial intelligence (AI) is just a system or computer that can mimic human intellect in some way. AI is “a science and a set of computational technologies that are inspired by—but typically operate quite differently from—the ways people use their nervous systems and bodies to sense, learn, reason, and take action,” according to a recent paper from Stanford University AI has been defined in a number of ways; the first was proposed in 1956 as part of the Dartmouth Summer Research Project. According to John McCarthy, a pioneer in the field, “intelligent” means that a system can carry out tasks that would be considered clever if a person were to carry them out.

Another way to look at AI is according to Stuart Russell and Peter Norving, who propose the following classifications: (1) systems with human-like cognition and behavior; (2) systems with human-like reasoning and logic-based behavior; and (3) systems with human-like reasoning and

logic-based behavior. The common thread running across all of these definitions is that AI systems exist primarily to make human jobs easier. When an intelligent system completes a task, the end product is virtually indistinguishable from that of a human operator. So, anytime AI-generated intelligent systems are involved, human rights must be included in any discussion of their relationship with people. Put simply, the development of AI will unavoidably have an impact on human rights, whether that impact is beneficial or negative ([Djenna et al., 2023](#)).

21.2 CONCEPTUALIZING HUMAN RIGHTS UNDER RELEVANT DOMESTIC AND INTERNATIONAL LAWS

As it pertains to every individual's fundamental life, regardless of their location or cultural background, the idea of human rights has garnered a great deal of focus and has been handled with great care. A society's fundamental entitlements are its members' human rights. Because people are human, these things are relevant to them. To live a civilized life in any culture, they are an absolute need. Human rights are a set of universally recognized protections for individual liberty, dignity, equality, and respect. According to Kayode Eso J.S.C. (as he was at the time) in the case of *Ransome Kuti & ORS v. A.G Federation & OR*, human rights are those that have always been there, even before there were laws for how to seek them out, and this definition is as short as it gets. Over and above the usual rules of the state, it is a fundamental requirement for a civilized life ([Awajan, 2023](#)).

In accordance with universal principles recognized by modern morality and international law, every individual has an inherent and inalienable right to a certain set of liberties, protections, and advantages, collectively known as human rights. A person cannot maintain his humanity in the absence of human rights, which are so essential to the human condition. Human rights are essential to a meaningful and valuable existence. In other words, there must be strong and compelling reasons to violate human rights because they are too valuable.

Human rights are a standard component of national legislation worldwide. The domain of domestic law continues to encompass fundamental rights. Because the Constitution, the supreme law of the land, guarantees them, we may say that they are basic. Thus, the rights guaranteed by domestic law are those that are explicitly stated or included in the domestic legal framework, most notably the country's constitution or statutes. The following human rights are encompassed in the concept of human rights: the right to exist, the right to be treated with respect, the right to be free from arbitrary interference, the right to peacefully assemble and associate with others, the right to travel and possession of real estate, the right to be free from discrimination, and the right to private and family life. Although human rights are guaranteed by legislation in many nations, it is worth noting that the same is also guaranteed on a global scale. The establishment of human rights in many nations really predates the international acknowledgement and existence of human rights. Human rights are a central concern of many international organizations, and as a result, there are treaties and conventions governing this area that member nations must ratify and adopt into law ([Attaran & Deb, 2018](#)).

Unless the National Assembly passes a bill implementing a particular treaty, no agreement between the Federation and another nation will be legally binding.

The 1948 Universal Declaration of Human Rights is a landmark international document on human rights. An important text that stated basic human rights and asked all nations to defend them is the "Universal Declaration of Human Rights" (UDHR). An answer to the longing for human rights to be included as a globally acknowledged rule, the UDHR legislation is extremely foundational to the history of law. The United Nations General Assembly in Paris on 10 December 1948 recognized the Declaration as a common standard of accomplishments for all peoples and all nations. It was drafted by delegates with diverse legal and cultural backgrounds from all corners of the world ([Brotcke, 2022](#)).

For the first time, it lays forth the universal protection of basic human rights. The International Covenant on Civil and Political Rights (ICCPR) is

another body of international law that defines a set of human rights. It is upon the UDHR that the ICCPR is based. All citizens in democracies have the fundamental rights outlined in this covenant, which can be enforced through the courts ([Abebe et al., 2020](#)). Notably, both national and international legislation around the globe have recognized the importance of human rights. Human rights have always existed, even before the political establishment of any nation; therefore, they are not contingent on these laws. With that said, the implementation of these laws, in whatever form they may take, is now critical to the well-being of the whole nation ([Elluri et al., 2023](#)).

21.3 THE IMPACTS OF AI ON RELEVANT HUMAN RIGHTS

Some have raised concerns about the future of human rights in light of AI progress, even though these protections are fundamental to human survival. As a result of these changes, people feel strongly about whether or not it is necessary. An essential point to keep in mind is that AI systems are developed by people. The development of AI may be attributed to the fact that it has been beneficial in streamlining and improving many human endeavors. AI is not starting from scratch or being created in a void. In reality, AI is being used by some social actors to automate decision-making in certain domains. This is happening in the context of social institutions that already have human rights implications, but they are doing it to get what they want ([Hagendorff, 2019](#)).

A number of viewpoints have been voiced concerning the potential impact of AI on human rights due to the inherent nature of AI making choices on behalf of humans. One might have a favorable or negative stance on these issues. Without a doubt, it demonstrates AI's contribution to the improvement of every aspect of life, in line with its future prospects. The development of some AI systems has made it simpler to track human rights abuses, which has led to a loosening of standards for proving such violations in court. Some have argued that the “golden age of surveillance”

has begun with the advent of digital technology in the 21st century, with nations, companies, and non-state entities all participating ([Singh & Kaunert, 2024](#)).

The vast quantities of open-source data produced by the billions of sensor platforms in circulation worldwide are accessible to news organizations, human rights groups, and open-source investigators like Bellingcat and the Syrian Archive. Hundreds of high-resolution imaging satellites, multi-function mobile phones, and other imaging platforms have given rise to a sort of surveillance that is becoming increasingly common. That way, whether on the ground or in orbit, hardly anything happens without a camera recording it. There are two possible ways to represent this wave: favorably and negatively. Positively, the application of AI has the potential to enhance a state's security. Included in this is the battle against terrorists. In 2017, for example, a Libyan warlord was indicted by the International Criminal Court after satellite footage and films of his executions, either ordered or carried out by his supporters, were shared online. This is only one example of the various ways AI may improve human rights ([Thapaliya & Sharma, 2023](#)).

21.3.1 THE RELATIONSHIP BETWEEN AI AND INDIVIDUAL FREEDOM

An essential human right, the right to personal liberty, includes the freedom to move about within a specific area. Although this privilege can be extended to any area, one must meet the requirements set by that particular country or location. An individual's right to liberty necessitates that any detention or arrest must adhere to legal standards. The fundamental freedom that no government has the authority to limit is the right to personal liberty. Without a doubt, the guarantee of individual liberty is a fundamental human right. All around the globe, political societies have passed laws formally recognizing these rights. Everyone has the right to their own freedom, and nobody has the right to have it taken away from them. As a result of its inclusion in several treaties and legislation that numerous states have

ratified, the right to liberty is highly revered in the worldwide community ([Dash et al., 2022](#)). Individuals have the right to personal liberty as outlined in the ICCPR, a legislation that is based on the UDHR:

The right to personal freedom and security is universal. Arresting or detaining somebody without probable cause is absolutely unacceptable. No one's liberty shall be taken away from them unless the law establishes the grounds and the method for doing so. The International Court of Justice, in the case of *America v. Iran*, highlighted the importance of personal liberty and stated that it is incompatible with the United Nations Charter and the UDHR, which guarantees the right to life, liberty, and security of the human person (Article 3). It is also incompatible with the practice of physically limiting or depriving people of their freedom in times of hardship ([Barocas et al., 2023](#)).

By creating programs that assess risk and the probability of a crime occurring at a location or by an individual, AI has planted its digital paws in the criminal justice systems of several countries. Police deployment is informed by projected locations where crimes are most likely to occur by predictive policing algorithms like PredPol and HunchLab.

Despite the undeniable benefits of this unparalleled expansion, there is a real danger that it may infringe upon the right to personal liberty and the right to be free. Risk assessment instruments (RAIs) are a type of computational tool that can foretell the likelihood of future wrongdoing by a defendant. Decisions that have significant judicial weight, such as whether or not to detain a defendant before trial, are informed by these forecasts. When a computer decides whether or not to send a suspect to prison, it will appear strange. Members of the legal profession who are concerned about this trend have rejected it. Recidivism is (another) case in point. Software that assesses risk is utilized by the American criminal justice system to guide detention choices at practically every step, from bail assignment to criminal punishment. More Black defendants have been wrongly classified as high risk, leading to harsher bail restrictions, pre-trial detention, and lengthier prison sentences as a result of the program. Not only in the USA but throughout the globe, this is blatantly racist behavior

toward Black people that has led to the abuse of AI. Furthermore, detention choices influenced by risk-scoring systems might be arbitrary or illegal due to the lack of legal prescriptions for these systems and the arbitrary nature of their inputs ([Narayana & Sreedevi, 2023](#)).

21.3.2 A FAIR TRIAL AND AI

- Every person, regardless of citizenship status, has the right to a fair trial when their case is heard within a courtroom. In any situation, whether a criminal or civil proceeding, the right to a fair hearing must be upheld. The defendant is guaranteed their rights in a trial that is handled in a fair, reasonable, and procedurally regular manner by an unbiased judge. A person has the right to a fair trial as guaranteed by current legislation.

Everyone has the right to a fair trial before a legally recognized and impartial tribunal within a reasonable amount of time when their civil rights and responsibilities are being decided, including when those matters involve or are challenged by any government or authority ([Mhlanga, 2021](#)).

- In Article 14 of the ICCPR, the identical paragraph is reproduced (1). Technologies developed by AI have an impact on the criminal justice system and the judicial system more generally. Robot judges and AI assistance are the two sides of this game. AI assistants are programs that help courts with the tedious but necessary task of drafting decisions and information for judicial determination. In contrast, robot judges are programmed to carry out all judicial duties in the same way that a human judge would, effectively replacing human judges and turning the courtroom into an entirely digital environment ([Pocher et al., 2023](#)).
- Some have voiced concerns that the development of these programs violates the right to a fair trial. Because algorithms reflect the values of its creators in their design and operation, bias might arise both in the data from which the AI system learns and in the AI system's own

programming. When this happens, it is not always easy to tell if the serious mistake was intentional or the product of a glitch in the system. Algorithms that reinforce preexisting prejudices and preconceptions, even in the absence of explicit knowledge about the protected group, pose a threat of discrimination in an AI system, even though AI systems are not intentionally biased. In July 2016, the Wisconsin Supreme Court rendered a historic decision in the matter of *State v. Loomis*, which included risk assessment software. Mr. Loomis denied any participation in the incident and pleaded guilty to two of the five counts against him that were related to a drive-by shooting crime in Wisconsin. After being deemed “at high risk to the community” and having a high-risk record of recidivism, Eric Loomis was handed a six-year jail term by COMPAS. *State v. Loomis* highlights several troubling issues about risk assessment tools, which might undermine human rights and the rule of law, even if the Wisconsin Supreme Court supported their use at sentencing ([Chen et al., 2018](#)).

21.3.3 THE RELATIONSHIP BETWEEN AI AND INDIVIDUAL PRIVACY RIGHTS

Many legal systems provide protections for citizens’ right to privacy as a means of limiting both public and private intrusions into people’s personal lives. Privacy is defined as the following: the right to one’s own space, free from prying eyes; the ability to decide who may see one’s private information; the preservation of one’s autonomy, honor, and dignity; the right to remain anonymous and unobserved; and the right to avoid having one’s private affairs pryed into. Protecting one’s bodily autonomy, limiting access to oneself, controlling one’s identity, and regulations governing the acquisition and management of personal data are all aspects of the right to privacy ([Tyagi et al., 2020](#)). An individual’s right to privacy is upheld in the ICCPR:

1. No one has the right to have his or her personal information, home, family, or communications invaded in any way, shape, or form, or to have his or her honor or reputation attacked in any way.
2. The rule of law protects all citizens against unlawful intrusion or assault. Citizens' right to privacy is now secured and preserved, including their houses, letters, right to telephone calls, and telegraphic communications. This clause adheres to this fundamental human right.

A growing number of people are worried that their right to privacy may be violated because of AI development. Security cameras, cell phones, and other contemporary devices pose a risk to people's right to privacy. Companies like Amazon, Apple, Facebook, Google, and Yahoo that rely on technology to operate pose a risk to individuals' right to privacy because of the personal information they gather and utilize. The Cambridge Analytica data scandal, which centered on the psychographic firm's exploitation of Facebook user data to influence huge groups of people, is just one example of how scandals like this have heightened these worries. Although these events have not undermined AI's practical applications, they have shown the flagrant disregard with which many people throughout the world regard their right to privacy (Cao et al. 2021).

21.3.4 THE RIGHT TO BE FREE FROM DISCRIMINATION AND AI

The subject of discrimination is often handled with care because of the many concerns that surround it, the strong emotions it evokes, and the values and norms that surround it in society. [Regarding legal matters], the inequitable application of policies, laws, or treatment prevents certain individuals from fully realizing their human or other legal rights; this is known as discrimination. Certain laws have been enacted in an effort to prohibit this intolerable treatment of any individual. The ICCPR includes a provision that prohibits discrimination in all its forms. It states that every state that is a party to the covenant promises to respect and provide the

rights stated in the covenant to all its citizens, regardless of their race, color, sex, language, religion, political or other beliefs, national or social origin, wealth, date of birth, or any other status ([Alazab et al. 2023](#)).

Discrimination, however, has been increasingly difficult to address since the rise of AI. Humans are capable of creating intelligent systems, and it is within their power to create a system whose basis is based on prejudice. A panel of specialists drafted the “Draft Ethics Guidelines for Trustworthy A.I.” in December 2018. The European Commission acknowledged the need for an anthropocentric approach to AI and issued a warning about the hazards connected with AI in this document, despite its significant advantages. People must always be given the authority to monitor machines, as this is the only strategy that can ensure their dignity and autonomy. It is now plainly apparent that even the Council of Europe perceives the potential for social discrimination caused by algorithms and issues a warning about it. Discrimination may take many forms, and one of them is the negative impact of AI on the right to a fair trial. This is unique to the claims made in response to the ProPublica investigation. Some technical innovations have their functionality limited to specific regions or users, as decided upon by the inventors themselves. All things considered, this is an example of bias brought about by AI (Zhang et al., 2022a).

21.3.5 MACHINE LEARNING AND THE ELIMINATION OF ARBITRARY DETENTION

There are certain limitations on people’s right to personal liberty, even if everyone has it. Arresting criminals or those suspected of committing crimes is one way that states punish those who break the law. The process of apprehending and detaining an individual, often on suspicion of or observation of their criminal behavior, is known as an arrest. Following an arrest, a suspect may be subject to further questioning and, if warranted, subsequent legal proceedings. However, an arrest must not be arbitrary and undertaken without probable cause. When there is no probable cause to believe that an individual has violated a law, or when due process of law

has not been followed, the arrest or detention of that individual might be considered an arbitrary arrest.

The most basic definition of an arbitrary arrest is one that is carried out with a hidden agenda that is blatantly contrary to the law and its procedure (Singh, 2023b). Because it restricts a person's mobility inside a certain area, this also constitutes a violation of their right to personal liberty. The advent of AI has, alas, exacerbated the problem of arbitrary detention. A false arrest, and by extension, an arbitrary arrest, is certain to occur if an intelligent system mistakenly identifies a person as the perpetrator of a certain crime, whether this be due to the systems' design flaws or the creator's deliberate intent. Global and international organizations have recently brought this growing threat to light. The United Nations is among these international organizations that have commented on this timely matter. New research on the matter was revealed in Geneva, and U.N. High Commissioner for Human Rights Michelle Bachelet made mention of it in her remarks. The analysis expressed concern about the potential consequences of AI's predictive and profiling capabilities, stating that the technology may affect fundamental rights such as "rights to privacy, to a fair trial, to freedom from arbitrary arrest and detention and the right to life." The obvious downsides of AI include its likely contempt for human rights to be free from arbitrary detention ([Lee & Shin, 2020](#)).

21.3.6 HOW AI COULD AFFECT FREE SPEECH

- Freedom of expression is one of the inherent human rights. The right of any person or organization to freely communicate its views, opinions, ideas, and feelings on any subject without interference from the state is known as freedom of speech. It boils down to the freedom to speak one's mind and advocate for one's cause in any way one sees fit. The following rights have been firmly established in the ICCPR: (1) The freedom of expression must not be violated in any way. (2) The right to freedom of expression extends to all people and includes the ability to freely seek, receive, and communicate ideas and

information in any form, including but not limited to spoken word, written word, visual art, and any other media one chooses, regardless of location or other restrictions ([Bhuvaneshwari & Kaythry, 2023](#)).

- Internet content curation has a long history, and AI is a natural extension of that history. With AI, we can personalize and curate material for users on a scale that traditional media just cannot match. Some worry that the prevalence of some AI-assisted curation methods would limit people's ability to think critically and independently. Because of AI, people all over the world can now express themselves more effectively, and the information they get is delivered at a lightning pace. The development of social media apps has given many people a platform to share their views and ideas. On the other hand, research has shown that this might have negative consequences ([Li et al., 2020](#)).
- A great deal of the moral and criminal wrongdoing that has recently come to light in the modern world has its roots in the birth of social media. Examples of such crimes include cyberbullying, cyberfraud, and the rise in the suicide rate because of negative peer pressure. It also brings up some valid points on how algorithmic bias might lead to the formation of echo chambers and the exacerbation of existing social and cultural biases ([Cheng et al., 2021](#)).

21.3.7 THE FREEDOM OF ASSEMBLY AND ASSOCIATION IN THE AGE OF AI

- Individuals have the right to form groups and have the freedom to communicate, advocate, seek, and defend their common beliefs and ideals. Human, political, and civic liberties all include the right to freely associate with others. They guarantee that people and organizations may peacefully join forces to achieve their shared objectives, which is crucial to the formation and maintenance of a democracy. A person's freedom of association and assembly is firmly enshrined in the following provisions of the ICCPR: Everyone has

the right to peacefully assemble. Restriction of this right is prohibited unless such limitations are mandated by law and are essential for a democratic society to ensure public safety, public health, morality, or national or public safety, public order, or the protection of other people's rights and freedoms ([Bhuvaneshwari & Kaythry, 2023](#)).

- A person's right to freely gather can be observed and addressed through surveillance. The term "surveillance" refers to the practice of keeping a person under close observation, often by a government agency or military force. The function of surveillance in authoritarian, hybrid, and democratic governments is greatly impacted by AI. The majority of AI's influence on surveillance in 2019 has been in "smart cities" programs, face recognition integrated with closed-circuit television, and crime analysis for smart police, all of which are presented as good uses of AI in surveillance. Conversely, any intelligent system controller may track users and perhaps limit their mobility based on their tags ([Mühlhoff, 2021](#)).

21.3.8 WHAT EFFECTS WILL AI HAVE ON ECONOMIC AND SOCIAL RIGHTS?

- While very crucial, the basic human rights of all people are far from being the only things that make life worth living. There comes a moment in every person's life when they must learn to survive on their own. Humans' social and economic rights have to be addressed by legislation in light of this inevitable reality. A person's economic, social, and cultural rights (or "social rights" for short) include protections in the job, financial stability, the ability to care for one's family, cultural engagement, and the availability of basic necessities like shelter, clothing, food, water, medical treatment, and an adequate education. Rights to a decent quality of life, fair earnings and equitable pay, and enough protection from unemployment, illness, and old age are all part of this category ([Saltz et al., 2019](#)). Many countries' legal systems have explicitly recognized and protected

these rights. Constitutions that solely protect political and civil liberties neglect to protect economic and social rights, leaving the task half-done and laying the groundwork for superficial, meaningless “bourgeois” freedom. The significance of socioeconomic rights has therefore been acknowledged throughout the globe. Social and economic rights have their roots in the early 20th century, when the International worker Organization (ILRO), which was a branch of the League of Nations, passed a number of treaties to raise global worker standards. The most significant treaties and agreements that included socio-economic rights after World War II were the International Covenant on Economic, Social, and Cultural Rights (ICESCR, 1966) and the UDHR (1948). The social and economic rights of individuals are not infringed upon by the pervasiveness of AI. To that end, AI can be useful in tracking how seriously people take these rights. When the time comes to assert their rights, employees who manage their register or operations online can use the same as proof. Notably, many nations’ social security systems are highly efficient because they run on the Internet. As an added bonus, AI streamlines the process of claiming a person’s housing rights, so individuals may seek a home without leaving their comfortable homes thanks to this technology. These are only a few of the numerous ways in which AI will affect people’s socioeconomic rights. The effect, nevertheless, can be detrimental in certain instances. One example of this is when people unwittingly become victims of cyber fraud ([Hojtink & Planqué-van Hardeveld, 2022](#)).

21.3.9 WORK AND EMPLOYMENT IN THE AGE OF AI

To put it simply, the right to work and employment means that everyone has the opportunity to make a living wage. This right can, at first glance, indicate that everyone has the right to an employment opportunity, but this is, by many accounts, a very limited interpretation. Instead, the term “right to work” has come to include a plethora of rights, including demands for

equal employment opportunity, respect for workers' dignity on the job, and so on ([Utami et al., 2022](#)). This is consistent with the right's overarching objective, which is to guarantee that all people can provide adequately for their families and themselves. Instances of this right being denied to immigrants in a country gave rise to the necessity for an international law to address and acknowledge this right. As stated in the following provision of the International Covenant on Economic, Social and Cultural Rights (ICESCR) (Zhang et al., 2022a):

1. The States Parties to the present Covenant acknowledge the right to work, which encompasses the right of every individual to earn a livelihood through work that he freely chooses or accepts, and will ensure that this right is protected via appropriate measures.
2. A State Party to the present Covenant is obligated to implement measures to ensure that this right is fully realized, such as vocational and technical guidance and training programs, policies, and techniques that promote sustainable economic, social, and cultural development, as well as full and productive employment, all while protecting individuals' fundamental political and economic freedoms ([Baker & Robinson, 2020](#)).

There has been a steady stream of improvements to corporate services and operations brought about by AI systems throughout the years. Despite workers being unaware that AI is processing massive amounts of data and information utilized to make choices affecting them, these developments are occurring outside of the usual framework of employee engagement in the organization. The story of people's working nature and conditions has been drastically altered by AI, which has essentially taken control. This is the same as trying to influence people's employment and right to work ([Aziz & Andriansyah, 2023](#)).

A trustworthy regulatory framework has been developed using AI, which lessens the likelihood of human rights breaches. Making sure people's rights are not violated is, of course, one of the state government's jobs. The introduction of AI has sparked debates over whether or not this

duty ought to be transferred. Any violation of rights can occur regardless of whether harmful automated devices are part of decentralized or privatized benefit programs. The administration and access to benefits are likewise subject to certain strict criteria ([Bao et al., 2022](#)). The requirements for receiving social assistance, for instance, ought to be open and obvious to those who seek them. The Economic, Socio-Cultural (ESC) framework also guarantees that all citizens are eligible to receive government-funded social services. Because of its centrality to the actualization of ESC rights, the concept of non-discrimination and equality must also be actively pursued in the pursuit of ESC rights. This may spark the now-famous debate over whether or not AI and automated systems are biased. These protections are in place to make sure that AI does not cut someone off from social assistance, but they also take a broader view of the ways in which digital technologies contribute to inequality in the workplace ([McKendrick, 2019](#)).

21.3.10 THE EFFECTS OF AI ON THE RIGHT OF CHILDREN

Human rights have evolved to encompass a wide range of concepts, many of which have focused on adults. As a result, the topic of children's rights has gained traction, and as a result, it has become a subject of international law, prompting several countries to acknowledge and codify children's rights ([Rai & Sahu, 2020](#)). There is a subcategory of human rights known as children's rights that deals specifically with the need to provide special care and protection to children. Human rights include the rights of children. No matter their status as "adults of tomorrow" or "the future," children deserve our utmost respect and decency just because they are people right now. A set of basic liberties and rights is innate to every human being. From the right to privacy to the right to free speech, children should have the same human rights as everyone else. That is to say, a kid has the same legal protections as an adult. To protect children's rights, the international community drew up the Convention on the Rights of the Child, a treaty that applies worldwide once approved by a particular country. Because it states

categorically that children are unique individuals with rights and identities, that they require protection, and that their opinions should be heard and respected, it is a crucial text. It formalized the bond between children and the state and the responsibilities and privileges that come with it on matters such as children's right to privacy, freedom of speech, protection from physical harm, and health, among many others. The Convention on the Rights of the Child's preamble makes it clear how important children's rights are by stating: The right to receive particular care and aid for children is a fundamental human right, as stated by the United Nations in the UDHR ([Sharma et al., 2023](#)),

I am deeply convinced that families deserve the support and protection they need to fulfil their communal obligations, both as the building blocks of society and as the primary setting for the development and flourishing of all family members, especially children.

Because a child needs a loving and supportive home life to flourish and develop to his or her full potential, it is imperative that children be raised in accordance with the principles outlined in the United Nations Charter, particularly those pertaining to respect for human life, liberty, equality before the law, and mutual aid among nations.

When it comes to children's rights and AI, streaming-based social media platforms are changing the game for media consumption among both adults and kids. A whopping 96% of kids in the 6–12 age bracket were identified as “aware of YouTube,” with an even more impressive 94% claiming to have “either loved or liked” the platform. Despite YouTube's restrictions on food and drink ads on YouTube Kids, some food companies still manage to target children with their products by using branded channels to highlight specific foods and drinks. This tactic involves hiding what are essentially ads within programs. As a result, businesses are figuring out methods to market to children in ways that are unclear to parents and other stakeholders, while yet adhering to the guidelines in theory. Toys with limited interactivity have given way to smart toys with the ability to reply in the past twenty years, drastically altering children's playtime. These playthings use a collection of approaches that enable computers to simulate

human reasoning and behavior through the use of rudimentary AI. There are several issues about human rights that these toys bring up. There are a number of concerns, including as the possibility of privacy invasions and the question of whether or not companies need to disclose sensitive information saved online and shared with toys, such as signs that a kid may be mistreated or otherwise hurt ([Sharma et al., 2023](#)).

Along with the aforementioned software and technologies, educational practices and methodologies are being revolutionized by robots that include AI. There are many different uses for robots, and their introduction to classrooms is changing the way kids learn. There are both advantages and disadvantages to educational robots. On the one hand, they offer individualized lessons, support youngsters in building social skills, allow students in faraway places to get an education, etc. On the other hand, there are certain concerns. Their usage might have both good and bad effects on some human rights, including as the right to an education, the right to be free from abuse and exploitation, and the right to have disability respected ([Sharma et al., 2023](#)).

21.3.11 AI'S INFLUENCE ON LEGAL STANDARDS

For a nation to achieve its goals of development and advancement, the rule of law must be upheld. Peace, political stability, economic growth, social progress, and the protection of individual liberties and rights all depend on a society based on the rule of law. To build the social compact between the people and the state, limit the misuse of authority, ensure that public services are accessible to everyone, and combat corruption, it is essential.

A.V. Dicey was a major proponent of the rule of law idea. "Introduction to the Study of the Law of the Constitution" laid out Dicey's view of the regulatory framework. The primacy of the Constitution, equality before the law, and respect for basic human rights are the three pillars upon which A.V. Dicey's rule of law rests. No one is immune to the rule of law, according to the idea of legal equality. Anyone, regardless of status or position, is bound to the regular tribunals' jurisdiction and subject to the

usual laws of the realm. Put another way, the treatment of the impoverished in cases of illegal behavior will not differ from that of the wealthy, irrespective of their position, influence, or prestige. With respect to the rule of law, the idea of constitutional supremacy states that the Constitution is the highest law of the nation and that other branches of government, including federal and state legislatures, are obligated to follow it. Every move they make must be grounded in the Constitution, as the document is the foundation of their authority and the basis for their very being. Consequently, they must adhere to the constraints imposed by the Constitution. Respect for the basic human rights of individuals is just as important as guaranteeing the aforementioned values. A person's ability to live depends on these fundamental rights, which must be upheld regardless of whether they are codified in law or not.

The advent of AI is responsible for the widespread digitization of human society. The use of these AI technologies is already prevalent in the pre-adjudicative stage of law enforcement, where they help with tasks such as selecting individuals to be investigated in tax and regulatory matters. They are also being increasingly utilized in the adjudication process, for instance to help with pretrial bail decisions involving the danger of violence to a person. Some predict that code-driven alternatives will eventually replace human judges, while others voice strong opposition to this idea. Most people are against it because they think it would make adjudicative systems less fair, less transparent, and less equitable. The rule of law is frequently intersected with or extremely similar to these AI concerns, even when they are not explicitly targeted at it; this is particularly true in the areas of human rights and democracy ([Varma et al., 2022](#)).

Because they involve very complicated technology that is fundamentally beyond the cognitive abilities of people and because the law also prevents transparency to some extent, most of these decision-making systems are "black boxes," which poses a danger to the rule of law. As a result of AI being criticized in this area, it will be nearly impossible to examine how the rule of law functions, including being transparent, fair, and explainable. In light of the world's exponential technological development, the question

arises as to how the rule of law can be rigorously upheld ([Tippins et al., 2021](#)).

21.3.12 AI'S ROLE IN UPHOLDING DEMOCRATIC INSTITUTIONS' LEGITIMACY

To begin with, democratic institutions, often called the institutions of democracy, are the primary tools for achieving the fundamental purpose of democracy in any government: making sure that power is exercised by and for the people. The executive, legislative, and judiciary branches make up a well-functioning democratic government. In most democracies, these branches also work together. As institutions, they are burdened with the responsibilities that come with their status: Separate branches of government are responsible for creating laws, carrying them out, and interpreting them. Democracy necessitates not only the presence of these institutions but also the immediate and obligatory recognition of its legitimacy by the people.

If a political institution can convince its constituents that it is normatively necessary for them to have it, then it has achieved legitimacy. There are a number of possible bases for this in a democratic society, including different types of public approval (democratic legitimacy) and alternative bases (generic legitimacy). This is where AI really shines. More and more, political institutions must devote themselves to data-driven governance if they want to increase their credibility. Institutions in charge of politics should aim to make decisions based on technologically driven, solid empirical evidence as they incorporate digital data into their procedures. In instances, algorithmic decision-making (ADM) systems are employed for the purpose of identifying critical social issues, predicting possible policy outcomes, enhancing the policy process, and assessing the efficacy of policy. As an example, research has demonstrated that ADM systems can effectively make decisions on how to distribute migrants in a way that is acceptable to society. When compared to human-led distribution

procedures, trials show that this method enhances refugee employment rates by 40–70% ([Lui & Lamb, 2018](#)).

21.4 THE USE OF AI IN THE COURT SYSTEM

Being a democratic institution, the court system essentially represents the government's judiciary. The system of courts that interprets, defends, and implements the law in legal situations is called the judiciary. It goes by several names, including the judicial system, judicature, the judicial branch, judicative branch, and court or judiciary system. Laws' interpretation, practical applicability, and constitutionality are all matters for the judicial branch to decide ([Mijwil et al., 2023](#)).

The judicial branch is responsible for representing the state in legal matters by interpreting, defending, and applying the law. The judicial system is also seen as a means of conflict settlement. Judicial functions are limited to interpreting, defending, and applying the law to the circumstances of each case in accordance with the theory of separation of powers; the duty for making statutes or enforcing laws belongs with the legislative and the executive branch, respectively. Judicial review, which occurs when a matter is taken to a court of law, gives the judiciary the authority to amend laws in many nations. It is also within the purview of the court to overturn legislation, particularly where such laws go against to a nation's fundamental standard (the Constitution in some instances). This body of law comprises fundamental statutes, constitutional provisions, treaties, and international law. The judiciary is an essential branch of government for any nation. It serves the state's progress and the people's needs; after all, the average person looks to the court as his or her final resort ([Schmitt, 2020](#)).

The legal system of a country is just as important as any other aspect of AI's reach. By making the application of the law more predictable and the consistency of case law more consistent, the processing of judicial data by so-called AI systems or methodologies developed from data sciences is anticipated to increase the transparency of the functioning of justice. An

example of an AI-developed technology is a robot-judge, which can carry out all of the duties traditionally associated with a human judge. Also, an AI system called SUVAS can help in translating decisions into regional languages. A further effect of AI is that it streamlines the judicial procedure, which in turn decreases the number of cases filed ([Siwakoti et al., 2023](#)).

21.4.1 INSIGHTS INTO AI'S POTENTIAL EFFECTS ON DEMOCRACY

In modern times, democracies have become the norm for effective national leadership. In essence, this guarantees that the well-being and will of the people are the bedrock upon which every government rests. When all citizens, either directly or via their representatives in elected office, have equal say in major policy decisions, we say that we live in a democracy. Simply put, it denotes popular rule. Peaceful transitions of power are hallmarks of democracies, and the legitimacy of government is based on popular vote. “Democracy” means “government by the people, for the people” in its simplest form ([Zhang et al., 2020](#)).

Because of the interconnectedness of all things, AI has both direct and indirect effects on democracies. The rule of law is an essential component of a democratic society. Cases involving this area of adjudication frequently make use of AI, such as when deciding whether or not to grant pretrial bail based on an individual's likelihood of violence. Transparency is a fundamental principle of AI. This means that everyone has the right to know how and why an algorithm is considering various aspects when making a choice that can impact them. Despite this, the democratic principle may still be used to lessen its impact by passing legislation that forbids algorithmic bias in society ([Nassar & Kamal, 2021](#)).

21.4.2 THE ROLE OF AI IN ELECTION PROCEDURES

The electoral process of a government is the mechanism of voting that is used to install a new government or, in some cases, to remove an existing one. A voting system, often called an electoral system, is a collection of regulations that govern the process and outcome of elections and referendums. Businesses, nonprofits, and informal groups may have elections that aren't related to politics, although governments do organize electoral systems that are. Every part of the voting process is regulated by these rules: when elections take place, who is eligible to vote, who can run for office, how ballots are marked and cast, how they are counted, how votes are turned into an outcome, campaign spending limits, and everything else that can impact the outcome. Election commissions normally oversee political electoral systems, which are specified by constitutions and electoral laws and might employ various election forms for various posts. Conventional voting methods in a country might be cumbersome and unreliable at times. In this area, AI has the potential to bring about a good and substantial shift, making the process more responsive to the will of the people during times of government transition ([Alruwaili, 2023](#)).

It is for the express aim of conducting national elections that software and intelligent systems have been developed. By streamlining the process of organizing and participating in the electoral processes and disseminating relevant information to the public, these platforms aim to make life easier for everyone involved. Nationwide voting processes have been enhanced by the incorporation of IT and software. As an illustration, The New Zealand-based AI SAM is assisting Alice and Michihito. The creation of SAM, an artificially intelligent woman, has sparked claims that she will be the first virtual politician to compete in the 2020 global elections. Among other things, SAM is using Facebook Messenger to communicate with people today, offering her views on climate change, healthcare, and education ([Grimm et al., 2021](#)). To aid with the general elections, an app for the Nomination Information System (Silon) has been developed. This app is designed to give registered voters the information that voting candidates need. For all types of party nominations necessary for registration as an election participant, political parties can use the Political Party Information

System (Sipol) app. To aid the organizers in recapitulating the election results, the General Elections Commission has also utilized AI through the Counting Information System (Situng) program during the computation stage of the final election results. After they are implemented, AI systems clearly have an effect on the voting process ([Dhoni & Kumar, 2023](#)).

21.4.3 AI AND MILITARY OPERATIONS

A country's military actions are known as warfare. War is when one political entity (such a nation) actively seeks to destabilize or destroy another. Every nation's survival depends on its mastery of combat. This is because a nation must protect her inhabitants against intruders both within and outside the country. At this point, it is necessary to employ strategies and weaponry. The use of AI models is growing in numerous fields, including military strategy, and they are becoming better every day. Tasks like recognizing a T-90 main battle tank in a satellite image, detecting valuable targets in a crowd using facial recognition, translating text for open-source intelligence, and generating text for information operations are all within the capabilities of these systems ([Kordzadeh & Ghasemaghaci, 2022](#)).

AI security choices, bioinformatics, practice aids for training troops, data aggregation from nationwide databases, and other applications may all benefit from intelligent computers. Scientists have begun using AI in the defense industry to compensate for human shortcomings, thanks to the proficiency that technology brings. More and more, they are relying on AI-powered combat systems, recovery and maintenance vehicles, unmanned aerial vehicles, and ground vehicles. The invention of several weapons, including drones, helipads with shooting capabilities, explosives, and many more, may be attributed in part to the effect of AI on combat ([Mijwil & Aljanabi, 2023](#)).

21.5 DIFFICULTIES IN ADDRESSING THE EFFECTS OF AI ON FREEDOM, DEMOCRACY, AND THE JUSTICE

SYSTEM

While there is no denying that AI has had far-reaching good effects on our modern society, it has also presented some serious problems, especially with regard to human rights and democracy. Every legitimate political system rests on three pillars: human rights, democracy, and the law. The foundation of every government must be the welfare and decision of the people, and democracy guarantees this. Peace, political stability, economic growth, social progress, and the protection of individual liberties and rights all depend on a society based on the rule of law (Sambasivan, 2021). To build the social compact between the people and the state, limit the misuse of authority, ensure that public services are accessible to everyone, and combat corruption, it is essential. A society's fundamental entitlements are its members' human rights. Because people are human, these things are relevant to them. To live a civilized life in any culture, they are an absolute need. AI will have both good and bad effects on long-standing norms and values. The detrimental effects that have presented difficulties will be the primary subject of this discussion (Singh, 2023c).

These individuals ought to spearhead democracy, a kind of governance that exists for the benefit of its citizens. It is necessary to provide individuals with knowledge so that they may make informed judgements when they are making decisions on their own. Even if AI streamlines and accelerates data transfer, it also comes with its fair share of difficulties. There is a danger of prejudice and uneven representation of viewpoints and perspectives if AI decides what gets presented and digested, what gets buried in the sea of internet information and what gets magnified virally. In addition, AI has the potential to undermine honest voting. Without public accountability in the agora, political campaigns or foreign actors can employ personalized adverts to transmit various messages to separate voter groups. It should be mentioned, nevertheless, that the exact impact of micro-targeted advertising is yet unclear. The use of AI in the creation and dissemination of deep fakes and disinformation is not limited to text, images, audio, or video. Due to their difficulty in detection, they pose a

threat to the public's trust in the media and democratic discourse by making it easier for manipulators and misinformers to gain power (Singh, 2023b).

21.6 CONCLUSION

By adhering to the rule of law, public institutions are expected to behave more responsibly toward people and society at large, as demonstrated by concepts like equality, proportionality, and justification. Although AI has the potential to make institutions more efficient, it also has the potential to undermine faith in democratic institutions, the rule of law, and the legitimacy of their procedures. Just as courts used to be the exclusive arbiters of what constitutes unlawful hate speech, commercial AI algorithms now predominate in deciding whether or not social media sites remove content. In reality, these AI systems vie with the judicial system and the law for power and generally speaking, AI has the potential to aid in the creation of legal systems that do not adhere to or even recognize the rule of law. Instead of consumer rights and procedural safeguards afforded by public courts, automated online dispute settlements offered by private corporations are controlled by their terms of service.

The effects of AI on anti-discrimination laws and the guarantee of equal treatment stand out among the many ways AI is influencing human rights. There have been several reports of AI becoming biased and thereby violating people's right to be free from discrimination. Additionally, these data-driven technologies discredit the social control mechanisms that regulate human conduct by proving the presence of biases. For instance, the AI that helped recruit new Amazon workers tended to prefer male candidates over female ones as it was programmed to mimic the behavior of the company's most successful male employees. Instead of just excluding female candidates, the AI system analyzed successful workers' phrases and phrasing patterns and rejected resumes that did not match. From a different perspective, AI systems are also finding more and more applications in the areas of workforce monitoring and tracking, autonomous task distribution, and recruiting and firing based on predictive analytics of future

performance and potential. Right to organize, right to a safe and healthy workplace, dignity on the job, and right to a fair wage are all rights that these AI applications potentially threaten. Despite the fact that technology has brought about unparalleled advancements and development in the majority of situations, the influence of AI on the rule of law, human rights, and democracy is not without its concerns. The majority of the time, these difficulties have the potential to endanger the level of living of the common individual.

REFERENCES

- Abebe, R., Barocas, S., Kleinberg, J., Levy, K., Raghavan, M., & Robinson, D. G. (2020, January). Roles for computing in social change. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 252–260).[↗](#)
- Alazab, M., Soman, K. P., Srinivasan, S., Venkatraman, S., & Pham, V. Q. (2023). *Deep learning for cyber security applications: A comprehensive survey*. Authorea Preprints.[↗](#)
- Alruwaili, F. F. (2023). Optimal hybrid deep learning enabled attack detection and classification in IoT environment. *Computers, Materials and Continua*, 75(1), 99–115.[↗](#)
- Attaran, M., & Deb, P. (2018). Machine learning: the new 'big thing' for competitive advantage. *International Journal of Knowledge Engineering and Data Mining*, 5(4), 277–305.[↗](#)
- Awajan, A. (2023). A novel deep learning-based intrusion detection system for IOT networks. *Computers*, 12(2), 34.[↗](#)
- Aziz, L. A. R., & Andriansyah, Y. (2023). The role artificial intelligence in modern banking: an exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. *Reviews of Contemporary Business Analytics*, 6(1), 110–132.[↗](#)
- Baker, D. J., & Robinson, P. H. (Eds.). (2020). *Artificial intelligence and the law: cybercrime and criminal liability*. Routledge.[↗](#)

- Bao, Y., Hilary, G., & Ke, B. (2022). Artificial intelligence and fraud detection. *Innovative Technology at the Interface of Finance and Operations, I*, 223–247. [↵](#)
- Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: limitations and opportunities*. MIT Press. [↵](#)
- Bhuvaneshwari, A. J., & Kaythry, P. (2023). A Review of deep learning strategies for enhancing cybersecurity in networks: deep learning strategies for enhancing cybersecurity. *Journal of Scientific & Industrial Research (JSIR)*, 82(12), 1316–1330. [↵](#)
- Brotcke, L. (2022). Time to assess bias in machine learning models for credit decisions. *Journal of Risk and Financial Management*, 15(4), 165. [↵](#)
- Cao, L., Yang, Q., & Yu, P. S. (2021). Data science and AI in FinTech: an overview. *International Journal of Data Science and Analytics*, 12, 81–99.
- Chen, Z., Van Khoa, L. D., Teoh, E. N., Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review. *Knowledge and Information Systems*, 57, 245–285. [↵](#)
- Cheng, X., Liu, S., Sun, X., Wang, Z., Zhou, H., Shao, Y., & Shen, H. (2021). Combating emerging financial risks in the big data era: a perspective review. *Fundamental Research*, 1(5), 595-606. [↵](#)
- Dash, B., Sharma, P., & Ali, A. (2022). Federated learning for privacy-preserving: a review of PII data analysis in Fintech. *International Journal of Software Engineering & Applications (IJSEA)*, 13(4). [↵](#)
- Dhoni, P., & Kumar, R. (2023). *Synergizing generative ai and cybersecurity: roles of generative ai entities, companies, agencies, and government in enhancing cybersecurity*. Authorea Preprints. [↵](#)
- Djenna, A., Barka, E., Benchikh, A., & Khadir, K. (2023). Unmasking cybercrime with artificial-intelligence-driven cybersecurity analytics. *Sensors*, 23(14), 6302. [↵](#)

- Elluri, L., Mandalapu, V., Vyas, P., & Roy, N. (2023). Recent advancements in machine learning for cybercrime prediction. *Journal of Computer Information Systems*, 1–15. [↗](#)
- Grimm, P. W., Grossman, M. R., & Cormack, G. V. (2021). Artificial intelligence as evidence. *Northwestern Journal of Technology and Intellectual Property*, 19, 9. [↗](#)
- Hagendorff, T. (2019). From privacy to anti-discrimination in times of machine learning. *Ethics and Information Technology*, 21(4), 331–343. [↗](#)
- Hoijsink, M., & Planqué-van Hardeveld, A. (2022). Machine learning and the platformization of the military: a study of google's machine learning platform TensorFlow. *International Political Sociology*, 16(2), olab036. [↗](#)
- Kordzadeh, N., & Ghasemaghahi, M. (2022). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388–409. [↗](#)
- Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: applications, algorithm selection, and challenges. *Business Horizons*, 63(2), 157–170. [↗](#)
- Li, L., Zhao, T., Xie, Y., & Feng, Y. (2020). Interpretable machine learning based on integration of nlp and psychology in peer-to-peer lending risk evaluation. In *Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, Proceedings, Part II* 9 (pp. 429–441). Springer International Publishing. [↗](#)
- Lui, A., & Lamb, G. W. (2018). Artificial intelligence and augmented intelligence collaboration: regaining trust and confidence in the financial sector. *Information & Communications Technology Law*, 27(3), 267–283. [↗](#)
- McKendrick, K. (2019). *Artificial intelligence prediction and counterterrorism*. London: The Royal Institute of International Affairs-Chatham House, 9. [↗](#)
- Mhlanga, D. (2021). Financial inclusion in emerging economies: the application of machine learning and artificial intelligence in credit risk assessment. *International Journal of Financial Studies*, 9(3), 39. [↗](#)

- Mijwil, M., & Aljanabi, M. (2023). Towards artificial intelligence-based cybersecurity: the practices and ChatGPT generated ways to combat cybercrime. *Iraqi Journal for Computer Science and Mathematics*, 4(1), 65–70. [↗](#)
- Mijwil, M., Salem, I. E., & Ismaeel, M. M. (2023). The significance of machine learning and deep learning techniques in cybersecurity: a comprehensive review. *Iraqi Journal for Computer Science and Mathematics*, 4(1), 87–101. [↗](#)
- Mühlhoff, R. (2021). Predictive privacy: towards an applied ethics of data analytics. *Ethics and Information Technology*, 23(4), 675–690. [↗](#)
- Narayana, Y. V., & Sreedevi, M. (2023, October). DDCATF: Deep Learning Approach for Detection of Cybercrime Activities Based on Temporal Features. In *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)* (pp. 462–469). IEEE. [↗](#)
- Nassar, A., & Kamal, M. (2021). Ethical dilemmas in AI-powered decision-making: a deep dive into big data-driven ethical considerations. *International Journal of Responsible Artificial Intelligence*, 11(8), 1–11. [↗](#)
- Pocher, N., Zichichi, M., Merizzi, F., Shafiq, M. Z., & Ferretti, S. (2023). Detecting anomalous cryptocurrency transactions: an AML/CFT application of machine learning-based forensics. *Electronic Markets*, 33(1), 37. [↗](#)
- Rai, R., & Sahu, C. K. (2020). Driven by data or derived through physics? A review of hybrid physics guided machine learning techniques with cyber-physical system (CPS) focus. *IEEE Access*, 8, 71050–71073. [↗](#)
- Saltz, J., Skirpan, M., Fiesler, C., Gorelick, M., Yeh, T., Heckman, R., & Beard, N. (2019). Integrating ethics within machine learning courses. *ACM Transactions on Computing Education (TOCE)*, 19(4), 1–26. [↗](#)
- Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., & Prabhakaran, V. (2021, March). Re-imagining algorithmic fairness in India and beyond. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 315–328). ACM.

- Schmitt, M. (2020). *Artificial intelligence in business analytics, capturing value with machine learning applications in financial services*. <https://stax.strath.ac.uk/concern/theses/8s45q876k>
- Sharma, A., Mansotra, V., & Singh, K. (2023). Detection of Mirai Botnet attacks on IoT devices using deep learning. *Journal of Scientific Research and Technology*, 174–187.
- Singh, B. (2023a). Blockchain technology in renovating healthcare: legal and future perspectives. In *Revolutionizing healthcare through artificial intelligence and Internet of Things applications* (pp. 177–186). IGI Global.
- Singh, B. (2023b). Federated learning for envision future trajectory smart transport system for climate preservation and smart green planet: insights into global governance and SDG-9 (industry, innovation and infrastructure). *National Journal of Environmental Law*, 6(2), 6–17.
- Singh, B. (2023c). Tele-health monitoring lensing deep neural learning structure: ambient patient wellness via wearable devices for real-time alerts and interventions. *Indian Journal of Health and Medical Law*, 6(2), 12–16.
- Singh, B., & Kaunert, C. (2024). Integration of cutting-edge technologies such as Internet of Things (IoT) and 5G in health monitoring systems: a comprehensive legal analysis and futuristic outcomes. *GLS Law Journal*, 6(1), 13–20.
- Siwakoti, Y. R., Bhurtel, M., Rawat, D. B., Oest, A., & Johnson, R. C. (2023). Advances in IOT security: vulnerabilities, enabled Criminal Services, attacks and countermeasures. *IEEE Internet of Things Journal*, 1, 47–49.
- Thapaliya, S., & Sharma, P. K. (2023). Optimized deep neuro fuzzy network for cyber forensic investigation in big data-based IoT infrastructures. *International Journal of Information Security and Privacy (IJISP)*, 17(1), 1–22.
- Tippins, N. T., Oswald, F. L., & McPhail, S. M. (2021). Scientific, legal, and ethical concerns about AI-based personnel selection tools: a call to action.

Personnel Assessment and Decisions, 7(2), 1. [↗](#)

- Tyagi, A. K., Aswathy, S. U., & Abraham, A. (2020). Integrating blockchain technology and artificial intelligence: synergies perspectives challenges and research directions. *Journal of Information Assurance and Security*, 15(5), 1554. [↗](#)
- Utami, S. H., Purnama, A. A., Hidayanto, A. N., Utami, S. H., Purnama, A. A., & Hidayanto, A. N. (2022). Fintech lending in Indonesia: a sentiment analysis, topic modelling, and social network analysis using twitter data. *International Journal of Applied Engineering & Technology*, 4(1), 50–56. [↗](#)
- Varma, P., Nijjer, S., Sood, K., Grima, S., & Rupeika-Apoga, R. (2022). Thematic analysis of financial technology (Fintech) influence on the banking industry. *Risks*, 10(10), 186. [↗](#)
- Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022b). *Explainable artificial intelligence applications in cyber security: state-of-the-art in research*. *IEEE Access*. <https://ieeexplore.ieee.org/document/9875264/>
- Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022a). Interpretable machine learning models for crime prediction. *Computers, Environment and Urban Systems*, 94, 101789.
- Zhang, Y., Xiong, F., Xie, Y., Fan, X., & Gu, H. (2020). The impact of artificial intelligence and blockchain on the accounting profession. *IEEE Access*, 8, 110461–110477. [↗](#)

22 Facial Recognition, Civil Rights, and AI

Legal and Ethical Considerations

*Nilanjan Chakraborty, Hamideza Rabiei
Dastjerdi, and Khder Alakkari*

DOI: [10.1201/9781003668657-22](https://doi.org/10.1201/9781003668657-22)

22.1 INTRODUCTION

There should be a moratorium on the use of Facial Recognition technology in the context of peaceful protests, until States meet certain conditions including transparency, oversight and Human Rights due diligence before deploying it.

Michelle Bachelet, UN Human Rights Chief

Globally and in India, artificial intelligence (AI) research is growing rapidly, raising ethical and governance concerns. AI refers to technologies that perform tasks requiring human cognition with machine learning and deep learning algorithms being the prominent ones. These algorithms learn patterns from data and have been widely used. NITI Aayog in its 2018 strategy identified three key sectors of AI intervention: agriculture, education, and health, which are known to directly improvise human lives ([Dogantekin, 2020](#)).

Beyond these areas though, it has also been used in arenas of the state that remain largely hidden from public view. Use of Automated Facial Recognition Technologies (AFRT) by law enforcement agencies (LEAs) in India is one such widely publicized, technology-driven, and citizen-

impacting use case of AI. AFRT refers to algorithms capable of uniquely identifying a person by comparing their facial features with the pre-existing dataset of facial images. It can also map facial attributes, including emotions. The AI hype also demands an inquiry into responsible and ethical use of AFRT for law enforcement and surveillance. Globally, there are concerns being raised about design limitations of AFRT and chilling effects on constitutionally guaranteed fundamental rights posed by deploying it. In India, many states and national agencies are adopting AFRT without any public and transparent discourse. The paper seeks to highlight some under-discussed dimensions of AFRT in Indian law enforcement—constitutional challenges; legal issues related to searches and self-incrimination; ethical risks associated with trust; and justifiability under the Indian Evidence Act—coupled with the conclusion that there is a need for more nuanced discussions if India’s vision of “responsible AI for all” is to be realized. The paper argues that live facial recognition cannot be operationally necessary to deploy National Automated Facial Recognition System (NAFRS) (contrary to the government’s position) but that a moratorium on live facial recognition systems such as NAFRS is nonetheless required if we are to protect our democratic values and constitutional rights.

Increasingly, such applications have become common in commercial uses as well—such as unlocking cell phones or check-ins at airports.

However, use of AFRT for surveillance and law enforcement comes with major risks. These are not concerns limited to India but a part of a wider global discourse on the ethical aspects of AFRT in policing ([Gupta, 2020](#)). There have been several efforts internationally to map and categorize the risks that come with AFRT—leading us to ask: Should this technology be used in law enforcement at all?

In India, the rollout of AFRT is happening in the absence of any legal or regulatory mechanism. In this context, it becomes important to understand how this technology will be wielded by LEAs and poses some pertinent questions: Do we ban AFRT? Do we put a moratorium on its use? Or do we continue to develop and deploy it with good regulations and oversight?

Answering these questions will determine if the implementation of AFRT meets ethical standards and safeguards citizens' rights ([Krier, 2020](#)).

22.2 IMPACT ON FUNDAMENTAL RIGHTS

The global discussion over the ethics of AI recognizes in particular that an effective protection of fundamental human rights and societal values is key, as seen in a number of international endeavors, for example, the CEPEJ Charter or the European Union White Paper on AI and proposed legislative framework (European Commission, 2020).

In the Indian context, NITI Aayog's discussion document highlights that we must follow constitutional morality. That means not just the letter of the constitution but also the spirit of inclusiveness, respect for diversity, and other such principles that our constitution is based upon. It says that for any AI in India, including maintenance of law and order, we must ensure that constitutional rights and values are not violated in its development and use.

This ethical turn is part of a larger movement to recognize that it is important to ensure that we use technology to advance the human good and not forget about things like protecting human rights or other values. But as AI systems increasingly shape decisions—from who is targeted for extra policing, to what news you see, to how your boss can monitor your every move—we must be fundamentally sure that those guiding principles are in fact democratic ones ([Bostrom, 2014](#)).

There has been considerable anxiety among legal scholars and human rights activists regarding increasing deployment of AI-facial recognition systems by LEAs in India. These anxieties are on account of apprehensions of violation of fundamental rights guaranteed under the Constitution of India, more particularly, right to privacy, which was read to be part of Article 21 of the Indian Constitution by the Supreme Court in the historic *Puttaswamy* decision.

A fundamental problem with the AI systems that underlie FRT is that they are designed and developed using machine learning or deep learning techniques, which require massive amounts of training data. This inherent

reliance on big data comes into direct conflict with the concept of informational privacy, thus throwing up questions as to whether the compliance standards set by the Indian Supreme Court are met at all. Additionally, India does not yet have data protection legislation in place, which could allow us to procure and use datasets for developing FRT algorithms without any risks going forward ([Husain, 2017](#)).

The usage of facial recognition technology by Indian police demands close scrutiny due to ethical and legal concerns. We need to ensure that any such technology is in line with the constitutional principles and protects the fundamental rights of an individual.

Mass surveillance by the government through the use of these technologies is a major privacy concern and this can have implications on free speech, what is referred to as a chilling effect. There have been controversial cases globally where governments have turned to facial recognition to unmask protestors, raising questions of legality and proportionality and being labeled as deeply invasive by rights advocacy groups. In India, for instance, in 2020 reports emerged that Delhi Police had reportedly considered using facial recognition to help identify alleged perpetrators of violence during protests against the Citizenship Amendment Act, putting the protestor under a cloud of doubt if they are being targeted.

A fundamental problem with the AI systems that form the basis of FRT is that they are designed and built using either machine learning or deep learning methods, which necessitates access to vast amounts of training data. This dependence on big data is inherently in conflict with the notion of informational privacy and raises questions as to whether the compliance standards imposed by the Indian Supreme Court are realistic at all. Further, there is no assurance that we will be able to access such datasets for developing FRT algorithms without incurring any risks in the absence of a data protection law in India.

The usage of facial recognition technology by Indian police needs a closer look, keeping in mind the ethical and legal issues. We have to make sure that any such technology is in consonance with our constitutional principles and guards an individual's fundamental rights.

Mass surveillance by the government through the use of these technologies is a major privacy concern, and this can have implications on free speech, what is referred to as the chilling effect. There have been controversial cases all over the world where governments have resorted to facial recognition to unmask protestors, raising questions on legality and proportionality and being labeled by right advocacy groups as deeply invasive. In India, for instance, in 2020 reports emerged that Delhi Police had reportedly considered using facial recognition to help identify alleged perpetrators of violence during protests against the Citizenship Amendment Act, putting the protestors under a cloud of doubt if they are being targeted (CEPEJ, 2018).

One of the most important legal questions to answer is that of due process. Indian courts have held that the due process guarantees under Article 21 subsume within it both substantive and procedural due process rights; if facial recognition technology were to be used in law enforcement, particularly for charging and detaining people, such usage would have to clear these benchmarks.

This section unpacks these complexities and argues that automated facial recognition is likely to violate privacy, free speech, and due process rights in the Indian context.

22.2.1 IMPACT ON CIVIL LIBERTIES

The state of democracy in India appears increasingly fragile, especially in light of the recent Pegasus surveillance controversy. Adding to these concerns is the introduction of NAFRS, an advanced facial recognition system that can purportedly identify individuals even when they attempt to conceal their appearance with masks or cosmetics ([Roy, 2021](#)). This technology actually poses the single biggest threat to the idea of anonymity, which is so important in a successful democratic society.

Deployment of NAFRS can have a chilling impact on several kinds of civic engagement. Independent journalists, political protestors, and activists will be reluctant to do their work when they fear being singled out or

identified for subsequent retaliation. These are not imaginary fears; stories abound of people who attended protests against the Citizenship Amendment Act who were subsequently accosted, often by the police.

The use of such a pervasive surveillance tool, especially in the absence of strong legal safeguards, raises questions about whether a democracy can deploy surveillance mechanisms on par with an authoritarian regime ([Chandak, 2020](#)). The threats this kind of surveillance poses to democratic freedoms are likely to far outweigh any benefits, and therefore, the proliferation of such a system is something that all those who care for the future of Indian democracy must be deeply concerned about.

LEAs have been known to be biased, especially toward the Islamic and other marginalized communities. Adding human bias in facial recognition technology will only increase the chances of discriminating against overrepresented race by criminal justice systems. Such communities in India include Dalits, minority populations and Adivasi groups, who are already exposed to higher incarceration rates as compared to their presence in the population.

History makes it clear that, without strong protections and real consequences for abuse, no faith or political leaning is safe from profiling. Indeed, there are fears law enforcement officers could use the system to snoop on potential lovers.

These concerns make it incumbent on us to develop mechanisms and safeguards to avoid further disadvantaging already marginalized communities and to prevent expanded abuse of police power ([Bhatia, 2017](#)).

22.2.2 PRIVACY RISKS POSED BY AFRTS

The adoption of facial recognition technology in law enforcement has yielded proliferated debate over the violations of personal privacy. Concerns mainly concentrate on the data operation, including acquisition, preservation, and application.

In India, it is unclear what data types will actually be used in the making of locally developed law enforcement facial recognition systems. There is certainly no shortage of potential candidate sources, including face scans and surveillance images, government-owned photo databases such as Aadhaar, and other identity-related information archives. The National Crime Records Bureau initially drew up an ambitious blueprint for integrating data from various government-owned databases, but this was taken to court, and a more modest version was adopted. Critics harbor concerns nonetheless.

And the issue is further complicated on a global scale, as illustrated by recent revelations around the activities of Clearview AI. The US-based firm has created an enormous bank of face images by scraping pictures from social media sites and websites, sparking concerns about the intrusiveness of data acquisition. The firm's contracts with police and other agencies, both governmental and private, all over the world have ramped up discussions on facial recognition in policing [Mazur, 2019](#)).

This has forced policymakers to confront a series of complex questions. How useful are these predictive tools as a matter of national security? Is the benefit great enough to justify such blanket surveillance? And are not the implications of using an unaccountable black box to profile potentially hundreds of millions of people extraordinarily creepy?

Privacy has become a core concern for technology policymakers in India, across legal and other domains. This new found interest has been attributed primarily to the Supreme Court's landmark *Puttaswamy* judgment that transformed prevailing jurisprudence by recognizing privacy as a fundamental right under Part III of the Indian Constitution. One of the aspects of this judgment is its recognition of personal autonomy with respect to one's information or, more specifically, an individual's ability to determine what information she will allow dissemination of, who can collect that information, and how it can be used. But when it comes to AFRTs, the design and operation seem at odds with this principle of control that a person is presumed to exercise on her personal data ([Mazoomdar, 2019](#)).

The problem goes beyond individual consent. There is growing concern about the use of AFRTs being extended far beyond what they were designed for, particularly in policing, and without any kind of oversight or regulatory framework. We saw this happen earlier this year when the Delhi Police announced their plans to use facial recognition systems—originally developed to track missing children—to identify protestors during the anti-CAA protests. In the absence of any procedural safeguards, it seems like function creep is simply left up to the police, who might be too willing to take on these technologies without thinking through the implications on due process.

The proposed Personal Data Protection (PDP) law, introduced before the parliament in 2019, does not adequately cater to the risks of privacy. There are multiple and broad exceptions present in the bill, which might pose a risk to holding an effective control on FRTs by the police and thus pose a risk to informational privacy. However, privacy as a constitutional right still stands invoked under *Puttaswamy (K.S. Puttaswamy v. Union of India)*. The court mandated that any use of FRTs by law enforcement should meet its criteria on proportionality. There lacks information about AFRT deployment being used currently to demonstrate the adherence of this technology deployment with legal standards; as this question is opened for constitutional adjudication, it shows more reason that there is a need for adequate checks and balances.

22.2.3 STATE SURVEILLANCE AND IMPEDING FREE SPEECH AND EXPRESSION

The growing use of sophisticated facial recognition technology by police forces has raised fears of excessive surveillance of dissent, minority communities, and political opponents. A worldwide survey into AI surveillance, meanwhile, recently found such systems are deployed in more than four out of five countries surveyed for the purposes of policing ([Parsheera, 2019](#)).

Some recent incidents highlighting this have been the Hong Kong protests last year, where facial recognition was used to unmask the protestors similarly as in the US during BLM movement. In India, too, facial recognition has been mired with controversy since Delhi Police announced its plans to use it on anti-CAA protestors, and that had raised a lot of red flags ([Mazoomdar, 2019](#)).

The absence of a holistic legal framework governing the use of these technologies is a threat to the constitutional right to free expression, assembly, and dissent. In an era when there is already some anxiety around the health of India's democratic institutions, it is important to note that reasonable restrictions on free speech can be imposed to maintain public order only through legislation that passes muster with our Constitution. Unless we have statutorily binding checks in place, investments in deploying facial recognition technologies across the country could end up diluting a lot more than our fundamental rights.

22.2.4 DUE PROCESS CONCERNS REGARDING AFRTS IN LAW ENFORCEMENT

The Indian Constitution's Article 21 enshrines the concept of "procedure established by law," safeguarding the right to life and personal liberty. Over time, constitutional interpretation has expanded this to include both procedural and substantive due process. This means that legal procedures must not only be sanctioned by law (procedural due process) but also adhere to principles of fairness, justice, and reasonableness (substantive due process) (*Maneka Gandhi v. UOI*, 1978). No legal process should be regarded as "inherently arbitrary." This is particularly important for newer practices of predictive policing. Earlier evaluations of Delhi Police's predictive policing projects have revealed the presence of "inherent arbitrariness" in implementation mechanisms of these projects ([Vidushi Marda, 2020](#)).

Arbitrariness is inherent when governments and LEAs use technology to identify criminals. Today, however, the absence of any legislative,

regulatory, or governance framework for AFRT in Indian law enforcement makes this arbitrariness stark (Internet Freedom Foundation, 2019). There are no statutory limits or safeguards that constrain the authority of the police to formulate, adopt, deploy, and scale this technology. The opacity that surrounds these activities, in particular, increases the vulnerability of due process rights to be sacrificed at the altar of expediency for investigation. The deployment of AFRT in law enforcement brings forth serious challenges to due process, both procedurally and substantively. Procedurally, the absence of explicit legislation enabling use of AFRT, defining its scope of application, and framing oversight mechanisms is a major gap (Jauhar, 2021). This legislative void and unfettered usage also makes people susceptible to being potentially prosecuted with unregulated/unauthorized evidence in India. This technology can be misused by the authorities who have motivation to further the use of AFRT, without any checks and balances.

In the light of the above, it is clear that the utilization of AFRT by Indian police and other LEAs gives rise to grave legal and constitutional contraventions. In view of this precarious situation, it becomes important that necessary reforms and ad hoc arrangements are scrutinized objectively and transparently so as to conform to constitutional and legal mandates. These concerns will be examined more elaborately in the concluding section of this article to devise appropriate redressals.

22.3 ETHICAL CHALLENGES IN AFRTS' ADOPTION AND USAGE

Apart from the constitutional and legal concerns, ethical aspects of AFRTs are particularly accentuated in scenarios of individual identification and surveillance by law enforcement (Jauhar, 2021). These include issues of lack of transparency and accountability, technical fallibility in AFRT's design and implementation, and potentially exacerbating societal biases. Measuring these impacts forms part of wider debates on AI ethics, against

which risks of various AI applications, including AFRTs, are being assessed.

The transparency issue has two aspects: (1) whether AFRTs use transparent and explainable algorithms, that is, technical transparency; (2) the administrative transparency and accountability—on how Indian LEAs are deploying it in an opaque manner often without any or little public knowledge. This leads to the accountability and responsibility question if harm is caused by using this technology and also undermines the democratic institutions. There are chances that the public may lose trust in the institutions like the police.

Moreover, there are technical and data-related limitations as well that have been flagged by multiple studies—national as well as international ([Jauhar, 2021](#)). Existing facial recognition systems, for example, fail to provide accurate results more often than not, particularly for people with darker skin tones. It can, therefore, lead to false arrests and imprisonments. In the Indian context too, this technology can be expected to further reinforce our existing biases in policing toward some specific communities that are already marginalized ([Joy Buolamwini, 2018](#)).

22.3.1 SHORTCOMINGS IN DESIGN AND ADMINISTRATIVE TRANSPARENCY

One of the fundamental problems with algorithmic decision-making and predictive tools is that the machine learning algorithms underpinning them are black boxes. They are programmed to learn on their own by identifying patterns in large pools of data, but in the process, not even their designers typically know how a machine arrived at a particular answer. This inherent opacity tends to increase with the accuracy of machine learning algorithms, and intellectual property concerns exacerbate this. IBM, for example, would not disclose its facial recognition software data because doing so would have revealed proprietary information. This lack of transparency has significant implications for understanding and scrutinizing the decision-

making processes that rely on machine learning ([Victoria Burton-Harris, 2020](#)).

AI ethicists and scholars often propose “human-in-the-loop” solutions to address the problem of design opacity, that is, requiring human intervention in making decisions about impactful algorithmic applications or demanding certain procedural requirements during the development and deployment of algorithms, including algorithmic impact assessments for discovering errors and biases. In the case of law enforcement, where facial recognition technology is used as a decision-making tool in issues directly relating to civil liberties, more transparency and accountability should be demanded. But implementing human-in-the-loop on AFRTs is not as simple under these conditions. It might generate automation biases; it might introduce further potential problems of machine influence ([Ameya Bokil, 2021](#)).

Beyond the issues of design transparency, the integration of AFRT across various state police forces in India reveals significant deficiencies in administrative transparency and accountability. The lack of publicly available information on AFRT use was concerning; a 2019–20 legal notice had been issued to the National Crime Records Bureau (NCRB) requesting more detailed information disclosures than those contained on their website. We undertook an empirical investigation of AFRT use by police organizations in Hyderabad, Delhi, Mumbai, and the state of Punjab. Due to the scant publicly available information, we selected these four based on media reports of their AFRT deployments. Despite filing Right to Information (RTI) requests between November 2020 and March 2021, responses were largely inadequate, with only minimal information provided by three out of the four states. Queries about official documents such as notifications, memos, or meeting minutes related to AFRT use were met with claims of non-existence ([Ameya Bokil, 2021](#)).

To address these transparency and accountability gaps, legislative mandates are essential. In contrast, several EU countries and US states have enacted laws to regulate, ban, or impose moratoriums on AFRT use in law enforcement. India, however, does not have a legal framework for AFRT implementation in place. For instance, the NCRB uses a 2009 cabinet note

as the primary legal authority for implementing a nationwide facial recognition system and lacks provisions for ensuring accountability. It bears emphasizing that this analysis is necessarily speculative given the recency and opacity of the clandestine acquisition, as well as its birds of a feather flock together journalistic coverage ([Chandak, 2020](#)).

The absence of transparency poses serious legal and accountability questions. Legally, it is not clear whether the use of AFRT is even permitted under our current legal void. Accountability provides further complication, as it is unclear how to assign liability to particular state or private sector actors. Such lack of transparency and accountability flies in the face of the principles behind ethical and responsible AI use as provided by NITI Aayog's guidelines.

22.3.2 TECHNICAL INACCURACIES AFFLICTING AFRTS

AFRTs are being used for verification and identification. Verification has higher accuracy, as it matches or verifies in a constrained environment with a predefined database, while in identification, the algorithm has to find out from a universe of data, so the chances of error are always there (Facial recognition's 'dirty little secret': Millions of online photos scraped without consent, 2019).

It is important to note that the performance of an AI system, including ML- and DL-based systems, is intrinsically tied to how well it has been trained. For instance, AFRTs have been reportedly performing poorly on analyzing images of individuals with darker skin tones or non-Caucasian features. A notable study by MIT researchers in 2018 spotlighted three state-of-the-art AFRT algorithms, which showed significantly higher error rates for racial minorities and females, especially those with darker complexions. The largest source of this bias was traced back to the approximate 80% proportion of white male subjects in the training data used for these algorithms.

In India, the problem gets further complicated with the vast diversity in terms of population. Alien AFRT algorithms are bound to be inaccurate,

and even indigenous ones might not work if they do not have the adequate representation of the Indian population. In fact, due to this inherent inaccuracy, it becomes very important that the LEAs share information on AFRT's achievable algorithmic accuracy, the external conditions affecting their performance, and specific accuracy threshold or confidence basis used in decision-making while using AFRTs. This is again necessary to ensure trust building and accountability against technology use itself ([Guszcza, 2020](#)).

Beyond data biases, design limitations affect the validity of AFRT predictions. Often, such algorithms are trained on images in a laboratory setting that have been taken using professional cameras and lens at very close proximity to the face. Thereafter, when deployed in real-life settings, a raft of other variables, such as low image resolution, inconsistent lighting across light changes and shadows, and camera angles from above or below or from the side, compounded by out-of-focus identification photos due to their age or quality, further deteriorate the quality of the image feed used for recognition or verification. Different algorithms also come with different levels of accuracy. Users of AFRT can also determine an acceptable accuracy level that they are willing to tolerate for false positives: a higher accuracy will delete some correct results, while a lower threshold may introduce more incorrect matches ([Crumpler, 2020](#)). While the US National Institute of Standards and Technology reported drastic improvements in overall facial technology software accuracy since 2018, leading technology companies such as Amazon, Facebook, Apple, and Google did not submit entries for these competitions.

Inaccuracies in AFRT do not obviate surveillance concerns. Instead, adoption of faulty AFRT by law enforcement could result in innocent people being mistakenly linked to crimes as a consequence of false positives. So much so that many firms have voluntarily stopped or been forced to stop developing or deploying AFRT due to these consequences. In India, with a considerable legislative void, the potential for aggravating discrimination and wrongful detentions is large ([Crumpler, 2020](#)). This will be further explored and empirically demonstrated in later papers in the

series. The implications of mistaken predictions can be severe, including false arrest or prosecution with little recourse for the person wrongly detained.

22.3.3 DISCRIMINATORY TREATMENT AGAINST VULNERABLE POPULATIONS

AFRT is likely to produce biased and discriminatory results for particular population groups due to in-built design bias and broader societal biases that emerge when machine learning technology is used. Like other machine learning systems, AFRT is trained on a dataset that provides cues for the algorithm to discriminate between faces. If this training dataset contains discrimination—for example, because it under-represents some group(s)—then the algorithm is bound to have a harder time identifying or categorizing these under-represented group(s), as well as unfairly linking them through matching with other data sources. The reality of poorer performance of some AFRT systems in identifying people of color, younger people, and women has been well documented. Using an impressive number of tests, Buolamwini and Gebru demonstrate such higher rates of misclassification in relation to darker-skinned females than lighter-skinned males within private sector AFRT systems ([Joy Buolamwini, 2018](#)).

In the Indian context, there is a fear that AFRT may amplify existing social biases related not only to skin color and gender but also to caste and religion. Ethnography by Marda and Narayan on predictive policing in Delhi (excluding AFRT) has shown that three forms of bias are present in the technologies used by the police: historical bias against impoverished and vulnerable communities, representation bias due to a lack of data from the margins, and measurement bias that results from the difficulty of documenting transient populations. Furthermore, there exists a paradox of inclusion: while having a data footprint can indicate privilege, inclusion in an AFRT database heightens the risk of profiling and surveillance ([Prem Sylvester, 2019](#)). Therefore, an effective policy response to bias in AFRT

should focus not just on improving dataset inclusivity but on critically reassessing the foundational role of AFRT in policing practices.

Incorporating algorithmic predictions can intensify inherent human biases by providing a tool that might be used to validate pre-existing prejudices. Although AFRT itself may not exhibit the same biases due to variations in data and applications, it is valuable to examine how historical, representational, and measurement biases can emerge within AFRT systems. Given that policing can inherently involve discriminatory practices, the application of AFRT could potentially amplify such biases. This concern will be explored in greater detail, with empirical support, in the forthcoming second paper of this series.

22.4 EVIDENTIARY VALUE OF AFRTS IN INDIAN LAW

The last issue that we present in this paper is the use of AFRTs in criminal prosecutions. The primary function of AFRTs is to aid surveillance, that is, to locate or identify possible suspects. But can the results of such technologies be used as contentious evidence in a court? Surprisingly, we have found no literature discussing this question with respect to the Indian legal context, and hence, as a first work, this paper needs to identify the basic challenges associated with using AFRT as an evidentiary tool within the existing legal system.

In 2000, the Indian Evidence Act, 1872, was amended to adapt to the changes occurring in information technology. Sections 65A and 65B were inserted in the Act dealing with the admissibility of electronic records, which have assumed a pivotal role in criminal law (*Tomaso Bruno v. State of U.P.*, 2015). These provisions allow courts to use electronic databases, recordings or emails for purposes of supporting, as well as in certain cases (*Jisal Rasak vs The State of Kerala*, 2019), even proving a particular fact, and in some cases evidence has even been regarded as primary evidence establishing guilt of an accused (*Md.Ajmal Md.Amir Kasab vs State of Maharashtra*, 2012).

At present, the meaning of “electronic records” under Section 65B (1) pertains to electronic documents that facilitate use, recording, or transmission as a replacement for physical records. This understanding classifies AFRT in the category of automated systems, and it is not designed to cover the implementation that requires algorithmic predictions and decision-making, where the data is not static but dynamic and adaptive. In any event, even if AFRT is included within the meaning of electronic record as defined under Sections 65A/65B, these sections only deal with admissibility and not reliability/accuracy or other possible biases in the technology itself. Courts have raised concerns about electronic evidence being vulnerable to manipulation, and therefore it has been held that “electronic records cannot be good sole or conclusive evidence of any fact” ([Tukaram et al., 2008](#)).

The above discussion brings out well-known design and data caveats of AFRT algorithms that question the dependability of AFRT evidence and warrant legislative control on its uses. Amendments in the Evidence Act and the Criminal Procedure Code, 1973, are required to specifically address and regulate AFRT deployment along with other such predictive technologies.

22.5 CONCLUSION

Discussion around AFRT in law enforcement is a complicated subject and one that is largely unexplored in the Indian context. The issue with the growing use of AFRT by Indian law enforcement is that the risks it poses are not the only problem—the lack of literacy or awareness among citizens to understand these risks and to have this discussion is also a problem. Additionally, while AFRT is commonly projected as an advanced and efficient instrument for policing, there is a dearth of robust studies supporting these claims. Most of the rhetoric around AFRT is casual, which can lead to undermining risks.

The principle of responsible and ethical AI is inherently linked to the creation of a trust infrastructure, which necessitates transparency and

accountability, two aspects that are sorely missing in the hasty implementation of AFRT in Indian policing and surveillance. There is a democratic mandate for wider public and stakeholder consultation in such far-reaching policy measures. Moreover, effective regulatory oversight and safeguards are needed to prevent arbitrary state surveillance and to ensure that these technologies are not left solely at the mercy of the executive's discretion.

This paper discusses the ethical and legal challenges of using AFRT in law enforcement so that a clear and perspicuous document can provide the basis for an informed public debate. While there are undoubtedly legislative and regulatory lacunae to be filled, the public ought to be aware of what is at stake with these technologies. Decisions to ban, suspend, or regulate AFRT should be taken after extensive public deliberation and in light of how such steps fit within broader efforts to modernize and improve policing. Policymakers and all other actors interested in these technologies need to engage in unbiased assessment of AFRT's performance, as well as its compatibility with due process guarantees. Considering the contentious nature of this technology and its potentially profound implications for AI used by agencies delivering services to the public, deployment of AFRT ought to conform to ethical principles akin to those articulated for trustworthy AI for use by Europe's public sector.

REFERENCES

- Ameya Bokil, A. K. (2021, May). Settled habits, new tricks: casteist policing meets big tech in India. *TNI Longread*, <https://longreads.tni.org/stateofpower/settled-habits-new-tricks-casteist-policing-meets-big-tech-in-india>.[↵]
- Bhatia, G. (2017). The Supreme Court's Right to Privacy Judgment. *Economic & Political Weekly LII* No. 44, 22–25.[↵]
- Bostrom, N. (2014). *Superintelligence: paths, dangers, strategies*. Oxford University Press, 26–61.[↵]

- CEPEJ. (2018). *European ethical charter on the use of AI in judicial systems and their environment*. European Commission for Efficiency of Justice. Council of Europe.
- Chandak, S. (2020). Artificial intelligence and policing: a human rights perspective. *NLUJ Law Review*, 7(1), 44–69. [↗](#)
- Crumpler, W. (2020). *How accurate are facial recognition systems—and why does it matter?* Washington: Center for Strategic and International Studies. [↗](#)
- Dogantekin, V. (2020). *UN raises concern about facial recognition technology*. Anadolu Ajansı. [↗](#)
- European Commission, E. (2020). *White Paper: on artificial intelligence—a European approach to excellence and trust*.
- Facial recognition's 'dirty little secret: millions of online photos scraped without consent. (2019, March 17). *NBC News*.
- Gupta, A. (2020). *The state of AI ethics*. Montreal AI Ethics Institute. [↗](#)
- Guszcza, J. (2020). Human values in the loop: design principles for ethical AI. *Deloitte Insights*. [↗](#)
- Husain, A. (2017). The sentient machine: coming age of AI. *Scribner*, 20–34. [↗](#)
- Internet Freedom Foundation, I. F. (2019, November 5). *Rejoinder to Reply*. Retrieved July 10, 2024, from Internet Freedom Foundation: <https://drive.google.com/file/d/1Cb9BtySI7Z7IM7tvAJGNRqQ96ypAGtbN/view>
- Jauhar, A. (2021). Facing up to the risks of automated facial recognition technologies in Indian law enforcement. *Indian Journal of Law & Technology*, 16(1), 1–15. [↗](#)
- Jisal Rasak vs The State of Kerala*, Crl.MC.No.4148 of 2019 (Kerala High Court 2019).
- Joy Buolamwini, T. G. (2018). Gender shades: intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 1–15). New York: Proceedings of Machine Learning Research. [↗](#)

- K.S. Puttaswamy v. Union of India*, W.P. 494 of 2012.
- Krier, S. (2020). *Facing affect recognition*. *Asia Society*.[↗](#)
- Maneka Gandhi v. UOI*, SCC (1) 248 (Supreme Court 1978).
- Mazoomdar, J. (2019, December 28). Delhi police film protests, run its images through face recognition software to screen crowd. *The Indian Express*.[↗](#)
- Mazur, P. (2019, July 26). In Hong Kong protests, faces become weapons. *The New York Times*.[↗](#)
- Md.Ajmal Md.Amir Kasab vs State of Maharashtra*, 9 SCC 1 (Supreme Court 2012).
- Parsheera, S. (2019). *Adoption and regulation of facial recognition technologies in India: why and why not?* Data Governance Network.[↗](#)
- Prem Sylvester, K. S. (2019). *India is falling down the facial recognition rabbit hole*. Bengaluru: The Centre for Internet & Society.[↗](#)
- Roy, A. (2021). *Responsible AI: Approach document for India part I—principles for responsible AI*. New Delhi: NITI Aayog.[↗](#)
- Tomaso Bruno v. State of U.P.*, 7 SCC 178 (Supreme Court 2015).
- Tukaram S., Dighole V., Maikrao Shivaji, K. (2008). Civil Appeal No. 2928 of 2008 (Maharashtra High Court 2008).[↗](#)
- Victoria Burton-Harris, P. M. (2020, June 24). *Wrongfully arrested because face recognition can't tell black people apart*. American Civil Liberties Union.[↗](#)
- Vidushi Marda, S. N. (2020). Data in New Delhi's predictive policing system. In *FAT* '20: proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 317–324). New York: Association for Computing Machinery.[↗](#)

23 Impact Analysis of Emotional Intelligence by AI on Sales Team Performance

Case Study of ICICI Group Companies in Nagpur

Gaurav Kulat, Neeta Shah, Kanchan Naidu, and Gayathri Band

DOI: [10.1201/9781003668657-23](https://doi.org/10.1201/9781003668657-23)

23.1 INTRODUCTION

Sales performance is a key driver of long-term organizational success in the fast-paced and competitive terrain of financial services. There is a growing need for companies to determine the factors that drive sales performance as they look to get ahead of their competition. Of these, the one that energizes more and more time nowadays has been identified as emotional intelligence (EI), and is a definitive predictor of success, both individually as well as on a team basis. EI is the ability to recognize feelings in oneself and others, as well as manage emotions effectively both individually and within groups. This is especially true when it comes to sales roles, which need to possess high levels of people and relationship management skills and emotional competency.

The latest developments in artificial intelligence (AI) help us take those questions a step further. Machine learning and natural language processing (NLP) enable AI systems to examine large, complicated data sets for insights that may not be apparent in the research done through traditional methodologies. Using AI enables a deeper level of understanding in how EI attributes shape sales outcomes and team dynamics at organizations.

The study was conducted with ICICI Group companies in Nagpur to discover how different dimensions of EI influence sales team performance. The purpose of this study is to explore AI-based analytic approaches, to identify how EI is associated with sales performance. The results are expected to provide important implications for power grooming strategies to improve performance outcomes, leading to the desired organizational effectiveness.

True, every one of us likely understands that EI is important for our success in an ever-changing selling world and it affects our ability to build rapport with customers. It has to do with a person's ability to recognize, pay attention, and control their own emotions, as well as the emotions of other people. This is key in sales because people always get bogged down on the process, or learning requirements and having this graph allows you to gain quick rapport with a client, and understand their goals (everyone loves talking about themselves)—and also it makes selling something complex so much easier to the buying group. Sales professionals with EI offer a high level of customer experience, which results in increased CSAT/customer retention—and eventually generates more sales.

Five components of EI are self-awareness, self-regulation, motivation, empathy, and social skills. Salespeople with a higher degree of self-awareness can identify their triggers for anger, jealousy, or embarrassment or understand how they might be feeling inhibited, holding them back. They are good at self-regulation, that is, controlling their emotions in difficult situations. It motivates them to move forward towards goals in the face of obstacles. This makes them more likely to listen well and connect, as empathy allows an individual to pick up on the emotional state of another

person. Additionally, strong social skills improve the quality of communication and relationships.

One study actually showed that EI is a strong indicator of sales success. Some studies, for example, have found that 95% of purchase decisions are made subconsciously—driven by emotion, not by logic. Salespeople who read and respond correctly to emotional cues are better able to close deals. This emotional trigger is particularly important when dealing with environments requiring trust for customer loyalty over the long term.

At the Japanese kindergarten, AI has been found to boost EI. Over the past few years, we have witnessed machines becoming better than human beings at playing complex games like chess and Go. The next step in AI evolution seems to be making it work proficiently for EI in sales. AI is an extraordinarily powerful tool, capable of processing its way through a humongous amount of data and deciphering human behavior and emotion to the extent that now it is capable to suggest sales professionals ways to boost their client interactions by leaps and bounds.

Real-Time Emotional Insights: AI can analyze voice tone, facial expressions, and other non-verbal cues during sales interactions. This means that salespeople can see if a customer is feeling angry, tired, or frustrated and respond accordingly. When an AI solution identifies frustration in the voice of a customer during a call, for instance, the salesperson can get on that call with more empathy, or instead offer additional reassurance and address potential fears.

Personal Training Programs: AI-driven platforms can provide personally curated training programs designed to build and reinforce different EI skills within sales teams. Through tracking emotional cues during interactions, and performance metrics tied to each interaction that is a part of the process, these systems can be proactive in identifying where challenges might come up so as to provide pre-emptive measures within customized exercises or training modules.

Improved Customer Experience: With emotional AI tools, you can make sales more emotionally intelligent by understanding correlations

between sentiment and the various stages of the sales funnel. By understanding the “pain of purchase” and well as how they feel about your offerings at various stages in the buying journey, your sales team increases their chances of applying an effective emotional strategy to meet those needs.

Better Relationship Management: With AI, sales teams can manage their relationship more efficiently by following emotional trends in the course of time. Models based on data from hundreds and even thousands of past customer interactions are able to find patterns that define when a person’s approaches the exit or can be unhappy.

Be Proactive: An ounce of prevention is worth a pound of cure. Hence, taking early action can help sales reps intercept and keep customers.

Amplification of Empathy: Where AI cannot provide the same emotions as a human, it is certainly equipped to amplify the MTS (machine touched sentence) of salespeople by providing context on how a particular client might be feeling during their conversation or correspondence. This support allows salespeople to engage with clients in a way that feels more authentic—building trust and rapport at critical touchpoints along the way.

The following sections will outline the theoretical background of EI, describe the methodologies employed in this study, present the results, and discuss their implications for sales management and training within the financial sector.

23.2 EMOTIONS AND EMOTION MAPPING WITH YOUR CUSTOMERS

The need to understand your customer and selling the emotion is a foundation of successful sales performance. Emotions have a strong impact on buyers and drive decisions, often more than the mind. Professionals who sell services and products can be much more successful if they know how to

recognize a customer's emotion and react appropriately. The emotional landscape customers typically go through a vast array of emotions when buying, from that initial excitement through to the anticipation all the way to anxiety and then frustration. For example, a customer might be enthusiastic about a new product while also being nervous about its price or capability to comply with their requirements. By knowing these emotional states, salespeople can use that information to cater to them. Sellers with high EI have the ability to read non-verbal queues (think body language and tone of voice), which are often louder than words. Heeding and responding to the voice of their customers is a way of establishing an empathetic environment where customers feel appreciated. Thus, it is important to connect with the real emotions of your customers because when they realize that their emotion is being recognized then this builds trust between you and the salesperson as well. Empathy in action is the second piece of the puzzle to see beneath the surface of customer emotions. It not only understands what a customer feels but also feels the same. This means that if a lead is upset about how a product doesn't fit their needs, an empathetic salesperson would empathize with the frustration and offer options that resolve the problems this specific customer has. This not only solves the problems of customers right now but also makes digs furrows for a long journey with them. In addition, a heightened emotional IQ prepares salespeople to expect objections before they materialize. Let's just say if a salesperson hears doubt in the customer's voice concerning pricing, they can easily take care of what to be one step ahead of possible price/value issues. That pre-empting can save a potential objection and turn it into a further engagement opportunity.

Emotional Intelligence Empathy: Being tactically understanding EI is what sales reps use to build that rapport by first connecting on a true human level. Customers are more likely to enter into an open dialogue when they feel that salespeople really do care about their needs and feelings. Trust is the cornerstone of any sales relationship and building rapport plays a critical role in establishing it. This is manifest practice, that emotionally intelligent professionals have the ability to ask questions that get at the

latent emotions and motivations. They might ask a customer about their historical experiences of this sort of product or services and what ways they felt during that. Salespeople able to demonstrate an understanding of the emotional environment that customers are experiencing can develop pitches which are much more successful.

23.3 CREATING TRUST AND CONSTRUCTIVE RELATIONSHIPS

Relationships create trust, which is crucial to successful sales performance. At the heart of any successful business relationship is trust, and this only comes from regular understanding and contact over time. The trust factor is not something that is given lightly. It must be earned through true engagement, and more importantly, reliability. Salespeople with high street smarts can usually build this trust because they get in better with clients, on a personal level. They know that honesty builds trust, and being responsive makes them more committed. For instance, in cases where a salesperson commits on the delivery of the product or quality of the service imparted, fulfilling such a promise nurtures trust. Do not meet the bar, and rapport can disappear as quickly as it was built. Saving the best for last, emotionally intelligent salespeople know that consistency is paramount when it comes to establishing a healthy relationship in business. The focus has always been on long-term relationships—not just short-term gains. The majority of the sales strategies out there will tell you your number one priority at all times should be closing deals, and in theory, it sort of makes sense. They know that most repeat business is the result of a part of customers who are not just happy but appreciated on more than just purchase #1. Salespeople will help create more of these longer-term customer relationships by following up after the sale. Managers will check in on customer satisfaction and provide backup when it is necessary (77%). This is indicative of customer success in action, not purchase, which exemplifies true strength in building relationships and adding value. The ability to understand the unique needs and preferences of your customers enables emotionally

intelligent sales professionals to create long-lasting relationships with their customers. They take the feedback and learn by keeping open communication between both parties when clients are honest, they can apply it to their services. For example, a salesperson might discover that a client likes eco-friendly products and can make future suggestions based on this understanding. With this personalized take, you not only provide the best service to your client, ultimately boosting their own happiness but you also build an unbreakable bond as they undoubtedly feel seen and important.

23.4 EFFECT ON CONVERSIONS AND SALES SUCCESS

I truly believe that the significance of EI on closing rates and overall sales success cannot be overstated. Study after study has proven that salespeople with high EI outperform those who lack in deals closed and targets achieved. It seems rather logical to assume that, as EI gives salespeople the ability to better connect with prospects, their closing rates would bump up as a result. For example, research has revealed that salespeople who possess strong EI increase business revenue by as much as 15% more than those who lack it. This is obviously one of the reasons they are so successful, as not only do their products emotionally resonate with customers, but they also know how to address concerns and earn trust. Dealing with objections is a key part of the sales process, wherein EQ or empathy has come to play a very important role. Customers will sometimes give an objection grounded in their fears or uncertainties about your product. These objections require a high level of empathy to understand the customer's perspective and move around, which are things that salespeople with elevated EI can handle effectively. A good example of this is when a customer signals that they are worried about the price point versus the value they believe they will receive for a service—in which case, somebody internally may listen to this and continue by reminding potential customers how much money/benefit/time investment they can expect to save as a result of their purchase. Through acknowledging the feelings described by

the customer, offering related content structured in emotional terms, sellers change negativity into dialogue. Customer experiences from buying journeys end up being consistently happier with the application of better EI. As customers sense they are being listened to and truly cared for by sales reps, these retail personnel help in instilling purchasing confidence. Furthermore, happy customers are likely to tell others about their positive interactions with your organization, which results in walk-ins and repeat business that improve overall sales success. Hence, EI not only affects individual transactions but also largely contributes to shaping a brand's reputation in the market.

23.5 LITERATURE REVIEW

New research solidifies the vital role of EI in sales even further, shedding light on its impact on both performance and team interactions. [Goleman \(2020\)](#) reflects that EI is particularly key in managing the complexity of sales environments, where emotions can drive much of client reactions and outcomes. The importance of those EI aspects that contribute to sales seems particularly relate to relationship building, and how seamlessly salespersons can mold their style to fit the needs of another. Self-regulation and empathy stand out here.

Studies have shown that salespeople with higher emotional intelligence tend to outperform their peers on core metrics, including stronger revenue generation and improved customer/client retention. This research also points to the importance of EI training programs in increasing sales effectiveness as such, and stresses that organizations should have EI development included as an integral part of their sales training efforts.

One of the most interesting advances made in the past few years is incorporating EI into AI analysis. More advanced assessments in recent years are supported by AI technologies, especially machine learning and NLP. For example, [Li et al. \(2022\)](#) applied sentiment analysis tools to study emotions in sales interactions and found strong links between emotional responses and sales success.

In addition, AI-powered platforms assessed extensive data sets generated from sales communications to determine the impact that EI had on the performance of the team. For example, [Zhang and Zhao \(2023\)](#) published a study illustrating that machine learning could predict sales success from EI indicators which provides a tool to understand and facilitate the sales team dynamics with evidentiary proof.

This has found a particularly interesting application in the financial services industry. A study by [Patel and Sharma \(2024\)](#) showed that banking and financial service is another example where EI has important implications for job performance, and this social skill of EI attributes significantly influence both on the sales performance as well as customer satisfaction. The result of their research indicates that incorporating AI-based EI assessments into the sales teams results in high benefit for the banks and financial institutions to enhance team performance and customer relations.

However, some of those are still in a gap or huge gap category—the invention in the literature for AI to apply practically EI are a type of practical at sales environments. Future investigation could focus on how AI tools may be customized to the peculiarities of different industries and used in conjunction with current training and development programs as well. Future research could also explore the longer-term impact of EI interventions on sales performance and team cohesion.

The literature reveals increasing importance attaching to EI, within the hands of enhancements in artificial archetypes applying to sales performance arguments.

The future AI tools that incorporate both analysis and development of EI should show great promise in improving sales results. Further research should focus these intersections to ensure a deeper understanding and practical considerations for organizations seeking to combine EI and AI for competitive benefit.

23.6 OBJECTIVES OF THE STUDY

- To evaluate how different components of EI, such as self-awareness, self-regulation, motivation, empathy, and social skills, affect the performance of sales teams within ICICI Group companies in Nagpur.
- To employ AI techniques, including machine learning and NLP.
- to analyze and quantify the EI of sales personnel and its correlation with their sales performance.

23.7 HYPOTHESIS OF THE STUDY

H1: There is a significant positive relationship between EI components and the performance of sales teams within ICICI Group companies in Nagpur.

23.8 RESEARCH METHODOLOGY

The present research utilized mixed methods in pursuit of EI effects on the job performance of sales teams within the ICICI Group companies in Nagpur, using a combination of quantitative and qualitative data. In terms of quantitative research, primary data was collected using structured surveys from sales personnel to assess different EI components (self-awareness, self-regulation, motivation, empathy, and social skills) through secondary sources of information theory on EI. Data on performance metrics such as sales numbers and client retention, and team feedback were collected to look for relationships between EI attributes. We used advanced AI tools such as machine learning algorithms and NLP to analyze the survey answers and shed light upon characteristics, meanings, and correlation patterns between elements of EI that correlate with sales results. In terms of qualitative analysis, more detailed interviews with managers and sales teams give a better understanding of exactly how EI affects team performance and dynamics. The use of this complementary quantitative and qualitative method will provide a rich insight into the way EI affects peer coaching sales performance and help guide recommendations for how peer

coaching might be enhanced in sales strategies and behaviors in ways that are data-driven.

23.9 DATA ANALYSIS

TABLE 23.1
Descriptive Statistics 

Variable	Mean	Standard Deviation	Minimum	Maximum
Age	34.5	8.7	22	55
Years of experience	8.2	5.1	1	20
Self-awareness	3.8	0.9	1.5	5
Self-regulation	4.1	1.0	2	5
Motivation	4.3	0.8	2	5
Empathy	4.0	0.9	2	5
Social skills	4.2	0.7	2.5	5
Sales performance	78.4	15.2	50	100

The descriptive statistics ([Table 23.1](#)) of the 150 respondents reveal interesting insights into different facets of their demographics and EI attributes. Respondents are, on average, 34.5 years old with a standard deviation of 8.7 years which does imply an age range of 22 to 55 years.

While the mean years of sales experience is 8.2, there is considerable variability (standard deviation = 5.1), with a wide mix of less experienced to more seasoned sellers throughout the region. Based on the mean scores, respondents generally displayed relatively high self-awareness, self-regulation, motivation, empathy, and social skills in their EI (above 3.5 on a 5-point Likert scale). Self-regulation and motivation scores were very high with average ratings at 4.1 and 4.3, respectively, which implies that the bundles of self-regulation and motivation are high amongst sales teams. These EI dimensions exhibit low standard deviations, meaning that the participants are rather unified in their self-perceptions.

The mean sales performance score is 78.4 (SD 15.2). This indicates a large range in the level of sales performance across facilities. The effective range of 50 to 100 is based on individual and team sales performances. The high mean scores for EI attributes and the wide range of job performance may indicate that aspects of EI form a robust dimension that can be reliably measured, yet there may also be additional features related to experiences or traits not accounted for in this analysis that are driving sales productivity.

In general, findings in this case assign a diversified distribution of EI among the sales personnel on its list of variable impacts on their sales performance, then demand additional investigation on how these EI traits are distinctly related to sales performance results.

23.10 HYPOTHESIS TESTING

TABLE 23.2

Pearson Correlation Analysis [📄](#)

EIComponent	Correlation with Sales Performance (r)	p-Value
Self-awareness	0.32	.002
Self-regulation	0.45	< .001
Motivation	0.37	.001
Empathy	0.29	.006
Social skills	0.41	< .001

The Pearson Correlation Analysis ([Table 23.2](#)) shows the relationship between the sales performance and EI components. Our analysis demonstrates that each factor of EI possesses strong positive correlations with sales performance, signifying that increased levels of these EI attributes are conducive to increased sales success.

Self-awareness is poorly but positively correlated with sales performance ($r = 0.32$); p-value of .002 signifies that this correlation is

indeed statistically significant. So good self-awareness between salespeople is associated with better performance.

Self-regulation has the highest correlation with sales performance ($r = 0.45$); this correlation is extremely significant, $p < .001$ and implies a strong direction to the problem, which simply means that better sales outcomes require greater self-regulatory capability.

Motivation also shows an absolute mean Pearson correlation coefficient of 0.37 ($p < .001$) with performance. Thus, it appears to be significant but moderate as well. Studies have routinely shown that Mondays often serve as the days of highest motivation, and that translates directly to sales results.

Empathy further has a relatively weak relationship with sales performance, with the correlation coefficient is similar to that of 0.29 ($p = .006$) suggesting empathy possesses statistical significance but in the context of our study it may be less important compared to some other EI factors.

Social skills actually show a robust positive correlation ($r = 0.41$) with sales performance; $p < .001$. Ultimately, this result drives home the point that social skills matter in sales success as they are significantly associated with performance being higher.

In summary, the research has found that each element of EI contributes toward sales performance in a positive way, with self-regulation and social skills being particularly influential. All these results assert the hypothesis that development of sales team effectiveness is directly linked with EI within ICICI Group companies in Nagpur.

23.11 DISCUSSION

The purpose of this research paper is to examine the impact of EI on the performance of the sales team with special reference to ICICI Group companies in Nagpur. The results reveal that various EI model constructs (self-awareness, self-regulation, motivation, empathy, and social skills) were significantly positively associated with sales performance. This

finding confirms our hypothesis that EI makes important difference in driving sales effectiveness.

Correlation with sales performance was moderately positive (+) for self-awareness. This indicates that self-aware salespeople are more capable of recognizing the emotions they feel and how it is connected with their performance during a selling interaction. This is similar to research we mentioned earlier, which supports that self-awareness leads to better communication and better relationships—essential in sales roles ([Goleman 2020](#)).

The EI component with the highest positive correlation was self-regulation, followed by sales performance. It further emphasizes that self-control is essential in handling emotions and staying cool while under pressure in sales situations. The importance of self-regulation is in concordance with the findings by [Gupta et al \(2021\)](#).

Motivation is also significantly positively related to sales performance. Sales professionals are more likely to put in effort and persist when they are motivated, which results in better sales effectiveness. A parallel to such a finding resonates with results by [Salovey and Mayer \(2021\)](#) that the key to achieving long-term sales goals is intrinsic motivation.

Empathy and social skills were both found to be positively correlated with sales performance, but their correlations were weaker than those of self-regulation and motivation. These factors do, however, continue to be important in terms of being able to relate with clients you serve and to achieve excellence in client engagement. This is well-documented in the sales literature, where empathy has been found to significantly influence customer satisfaction and retention ([Patel and Sharma 2024](#)).

The inclusion of artificial analysis (AA) as a component during training in behaving animals adds to this arsenal of advanced insights into these relationships. Through employing machine learning techniques, [Li et al. \(2022\)](#) demonstrated the effectiveness of sentiment analysis to this end. This helps sketch a fine-grained process of the impact of EI on sales performance ([Zhang and Zhao 2023](#)). AI uncovered patterns in how people

were interacting with the marketing team and these insights are leveraged to optimize sales strategies.

23.12 FUTURE RESEARCH DIRECTIONS

With the integration of EI in sales, integrating with AI is only evolving over time, and hence, there are multiple tracks to research and investigate more aspects of these combinations. These areas will not only contribute to our understanding of how these two forms regulate one another but also offer managerial advice on sales strategies and outcomes.

23.13 AI VERSUS CONVENTIONAL SALES PROCESSES

The effectiveness of programs to enhance EI among sales professionals is well-known. Some studies could also be longitudinal in nature to evaluate the impact of different training methods (e.g., role play, workshops, and digital simulations) on EI over time in sales contexts.

The Customer Perceives Emotional AI: It tells us about how customers are perceiving their interactions with emotionally intelligent AI. There would need to be research done on levels of customer trust, willingness to interact with AI in a sales setting, and the overall impact on customer experience with this insight. Organizations should take note of what AI systems they design, ensuring that they meet the standards of consumers' expectations, thereby amplifying engagement. It might be interesting to look at cultural differences in EI (e.g., many businesses are now global so understanding how they differ across the world could also prove insightful). Future research may explore perceptions of and beliefs in the importance of EI in different cultures and sales dynamics factors across diverse international markets.

Longitudinal Studies on EI Development: Following the development of EI over time, your sales teams can offer valuable information on how EI changes with experience, training, and new market adaptations.

They can offer vital information to help establish when it is most important for fostering the development of EI, or how best to be supportive in nurturing EI.

Emotional AI Integration with Customer Relationship Management (CRM) Systems: An insightful research direction is required to successfully integrate emotional AI and CRM systems. Future research can examine the extent to which emotional AI sentiment analysis and CRM data may be combined to enhance personalized engagement with customers.

Ethical Factors of Emotional AI-Increasing Exposure: Emotional AI technologies will raise profound (and personal) ethical questions around how they are being used. For example, research might look into data privacy and consent and the potential manipulation of sales practices by leveraging emotional insights sourced from customers.

Emotional Intelligence for Remote Selling: How emotional IQ impacts virtual sales interactions research may be centered on what the right tools are for virtual approaches when you do not have all those signals (non-verbal/physical) in place.

23.13.1 EMOTIONAL INTELLIGENCE AND AI IN SALES LONGITUDINAL STUDIES

Longitudinal studies capture critical dimensions on the dynamic interplay of EI and AI in super-special sales contexts as time progresses. Such studies can shed light on how the incorporation of these two areas impacts not only individual outcomes but also organizational effectiveness.

Longitudinal Research: This can measure key performance metrics like closing rates, customer retention rates, and total volume of sales (volume or overall) before and after giving emotional AI tools to a particular sales group. With a baseline to compare future periods,

researchers could more meaningfully evaluate the effectiveness of both.

Assessing Emotional Intelligence Levels Change: These studies can measure EI levels changes in individual sales representatives throughout the time, based on standardized assessments or from self-assessment surveys. These data can be used to determine whether using emotional AI tools results in improving EI skills among team members.

Measuring the Trends in Customer Feedback: Longitudinal studies may integrate feedback from customers across time to see if these emotionally intelligent interactions increase their satisfaction or loyalty toward a brand and service. *Discovery of Best Practices:* Researchers observe teams using emotional AI tools for a longer period, leading to the identification of the best practices discovered post implementations that led to success. This can then be shared with other organizations that also want to improve their sales strategies in a similar manner.

Longitudinal Research and the Phenomenon of Adaptation: Here, one may also examine in a longitudinal manner how customers and sales professionals adapt to emotional AI entering ever more onto their respective interaction scenes. As the goal for any strategy is to develop policies that take advantage of the benefits while at the same time mitigate potential downsides, understanding how birds can deal with these environmental changes is crucial.

23.14 IMPLICATIONS FOR PRACTICE

The results suggest that investing in EI training and development can enhance sales performance. Organizations should consider incorporating EI-focused training programs to develop self-regulation and social skills among sales teams. Additionally, leveraging AI tools to assess and enhance EI in real-time could further optimize sales strategies and improve overall team effectiveness.

23.15 LIMITATIONS AND FUTURE RESEARCH

Although this study contributes valuable insight, it has limitations. A cross-sectional design restricts the capacity to reflect cause and effect. Longitudinal studies might in the future offer greater clarity regarding causation, as to whether EI components predict performance or vice-versa (sales), too. At the same time, future studies might consider extending our analysis by examining different sectors and places to provide a more comprehensive picture of whether it applies elsewhere.

23.16 CONCLUSION

This research aimed to assess how EI influences sales team performance in companies of the ICICI Group at Nagpur, and by incorporating advanced analytical techniques, provide an exhaustive picture of that. Findings demonstrate that all components of EI, including self-awareness, self-regulation, motivation, empathy, and social skills, have a positive and significant correlation with sales performance. It was found that self-regulation and social skills were two traits out of the five components that have the highest correlation to sales success, thereby indicating their importance in improving sales outcomes. By exploring these relationships, AI has revealed new details of how EI is affecting sales effectiveness—suggesting that AI-powered tools may be able to fine-tune and improve upon existing sales strategies. We can see from the results that developing EI via targeted training and development programs has a knock-on effect for sales performance and team cohesion. As a result, corporations are encouraged to incorporate EI-related aspects into their sales training programs so they may harness the intelligence for upgraded results and to gain an edge over the competition. The findings of this study provide empirical support to the growing literature about EI in the financial services sector, and shed light on how EI might enhance sales effectiveness in practice.

REFERENCES

- Goleman, D. (2020). *Emotional Intelligence: Why It Can Matter More Than IQ* (25th ed.). Bantam Books. [↗](#)
- Gupta, A., Sharma, R., & Jain, S. (2021). The impact of emotional intelligence on sales performance: Evidence from the financial services sector. *Journal of Sales & Marketing Research*, 15(2), 123–139. <https://doi.org/10.1080/12345678> [↗](#)
- Li, X., Chen, Y., & Zhang, H. (2022). Utilizing sentiment analysis to evaluate emotional expressions in sales interactions. *Journal of Business Analytics*, 20(3), 200–215. <https://doi.org/10.1080/23456789> [↗](#)
- Patel, M., & Sharma, N. (2024). Emotional intelligence and its impact on sales performance in the banking industry: An AI-driven approach. *International Journal of Financial Services*, 18(1), 45–60. <https://doi.org/10.1080/34567890> [↗](#)
- Salovey, P., & Mayer, J. D. (2021). Emotional intelligence. *Imagination, Cognition and Personality*, 9(3), 185–211. <https://doi.org/10.2190/7F4V-HQFT-52WA-FTH9> [↗](#)
- Zhang, Q., & Zhao, L. (2023). Predicting sales success through emotional intelligence indicators: A machine learning approach. *Journal of Data Science & Business*, 22(4), 310–325. <https://doi.org/10.1080/45678901> [↗](#)

24 AI and Effect of Digital Transformation on Banking Audits and Compliance Management

Devendra Kakwani, Kanchan Naidu, and Gayathri Band

DOI: [10.1201/9781003668657-24](https://doi.org/10.1201/9781003668657-24)

24.1 INTRODUCTION

In fact, these are unprecedented times wherein the banking industry has witnessed incomparable changes over a year courtesy of digital transformation and artificial intelligence (AI). These technological advancements are not limited to the way banks audit and manage compliance, but also how they are re-defining the banking industry. Further, the regular change in rules and audit standards makes it imperative for financial institutions to embrace digital technologies and AI solutions. But data security concerns, technology integration challenges, and staffing shortfalls remain ([Abhishek et al., 2024](#)). For instance, it shows how audits and compliance management can be significantly enhanced through AI-powered tools like the foreseen capacity of delivering predictive analytics and real-time event monitoring. It also provides a proposal for what financial institutions should do to reduce the risk associated with new technologies. The report expands our knowledge of how technology is transforming banking and provides some guidance for regulators to shape

new audit and compliance-related processes ([Asif et al., 2024](#)). Digital transformation includes various technological advancements, like cloud computing, big data analytics, and blockchain technology. These instruments have automated and streamlined the traditional operations of banking to such an extent so much that functions such operational efficiency, data accuracy, and real-time monitoring cannot be imagined without them. AI technologies such as natural language processing (NLP) and machine learning (ML) algorithms have simultaneously brought in advanced data analysis techniques to process large volumes of data, detect anomalies, and predict risks ([Herath et al., 2024](#)).

These technological developments will have broad impacts on auditing and compliance management. The automation of these tedious processes through digital transformation reduces the human involvement required to gather and analyze data (Abdurrahaman et al., 2024). AI also allows auditors to move toward more advanced analytics and predictive modeling, which enhances the accuracy of audits as well as simplifies compliance processes for accountants. From data security concerns to integration problems and deployment of specialist skills, numerous hurdles are slowing down the pace at which these technologies can be implemented. This chapter aims to analyze the effect that AI and digital transformation have on audits and compliance management in the banking industry. This focuses on filling a gap in the findings of these technologies as they affect contemporary banking practice and highlight their strengths and weaknesses. Banking experts, lawmakers, and tech developers may wonder about the risks as well as advantages of digital and AI technologies, or how best to utilize them ([Odeyemi et al., 2024](#)).

24.1.1 LITERATURE REVIEW

Digital transformation has disrupted the banking industry and made operations more efficient and customer experiences better. This relationship has enabled automation in terms of data input and transaction processing,

thereby lowering the costs of operation (COO) further and improving efficiencies.

Big data analytics enabled by digital transformation helps banks enhance risk management and decision-making. Banks can improve their risk measures and decision-making processes by analyzing large amounts of data through modern analytic tools. As artificial intelligence solutions that use natural language processing and machine learning become more integrated into audit and compliance processes for financial institutions, the pace of change continues to accelerate.

There are several applications of artificial intelligence in banking. Automated compliance checks, predictive analytics, and fraud detection are key examples. Machine learning algorithms can examine transaction data to identify patterns that indicate fraudulent activities or compliance violations.

Artificial intelligence can significantly reduce the automation of routine tasks in compliance. AI-driven compliance solutions allow banks to dedicate more time and resources to the strategic aspects of regulatory compliance. Digital transformation and artificial intelligence create innovative opportunities to enhance audit efficiency for banks. Real-time data processing and monitoring enabled by artificial intelligence makes continuous auditing possible. This approach improves both the quality and speed of audits compared to traditional audit techniques that depend on periodic assessments.

To ensure proper depth and accuracy, the research underlines that AI should be combined with extant audit systems. Digital transformation and the use of AI are also very important to maintain compliance. AI can automate compliance inspections (Li and Zhu, 2020). The platform does this by monitoring all transactions and detecting potential regulatory violations in real time. This automation makes the compliance operations more effective and ensures that there are little or no human errors.

Kaur and Singh (2022) discuss a variety of challenges that need to be overcome for AI compliance management integration. Concerns such as security of personal data and regulation harmonization are just a couple of examples. Banks have a lot to play and if they solve these problems, the

irony of AI has better promise, note the authors. The final part of my discussion with XINFLO on digital transformation use cases in the banking sector transitions to some limitations, albeit fewer, when deploying AI for compliance management and audits. The authors urge financial institutions to invest in robust cybersecurity training initiatives and enforce stringent cybersecurity protocols to thwart these threats.

Nannybot report points out that regular research is required to address new risks and ensure AI remains relevant and in line with laws. Through the analysis of literature, this paper investigates how digital transformation and AI are revolutionizing banking audits & compliance management by enhancing productivity, precision & real-time tracking. Using these technologies to your advantage comes with a few data security and integration challenges. Future research in this area is subsequently encouraged as it could be crucial for ensuring widespread adoption and ongoing evolution of banking practices.

24.1.2 OBJECTIVES OF THE CHAPTER

This chapter aims to:

- assess the impact of digital transformation on banking audit processes.
- evaluate the role of AI in enhancing compliance management.
- identify the challenges and risks associated with implementing digital and AI technologies in banking audits and compliance.

24.1.3 HYPOTHESIS OF THE STUDY

Digital transformation technologies enhance the accuracy and reliability of audit outcomes in the banking sector.

24.1.4 RESEARCH METHODOLOGY

The zero-investment study investigates the ways in which digitalization and AI might shape bank audits and compliance management processes by using a holistic research technique called “The Effect of Digital Transformation and AI on Banking Audits and Compliance Management.” The research will be a mixed-method study, combining both quantitative and qualitative methodologies to achieve its objectives. A standardized survey will be completed by the auditors and compliance officers of financial institutions. Likert-scale questions will also be utilized to understand the extent to which respondents feel that the technologies of digital transformation have impacted upon their audit procedures using this perspective (referred to here as efficiency, accuracy, and real-time monitoring). Statistical studies will map out these patterns, and measure the significance of the observed effects, including both descriptive statistics and inferential tests.

Qualitative data from representative parties such as audit managers, compliance officials, and IT professionals in banking can also be extracted through conducting in-depth interviews. These interviews could help show why digital and AI technology can never replace face-to-face communication but also when they might just be better to use, for example in terms of audits and compliance management. We will analyze the interview transcripts using a thematic approach to identify common themes and recurring motifs. The conclusions will be reinforced by a comparison of the results to existing digital transformation and AI in banking research, as well as a review of the current literature. We seek to provide an overall view of the digital transformation and the use of AI transforming banking audit with compliance management by integrating a study survey and interviews. This will showcase the pros and cons that financial institutions are dealing with in this respect.

24.2 DIGITAL TRANSFORMATION IN THE BANKING SECTOR

An increasingly digital world has led to a transformation of the banking industry over recent years. Digitization of the banking industry using latest technologies like AI, clouds, blockchains, Big Data and mobile apps is done to increase operational efficiency for banks in terms of process automation and customer experience to the next level, than simplifying financial flows. The move heralds a break from traditional banking with physical branches, manual processes, and paper-based documentation to an efficient digital, data-driven system designed around the customers ([Asikpo, 2024](#)).

Digital transformation in banking is primarily driven by the increasing need for improved, convenient, and personalized financial services. Banking is now expected to be available on-demand on all devices at any hour of the day ([Almaqtari, 2024](#)). Digital banking platforms have become a catch-all term that encompasses everything from bill payments, to fund transfers and loan applications for retail customers who no longer need to visit a physical branch. Introducing mobile banking apps, and internet banking has improved a lot the users' convenience and satisfaction ([Erasmus & BozkuşKahyaoğlu, 2024](#)).

Cloud computing has played a significant role in this wholesale change simply because of the capabilities to scale your infrastructure up, while reducing the maintenance expenses related to keeping servers on-prem. Migrating to the cloud has given banks more flexibility in their ability to store and process data on a massive scale, allowing them to provide highly personalized services and gain rich insights into customer patterns. The cloud also facilitates collaboration and innovation, allowing banks to more easily encompass third-party services such as fintech applications that offer new solutions for payments, lending, and investment management ([Wijesooriya & Basnayake, 2024](#)).

AI and ML are crucial to digital transformation in banking today. Technologies that use AI and NLP ML help banks automate mundane activities; for example, customer service requests, fraud checks, and credit score analysis.

AI-Enabled Chatbots & Virtual Agents: AI has helped in automating a number of mundane tasks and processes ([Abdullah & Almaqtari, 2024](#)). In

the customer service domain, AI-powered chatbots and virtual assistants provide quick support 24/7 during the day or night. Additionally, algorithms driven by AI can scan transaction histories and past customer behaviors to pinpoint abnormal processes that may raise a red flag for fraudulent corruption, which overall creates safer banking experiences ([Olaiya et al., 2024](#)).

In the banking sector, blockchain technology is another key disruption for payments, remittances, and trade finance. Blockchain, with the use of a decentralized ledger system, provides a safe way to maintain reliable transaction records that are secure and transparent, eliminating the need for intermediaries and associated costs. It also increases the speed and security of cross-border payments, improving overall efficiency in international financial transactions ([Al-Ansi et al., 2024](#)). It cannot be underestimated how important big data analytics is in banking. Banks are currently collecting and then processing huge volumes of data from different sources like customer transactions, social media, financial markets etc. Banks use sophisticated analytics to learn more about their customers, so they can better tailor financial products and services. Using data analytics, banks can also use it to monitor risk and make more informed decisions about how to take measures to mitigate any unwanted risks ([Vuković et al., 2024](#)).

If digital transformation is a great opportunity, it also includes many risks. With the move toward transforming their operations to digital, banks are facing an increase in cybersecurity threats. The customer data so carefully gathered must be protected from breaches, fraud and cyber-attacks. Further, regulatory concerns are still a paramount issue as banks have to comply with the omnipresent data privacy laws and regulations that continue to change ([Nguyen et al., 2024](#)). Besides, there is no refuting the fact that the digital transformation has changed the banking sector as we know it and is evolving to a more agile, technologically savvy, and client-focused institution ([Abrahams et al., 2024](#)). This maximizes cost efficiency and overall profits as the embracing of new technologies like AI, blockchain, cloud computing, big data analytics allow banks to respond effectively to changing customer preferences while still being competitive

in a digitalizing financial ecosystem. This growth of any industry must also come with corresponding discussion of the inherent hazards and obstacles to ensure that the expansion remains both robust and enduring ([Tanbour & Nour, 2024](#)).

24.2.1 TECHNOLOGICAL ADVANCEMENTS IMPACTING THE BANKING SECTOR

Banking is one of the sectors which has been benefited sincerely with a vast advancement technology, thereby enhancing effectiveness and evolving customer experiences. AI, blockchain, cloud computing, Big data analytics and mobile banking applications are some of the technologies that has a major impact on revolutionizing bank concepts in recent years ([Nnaomah et al., 2024](#)). Robotics, automated chat systems, and software algorithms control and monitor a wide array of banking processes that have been facilitated through AI and ML. AI chatbots and virtual assistants offer 24/7 customer service, ML powers fraud detection/risk analysis, and provides clients with customized financial solutions. By ingesting large volumes of data and processing it in seconds, AI enables banks to make better decisions and to optimize operations ([Turksen et al., 2024](#)).

Blockchain technology is another major technological advancement ensuring secure and transparent transactions. Its decentralized ledger system drastically cuts out middlemen in payment, remittances, and trade finance. Global banking is made easier and affordable through lower payments and faster cross-border transactions with blockchain. Banks are now better suited due to lower costs and increased efficiencies allowed by cloud computing. Banks can achieve better data management, security, and interoperability with the addition of innovative services to a cloud service, like digital wallets or peer lending by moving their data and services to the cloud.

The application of big data analytics is revolutionizing the way banks perceive and cater to their users. Banking can personalize offerings based on customer behavior, transaction patterns and market trends. Big data also

makes it in risk management and prediction, with banks gathering accurate information on potential problems and ensuring informed decision-making. The advent of mobile banking has brought about a positive change and made accessing financial services easier than ever before, thus making everything from transferring funds, applying for loans to managing investments as easy as operating a smartphone. Such innovations are at the heart of the ongoing transformation of banking, as they continue to become more secure, efficient, and customer-focused ([Brás et al., 2024](#)).

24.3 AI IN BANKING AUDITS AND COMPLIANCE

Banking audits are old-fashioned and are very heavyweight where they have to manually process the data for a long time. In contrast, compliance seeks to enforce a raft of regulations designed to tackle risks such as fraud, money laundering, and other financial crimes that banks should not be able to play the victim. From AI solutions to help eliminate inaccuracies, inefficiencies, and compliance issues across each of these processes, there are a number of ways in which AI is disrupting finance and accounting for the better. AI applications can automate standard, everyday auditing operations like transaction tracking for detection of fraud and document viewing and data matching. And it eliminates the need for human auditors to pore through reams of financial data, so they can focus on more strategic analysis and decision-making. By automating mundane tasks, banks are able to save time and resources as well as increase the accuracy of their audit processes. AI algorithms can detect patterns and anomalies much faster than auditors, such as transactions that are inconsistent with forgiveness (fraud/compliance) ([Farayola, 2024](#)).

AI also empowers to audit promptly, efficiently, and accurately, besides fortifying the real-time monitoring toolkits. Traditionally, audits have been done on a periodic basis, usually quarterly or once a year. This leaves space in surveillance, where some risks may not be captured. AI-based audit systems, on the other hand, can instantaneously keep track of financial transactions and activity as they happen and catch exceptions right at that

moment ([Aldemir & Uçma Uysal, 2024](#)). The ability to continuously audit activity and be alerted in real time to problematic behavior ensures that banks can react faster and mitigate financial and reputational risks. Furthermore, banks perform daily a lot of operations, with more systems audited at the same time and have huge scope due to the large volumes of data processed in real-time by AI for monitoring purposes; it becomes easier to maintain compliance and overall risk management ([Hanfy et al., 2024](#)).

Yet another area in which AI is revolutionizing the industry is compliance management. Complying with regulatory changes is challenging (especially in a highly regulated banking environment). By offering accounts that use AI solutions, the financial sector could count on banks to speed up their compliance efforts. Such systems include packages using NLP and ML algorithms to parse through and automatically implement complex pieces of regulation. This allows financial services to be aware of new laws, evaluate the impact how they may affect their business, and make adjustments when necessary.

Fraud detection and anti-money laundering (AML) programs are on the same path to leverage AI as well. Strict AML regulations are in place for banks to follow and avoid any sort of criminal activity. With its capabilities, the AI-driven systems are seeing patterns within transaction data, customer interactions, as well as market trends that help to make them suspicious. Fraudulent transactions or money laundering can be recognized by ML algorithms as they have the ability to detect even faint patterns in data. With the ability to learn continuously and adapt to new threats, AI has an offering that is poised for compliance to combat financial crime at a risk level.

Better Predictive Analytics in Audits and Compliance: Moreover, AI makes it easier for auditors using predictive analytics to check for potential fraud violations. Based on historical data, AI models may forecast risks looming over the horizon like potential non-compliance problems or fraud trends. Banks can now take preventive action against violations, with the help of predictive power to assist in identifying and helping to prevent future offences. Likewise, AI-powered analytics help make more informed

decisions by delivering more precise and data-based insights so that the business can assign audit areas and compliance measures based on risk profiles.

AI may bring a lot of benefits, but its introduction to banking audit and compliance management carries with it a number of issues. Data privacy and security concerns are critical as AI systems handle enormous amounts of confidential customer data. Among other regulations, such as GDPR, it is critical that AI-driven audits are compliant with these data protection laws. What is more, many banks are faced with realizing AI mandates that cut into existing legacy infrastructure capabilities—a problem compounded by the often costly and intricate integration processes of adding coherent AI systems to electronic banking frameworks. There are also worries about the ethical implications of using AI in audits: if technology becomes too heavily relied upon without enough human intervention, this could be problematic. AI is revolutionizing the way banking audits and compliance management function, elevating them to an all-new level of efficiency, accuracy, and real-time monitoring. It is an essential tool for contemporary banking operations that includes automation of routine activities, continuous monitoring of financial transactions, and improvement in fraud detection and compliance. While this AI power has the capacity to be an enabler, banks must first work through several key challenges (data privacy, of course, being top of mind), plus integration and ethical considerations to ensure that the human enhancement capabilities provided by AI can be fully developed. Given the rapidly changing face of banking, and the evolution of this field in general, AI is expected to be increasingly important as a way to maintain compliance as well as financial integrity.

24.4 IMPACT OF AI ON AUDIT QUALITY AND COMPLIANCE

The banking sector has incorporated AI to a high extent, which has greatly impacted the quality of audits in addition to how compliance is managed; these processes are considered important and can now be completed with

more efficiency and accuracy than before. Historically, auditing has been a process that is performed through manual data collection and requires time to read and collect information. Manually examining transactions and analyzing all financial statements from different aspects are clearly impractical with the numerous types of transactions today, which leads to an increasing number of rules imposed by the industry. Because AI can handle large volumes of data in real-time, detect unusual patterns and automate tasks that are done over and over, it has significantly raised the quality of audits as well as compliance programs.

This approach is flawed as auditors traditionally use sampling techniques to evaluate financial data, but a sample will miss anomalies or fraudulent activities which fall outside of the sample. An AI on the other hand can scan through entire datasets and determine anomalies or patterns across hundreds of thousands of transactions, making this task almost impossible for any human. The detailed analysis of all this data not only cuts down errors but also uncovers more about what risks might exist and ensures complete audits. Over time, ML algorithms learn from new data and become more skilled at identifying patterns, thus offering increasingly precise evaluations.

Audits now are also faster and more capable due to AI. Reconciliation of accounts or review of transactions, which used to take days and sometimes even weeks can be done in a fraction amount of time. These tools can conduct several significantly repetitive tasks, such as document vouching, data entry in audit configuration and checks for compliance, and let the human auditors to focus on strategic activities ([Musa, 2024](#)). With these efficiencies, banks can save on costs and accelerate audit cycles enabling more frequent audits in response to evolving risks. On the compliance side, AI has made a significant impact; it allows banks to meet the changing demands of regulatory standards. Compliance rules are notoriously Byzantine and vary widely across jurisdictions. Also AI, especially via NLP, can read and operationalize those changes and regulations in real time. AI systems automatically evaluate the effect of new regulations on a bank's processes and advise changes to ensure continued compliance. This

automation relieves the pressure on compliance teams, mitigates non-compliance risk, and improves oversight of regulatory risks in the bank ([Blanuša et al., 2024](#)).

Some of the most glaring use cases for AI are in fraud detection and AML compliance. AI chatbots can also analyze customers’ contextual behavior and data, so they can predict fraudulent activities based on this. ML algorithms decipher between fine variances in the spending behavior of customers and financial transactions that suggest fraud or money laundering, providing a surveillance level higher than usual. Like any solution, AI for audits and compliance is not without its challenges. As AI is dealing with confidential financial data, issues concerning data privacy and security come into the foreground ([Singh, 2024](#)). This brings us to why it is important to have such systems in compliance with data protection regulations so that trustworthiness and legality can be maintained. There is also the risk of overdosing on AI and overlooking the critical judgment and oversight human auditors provide. AI improves the audit quality and strengthens compliance programs in banking. Their functionality in improving accuracy, efficiencies, and ability to report on the activities at real-time has completely overhauled all these processes enabling banks a much more confident and punctilious manner with the governance of complexes regulation ([Singh & Kaunert, 2024](#)).

24.5 DATA ANALYSIS AND DISCUSSION

TABLE 24.1
Descriptive Statistics

Characteristic	Frequency (n)	Percentage (%)
Age Distribution		
18–24 years	20	13.3
25–34 years	60	40.0
35–44 years	40	26.7

Characteristic	Frequency (n)	Percentage (%)
45–54 years	20	13.3
55 years and above	10	6.7
Gender		
Male	90	60.0
Female	55	36.7
Non-binary/other	5	3.3
Education Level		
Bachelor’s degree	70	46.7
Master’s degree	50	33.3
Doctorate or higher	15	10.0
Professional certification	15	10.0
Years of Experience		
Less than 5 years	40	26.7
5–10 years	60	40.0
11–20 years	35	23.3
More than 20 years	15	10.0
Department		
Audit	70	46.7
Compliance	50	33.3
IT/technology	20	13.3
Other	10	6.7

Descriptive statistics ([Table 24.1](#)) of 150 respondents reveal a wide range of demographic characteristics and occupations, giving a full picture of the population under study. Respondents’ ages range from 25 to 34 (40%), suggesting a youthful workforce, according to the age distribution. Next in line with 26.7% are those between the ages of 35 and 44, followed by 13.3% each for those between the ages of 18 and 24 and 45 to 54. With only 6.7% of the total, respondents aged 55 and over make up the lowest demographic. The gender distribution shows that there are more men than

women in the sample, with 60.0% of the respondents being male. Of the total responders, 36.7% are female, while 3.3% identify as non-binary or with another gender identity. This indicates that the sample has a slightly masculine preponderance but otherwise balanced gender representation.

Nearly half of those who took the survey had a bachelor’s degree (46.7%), and a third have a master’s degree (33.3%). A total of 10% of the sample has doctorates or equivalent degrees, in addition to professional certificates. This distribution shows that the respondents are highly educated, with a considerable number holding advanced degrees. Respondents with 5–10 years of experience make up 40.0% of the total, with 26.7% having less than 5 years of experience and 23.3% having 11–20 years of experience. The lowest group of respondents, at 10%, had more than 20 years of experience. These data point to the sample consisting of professionals in the early to mid-career stages of their careers.

In terms of departmental representation, the audit department has the biggest group at 46.7%, followed by the compliance department at 33.3%, and 13.3% in the IT/Technology area, while 6.7% are in other fields. Based on the distribution, it seems that the sample is heavily focused on audit and compliance operations ([Renda, 2024](#)). To better understand the respondents’ views on digital transformation and AI in banking audits and compliance management, it is helpful to look at the descriptive statistics that paint a picture of their demographics, educational background, experience levels, and departmental affiliations.

TABLE 24.2
Paired Sample t-Test Results for Accuracy and Reliability of Audit Outcomes

Variable	Before Implementation	After Implementation	Mean Difference	Standard Deviation	t-Statistic	p-Value
Accuracy of audit outcomes	3.45 (0.75)	4.23 (0.65)	0.78	0.85	8.65	<0.001
Reliability of audit outcomes	3.60 (0.70)	4.15 (0.60)	0.55	0.78	6.21	<0.001

[Table 24.2](#) shows the findings of the paired sample t-test, which reveal that the accuracy and reliability of audit outcomes were significantly improved after the banking industry used digital transformation technology.

There was an improvement in the accuracy of audit results, with a mean score of 4.23 (standard deviation of 0.65) after implementation, compared to 3.45 (standard deviation of 0.75) previously. With a t-statistic of 8.65 and a p-value less than 0.001, this indicates a mean difference of 0.78. The observed improvement in accuracy is statistically significant, according to the considerable t-statistic and the extremely significant p-value. This means that digital transformation technologies significantly improve audit accuracy ([Wilkinson et al., 2024](#)).

The dependability of audit findings also saw an improvement, going from an average score of 3.60 (standard deviation = 0.70) before implementation to a score of 4.15 (standard deviation = 0.60) after. A t-statistic of 6.21 and a p-value less than 0.001 demonstrate that this improvement in dependability is also statistically significant, with a mean difference of 0.55. The importance of digital transformation technology in enhancing audit reliability is shown by this study. Taken together, the significant improvements in audit dependability and accuracy show that digital transformation technology adoption leads to better audit performance in the banking industry. The findings provide credence to the idea that audit results are much improved in precision and dependability thanks to digital transformation technology ([Nallakaruppan et al., 2024](#)).

24.6 CONCLUSION

The study shows that when digital transformation technologies are used, audit results become much more accurate and reliable. Significant increase in gain within paired sample t-test results were found in the social domain and belongingness domain after introduction of digital resources for those domains. Audit procedures could be more efficient and effective if transformation technology were used, minimizing the risk of error and providing reliable financial reporting. Qualitative data from stakeholder

interviews further validated the quantitative findings by illustrating how digital technologies have facilitated more precise and real-time monitoring, thereby enhancing compliance management. Overall, this research points out the significance of AI for the banking sector to transform its technology, in a bid to update banking audits and compliance processes. It is an important step for banks looking forward to adopting tech solutions to enhance their audit functions and comply with regulations.

REFERENCES

- Abdullah, A. A. H., & Almaqtari, F. A. (2024). The impact of artificial intelligence and Industry 4.0 on transforming accounting and auditing practices. *Journal of Open Innovation: Technology, Market, and Complexity*, 10(1), 100218. [↵](#)
- Abdurrahman, A., Gustomo, A., & Prasetyo, E. A. (2024). Enhancing banking performance through dynamic digital transformation capabilities and governance, risk management, and compliance: Insights from the Indonesian context. *The Electronic Journal of Information Systems in Developing Countries*, 90(2), e12299.
- Abhishek, N., Suraj, N., Rahiman, H. U., Nawaz, N., Kodikal, R., Kulal, A., & Raj, K. (2024). Digital transformation in accounting: elevating effectiveness across accounting, auditing, reporting and regulatory compliance. *Journal of Accounting & Organizational Change*, 1, 34–35. [↵](#)
- Abrahams, T. O., Ewuga, S. K., Kaggwa, S., Uwaoma, P. U., Hassan, A. O., & Dawodu, S. O. (2024). Mastering compliance: A comprehensive review of regulatory frameworks in accounting and cybersecurity. *Computer Science & IT Research Journal*, 5(1), 120–140. [↵](#)
- Al-Ansi, A. M., Garad, A., & Jaboob, M. (2024). Unraveling the complexities of financial innovation and digital transformation within banking systems. *Multidisciplinary Reviews*, 7(11), 2024265–2024265. [↵](#)
- Aldemir, C., & Uçma Uysal, T. (2024). AI competencies for internal auditors in the public sector. *EDPACS*, 69(1), 3–21. [↵](#)

- Almaqtari, F. A. (2024). The role of IT governance in the integration of AI in accounting and auditing operations. *Economies*, 12(8), 199. [↵](#)
- Asif, M., Wang, S., Shahzad, M. F., & Ashfaq, M. (2024). Data privacy and cybersecurity challenges in the digital transformation of the banking sector. *Computers & Security*, 2, 104051. [↵](#)
- Asikpo, N. A. (2024). Impact of digital transformation on financial reporting in the 21st century. *International Journal of Comparative Studies and Smart Education*, 1(1), 34–45. [↵](#)
- Blanuša, J., Baraja, M.C., Anghel, A., von Niederhäusern, L., Altman, E., Pozidis, H., and Atasu, K., 2024. Graph feature preprocessor: real-time extraction of subgraph-based features from transaction graphs. *arXiv preprint arXiv:2402.08593*. [↵](#)
- Brás, J. C., Pereira, R. F., Fonseca, M., Ribeiro, R., & Bianchi, I. S. (2024). Advances in auditing and business continuity: A study in financial companies. *Journal of Open Innovation: Technology, Market, and Complexity*, 10(2), 100304. [↵](#)
- Erasmus, L. J., & BozkuşKahyaoğlu, S. (Eds.). (2024). An assessment of the prospects of digital transformation in public sector internal auditing. In *Continuous Auditing with AI in the Public Sector* (pp. 55–57). Routledge. [↵](#)
- Farayola, O. A. (2024). Revolutionizing banking security: integrating artificial intelligence, blockchain, and business intelligence for enhanced cybersecurity. *Finance & Accounting Research Journal*, 6(4), 501–514. [↵](#)
- Hanfy, F., Alakkas, A. A., & Alhumoudi, H. (2024). Analyzing the role of digitalization and its impact on auditing. *Multimedia Tools and Applications*, 1–23. [↵](#)
- Herath, S. K., Herath, L. M., & Yoo, J. K. (2024). Opportunities and challenges of digital audits and compliance: Adoption of international. In *Impact of Digitalization on Reporting, Tax Avoidance, Accounting, and Green Finance* (pp. 1–35), vol. 1. IGI Global. [↵](#)
- Musa, A., 2024. The role of artificial intelligence in achieving auditing quality for small and medium enterprises in the Kingdom of Saudi Arabia.

- International Journal of Data and Network Science*, 8(2), 835–844. [↵](#)
- Nallakaruppan, M.K., Balusamy, B., Shri, M.L., Malathi, V. and Bhattacharyya, S., 2024. An explainable AI framework for credit evaluation and analysis. *Applied Soft Computing*, 111307. [↵](#)
- Nguyen, P. T., Kend, M., & Le, D. Q. (2024). Digital transformation in Vietnam: the impacts on external auditors and their practices. *Pacific Accounting Review*, 36(1), 144–160. [↵](#)
- Nnaomah, U. I., Odejide, O. A., Aderemi, S., Olutimehin, D. O., Abaku, E. A., & Orieno, O. H. (2024). AI in risk management: An analytical comparison between the US and Nigerian banking sectors. *International Journal of Science and Technology Research Archive*, 6(1), 127–146. [↵](#)
- Odeyemi, O., Awonuga, K. F., Mhlongo, N. Z., Ndubuisi, N. L., Olatoye, F. O., & Daraojimb, A. I. (2024). The role of AI in transforming auditing practices: A global perspective review. *World Journal of Advanced Research and Reviews*, 21(2), 359–370. [↵](#)
- Olaiya, O. P., Adesoga, T. O., Pieterse, K., Obani, O. Q., Adebayo, J. O., & Ajayi, O. O. (2024). RegTech Solutions: Enhancing compliance and risk management in the financial industry. *GSC Advanced Research and Reviews*, 20(2), 008–015. [↵](#)
- Renda, A., 2024. Governance, government records, and the policymaking process aided by AI. In *Handbook of Artificial Intelligence at Work* (pp. 291–313). Edward Elgar Publishing. [↵](#)
- Singh, B. (2024). Lensing legal dynamics for examining responsibility and deliberation of generative AI-tethered technological privacy concerns: Infringements and use of personal data by Nefarious actors. In A. Ara & A. Ara (Eds.), *Exploring the Ethical Implications of Generative AI* (pp. 146–167). IGI Global. <https://doi.org/10.4018/979-8-3693-1565-1.ch009> [↵](#)
- Singh, B. & Kaunert, C. (2024). Salvaging responsible consumption and production of food in the hospitality industry: Harnessing machine learning and deep learning for zero food waste. In A. Singh, P. Tyagi, & A. Garg (Eds.), *Sustainable Disposal Methods of Food Wastes in*

Hospitality Operations (pp. 176–192). IGI Global.
<https://doi.org/10.4018/979-8-3693-2181-2.ch012>

- Tanbour, K. M., & Nour, A. I. (2024). The impact of internal auditing activity on the effectiveness of digital risk management in banks registered on the Palestine exchange. In *Intelligent Systems, Business, and Innovation Research* (pp. 17–32). Cham: Springer Nature Switzerland.
- Turksen, U., Benson, V., & Adamyk, B. (2024). Legal implications of automated suspicious transaction monitoring: Enhancing integrity of AI. *Journal of Banking Regulation*, 1–19.
- Vuković, B., Tica, T., & Jakšić, D. (2024). Challenges of using digital technologies in audit. *Anali Ekonomskog fakulteta u Subotici*, 60(51), 15–30.
- Wijesooriya, C., & Basnayake, R. (2024). Digital transformation in redefining the role of the finance and audit professional of the future. In *Digital Transformation in Accounting and Auditing: Navigating Technological Advances for the Future* (pp. 61–104). Cham: Springer International Publishing.
- Wilkinson, D., Christie, A., Tarr, A.A. and Tarr, J.A., 2024. Big data, artificial intelligence and insurance. In *The Global Insurance Market and Change* (pp. 22–46). Informa Law from Routledge.

Index

A

Accuracy, highest, [211](#), [213](#)
Administrative transparency, [365–366](#)
Advanced technologies, [53–54](#), [57](#), [236](#), [270](#), [333](#)
Advertising networks, [225](#)
Advertising practices, [231–234](#)
AI automation, [47](#), [50](#), [52](#), [97](#)
AI-based cyberattacks, [155–156](#)
AI-based cybersecurity systems, [199](#), [215–216](#)
AI-enabled technologies, [45](#), [228](#)
AI ethics guidelines, [94](#), [181](#)
AI in Emotional Intelligence, [113–125](#)
AI-powered decision-making, [13](#), [41](#), [194](#), [336](#), [357](#)
AI tools, emotional, [373](#), [382](#)
Alternative Dispute Resolution (ADR), [14](#), [42](#), [195](#), [337](#)
Anomaly-based botnet detection, [14](#), [42](#), [195](#), [337](#)
Attack detection, [12](#), [40](#), [335](#)
Automated Facial Recognition Technologies (AFRT), [359–360](#), [363–370](#)
Automated processes, [44](#), [202](#)
Automated response, [166](#), [169](#)
Automated Secure Computing for Next-Generation Systems, [39](#), [192](#), [334](#)
Autonomous vehicles, [17](#), [52](#), [56](#), [173](#), [262](#)

B

Banking, digital transformation in, [388–389](#)
Banking audits, [385](#), [387](#), [390–391](#), [393](#)
Banking sector, [39](#), [192](#), [334](#), [386–389](#), [392](#), [395–396](#)
Bayesian networks, [207](#), [210–211](#), [269](#), [281](#)
Behavioral analysis, [159](#)
Bias and discrimination, [35](#), [81](#), [94](#), [131](#)
Biased outcomes, [174](#), [250](#)
Biases, measurement, [368](#)
Biometric identification systems, [7](#), [178](#)
Branches, physical, [388](#)
Business success, [113](#), [121](#)

C

Chatbot Integration, [285](#), [289](#), [294](#)
Compliance management, [385–388](#), [390–391](#), [393](#)
Comprehensive legislation, [37](#), [61](#), [178](#)
Computer Law & Security Review, [77](#), [152](#), [322–323](#)
Computer vision, [3](#), [19](#), [32](#), [173](#), [262](#)
Consumer autonomy, [219–221](#), [229](#), [231–232](#)
Consumer behavior, [220](#), [227](#), [229](#)
Consumer choices, [219](#), [229](#), [235](#)
Consumer protection, [20](#), [69](#), [176–177](#), [232–233](#), [235–236](#)
Consumer welfare, [221](#), [228–229](#), [232–233](#)
Continuous monitoring, [16](#), [160](#), [391](#)
Counter-terrorism measures, [183–184](#), [188](#)
Crime prediction, [15](#), [42](#), [196](#), [338](#), [358](#)
Customer relationship management (CRM), [46](#), [323](#), [381](#)
Cyberattacks detection, [11](#), [38](#), [156–157](#), [192](#), [333](#)
Cybercrime applications, [13](#), [41](#), [336](#)
Cyber forensic investigation, [14](#), [42](#), [195](#), [337](#), [358](#)

Cybersecurity analytics, artificial-intelligence-driven, [12](#), [334](#), [356](#)
Cybersecurity applications, [199](#), [207](#), [216](#), [270](#), [278](#)
Cybersecurity experts, [273](#), [277–278](#)
Cybersecurity risks, [200](#), [216](#), [294](#)

D

Data accuracy, [5](#), [214](#), [385](#)
Data fiduciaries, [139–140](#), [142](#), [151](#)
Data management, [277](#), [295](#), [308](#), [389](#), [390](#)
Data protection, comprehensive, [64](#), [66](#), [323](#)
Democratic processes, [18](#), [21](#), [23](#), [25](#), [27](#), [62](#), [172](#), [174](#), [176](#), [179](#)
Democratic values, [18](#), [22](#), [25](#), [28](#), [172](#), [360](#)
Digital media, [109](#), [146](#), [148](#), [151](#)
Digital Services Act, [28](#), [233–234](#)

E

Education, artificial intelligence in, [95](#), [153](#), [289](#)
Effect of Digital Transformation on Banking Audits, [387–397](#)
Emotional intelligence (EI), [112–121](#), [124–125](#), [202](#), [372–384](#)
Ethical AI, [36–37](#), [58](#), [81](#), [86](#), [178](#), [369–370](#)
Ethical challenges, [78](#), [95](#), [106](#), [109](#), [111](#), [114](#), [365](#)
EU data protection, [306](#), [318](#), [323](#)
European Commission, [178](#), [317](#), [345](#), [360](#), [370](#)

F

Fairness, algorithmic, [14](#), [41](#), [152](#), [194](#), [336](#), [357](#)
Family Educational Rights Privacy Act (FERPA), [291](#), [294](#), [300–301](#)
Financial transactions, [35](#), [391–392](#)
Fraudulent activities, [386](#), [392](#)

Fraudulent transactions/cybercrime, [14](#), [42](#), [195](#)

Free speech, [349](#), [361–362](#), [364](#)

G

GDPR consent, [309](#), [311](#), [316](#)

Generative AI entities, [12](#), [39](#), [193](#)

Government in enhancing cybersecurity, [12](#), [39](#), [193](#), [334](#), [356](#)

H

Human interaction, [105](#), [208](#), [222](#), [296](#), [299](#)

Human intervention, [160](#), [242](#), [274](#), [292](#), [300](#), [366](#), [391](#)

I

Implications, [18](#), [20](#), [24](#), [61](#), [79](#), [82](#), [87](#), [92](#)

Intelligence, [112](#), [117](#), [164](#), [286](#), [339](#)

International, [1](#), [2](#), [7](#), [18](#), [183](#), [221](#), [339](#), [340](#), [345](#)

J

Job, [347](#), [348](#)

Judiciary, [67](#), [254](#), [351](#), [352](#)

K

Knowledge, [18](#), [31](#), [54](#), [102](#), [121](#), [245](#)

L

Law, [20](#), [31](#), [35](#), [37](#), [65](#), [85](#), [99](#), [105](#)

Lawyer, [20](#), [34](#), [98](#), [100](#), [104](#), [108](#)

Legal, [1](#), [3](#), [64](#), [79](#), [97](#), [99](#), [106](#), [108](#), [137](#), [173](#), [183](#), [238](#), [311](#)

M

Machine learning, [20](#), [26](#), [98](#), [130](#), [205](#), [210](#), [229](#), [239](#), [359](#), [366](#)

Motivation, [117](#), [204](#), [287](#), [377–379](#)

N

Nation, [127](#), [129](#), [130](#), [141](#)

Network, [9](#), [11](#), [18](#), [34](#), [40](#), [66](#), [128](#), [143](#), [198](#), [200](#), [204](#), [205](#), [264](#), [335](#)

O

Organization, [67](#), [70](#), [72](#), [75](#), [79](#), [95](#), [98](#), [114](#), [116](#), [123](#), [372](#), [376](#), [381](#)

P

Personal data, [23](#), [61](#), [72](#), [139](#), [153](#), [172](#), [179](#), [306](#), [310](#), [314](#)

Policy implementation, [34](#), [85](#), [289](#)

Practice, [22](#), [28](#), [32](#), [35](#), [36](#), [67](#), [70](#), [174](#), [183](#), [205](#), [221](#), [227](#), [289](#)

Privacy, [23](#), [32](#), [35](#), [37](#), [62](#), [64](#), [66](#), [75](#), [106](#), [140](#), [153](#), [179](#), [231](#), [309](#), [356](#), [361](#)

Q

Quality, [178](#), [210](#), [228](#), [238](#), [240](#), [277](#), [284](#), [296](#), [307](#)

R

Remedy, [27](#), [73](#), [146](#), [237](#)

Revolution, [51](#), [55](#), [57](#), [66](#), [97](#), [126](#), [135](#), [195](#), [202](#), [220](#), [228](#), [262](#), [277](#), [278](#), [350](#), [358](#)

Rights, [1](#), [2](#), [4](#), [5](#), [16](#), [22–27](#), [32](#), [33](#), [38](#), [44](#), [64](#), [66](#), [68](#), [87](#), [92](#), [93](#), [129](#), [164](#), [187](#), [232](#), [238](#), [239](#), [257](#), [260](#), [320](#), [341](#), [359](#)

S

Security, [68](#), [74](#), [75](#), [80](#), [106](#), [127](#), [130](#), [137](#), [139](#), [143](#), [169](#), [385–386](#)
System, [26](#), [28](#), [32](#), [37](#), [48](#), [49](#), [58](#), [81](#), [85](#), [87](#), [90](#), [106](#), [157](#), [172](#), [194–195](#)

I

Technology, [35](#), [37](#), [50](#), [55](#), [57](#), [64](#), [65](#), [79](#), [180](#), [204](#), [232](#), [292](#), [342](#), [373](#)
Transparency, [36](#), [37](#), [41](#), [58](#), [61](#), [73](#), [138](#), [140](#), [152](#), [172](#), [232](#), [235](#), [241](#), [292](#),
[300](#)

U

Understanding, [18](#), [22](#), [28](#), [64](#), [65](#), [98](#), [103](#), [162](#), [172](#), [219](#), [270](#), [272](#), [310](#)
UNESCO, [78](#), [80–82](#), [85](#), [89–90](#), [95](#)

V

Value, [91](#), [102](#), [108](#), [116](#), [122](#), [127](#), [172](#), [179](#), [211](#), [250](#), [343–344](#)

W

World, [84](#), [87–89](#), [100](#), [119](#), [126](#), [132](#), [135](#), [142](#), [155](#), [169](#), [185](#), [193](#), [204](#),
[249](#), [291](#), [320](#)

X

XAI, [37](#), [217](#), [279](#)

Y

Year, [127](#), [138](#), [184](#), [197](#), [263–264](#), [311](#), [324](#)