Routledge Studies in Trust Research

BUILDING TRUST IN THE GENERATIVE ARTIFICIAL INTELLIGENCE ERA

TECHNOLOGY CHALLENGES AND INNOVATIONS

Edited by
Joanna Paliszkiewicz, Jerzy Gołuchowski,
Magdalena Mądra-Sawicka, and Kuanchin Chen

ROUTLEDGE

Building Trust in the Generative Artificial Intelligence Era

In an era where generative artificial intelligence (AI) is reshaping industries and daily life, trust has become a cornerstone for its successful adoption and application. *Building Trust in the Generative Artificial Intelligence Era: Technology Challenges and Innovations* explores how trust can be built, maintained, and evaluated in a world increasingly reliant on AI technologies. Designed to be accessible to a broad audience, this book blends theoretical insights with practical approaches, offering readers a comprehensive understanding of the topic.

This book is divided into three parts. The first part examines the foundations of trust in generative AI, highlighting trends and ethical challenges such as "greenwashing" and remote work dynamics. The second part provides actionable frameworks and tools for assessing and enhancing trust, focusing on topics like cybersecurity, transparency, and explainability. The final section presents global case studies exploring university students' perceptions of ChatGPT, generative AI's applications in European agriculture, and its transformative impact on financial systems.

By addressing both the opportunities and risks of generative AI, this book delivers groundbreaking insights for academics, professionals, and policymakers worldwide. It emphasizes practical solutions, ensuring readers gain the knowledge needed to navigate the evolving technological landscape and foster trust in transformative AI systems.

Joanna Paliszkiewicz works as a Full Professor at the Warsaw University of Life Sciences (WULS—SGGW). She is the Director of the Management Institute. She also works for University of Economics Ho Chi Minh City (UEH), Vietnam. She is also an Adjunct Professor at the University of Vaasa in Finland. She obtained academic title "Full Professor" from the International School for Social and Business Studies in Slovenia. She is well recognized in Poland and abroad for her expertise in management issues: knowledge management and trust management. She has published over 220 papers/manuscripts and is the author/co-author/editor of 21 books.

Jerzy Gołuchowski is a Full Professor at the University of Economics in Katowice. He has served as a Vice-Rector (2005–2009) and Dean (2009–2019; 2020–now) of the Faculty of Informatics and Communications and he is the Head of the Department of Communication Design and Analysis (since 2017).

Magdalena Mądra-Sawicka is an Assistant Professor at Warsaw University of Life Sciences (WULS) in the Institute of Economics and Finance. She obtained PhD degree in 2010 and started working at SGGW WULS in 2010 at the Department of Economics and Organization of Enterprises and later in the Department of Finance at the Faculty of Economic Sciences.

Kuanchin Chen is Professor of Computer Information Systems, Director of the Center for Business Analytics, and John W. Snyder Fellow at Western Michigan University. He has more than 25 years of research, teaching, and consulting experience.

Routledge Studies in Trust Research
Series editors: Joanna Paliszkiewicz and Kuanchin Chen

Available Titles in this Series:

Trust and Brand Management
The Role of Brand Heritage
Hanna Górska-Warsewicz

Trust and Supply Chains
Information Asymmetry in the Agri-Food Sector
Edited by Michał Pietrzak, Joanna Domagała and Aleksandra Chlebicka

Trust and Digital Transformation in the Public Sector
Barbara Kożuch and Katarzyna Sienkiewicz-Małyjurek

Trust, Media and the Economy
Mutual Relations
Edited by Joanna Paliszkiewicz, Jerzy Gołuchowski, Katarzyna Zdanowicz Cyganiak and Kuanchin Chen

Trust in Generative Artificial Intelligence
Human-Robot Interaction and Ethical Considerations
Edited by Joanna Paliszkiewicz, Ireneusz Dąbrowski and Leila Halawi

Trust, Sustainability, and Resilience
Management and Consumer Perspectives
Edited by Kuanchin Chen and Piotr Pietrzak

Building Trust in the Generative Artificial Intelligence Era
Technology Challenges and Innovations
Edited by Joanna Paliszkiewicz, Jerzy Gołuchowski,
Magdalena Mądra-Sawicka, and Kuanchin Chen

For more information about this series, please visit: www.routledge.com/Routledge-Handbooks-in-Religion/book-series

Building Trust in the Generative Artificial Intelligence Era

Technology Challenges and Innovations

**Edited by Joanna Paliszkiewicz,
Jerzy Gołuchowski,
Magdalena Mądra-Sawicka,
and Kuanchin Chen**

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2025
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging-in-Publication Data
Names: Paliszkiewicz, Joanna Olga editor
Title: Building trust in the generative artificial intelligence era : technology challenges and innovations / edited by Joanna Paliszkiewicz, Jerzy Gołuchowski, Magdalena Mądra-Sawicka, and Kuanchin Chen.
Description: New York, NY : Routledge, 2025. |
Series: Routledge studies in trust research | Includes bibliographical references and index.
Identifiers: LCCN 2025005359 | ISBN 9781032957999 hardback | ISBN 9781032958774 paperback | ISBN 9781003586944 ebook
Subjects: LCSH: Technological innovations—Management | Technological innovations—Moral and ethical aspects | Artificial intelligence | Trust
Classification: LCC HD45 .B7936 2025 | DDC 658.5/14—dc23/eng/20250425
LC record available at https://lccn.loc.gov/2025005359

ISBN: 978-1-032-95799-9 (hbk)
ISBN: 978-1-032-95877-4 (pbk)
ISBN: 978-1-003-58694-4 (ebk)

DOI: 10.4324/9781003586944

Typeset in Times New Roman
by codeMantra

To my friends, Dominika, Kasia, Lucyna, Magda, Marzena, and my family, thank you for your continuous support and inspiration.
JP

To my wife, Regina
JG

To my parents, whose constant support, love, and faith have been my greatest strength and motivation.
MM

To my wife Cathy and children for your inspiration, support, and love.
KC

Contents

Figures

Tables

Biographical Notes

Editors

Joanna Paliszkiewicz works as a Full Professor at the Warsaw University of Life Sciences (WULS—SGGW). She is the Director of the Management Institute. She is a Professor at the University of Economics in Ho Chi Minh City, Vietnam. She is also an Adjunct Professor at the University of Vaasa in Finland. She obtained the academic title "Full Professor" from the International School for Social and Business Studies in Slovenia. She is well recognized in Poland and abroad for her expertise in management issues: knowledge management and trust management. She has published over 250 papers/manuscripts and is the author/co-author/editor of 21 books. She has been a part of many scholarship endeavors in the United States, Ireland, Slovakia, Taiwan, the United Kingdom, and Hungary. She has actively participated in presenting research results at various international conferences. Currently, she serves as the Deputy Editor-in-Chief of the *Management and Production Engineering Review*. She is an associate editor for the *Journal of Computer Information Systems, Expert System with Applications*, *Issues in Information Systems, Przegląd Organizacji,* and *Intelligent Systems with Applications*. She serves as a Chair of the International Cooperation in European Business Club. She serves as the President of the International Association for Computer Information Systems in the United States. She has successfully supervised many PhD students, leading them to the completion of their degrees. She has also served as an external reviewer for several PhD students in Poland, India, and Finland. She is actively involved in participating in the scientific committees of many international conferences. She was named the 2013 Computer Educator of the Year by IACIS in the United States. She is an expert in the Polish Accreditation Commission and is a member of the Polish Academy of Science.

Jerzy Gołuchowski is a Full Professor at the University of Economics in Katowice. He has served as a Vice-Rector (2005–2009) and Dean of the Faculty of Informatics and Communications and he is the Head of the

Department of Communication Design and Analysis (since 2017). His scientific interests are concentrated on communication issues in decision-making processes, artificial intelligence, and the management of organizations, particularly the management of trust, knowledge, and communication in organizations. His publications include B. Filipczyk and J. Gołuchowski, *Cyfrowa Komunikacja Organizacji* (2020), and J. Liebowitz, J. Paliszkiewicz, and J. Gołuchowski (eds.), *Intuition, Trust, and Analytics* (Taylor & Francis, 2018). He has participated in international projects such as the "CALYPSO-Collaborative Fact-Checking" (2021–2022), "European Cooperation Project on Disinformation and Fact-Checking Training" (2019–2022), "Language Communication on eCommerce Websites" (2004–2007), and "European Master in e-Commerce and Cyber Communication" (2000–2003). He is a member of the Society for Social Communication and the Scientific Society for Business Informatics. He has also served as the President of the Scientific Society for Business Informatics and as a member of the Council of Scientific Societies of the Polish Academy of Sciences. From 2012 to 2016, he was a member of the Programme Council of TVP o/ Katowice, and from 2014 to 2016, a member of the Supervisory Board of Polish Radio Katowice.

Magdalena Mądra-Sawicka is an Assistant Professor at Warsaw University of Life Sciences (WULS) in the Institute of Economics and Finance. She obtained PhD degree in 2010 and started working at SGGW WULS in 2010 in the Department of Economics and Organization of Enterprises and later in the Department of Finance at the Faculty of Economic Sciences. Since 2019, she has been employed at the newly established Institute of Economics and Finance at SGGW in Warsaw. Since 2016, she completed two international internships at Budapest Business School and the University of Economics in Bratislava. She conducted research with the University of Vaasa in Finland and participated in three research projects concerned with Affordable Housing. Additionally, in 2017–2018, she participated in a research and development project titled "Inside Expert – development of methods and algorithms for codifying tacit knowledge in organizations and implementing them into an IT system supporting informal knowledge sharing." She also participated in many projects such as *SmartVillage project* (2023), the *Excellence in Teaching* (2023), and as part of the POWER project "*Success from Nature.*" She was a member of the team implementing the Syllabus system (2023). She actively participates in numerous national and international conferences, where she presents the results of her research. She is a reviewer of scientific articles at the national and international levels and two doctoral theses in Finland (University of Vaasa). Furthermore, in 2016, she received the second-degree team award from the Rector of SGGW in Warsaw for scientific achievements and a diploma of recognition for organizational activities. In 2020, she received the third-degree team award for organizational achievements. In 2021, she received the Bronze Medal for Long Service. In 2022, she received the first-degree team award for organizational achievements. In 2023, she received the first-degree team award for organizational achievements. Since 2016, she has been chairing the *Mazowieckie Polish Society for Production Management*. Since 2020, she has been the quality coordinator at the Faculty of Economics at Warsaw University of Life Sciences, Poland. Since 2019, she has been appointed to the Discipline Council for Economics and Finance at the Institute of Economics and Finance at SGGW in Warsaw. In the Department of Finance, Institute of Economics and Finance, she is the Head of the Department of Corporate Finance and Accounting. Currently, her scientific and research activities focus mainly on issues related to corporate financial management, including capital structure, profit distribution, and management of internal financing sources and AI, GenAI, in corporate and institutional finance. She is the author of over 130 scientific publications, including four books (three of which she edited). From 2014 to 2017, she served as a Deputy Editor-in-Chief of the *Journal of Corporate Finance Management and Accounting*.

Kuanchin Chen is a Full Professor of Computer Information Systems at Western Michigan University. He is the Director of WMU's Center for Business Analytics and also an associate editor of the *Behaviour & Information Technology* journal. His research interests include electronic business, analytics, social networking, project management, privacy, security and trust, online behavioral issues, data mining, and human-computer interactions. He has published in journals, such as *Information Systems Journal*, *Decision Support Systems*, *Information & Management*, *IEEE Transactions on Systems, Man, and Cybernetics*, *International Journal of Information Management*, *Journal of Database Management*, *Internet Research*, *Communications of the Association for Information Systems*, *Electronic Commerce Research and Applications*, *Journal of Global Information Management*, *DATA BASE for Advances in Information Systems*, *IEEE Transactions on Education*, *Decision Sciences Journal of Innovative Education*, *International Journal of Medical Informatics*, and *Journal of Computer Information Systems*. He has been an editor, associate editor, and editorial member of several scholarly journals. He is also the recipient of several research and teaching awards, including awards given by scholarly journals and conferences, departments, colleges, universities, and US Fulbright program. He has frequently been invited to present research talks at universities, government agencies, and other institutions.

Contributors

H. Nejat Basım is a Professor of Management at Baskent University in Turkey. He has articles concerning organizational citizenship, organizational learning, resilience, positive psychology, and conflict resolution published in *the European Management Review*, *the Journal of Computer Information Systems*, and *the Journal of Advanced Nursing.*

Michał Borowy, PhD**,** Economist, is an expert in industry innovation and an experienced project manager (with a proven track record of working in the research sector) and skilled in R&D business relationship management, change management, planning, and innovation management. Within 15 years of professional experience, his key achievements relate to technology transfer in life sciences, Spin-off Academy coordination at the University of Warsaw, and mentoring in acceleration programs, like Digital Health, EIT Manufacturing, or EIT Digital. He acts as an innovation coach for the industry sector within the Polish Ministry of Funds and Regional Development program. In the years 2020–2023, he serves as a member of the Board at the Polish Business and Innovation Centres Association. He developed, supervised, and managed projects related to business and science cooperation, technology audits, research, business offers, and implementation works.

Fatih Çetin is a Professor of Management in the Department of Business Administration at Baskent University in Turkey. His research focuses on the field of organizational behavior and knowledge management studies. His recent research has appeared in the *European Journal of International Management*, *International Business Review*, and *Decision Analytics Journal.*

Magdaléna Freňáková works as an Assistant Professor in the Department of Economics and Management at the Faculty of Business Economics of the University of Economics in Bratislava with a seat in Košice. She is a member of the Slovak Association of Finance and Treasury. Her research area focuses on financial and strategic management, financial performance, and intra-corporate controlling. She also deals with gender stereotypes, poverty issues, venture capital, and corporate social responsibility. She has published over 50 scientific papers and participated in several scientific research projects and international scientific conferences.

Marzena Ganc is an Assistant Professor at Warsaw University of Life Sciences – SGGW in the Institute of Economics and Finance. She obtained PhD degree in 2009. The doctoral thesis topic was the use and usefulness of cost accounting for making management decisions in dairy cooperatives. She is an author of over 100 peer-reviewed original scientific articles in corporate finance, financial and managerial accounting, and audit. She works in practice as an auditor of financial statements. She is a lead auditor of the organization's ICA, a reviewer of scientific articles in Poland and abroad, and an expert at the National Centre for Research and Development.

Jozefína Hvastová is an Associate Professor in the Department of Economics and Management at the Faculty of Business Economics of the University of Economics in Bratislava with a seat in Košice. She is a member of the Slovak Chamber of Certified Accountants. Her research is related to corporate social responsibility, application of ethical standards, and accounting information quality. She specializes in the integration of business management and ethical perspectives, which impact business decision-making. She has published over 60 scientific papers. She participates in project activities aimed at the creation of an education model to increase the skills of students in the field of using artificial intelligence tools.

Alexandra Ioanid, a doctor in Industrial Engineering with the distinction "Summa cum laude" in Social Media Business Strategies (2017), holds two Master's degrees in Business Administration – MBA (2011) and International Relations (2010) and has a BSc degree in Computer Science and Information Technology (2009) and Post-doctoral studies in Social Media and Neuromarketing. Currently, she is an Associate professor at the National University of Science and Technology Politehnica Bucharest, Faculty of Entrepreneurship, Business Engineering and Management, Department of Entrepreneurship and Management; she is teaching subjects such as marketing, entrepreneurship, project management, and business communication. She has published five textbooks and book chapters. Her research interest is in neuromarketing, social media, and entrepreneurship. She is active in internal and international education and research projects. Besides textbooks, she published more than 60 scientific reports and articles in the areas of research interest.

Rafał Kasprzak is a researcher and academic teacher at the Warsaw School of Economics in Warsaw, working at the Institute of Value Management, is the Head of Value-Based Marketing Unit, and an author of many research projects and economics devoted to the issue of innovation and stimulating practical cooperation between science and business. He has professional experience in the areas of business consulting for organizations implementing innovations and in the management of innovation projects. Scientifically, he is passionate about the issues of creative industries (cultural and creative industries), their relationship with innovation, and the measurement of culture's impact on the region's economy. His publication output focuses on Creativity, Design Thinking, Virtual Reality, Metaverse, and the determinants of Cultural and Creative Industries. He is an advisor and certified trainer in raising public funds and author of numerous project applications. He graduated with a Master of Business Administration from the Ecole Superieure de Commerce de Rennes (France).

Elżbieta Kacperska is an adjunct at the Institute of Economics and Finance of the Warsaw University of Life Sciences. Her research interests focus on international economic issues, including foreign direct investment, international migration, international management, European integration, and foreign trade. She is author of many publications in this field, as well as the supervisor of many bachelor's and master's theses in the field of Economics and Finance and Accounting, a member of the Editorial Board of the *Scientific Journal of the Warsaw University of Life Sciences* (SGGW), Problems of World Agriculture, coordinator of the Academy of Entrepreneurship at the Faculty of Economics at SGGW, coordinator for the Promotion and Cooperation with High Schools of the Faculty of Economics at SGGW.

Damian Kedziora works as an Associate Professor in the Software Engineering Department of LUT University in Lahti, Finland. He is also a Docent in Business Process Automation Software at LUT University in Finland and an Adjunct Professor at Kozminski University in Poland. His research concentrates on the low-code software, process automation, and commercial aspects of software business.

Paweł Kędziora works at the Administration and Administrative Law Department at Kozminski University in Poland. His research interests comprise the impact of technology and artificial intelligence on administrative proceedings and the digital transformation of legal processes.

Jakub Kraciuk is a Professor at the Warsaw University of Life Sciences (SGGW) and is employed at the Institute of Economics and Finance of this University. His research interests focus on the issues of international economics, with particular emphasis on economic globalization and its effects, such as concentration processes taking place in the world economy, as well as the processes of fragmentation and relocation of production. In his research work, he also dealt with the activities of transnational corporations, financial crises, and their impact on the world economy and the economies of individual countries, as well as the flow of goods, services, and production factors on an international scale. He is the author of many publications in this field, as well as the promoter of three defended doctoral theses and many bachelor's and master's theses in the fields of Economics, Finance, and Accounting. He is a member of the Editorial Committee of Scientific Journals of the Warsaw University of Life Sciences, Problems of World Agriculture. He was the Deputy Dean for Science and the Dean of the Faculty of Economics at the Warsaw University of Life Sciences. He is the Head of the Department of International Economics and Agribusiness. He was also the coordinator of the research of the members of the rural elite in Poland, carried out under the 6th Framework Program of the European Union, carried out by the Institute of Agricultural Economics at the University of Kiel.

Paulina Kramarz holds a PhD in Economics. She received her PhD at the Faculty of Economic Sciences of the Warsaw University of Life Sciences. Her specialization is focused on issues such as the economics of sustainable development, rural development, and health economics. Her scientific interests are related to the study of regions with specific conditions for economic development, the use of statistical and econometric methods in the study of regional differentiation of socio-economic development, the relationship between the level of the population's health resources and the directions of economic development. Currently, she is employed in the Department of Economics and Organization of Enterprises – Institute of Economics and Finance, Warsaw University of Life Sciences.

Jay Liebowitz is a Professor of Business Innovation and Industry Transformation and Director of a new AI-EDGE Center in the Crummer Graduate School of Business at Rollins College as of August 2024. He has recently served as the inaugural Executive-in-Residence for Public Service at Columbia University's Data Science Institute. He was previously a Visiting Professor at the Stillman School of Business and the MS-Business Analytics Capstone & Co-Program Director (External Relations) at Seton Hall University. He previously served as the Distinguished Applied Business and Finance Chair at Harrisburg University of Science and Technology. Before HU, he was the Orkand Endowed Chair of Management and Technology at the University of Maryland University College (UMUC) Graduate School. He served as a Full Professor in the Carey Business School at Johns Hopkins University. He was ranked one of the top ten knowledge management researchers/practitioners out of 11,000 worldwide and was ranked Number 2 in KM Strategy worldwide, according to the January 2010 *Journal of Knowledge Management*. At Johns Hopkins University, he was the Founding Program Director for the Graduate Certificate in Competitive Intelligence and the Capstone Director of the MS-Information and Telecommunications Systems for Business Program, where he engaged over 30 organizations in industry, government, and not-for-profits in capstone projects. Prior to joining Hopkins, he was the first Knowledge Management Officer at NASA Goddard Space Flight Center. Before NASA, he was the Robert W. Deutsch Distinguished Professor of Information Systems at the University of Maryland-Baltimore County, Professor of Management Science at George Washington University, and Chair of Artificial Intelligence at the US Army War College. He is the Founding Editor-in-Chief of *Expert Systems With Applications: An International Journal* (published by Elsevier; ranked as a top-tier journal; Thomson Impact Factor from June 2021 is 8.665). He is a Fulbright Scholar, IEEE-USA Federal Communications Commission Executive Fellow, and Computer Educator of the Year (International Association for Computer Information Systems). He has published over 45 books and a myriad of journal articles on knowledge management, analytics, financial literacy, intelligent systems, and IT management. He served as the Editor-in-Chief of Procedia-CS (Elsevier). He is also the Series Book Editor of the *Data Analytics Applications* book series (Taylor & Francis), as well as the Series Book Editor of the new *Digital Transformation: Accelerating Organizational Intelligence* book series. In October 2011, the International Association for Computer Information Systems established the "Jay Liebowitz Outstanding Student Research Award" to recognize the best student research paper presented at the IACIS Annual Conference. He was the Fulbright Visiting Research Chair in Business at Queen's University for the Summer of 2017 and a Fulbright Specialist at Dalarna University in Sweden in May 2019. He is in the Top 2% of the top scientists in the world, according to a 2019 Stanford Study. As of 2021, he is the Visiting Distinguished Professor at the International School for Social and Business Studies in Slovenia. His recent books are *Data Analytics and AI* (Taylor & Francis, 2021), *The Business of Pandemics: The COVID-19 Story* (Taylor & Francis, 2021), *A Research Agenda for Knowledge Management and Analytics* (2021), *Online Learning Analytics* (Taylor & Francis, 2022), *Digital Transformation for the University of the Future* (2022), and *Cryptocurrency Concepts, Technology, and Applications* (Taylor & Francis, April 2023), *Pivoting Government Through Digital Transformation* (Taylor & Francis, 2024), and *Developing the Intuitive Executive: Using Analytics and Intuition for Success* (Taylor & Francis, 2024). His newest book, *Regulating Hate Speech Created by Generative AI*, was published by Taylor & Francis in August 2024. He has lectured and consulted worldwide.

Katarzyna Łukasiewicz, PhD, Bachelor of Science in Engineering, is an Assistant Professor at the Institute of Management at the Warsaw University of Life Sciences. Research interests focused on the issues of tourism, management, quality management, trust, and customer relationship management. She is an author of many publications in this field, as well as the supervisor of many bachelor's and master's theses in Management and Tourism and Recreation (two of them awarded with the Minister of Sport and Tourism prize in 2009 and 2013), a secretary of the Scientific Discipline Board for Management and Quality Sciences, and an animal lover in everyday life.

Dorota Marquardt, PhD, conducts research in the field of media communication, computational linguistics, media discourse analysis, and new media. She wrote many publications about linguistic factors in chatbot dialogs, new media, and identity discourse. She is also a member of the Polish Communication Association.

Aneta Mikuła is an Assistant Professor in the Department of Economics and Economic Policy at the Institute of Economics and Finance at Warsaw University of Life Sciences – SGGW, Poland. She is a member of the Polish

Association of Agricultural and Agribusiness Economists. Her research activities focus mainly on issues related to sustainable development, changes in the agri-food sector, and digitalization. She has presented her research at many national and international scientific conferences. She is the author of over 60 scientific publications.

Giovanni Miragliotta (professor) graduated in Management Engineering, cum laude, in 1998 at Politecnico di Milano with a master's thesis about "Strategic Stocks Sizing through Uncertainty Deployment." After a visiting period at UCLA, he got his PhD in Management Engineering in 2003 from Politecnico di Milano, with a PhD dissertation about "Techniques and tools for multisite production management." He is a Full Professor, teaching Industrial Plants Design and Advanced Supply Chain Planning, respectively, to BS and MS students. He is a core faculty member of Polimi GSOM, co-director of three Observatories (Internet of Things, Industry 4.0, Artificial Intelligence), and co-director of the IoTL.

Małgorzata Wiktoria Paprocka graduated in Philosophy, Finance and Banking. Currently, she is continuing her studies at the Doctoral School of WULS in Warsaw in the discipline of Economics and Finance. She is professionally involved in banking since 2019, in the area of ESG risk and sustainable finance and interested in business ethics, sustainable finances, and new technologies.

Alexandru-Ionuţ Petrişor was born in Bucharest (1975), is an ecologist (BS, University of Bucharest, 1997; PhD, University of South Carolina, 2004), holds a Master's degree in Biostatistics (University of South Carolina, 2000), a second PhD in Geography (University of Bucharest, 2011), a Certificate Degree in Project Management (Bucharest University of Economic Studies, 2012), and a habilitation in urban planning (2014). Currently, he is a senior scientific researcher ranked I in biology (2013) and geography (2017), Full Professor, and doctoral advisor, coordinates over 30 journals, and has reviewed for over 100. His experience, including management positions, is mainly in research and academia. His work in over 70 research grants yielded over 700 publications (including over 300 articles in over 100 journals); over 500 presentations, lectures, and courses; and parts of over 35 urban/spatial plans, recognized internationally by over 2100 citations (including over 1700 citations of articles in over 550 journals), over 130 awards and distinctions, international memberships, including Sigma Xi. His h-index is 18 in the WoS, 16 in Scopus, 23 in Research Gate, and 28 in Google Scholar.

Casey Phillips is an accomplished Senior Technical Product Manager with over 10 years of experience, including 7+ years focused on AI/ML. This expertise spans Conversational AI, Natural Language Processing, and Generative AI, including work with Large Language Models. He has successfully launched

AI experiences in prominent products such as TurboTax, QuickBooks, eBay, and Automatic Data Processing (ADP) and also led innovative ML-based initiatives at Uber to enhance the quality of rider-to-driver matches at airports and large events.

With a strong technical background, he excels at translating complex technical concepts into clear product requirements. He holds a degree in Computer Information Systems from Western Michigan University and an MBA from Northern Michigan University. His passion lies in leveraging AI/ML to build products that improve lives and delight users, with a particular focus on ensuring transparency and ethical considerations in AI applications.

Piotr Pietrzak, PhD, is an Adjunct Professor at the Management Institute, Warsaw University of Life Sciences (SGGW), Poland. His research interests include several topics: strategic management, process management, sustainable development, and digital management. He is the author of over 100 articles and five books. He has published work in journals such as the *International Journal of Innovation and Learning*, *Human Systems Management*, *Foundations of Management*, and *Energies*. He is currently Head of the Rector's Rankings Committee.

Henryk Runowski is a Professor of Economics. He specializes in the issues of agricultural economics, particularly in the organization and economics of enterprises and farms. He is the Head of the Department of Economics and Organization of Enterprises at the Institute of Economics and Finance of the Warsaw University of Life Sciences. His main scientific interests include adjustment processes in farms, issues related to the measurement and assessment of agricultural income, and the dissemination of progress and innovation in agriculture. He conducted dozens of research and training projects for farmers, financed by the European Union and the state budget. He was an expert of the Senate of the Republic of Poland and President of the Association of Agricultural and Agribusiness Economists. Currently, he is the President of the Scientific Council of the Institute of Rural Development and Agriculture of the Polish Academy of Sciences, a member of the Scientific Council of the Agricultural and Food Economics, and a member of the Economic Sciences Committee of the Polish Academy of Sciences. He is the author or co-author of about 400 publications, including monographs. He has implemented several dozen projects of enterprise restructuring, recovery programs, and business plans. The main hobby is the restoration of rural monuments.

Cezar Scarlat is a Professor in the Department of Entrepreneurship and Management and PhD Supervisor within the Doctoral School of Entrepreneurship, Business Engineering & Management at the National University of Science and Technology POLITEHNICA of Bucharest (UPB).He is teaching Technology Entrepreneurship Business Development and Strategy, Project

Management, Marketing as well as related courses at universities in Romania and abroad (Finland, Portugal, Belgium) – as a Visiting/Exchange Professor, a Director of the Center for Business Excellence from UPB for almost two decades (1991–2009), and currently running a private consulting firm. He also acquired project management expertise by participating in over 40 international programs or projects funded by various donors (EU, World Bank, IBRD, USAID, British Know How Fund, GTZ a.o.). The author has two Master's degrees: Electronics Engineering from UPB – Faculty of Electronics and Telecommunication and International Business from the Academy for Economic Studies from Bucharest – Faculty of International Business. He got his PhD degree in Management and Industrial Engineering at UPB. International exposure by postgraduate studies in Romania and abroad (United States, Canada, France, Italy); research and study tours in Europe and North America. Publications: 35 books and manuals; about 200 published articles and scientific reports on technology entrepreneurship, entrepreneurial education, management and administration, project management, business development, marketing, and related areas. Over the years, he has membership in numerous international bodies as well as being a member of scientific/editorial boards of many international journals – among them are *Industrial Management & Data Systems* (IMDS), *Expert Systems with Applications* (ESWA), and *Advances in Related Business Scientific Research Journal* (ABSRJ).

Edyta Skarzyńska is an assistant professor at the Institute of Management at the Warsaw University of Life Sciences. Her scientific interests focus on topics related to intellectual capital, social media, and the use of artificial intelligence and trust in the field.

Monika Utzig is a lecturer and an Assistant Professor in the Department of Economics and Economic Policy at the Institute of Economics and Finance (former Faculty of Economic Sciences) at Warsaw University of Life Sciences since 2006. She has completed her Master's degree in the Faculty of Economic Sciences at Warsaw University and finished her doctoral studies in Economics at the Warsaw University of Life Sciences. Since 2020, she has been Associate Dean of the Faculty of Economics at WULS. Her areas of interest are household consumption and finance, sustainable development, the standard of living, and digitalization. She has presented her research at many national and international conferences and is an author of over 40 scientific publications.

Anezka Viskova-Robertson is a doctoral researcher focused on AI interaction and workplace well-being. Her research centers on applying socio-technical system theory to understand AI integration better and optimize it in organizations. She has a passion for innovation and system alignment and has written several papers on these topics. Her work explores how trust and integration within socio-technical systems can enhance the harmonious alignment of technology and human factors.

Hakkı Okan Yeloğlu was born in Ankara, Turkey. He earned his undergraduate degree in Statistics from the Faculty of Science at Anadolu University. He then completed a Master's degree in Business Administration at the Graduate School of Social Sciences, Başkent University. He subsequently obtained a PhD in Management and Organization Studies, with his doctoral research focusing on brokerage firms within Turkish capital markets and organizational ecology. His teaching portfolio encompasses technology management, knowledge management, strategic management, research methods, quantitative research methods, and statistics across undergraduate, graduate, and postgraduate levels. He has authored numerous books, peer-reviewed papers, and scholarly articles in these fields. His academic work reflects his expertise and ongoing contributions to technology and knowledge management. Additionally, he is fluent in English and has foundational German knowledge.

Marta Ziółkowska PhD is an Assistant Professor at the Institute of Management, Chair of Sustainable Business, SGH Warsaw School of Economics (Poland). She has a PhD in Management from SGH Warsaw School of Economics in 2009. She is an expert with 15 years of professional experience in franchising, small and medium enterprise management, internationalization of SMEs, design thinking, modern business models, and social entrepreneurship. She has gained international experience during many academic development programs, scholarships, and visiting professorships in several countries (New Zealand, United States, Malaysia, China, Kenya, Senegal, Canada, Iceland, Croatia, Portugal, France, Slovenia, Hungary, Norway, Finland, Jordan, UAE). Marta Ziółkowska has published numerous books, papers in peer-reviewed journals, working papers, and conference proceedings and participated in international research teams. She is a Polish ambassador of MAG Scholar Marketing in Asia Group and a mentor for the J&M Global Management Students' Association. She conducts the following courses for students: modern retail, management challenges, management in international business, modern business models, franchising business, cooperative strategies of enterprises, creative problem solving, design thinking, and lectures in the virtual world of Second Life. She is the Head of Postgraduate Studies in Franchise Management and Postgraduate Studies in Corporate Management in Retail.

Forewords

The advent of generative artificial intelligence (AI) is widely regarded as marking the onset of a transformative era that is redefining how we think, create, and innovate. For the first time in human history, machines have the ability not only to mimic human thought processes, but also to create and innovate on a previously unimaginable scale. However, like any disruptive technology, generative AI brings profound challenges, with the imperative of trust being the foremost among them. In a world increasingly shaped by AI systems that operate with remarkable autonomy and influence, trust is not a luxury, but an essential foundation.

Building Trust in the Era of Generative Artificial Intelligence: Technology Challenges and Innovations addresses this critical need with exceptional depth, clarity, and vision. Structured to guide readers from fundamental perspectives to actionable frameworks and real-world applications, this book provides a holistic understanding of trust in the AI era. Each chapter contributes to a broader narrative, highlighting the importance of ethical transparency, cybersecurity, informed decision-making, and the societal impact of AI adoption.

As someone who has spent years exploring the complex relationship between technology and society, I am deeply impressed by the insights and methodologies presented in this work. The editors and contributors have created a volume that not only bridges the gap between theories and tangible solutions but also serves as an indispensable guide to navigating the complexities of trust in generative AI.

This book is more than a reflection on the challenges of our time; it is a roadmap for the future. By equipping readers with cutting-edge research, practical tools, and compelling case studies, it ensures that trust becomes the cornerstone of this rapidly advancing technology. For academics, practitioners, and policymakers alike, this work is an invaluable resource for shaping an AI-driven world in a responsible, inclusive, and equitable way.

I am confident that this book will spark meaningful dialogue and inspire action to create a future where AI serves humanity responsibly, fulfilling its potential for innovation while honoring the values that unite us.

Prof. Seweryn Spałek
Silesian University of Technology
Forewords (cont.)

Building Trust in the Generative Artificial Intelligence Era: Technology Challenges and Innovations addresses one of the most pressing questions of our time: how to cultivate trust in generative AI systems that redefine creativity, decision-making, and ethics. As generative AI reshapes industries from education to government to business, this book offers a timely exploration of the challenges and opportunities associated with these transformative technologies.

Bringing together experts from diverse fields, this work provides a multidisciplinary perspective on the ethical, societal, and technical dimensions of trust in AI. Combining theoretical insights with practical case studies, it serves as a comprehensive guide for practitioners, academics, and policymakers navigating this complex landscape.

Amidst the promises and challenges of generative AI, this book equips readers with the tools to critically assess its implications and promote responsible innovation. It is an essential resource for anyone committed to advancing AI while ensuring its ethical and trustworthy development.

Ireneusz Dąbrowski, prof. SGH
SGH Warsaw School of Economics

Preface

In the rapidly evolving landscape of generative artificial intelligence (AI), trust has emerged as a critical factor shaping the adoption, utilization, and ethical implications of this transformative technology. As generative AI continues to influence various facets of society – from education and research to industry and governance – the need for robust frameworks and methodologies to build and sustain trust becomes ever more paramount. This book, *Building Trust in the Generative Artificial Intelligence Era: Technology Challenges and Innovations*, provides a timely and comprehensive exploration of these issues, offering insights from leading scholars and practitioners across the globe.

The chapters within this volume are organized into three sections. The first part lays the groundwork by examining the foundations of trust-building in the generative AI era, with a focus on emerging trends, technological challenges, and ethical considerations. Topics such as greenwashing, remote work dynamics, and the innovative potential of AI are explored to provide an understanding of the socio-technical relationships.

In the second part, readers will find practical frameworks and methodologies aimed at enhancing trust in AI systems. Contributions in this section delve into transparency, explainability, cybersecurity, and tools for evaluating confidence in generative models. These discussions are critical for ensuring that technological advancements align with ethical principles and societal needs.

The third and final part of this book presents case studies and comparative analyses, showcasing the real-world applications and implications of generative AI across diverse contexts. From university students' perceptions of ChatGPT to AI's role in agriculture and finance, these case studies illustrate the multifaceted challenges and opportunities that arise when fostering trust in generative AI technologies.

The contributions to this book reflect a multidisciplinary approach, bringing together expertise from fields such as economics, ethics, computer science, finance, and management. This diversity of perspectives underscores the complexity of trust-building in the generative AI era and highlights the collaborative effort required to address these challenges effectively.

Whether you are a technologist, academic, business leader, or policymaker, this book aims to equip you with the tools and insights needed to navigate the intricate relationship between trust and technology in the age of generative AI. By fostering transparency, prioritizing ethical practices, and developing actionable frameworks, we can ensure that generative AI serves as a force for societal good.

Joanna Paliszkiewicz, Warsaw University of Life Sciences, Poland,
University of Economics Ho Chi Minh City (UEH), Vietnam
Jerzy Gołuchowski, University of Economics
in Katowice, Poland
Magdalena Mądra-Sawicka, Warsaw University of
Life Sciences, Poland
Kuanchin Chen, Western Michigan University, USA

Part I

Foundations of Trust-Building in the Generative Artificial Intelligence Era

1 Trust-Building in the Generative AI – Future Perspectives and Emerging Trends

Magdalena Mądra-Sawicka, Jerzy Gołuchowski and Joanna Paliszkiewicz

1.1 Introduction

Trust is a cornerstone for developing, deploying, and adopting generative artificial intelligence (GenAI) technologies. As these systems become increasingly integrated into various aspects of daily life and decision-making, the trust ensures that users, organizations, and governments accept them. Trust enables users to interact with GenAI systems. Trust encourages GenAI's ethical and responsible implementation by ensuring fairness, inclusivity, and alignment with societal values.

Organizations and individuals are more likely to adopt and integrate GenAI when they trust its functionality, reliability, and ethical standards. Trust facilitates collaboration between AI developers, policymakers, and end-users by creating a shared understanding of the technology's capabilities and limitations. Trust reduces resistance to new technologies and facilitates smoother transitions.

GenAI is a branch of AI dedicated to developing models and systems that create new content, such as images, videos, music, or text. These models identify patterns and structures by analyzing large datasets to generate realistic outputs resembling the original data. Leveraging techniques like generative adversarial networks and variational autoencoders, GenAI drives creativity, facilitates data synthesis, and transforms industries like art, entertainment, and content production (*Generative AI – Worldwide, Statista,* 2024).

The GenAI market is experiencing rapid global growth, determined by the increasing adoption of digital technologies. This expansion is set to continue as businesses and industries increasingly embrace the potential of GenAI technology. Consumers increasingly prefer personalized and customizable GenAI solutions driven by cultural nuances and evolving lifestyles. Rising data privacy concerns drive demand for AI solutions, emphasizing user privacy and security. The GenAI market is seeing increased adoption of AI-powered chatbots and virtual assistants to improve customer service and automate processes, driving efficiency and cost reduction.

DOI: 10.4324/9781003586944-2

The GenAI market is expanding rapidly in China due to strong government support for AI adoption across industries. In Japan, the aging population drives demand for AI-powered medical devices and virtual assistants, addressing the need for personalized and cost-effective healthcare solutions. Meanwhile, the market thrives in the United States due to substantial investments in AI research and development, alongside widespread adoption across healthcare, finance, and entertainment sectors. The rising demand for intelligent automation across industries and the need for innovative solutions to boost business efficiency are accelerating the global adoption of GenAI (*Generative AI – Worldwide, Statista,* 2024). According to worldwide statistics, global market growth shows a growing trend of trust in AI and its users. However, there is still a primary concern for developers, policymakers, and users, especially potential users, that encompasses more than technical performance; it includes ethical transparency, data privacy, accountability, and alignment with societal values.

Developing and deploying assistive AI systems and ensuring their responsible implementation necessitate a multidisciplinary approach to thoroughly comprehend the factors influencing trust. The appropriate level of trust hinges on providing humans with clear explanations of an AI's predictions. Therefore, designing human-AI interactions (HMIs) requires carefully assessing and integrating algorithmic transparency and interpretability to foster effective collaboration and maximize potential synergies (Schmidt & Biessmann, 2020).

Trust in GenAI positively impacts users' attitudes toward the technology and their intentions to adopt and utilize it (Moon, 2024). Literature highlights the need for further studies to understand how AI-specific trust mechanisms can address transparency and ethical concerns, especially integrating GenAI into critical services. Applying insights from trust studies conducted in controlled environments to real-world contexts poses significant challenges.

This chapter analyzes the complex view of trends across trust-building within GenAI by synthesizing insights from existing research and theoretical frameworks. It aims to identify threats and challenges posed by GenAI, such as privacy concerns, ethical dilemmas, and biases, while exploring opportunities for fostering trust. Additionally, this chapter examines the implications of regulatory frameworks like the AI Act. It proposes future research directions and practical strategies to enhance transparency, accountability, and societal alignment in deploying GenAI systems.

This chapter systematically addresses the complexities of trust-building in GenAI, beginning with an introduction highlighting the importance of trust for adoption and implementation while outlining key challenges such as transparency, bias, and privacy. This is followed by a literature review, which examines trust through psychological, sociological, and economic lenses, analyzing existing trust mechanisms, algorithmic accountability, and the "black box" challenge. This chapter then delves into emerging challenges, exploring ethical issues like algorithmic bias, privacy concerns, and the trade-offs between complexity and interpretability. Next, sector-specific considerations focus on healthcare, education, and public administration trust dynamics. The conclusion synthesizes the findings and emphasizes stakeholder collaboration to build trust and ensure the responsible integration of GenAI into society. To conclude, this chapter outlines future research directions, proposing interdisciplinary approaches to enhance transparency, accountability, and ethical alignment. This chapter also acknowledges several limitations, including the reliance on existing literature and primary empirical data and the challenges of applying trust frameworks across diverse contexts and sectors.

1.2 Literature Review on Trust in GenAI

Trust in AI, specifically GenAI, involves expectations of system reliability, predictability, and alignment with human values. Psychological, sociological, and economic perspectives contribute to our understanding of trust. Trust can either be an innate aspect of one's personality or an emotional reaction shaped by social experiences, past interactions, and perceptions of fairness (Paliszkiewicz & Gołuchowski, 2024).

Many definitions of trust describe it as a willingness to rely on AI due to its anticipated benefits, AI technologies' rapid advancement, and inherent opacity that have led to growing skepticism. However, AI systems increase trust after incidents of trust loss, often relying on the perception that improvements have been made since the previous interaction (Dorton & Harper, 2022).

Trust in AI is not just about the technology itself but also about the user's perception and the context in which the AI is used. Trust in AI can be strengthened through effective communication, education, and training, enabling users to understand better and feel confident in interacting with AI technologies (Araujo et al., 2020; Li et al., 2024). However, developing trustworthy AI systems requires establishing a clear framework for defining the legal responsibility of AI-based systems, supported by a structured auditing process (Díaz-Rodríguez et al., 2023).

The widespread adoption of algorithmic decision-making systems streamlines processes, reducing time, transaction costs, and reliance on human resources, but it also introduces several risks. Algorithms are essential in decision-making and resource allocation across multiple sectors, utilizing large datasets to generate outputs through transparent, step-by-step processes similar to a formula. Despite their complexity, which often involves processing thousands of variables across millions of data points, limited consumer or civil rights protections govern the data types or require an audit of algorithmic decisions. Algorithmic accountability—assigning responsibility for harms caused by discriminatory or inequitable outcomes—remains essential to addressing these challenges (Donovan et al., 2018). However, algorithms are also employed to guide resource distribution and enhance economic decision-making by identifying critical factors and application areas, thereby ensuring the efficient allocation of resources (Li & Xiang, 2024; Terebukh, 2011).

The "black box" concept in AI refers to the opacity of complex systems like deep learning neural networks, making it difficult to understand their decision-making processes. AI algorithms' "black box" nature and intense learning limit transparency and can weaken user trust, particularly in high-stakes contexts such as finance and healthcare (Lukyanenko et al., 2022). Addressing the "black box" challenge requires bridging the gap between technical explanations and user comprehension to ensure that AI systems are trustworthy and transparent.

AI-specific trust mechanisms differ significantly from traditional trust mechanisms, requiring unique ethical considerations during design and implementation. These include ensuring transparency in AI processes, addressing algorithmic biases, maintaining accountability, and safeguarding user data privacy.

Trust in AI can recover over time, particularly when users perceive that the system has improved since their last interaction. This concept, known as "buoyant trust," suggests that trust can grow even after negative experiences, provided users observe meaningful enhancements in the AI's performance or behavior (Stanley & Dorton, 2023).

1.3 Emerging Trends in Trust-Building in GenAI

GenAI introduces unique challenges related to privacy and technical performance, ethical risks, data bias, and societal implications. AI systems, which rely on vast datasets, often reproduce and amplify biases present in training data.

1.3.1 Privacy and Technical Performance of GenAI Applications

Privacy concerns arise with AI applications in facial recognition, behavioral prediction, and data analytics. For instance, GenAI tools in the digital economy may inadvertently threaten individual freedoms if privacy safeguards are insufficient. AI systems frequently depend on large-scale data collection, raising risks of privacy breaches and unauthorized surveillance. These concerns are particularly pronounced in private organizations' use of AI for facial recognition and biometric identification, as such applications can infringe on individual privacy and lead to the potential misuse of sensitive information (Kouroupis, 2022; Penney, 2020).

Several key factors determine AI systems' effectiveness and user acceptance. These factors can be grouped into three main categories, each carrying significant weight: (1) functionality and reliability, (2) explainability and transparency, and (3) security and privacy. Users require reliable AI systems to perform their tasks consistently (Becker & Fischer, 2024; Qin et al., 2020; Tucci et al., 2022). This ensures that the systems can meet expectations and provide reliable outcomes to maintain functionality and reliability in various scenarios. The "black box" nature of AI systems often creates mistrust among users. Implementing explainable AI (XAI) and transparent decision-making processes is essential to foster trust and ensure that users understand decisions (Becker & Fischer, 2024; Salloum, 2024; Tucci et al., 2022). Safeguarding data security and privacy is of utmost importance. Users are deeply concerned about how their data is collected, managed, and protected, highlighting the need for robust security measures and transparent privacy policies (Panda et al., 2024).

The current trend in this field covers significant privacy risks, varying approaches to regulating GenAI, current legal framework changes, and the introduction of more innovative complex data processing modes (multimodal capabilities – like GPT, T5, and BERT – lead the field, leveraging extensive text datasets to produce text that closely resembles human writing) (Adamyk et al., 2023; Lombardi, 2023; Lucaj et al., 2023; Minssen et al., 2024; Wang et al., 2023; Xu et al., 2024). Integrating GenAI with technologies such as 6G and edge computing is anticipated to solve complex challenges and improve network performance (Celik & Eltawil, 2024).

1.3.2 Ethical Issues, Trends, and Risks as Trust-Building Issues

Ethical AI trust mechanisms must also prioritize fairness and inclusivity, fostering trust by aligning with societal values and user expectations (Omrani et al., 2022). Fairness in AI systems significantly affects user trust, mainly when fairness levels are low. Fairness in AI systems plays a crucial role in shaping user trust. Research indicates that low levels of fairness diminish trust, whereas higher levels can strengthen it (Angerschmid et al., 2022; Farayola et al., 2023).

Ethical transparency is essential in AI to ensure that these systems are trustworthy, align with societal values, and do not cause harm or perpetuate inequality. AI systems must be designed to minimize and eliminate biases that could result in unfair outcomes. Addressing and mitigating these biases is crucial for ethical and equitable AI deployment. These foster trust and ethically promote fairness in decision-making processes (Becker & Fischer, 2024; Moon, 2024; Panda et al., 2024; Salloum, 2024). Establishing precise accountability mechanisms is essential to determine responsibility and a new risk approach for the decisions and actions taken by AI systems. This ensures that any negative consequences or ethical concerns can be appropriately addressed and resolved (Moon, 2024; Salloum, 2024). Implementing trustworthy AI systems involves defining the responsibility of AI-based systems facing the law through a given auditing process (Stanley & Dorton, 2023).

Integrating GenAI with emerging technologies like 6G, edge computing, and IoT presents opportunities and risks, particularly in control and data security. Advances in GenAI have also facilitated the creation of highly realistic deepfakes, increasing the potential for disinformation, fraud, and a decline in public trust. Furthermore, the rapid progression of AI development often surpasses the creation of adequate regulatory frameworks, resulting in significant gaps in ethical oversight.

Trends of changes in GenAI cover the trust and ethical approach concerns: developing models that provide clear insights into their decision-making processes, reducing the “black box” nature of AI systems to build user trust and accountability, creating regulatory frameworks and global standards for AI ethics, detecting and preventing misuse of GenAI, merging with advanced technologies, reskill, and upskill workers for AI-driven economies, and addressing environmental challenges and sustainable solutions across industries.

1.3.3 Impact of Data Privacy and Data Bias on Trust-Building in GenAI

Data privacy is a cornerstone of building trust in AI systems, ensuring users feel secure about how their information is handled and protected. Proper handling and management of data are critical for maintaining user trust. This includes obtaining user consent, ensuring transparent data practices, and implementing data anonymization techniques to safeguard sensitive information (Emaminejad et al., 2024; Panda et al., 2024; Qin et al., 2020). Users’ concerns about privacy can significantly undermine their trust in AI systems.

AI systems tailored to accommodate diverse user profiles can enhance trust by providing personalized experiences that align with individual needs, preferences, and expectations (Riedl, 2022). The extensive data processed by AI systems amplifies the risk of data breaches, potentially exposing personal information and individual privacy (Burton et al., 2024; Tang, 2023). AI tools can expose sensitive data, create digital inequalities, and contribute to the monopolization of information, creating competitive disadvantages (Riedl, 2022).

The generating AI cover trend related to prioritizing data privacy and transparency for fostering trust in AI systems, as it addresses user concerns, ensures secure information handling, and supports personalized experiences that align with diverse user needs and expectations. Addressing these challenges through robust privacy policies and transparent communication is essential to maintaining confidence in the technology (Emaminejad et al., 2024; Moon, 2024).

1.3.4 Societal Values Expectations and Their Implications on Trust

Ensuring that AI systems align with societal values is essential for their acceptance and responsible integration into everyday life. AI systems should align with societal values and norms to GenAI wide acceptance (Collins & Jones, 2023; Pucelj & Bohinc, 2024; Rakowski & Kowaliková, 2024). This includes ensuring that AI applications do not harm societal interests and actively contribute to the well-being of communities (Cantens, 2024; Qin et al., 2020). Establishing and adhering to ethical guidelines is crucial for aligning AI systems with societal values. These guidelines provide a framework to ensure that AI development and deployment respect ethical principles and societal expectations (Panda et al., 2024; Salloum, 2024). Information bubbles and echo chambers created by AI-driven recommendation systems can limit users' exposure to diverse viewpoints, reinforcing existing beliefs and biases. Such phenomena pose significant risks to democratic processes by reducing public discourse quality and fostering societal divisions. A new trend in this area showed that GenAI is being applied across various fields, improving functionalities in natural language processing, content generation, and personalized service delivery.

1.3.5 Human-AI Collaboration Impact on Trust in GenAI

Incorporating human factors into AI design and deployment is vital for fostering trust and ensuring seamless interaction between users and AI systems. Educating users about AI, its capabilities, and its limitations is essential for building trust. Users must understand how AI works and its potential benefits and risks to interact with it confidently and responsibly (Moon, 2024; Panda et al., 2024; Tucci et al., 2022). Trust in AI systems can be strengthened by promoting effective human-AI collaboration. Ensuring that AI systems are designed to support rather than replace human decision-making helps create a balanced and cooperative relationship between humans and technology (Emaminejad et al., 2024; Paparic & Bodea, 2024; Tucci et al., 2022).

GenAI adoption in the US workplace varies by generation. Gen Z leads at 29%, Gen X at 28%, and Millennials at 27% (*Adoption Rate of Generative AI Adoption in the Workplace in the United States 2023, by Generation,* 2022). Broad acceptance of the technology suggests that GenAI is becoming a standard tool across all age groups in professional settings.

The trustworthiness of robots and autonomous systems can be assessed using diverse methods, with human-robot interaction serving as a key factor in building and maintaining trust (Devitt, 2018). HMI should be user-friendly and promote seamless communication with humans, enabling users to trust and depend on the system (He et al., 2022). Integrating empathy and social identity into systems can foster human-like trust relationships (Devitt, 2018).

As technology advances, it is poised to transform the workforce, potentially influencing traditionally safeguarded areas like legal and human resources. New trends should include that trust is pivotal in HMIs, shaping users' willingness to rely on and adopt AI systems.

1.3.6 Sector-Specific Considerations of Trust in GenAI

GenAI is currently utilized by researchers and developers in industries such as advertising and marketing, and it is anticipated that its adoption will soon expand among businesses and consumers for a diverse array of tasks. Furthermore, trust and acceptance of AI systems can vary significantly across different sectors, depending on each domain's unique needs, challenges, and expectations. Trust in medical AI systems is heavily influenced by their explainability, transparency, and the involvement of healthcare professionals in their development. These factors ensure that AI solutions align with medical standards and effectively support patient care (Asan et al., 2020; Tucci et al., 2022). In educational settings, trust in AI depends on its perceived helpfulness, interpretability, and the extent to which teachers are involved in integrating AI tools into the learning environment. Ensuring that AI enhances rather than detracts from the educational experience is critical (Amoozadeh et al., 2024; Qin et al., 2020). Using GenAI in public administration requires addressing specific risks, such as ensuring confidentiality and providing explainable outputs. These measures are vital for maintaining public trust and ensuring the ethical use of AI in governance (Cantens, 2024). The novel focus on GenAI in digital health underscores emerging research opportunities at the intersection of advanced AI technologies and healthcare GenAI innovation (Golda et al., 2024). This area holds promise for revolutionizing patient care, improving diagnostic accuracy, and personalizing treatment plans while addressing ethical, regulatory, and technical challenges.

1.4 GenAI Technologies – Emerging Challenges

The rising interest in XAI highlights the increasing awareness of its critical role in fostering trust, accountability, and transparency in AI systems. By making AI decisions more understandable to users, XAI helps bridge the gap between complex algorithms and human interpretability, ensuring that AI technologies are more accessible and trustworthy across various applications (de Zoeten et al., 2023; Kobrinskii, 2023). XAI helps build trust in AI systems by making their decision-making processes transparent and understandable to humans. By providing clear and understandable explanations for AI decisions, XAI enables users to hold AI systems accountable. This ensures that responsibility for AI-driven actions can be appropriately assigned and evaluated (De Brito Duarte et al., 2023; Srivastava et al., 2024; Thiruthuvanathan et al., 2023). Transparency through XAI helps identify and address biases within AI systems, promoting fairer and more ethical outcomes. This contributes to developing AI technologies that align with societal and ethical standards. XAI is crucial in promoting accountability and ethical use of AI systems. In finance, XAI helps in risk assessment and regulatory compliance by making AI-driven decisions transparent and justifiable. This transparency is crucial for GenAI user trust and ensuring ethical financial practices (Singh et al., 2024; Thiruthuvanathan et al., 2023; Varshney et al., 2024). Financial institutions can build trust and promote the responsible use of AI technologies by implementing XAI models and aligning them with international standards.

One of the main challenges in XAI is striking a balance between the complexity of AI models and the need for interpretability. Highly complex models, such as deep neural networks, often achieve superior performance but are difficult for users to understand. On the other hand, simpler models are easier to interpret but may sacrifice accuracy and efficiency. Achieving this balance requires innovative approaches that preserve model performance while offering clear and meaningful explanations to end-users (Mohammed, 2023; Praveenraj et al., 2023). XAI is pivotal in making AI systems more transparent, trustworthy, and ethical. By providing clear explanations of AI decisions, XAI fosters trust and accountability, essential for adopting AI in high-stakes domains like healthcare and finance.

AI-driven recommendation systems create "information bubbles," where users are repeatedly exposed to familiar content based on previous behavior. These bubbles can hinder diverse perspectives and informed decision-making, potentially impacting democratic processes. The ethical concerns surrounding AI-driven recommendation systems stem from their tendency to create information bubbles, which can limit exposure to diverse viewpoints. This restriction affects informed decision-making and poses potential risks to democratic processes by reinforcing bias and polarizing opinions (Hu et al., 2022; Magrani & da Silva, 2024).

1.5 Regulatory Challenges: AI Act and Legal Constraints

The AI Act and General Data Protection Regulation (GDPR), developed to protect user privacy and accountability in AI applications, offer critical protections. Both documents are part of the European Union's legal framework. Therefore, the EU and its institutions, including the European Parliament, the Council of the EU, and the European Commission, are responsible for their drafting, adoption, and enforcement. However, the operationalization of these regulations presents several challenges. Both the AI Act and GDPR emphasize the importance of XAI systems. GDPR is an EU regulation that came into effect in May 2018, governing the protection of the personal data of EU citizens and residents. It aims to ensure user privacy and organizational accountability in handling personal data. The GDPR mandates automated decision-making and profiling transparency, ensuring that individuals understand how their data is used and decisions are made (Hoxhaj et al., 2023; Nisevic et al., 2024). The AI Act is a proposed regulation by the European Union to establish harmonized rules for AI across EU member states. Its goal is to regulate the use of AI to ensure safety, transparency, and accountability. The AI Act, for instance, requires that organizations provide clear documentation and impact assessments, which can be resource-intensive and complex to implement. Users' trust can influence trust in AI in related entities, such as government institutions or parent companies. The GDPR's regulations on automated decision-making are frequently criticized for their ambiguity and lack of robust enforcement. The AI Act aims to strengthen these protections by establishing more precise and specific requirements for AI systems (Mougdir, 2020; Wulf & Seizov, 2024). When these entities are perceived as reliable and ethical, their association with an AI system can enhance user trust (Kim & Kwon, 2024; Park & Yoon, 2024; Wischnewski et al., 2024).

Users' trust in AI is profoundly influenced by their perceptions of fairness, accountability, transparency, and explainability. Users' confidence in the system increases when they perceive AI decisions as fair and unbiased. This perception reinforces the belief that AI operates ethically and equitably, fostering a positive relationship between users and the technology (Angerschmid et al., 2022). Accountability in AI systems is a fundamental factor in establishing and maintaining user trust. Users need assurance that mechanisms are in place to hold AI systems accountable for their decisions and actions. Clear accountability frameworks ensure that responsibility can be assigned, fostering confidence in the system's reliability and ethical integrity (Smit et al., 2022; Wang, 2023). According to the Technology-Organization-Environment framework, accountability, fairness, and transparency influence trust and AI adoption within organizations (Cath, 2018; Singh et al., 2024). These factors collectively help create an environment where AI systems can be integrated responsibly and effectively (Smit et al., 2022).

Clear and transparent AI processes build user confidence by making decision-making mechanisms easy to understand. This openness demystifies complex algorithms and assures users that the system operates fairly and ethically. In contrast, algorithms that lack transparency or exhibit bias can erode trust, as users may view them as unpredictable, unjust, or prone to errors (Becker & Fischer, 2024; Grimmelikhuijsen, 2023; Khosravi et al., 2022; Szczepanski et al., 2022). Tackling these challenges is crucial for preserving credibility and ensuring broad acceptance of AI systems. Biases present in training data can result in unfair and skewed model predictions. These biases often stem from historical data that reflect existing societal inequalities, perpetuating discriminatory patterns, and reinforcing inequities in AI system outcomes. Addressing these biases is essential to ensure fairness and ethical decision-making in AI applications (Ferrara, 2024). Bias in AI is a significant threat to trust, especially in areas where fairness and integrity are critical, such as recruitment, education, and criminal justice. Addressing these biases and developing systems prioritizing inclusivity and cultural sensitivity is vital to maintaining public trust. Promoting fairness and inclusivity in AI systems is crucial for reducing the risks of biased outcomes and ensuring equitable treatment across diverse sectors (Farayola et al., 2023).

In sensitive domains like healthcare, trust in AI systems heavily depends on their predictions' perceived accuracy and validity. Ensuring that AI tools meet high-reliability standards and are validated through rigorous testing is crucial for their acceptance in such critical applications. The GDPR requires organizations to implement measures that ensure compliance with data protection laws, including the appointment of Data Protection Officers and the conducting of Data Protection Impact Assessments (Hoxhaj et al., 2023; Vogel, 2024).

Legal limitations also restrict access to data necessary for training reliable and ethical AI systems, impacting innovation and algorithmic fairness (Kruse & Schöning, 2024). Striking a balance between fostering innovation and ensuring data protection is vital, but achieving this balance remains an ongoing challenge. Legal considerations include data protection, algorithmic bias, intellectual property, and liability (Akramov & Valiev, 2024; Aslan et al., 2022; Peng & Yu, 2024; Tamò-Larrieux et al., 2024). This complex legal landscape underscores addressing emerging risks and ensuring comprehensive protection.

As AI technologies evolve, new privacy, security, accountability, and ethical use challenges arise. Regulators must proactively update policies to safeguard users, promote transparency, and foster innovation while ensuring that AI applications align with societal values and legal standards (Brown et al., 2022; Nisevic et al., 2024). Evolving legal frameworks for digital technologies, such as the European Union's Artificial Intelligence Act, are being refined to tackle emerging ethical challenges in AI. These frameworks highlight the importance of data protection laws in safeguarding human rights, ensuring the need for adaptive policy frameworks to accommodate rapid AI advancements. Different regions are establishing regulatory frameworks aimed at promoting trustworthy AI systems. Furthermore, ambiguities in regulatory language may lead to inconsistent GenAI applications across industries, making it challenging for stakeholders to align with best practices uniformly.

Future research on GenAI and trust should focus on four primary areas:

1. Integrating perspectives from ethics, psychology, sociology, and technology can provide a holistic understanding of trust. Developing tools that facilitate greater transparency, explainability, and user control over AI interactions will also be critical. Recognizing the pivotal role of trust in the successful development and deployment of AI, it is crucial to synthesize existing research and identify strategies to address these issues (Paliszkiewicz & Gołuchowski, 2024).
2. Ethical considerations include addressing AI algorithm biases to prevent injustices and discrimination. Ensuring fairness and inclusivity in algorithmic design is crucial for building and maintaining trust in AI systems, as users are more likely to rely on technologies they perceive as equitable and unbiased. Organizations are leveraging AI to enhance and automate business processes, which raises ethical concerns about the potential for algorithms to perpetuate existing human biases. Trust issues arise when users perceive AI as inequitable or biased, potentially undermining its adoption and effectiveness.
3. Creating protocols to ensure transparency in algorithmic decision-making is essential for tackling the "black box" issue and boosting user trust and acceptance, particularly in areas where AI functions independently (Nešpor, 2024; Žlahtič et al., 2024). It enhances user trust by improving transparency and ensuring that the decision-making process is easier to understand. AI

validation enhances relevance and accuracy by aligning human trust with AI-assisted decision-making, particularly in scenarios where human AI expertise is complementary. The dynamics of trust development, erosion, and recovery in AI-assisted decision-making are explored through experimental tasks, emphasizing the key factors that impact the success of Trust Repair Strategies. The prevalent cognitive biases in AI-assisted decision-making, such as over-reliance and under-reliance, pose significant challenges for trust issues. Addressing these issues requires implementing solutions like XAI techniques or cognitive forcing functions to reduce and mitigate the impact of such biases.

4 Privacy preservation techniques, such as differential privacy and federated learning, should be integrated into GenAI systems to protect user data and increase trust. Additionally, research should address AI's role in environmental conservation, healthcare, and mental health to highlight ethical and secure applications.

1.6 Conclusion

Trust-building in GenAI is a critical factor influencing technology's acceptance and growth across multiple sectors. In this chapter, we expressed that building trust in GenAI involves addressing the ethical, legal, and technical challenges of transparency, data protection, and fairness. It was mentioned that Legal frameworks like the AI Act provide necessary safeguards but pose implementation challenges, especially concerning data access and regulatory compliance and the impact of building users' trust in GenAI. Moving forward, a collaborative approach among researchers, developers, and policymakers will be essential to developing a responsible and trustworthy AI ecosystem that aligns with societal values and user expectations. This trust-building approach, supported by interdisciplinary research and flexible regulatory frameworks, can facilitate AI's responsible development, ensuring that it remains beneficial to individuals and society. User education can mitigate the risks associated with over-trust or under-trust. Over-trust occurs when users rely on AI systems without questioning their outputs, potentially leading to biased or inaccurate information acceptance. Conversely, under-trust leads to skepticism and resistance, undermining the utility of AI technologies.

Educating users about AI algorithms will promote critical thinking and informed engagement with personalized content, which, in consequence, will impact the level of trust in this technology. Drawing from trust theories, this education helps establish cognitive-based trust, where users rely on knowledge and understanding of how AI operates rather than blind faith in the system.

Transparency and explainability in algorithmic processes enhance users' ability to evaluate the fairness, accuracy, and biases of personalized recommendations, which aligns with rational trust built on logical reasoning and evidence. An emerging theme in generative AI is trust, especially in the context of ongoing research into human-AI design collaboration—how people perceive and work with AI systems. This points to the importance of responsible AI design and the growing interest in human-centered AI within the field of human-computer interaction (Díaz-Rodríguez et al., 2023; Mehrotra, 2021; Schmidt & Biessmann, 2020; Yandrapalli & Sharma, 2024; Zerka et al., 2020).

There are some limitations of the study as the review primarily focuses on existing literature and theoretical insights, which might limit its applicability to emerging real-world contexts and rapidly evolving GenAI technologies. This chapter relies heavily on synthesized studies and does not present primary empirical findings, which could provide stronger validation for its conclusions and recommendations. The literature review broadly addresses trust-building in GenAI; it may not fully account for the variations in trust dynamics across different sectors, regions, and user demographics.

Acknowledgment

This research is partly funded by the University of Economics Ho Chi Minh City, Vietnam.

References

Adamyk, O., Chereshnyuk, O., Adamyk, B., & Rylieiev, S. (2023). Trustworthy AI: A fuzzy-multiple method for evaluating ethical principles in AI regulations. *Proceedings - International Conference on Advanced Computer Information Technologies, ACIT*, 608–613. https://doi.org/10.1109/ACIT58437.2023.10275505.

Adoption rate of generative AI adoption in the workplace in the United States 2023, by generation. (2022). *Statista* https://www.statista.com/statistics/1361174/generative-ai-adoption-rate-at-work-by-generation-us/

Akramov, J., & Valiev, B. (2024). The Level of Implementing AI and its Framework for Creating Structured Strategy for Corporations. *2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2024,* 1583–1587. https://doi.org/10.1109/ICACITE60783.2024.10617232.

Amoozadeh, M., Daniels, D., Nam, D., Kumar, A., Chen, S., Hilton, M., Srinivasa Ragavan, S., & Alipour, M. A. (2024). Trust in generative AI among students: An exploratory study. *SIGCSE 2024- Proceedings of the 55th ACM Technical Symposium on Computer Science Education,* 1, 67–73. https://doi.org/10.1145/3626252.3630842.

Angerschmid, A., Zhou, J., Theuermann, K., Chen, F., & Holzinger, A. (2022). Fairness and explanation in AI-informed decision making. *Machine Learning and Knowledge Extraction*, *4*(2), 556–579. https://doi.org/10.3390/make4020026.

Araujo, T., Helberger, N., Kruikemeier, S., & De Vreese, C. H. (2020). In AI, we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, *35*(3), 611–623.

Asan, O., Bayrak, A. E., & Choudhury, A. (2020). Artificial intelligence and human trust in healthcare: Focus on clinicians. *Journal of Medical Internet Research*, *22*(6), 15154. https://doi.org/10.2196/15154.

Aslan, A., Greve, M., & Lembcke, T. B. (2022). Let's Do Our Bit: How information systems research can contribute to ethical artificial intelligence. *28th Americas Conference on Information Systems, AMCIS 2022*, 1–10. https://www.researchgate.net/profile/Aycan-Aslan/publication/369480408_Lets_Do_Our_Bit_How_Information_Systems_Research_Can_Contribute_to_Ethical_Artificial_Intelligence/links/641d70bb66f8522c38cd02ec/Let-s-Do-Our-Bit-How-Information-Systems-Research-Can-Contribute-to-Ethical-Artificial-Intelligence.pdf.

Becker, C., & Fischer, M. (2024). Factors of trust building in conversational AI systems: A literature review. *Lecture Notes in Computer Science*, *14735*, 27–44. https://doi.org/10.1007/978-3-031-60611-3_3.

Brown, R., Truby, J., & Ibrahim, I. A. (2022). Mending lacunas in the EU's GDPR and proposed artificial intelligence regulation. *European Studies: The Review of European Law*, *Economics and Politics*, *9*(1), 61–90. https://doi.org/10.2478/eustu-2022-0003.

Burton, S. L., Burrell, D., White, Y. W., Nobles, C., Dawson, M. E., Brown-Jackson, K. L., Muller, S. R., & Bessette, D. I. (2024). An In-Depth Qualitative Interview: The Impact of Artificial Intelligence (AI) on Privacy Challenges and Opportunities. In A. Anand, A. Madaan, & A. Danielsson (Eds.), *Intersections Between Rights and Technology* (pp. 19–39). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-1127-1.ch002

Cantens, T. (2024). How will the state think with ChatGPT? The challenges of Generative artificial intelligence for public administrations. *AI and Society*, *1*, 133–144. https://doi.org/10.1007/s00146-023-01840-9.

Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180080. https://doi.org/10.1098/rsta.2018.0080.

Celik, A., & Eltawil, A. M. (2024). At the dawn of generative AI era: A tutorial-cum-survey on new frontiers in 6G wireless intelligence. *IEEE Open Journal of the Communications Society*, *5,* 2433–2489. https://doi.org/10.1109/OJCOMS.2024.3362271.

Collins, A., & Jones, J. J. (2023). Effect of artificial intelligence on social trust in american institutions. *Journal of Social Computing*, *4*(3), 221–231. https://doi.org/10.23919/JSC.2023.0022.

De Brito Duarte, R., Correia, F., Arriaga, P., & Paiva, A. (2023). AI Trust: Can explainable AI enhance warranted trust? *Human Behavior and Emerging Technologies*, *2023*(1), 4637678. https://doi.org/10.1155/2023/4637678.

De Zoeten, M., Ernst, C.-P. H., & Rothlauf, F. (2023). A matter of trust: How trust in AI-based systems changes during interaction. *29th Annual Americas Conference on Information Systems, AMCIS 2023*, 1–10. https://www.researchgate.net/profile/Marc_De_Zoeten/publication/374156709_A_Matter_of_Trust_How_Trust_in_AI-Based_Systems_Changes_During_Interaction_Completed_Research_Full_Paper/links/6511416461f18040c222c136/A-Matter-of-Trust-How-Trust-in-AI-Based-Systems-Changes-During-Interaction-Completed-Research-Full-Paper.pdf.

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy artificial intelligence: From

AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, *99*, 101896. https://doi.org/10.1016/j.inffus.2023.101896.
Donovan, J., Caplan, R., Matthews, J., & Hanson, L. (2018). *Algorithmic accountability: A primer* (Prepared for the Congressional Progressive Caucus). Data & Society Research Institute. https://datasociety.net/wp-content/uploads/2018/04/Data_Society_Algorithmic_Accountability_Primer_FINAL.pdf*
Dorton, S. L., & Harper, S. B. (2022). Self-Repairing and/or buoyant trust in artificial intelligence. *Proceedings of the Human Factors and Ergonomics Society*, *66*(1), 162–166. https://doi.org/10.1177/1071181322661098.
Emaminejad, N., Kath, L., & Akhavian, R. (2024). Assessing trust in construction AI-powered collaborative robots using structural equation modeling. *Journal of Computing in Civil Engineering*, *38*(3), 04024011. https://doi.org/10.1061/JCCEE5.CPENG-5660.
Farayola, M. M., Tal, I., Malika, B., Saber, T., & Connolly, R. (2023). Fairness of AI in predicting the risk of recidivism: Review and phase mapping of AI fairness techniques. *ACM International Conference Proceeding Series,* 1–10. https://doi.org/10.1145/3600160.3605033.
Ferrara, E. (2024). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*, *6*(3), 1–15. https://doi.org/10.3390/sci6010003.
Generative AI - Worldwide, Statista. (2024). *Statista*. https://www.statista.com/outlook/tmo/artificial-intelligence/generative-ai/worldwide.
Golda, A., Mekonen, K., Pandey, A., Singh, A., Hassija, V., Chamola, V., & Sikdar, B. (2024). Privacy and security concerns in generative AI: A comprehensive survey. *IEEE Access*, *12*, 48126–48144. https://doi.org/10.1109/ACCESS.2024.3381611.
Grimmelikhuijsen, S. (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review*, *83*(2), 241–262. https://doi.org/10.1111/puar.13483.
He, H., Gray, J., Cangelosi, A., Meng, Q., McGinnity, T. M., & Mehnen, J. (2022). The challenges and opportunities of human-centered AI for trustworthy robots and autonomous systems. *IEEE Transactions on Cognitive and Developmental Systems*, *14*(4), 1398–1412. https://doi.org/10.1109/TCDS.2021.3132282.
Hoxhaj, O., Halilaj, B., & Harizi, A. (2023). Ethical implications and human rights violations in the age of artificial intelligence. *Balkan Social Science Review*, *22*(22), 153–171. https://doi.org/10.46763/BSSR232222153h.
Hu, Y., Wu, S., Jiang, C., Li, W., Bai, Q., & Roehrer, E. (2022). AI Facilitated Isolations? The Impact of Recommendation-based Influence Diffusion in Human Society. *IJCAI International Joint Conference on Artificial Intelligence, 22*, 5080–5086.
Kate Devitt, S. (2018). Trustworthiness of Autonomous Systems. In: Abbass, H., Scholz, J., Reid, D. (eds) *Foundations of Trusted Autonomy. Studies in Systems, Decision and Control*, (Vol. 117, pp. 161–184). Springer, Cham. https://doi.org/10.1007/978-3-319-64816-3_9.
Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *3*, 100074. https://doi.org/10.1016/j.caeai.2022.100074.
Kim, B., & Kwon, S. (2024). Balancing Innovation and Regulation: The Impact of Government Policies and Perceived Risk on Public Trust in AI Applications. *IEEE*

International Workshop on Robot and Human Communication, RO-MAN, 33, 1802–1808. https://doi.org/10.1109/RO-MAN60168.2024.10731444.

Kobrinskii, B. A. (2023). Artificial intelligence: Problems, solutions, and prospects. *Pattern Recognition and Image Analysis*, *33*(3), 217–220. https://doi.org/10.1134/S1054661823030203.

Kouroupis, K. (2022). The AI Act in light of the EU digital agenda: A critical approach. *Journal of Data Protection and Privacy*, *5*(3), 216–229.

Kruse, N., & Schöning, J. (2024). Legal conform data sets for yard tractors and robots: AI-based law compliance check on the right to one's image. *Computers and Electronics in Agriculture*, *223*, 109106. https://doi.org/10.1016/j.compag.2024.109106.

Li, Y., Wu, B., Huang, Y., Liu, J., Wu, J., & Luan, S. (2024). Warmth, competence, and the determinants of trust in artificial intelligence: A cross-sectional Survey from China. *International Journal of Human-Computer Interaction*, 1–15. https://doi.org/10.1080/10447318.2024.2356909.

Li, Y., & Xiang, B. (2024). Reducing organizational inequalities associated with algorithmic controls. *Discover Artificial Intelligence*, *4*(1), 1–8. https://doi.org/10.1007/s44163-024-00137-0.

Lombardi, A. (2023). Data protection regulation and artificial intelligence regulation: Relationships, similarities and differences between GDPR and AI Act | Disciplina della tutela dei dati personali e regolazione dell'intelligenza artificiale: Rapporti, analogie e differenz. *European Journal of Privacy Law and Technologies*, *2023*(2), 240–252.

Lucaj, L., Van Der Smagt, P., & Benbouzid, D. (2023). AI Regulation is (not) all you need. *ACM International Conference Proceeding Series*, *1*, 1267–1279. https://doi.org/10.1145/3593013.3594079.

Lukyanenko, R., Maass, W., & Storey, V. C. (2022). Trust in artificial intelligence: From a foundational trust framework to emerging research opportunities. *Electronic Markets*, *32*(4), 1993–2020. https://doi.org/10.1007/s12525-022-00605-4.

Magrani, E., & da Silva, P. G. F. (2024). The ethical and legal challenges of recommender systems driven by artificial intelligence. In H. S. Antunes, P. M. Freitas, A. L. Oliveira, C. M. Pereira, E. Vaz de Sequeira, & L. B. Xavier (Eds.), *Multidisciplinary perspectives on artificial intelligence and the law* (Vol. 58, pp. 141–168). Springer International Publishing. https://doi.org/10.1007/978-3-031-41264-6_8.

Mehrotra, S. (2021). Modelling trust in human-AI interaction. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS, 3*, 1814–1816.

Minssen, T., Solaiman, B., Köttering, L., Wested, J., & Malik, A. (2024). Governing AI in the European Union: Emerging infrastructures and regulatory ecosystems in health. In B. Solaiman & I. Glenn Cohen (Eds.), *Research handbook on health, AI and the law* (pp. 311–331). Edward Elgar Publishing. https://doi.org/10.4337/9781802205657.ch18.

Mohammed, B. (2023). A review on explainable artificial intelligence methods, applications, and challenges. *Indonesian Journal of Electrical Engineering and Informatics*, *11*(4), 1007–1024. https://doi.org/10.52549/ijeei.v11i4.5151.

Moon, S. J. (2024). Effects of perception of potential risk in generative AI on attitudes and intention to use. *International Journal on Advanced Science, Engineering and Information Technology*, *14*(5), 1748–1755. https://doi.org/10.18517/ijaseit.14.5.20445.

Mougdir, S. (2020). Artificial intelligence in a privacy concerned world: Automated decision-making and the GDPR. *Journal of Data Protection and Privacy*, *3*(4), 393–402. https://doi.org/10.69554/bppn8465.

Nešpor, J. (2024). Automated administrative decision-making: What is the black box hiding? *Acta Universitatis Carolinae Iuridica*, *70*(2), 69–83. https://doi.org/10.14712/23366478.2024.23.

Nisevic, M., Cuypers, A., & De Bruyne, J. (2024). Explainable AI: Can the AI Act and the GDPR go out for a date?. *Proceedings of the International Joint Conference on Neural Networks,* 1–8. IEEE. https://doi.org/10.1109/IJCNN60899.2024.10649994.

Omrani, N., Rivieccio, G., Fiore, U., Schiavone, F., & Agreda, S. G. (2022). To trust or not to trust? An assessment of trust in AI-based systems: Concerns, ethics, and contexts. *Technological Forecasting and Social Change, 181*, 121763.

Paliszkiewicz, J., & Gołuchowski, J. (2024). Trust in artificial intelligence – future directions In J. Paliszkiewicz, J. Gołuchowski, *Trust and artificial intelligence: Development and application of AI technology*. Taylor and Francis https://doi.org/https://doi.org/10.4324/9781032633749.

Panda, R. S., Mishra, A., & Mohanty, A. (2024). Innovating reality: The pioneering applications of Generative AI. In R. Kumar, S. Sahu, & S. Bhattacharya (Eds.), *The Pioneering Applications of Generative AI* (pp. 85–105). IGI Global. https://doi.org/10.4018/979-8-3693-3278-8.ch004.

Paparic, M., & Bodea, C.-N. (2024). Building trust through responsible usage of generative artificial intelligence in projects: A case study. *Issues in Information Systems*, *25*(4), 143–157. https://doi.org/10.48009/4_iis_2024_112.

Park, K., & Yoon, H. Y. (2024). Beyond the code: The impact of AI algorithm transparency signaling on user trust and relational satisfaction. *Public Relations Review*, *50*(5), 102507. https://doi.org/10.1016/j.pubrev.2024.102507

Peng, W., & Yu, B. (2024). 5 Steps for Enterprise Artificial Intelligence Governance and Compliance. *International Conference on Big Data*, *15422 LNCS*, 88–104. https://doi.org/10.1007/978-3-031-77088-3_7.

Penney, J. W. (2020). Measuring surveillance chill and other regulatory impacts at scale. In R. Whalen (Ed.), *Computational legal studies: The promise and challenge of data-driven research* (pp. 146–165). Edward Elgar Publishing. https://doi.org/10.4337/9781788977456.00012.

Praveenraj, D. D. W., Victor, M., Vennila, C., Alawadi, A. H., Diyora, P., Vasudevan, N., & Avudaiappan, T. (2023). Exploring explainable artificial intelligence for transparent decision making. *E3S Web of Conferences*, *399*, 04030. https://doi.org/10.1051/e3sconf/202339904030.

Pucelj, M., & Bohinc, R. (2024). Balancing human rights, social responsibility, and digital ethics. In *Balancing Human Rights, Social Responsibility, and Digital Ethics*. https://doi.org/10.4018/979-8-3693-3334-1.

Qin, F., Li, K., & Yan, J. (2020). Understanding user trust in artificial intelligence-based educational systems: Evidence from China. *British Journal of Educational Technology*, *51*(5), 1693–1710. https://doi.org/10.1111/bjet.12994.

Rakowski, R., & Kowaliková, P. (2024). The political and social contradictions of the human and online environment in the context of artificial intelligence applications. *Humanities and Social Sciences Communications*, *11*(1), 289. https://doi.org/10.1057/s41599-024-02725-y.

Riedl, R. (2022). Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. *Electronic Markets*, *32*(4), 2021–2051. https://doi.org/10.1007/s12525-022-00594-4.

Salloum, S. A. (2024). Trustworthiness of the AI. In A. Al-Marzouqi, S. A. Salloum, M. Al-Saidat, A. Aburayya, & B. Gupta (Eds.), *Artificial intelligence in education: The power and dangers of ChatGPT in the classroom. Studies in big data* (Vol. 144). Springer. https://doi.org/10.1007/978-3-031-52280-2_41.

Schmidt, P., & Biessmann, F. (2020). Calibrating Human-AI Collaboration: Impact of Risk, Ambiguity, and Transparency on Algorithmic Bias. In A. Holzinger, P. Kieseberg, A. Tjoa, E. Weippl (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12279 LNCS* (pp. 431–449). https://doi.org/10.1007/978-3-030-57321-8_24.

Singh, J., Rani, S., & Srilakshmi, G. (2024). Towards Explainable AI: Interpretable Models for Complex Decision-making. *2024 International Conference on Knowledge Engineering and Communication Systems, ICKECS 2024*, 1, 1–5. https://doi.org/10.1109/ICKECS61492.2024.10616500.

Smit, D., Eybers, S., & Smith, J. (2022). A data analytics organisation's perspective on trust and AI adoption. In E. Jembere, A. J. Gerber, S. Viriri, & A. Pillay (Eds.), *Artificial intelligence research. SACAIR 2021. Communications in computer and information science*, (Vol. 1551, pp. 47–60). Springer. https://doi.org/10.1007/978-3-030-95070-5_4.

Srivastava, K., Sorathiya, A., Mehta, J., & Chotaliya, V. (2024). Enhancing Interpretability, Reliability, and Trustworthiness: Applications of explainable artificial intelligence in medical Imaging, financial markets, and sentiment analysis. *Proceedings of the 16th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2024,* 1–9. https://doi.org/10.1109/ECAI61503.2024.10607516.

Stanley, J. C., & Dorton, S. L. (2023). Exploring trust with the AI incident database. *Proceedings of the Human Factors and Ergonomics Society*, *67*(1), 489–494. https://doi.org/10.1177/21695067231198084.

Szczepanski, M., Pawlicki, M., Kozik, R., & Choras, M. (2022). Fast hybrid oracle-explainer approach to explainability using optimized search of comprehensible decision trees. *Proceedings -2022 IEEE 9th International Conference on Data Science and Advanced Analytics, DSAA 2022, 1–10. https://doi.org/10.1109/DSAA54385.2022.10032372*.

Tamò-Larrieux, A., Guitton, C., Mayer, S., & Lutz, C. (2024). Regulating for trust: Can law establish trust in artificial intelligence? *Regulation and Governance*, *18*(3), 780–801. https://doi.org/10.1111/rego.12568.

Tang, A. (2023). *Privacy in practice: Establish and operationalize a holistic data privacy program* (1st ed.). CRC Press. Boca Raton. https://doi.org/10.1201/9781003225089.

Terebukh, A. A. (2011). Components in mechanism of economic decision-making. *Actual Problems of Economics*, *125*(11), 49–57.

Thiruthuvanathan, M. M., Pradeep Kumar, K., Nasar, N., K.K, S., & Joseph, P. M. (2023). Security and privacy aspects in intelligence systems through blockchain and explainable AI. In F. Al-Turjman, A. Nayyar, M. Naved, A. K. Singh, & M. Bilal (Eds.), *XAI based intelligent systems for society 5.0* (pp. 365–400). Elsevier. https://doi.org/10.1016/B978-0-323-95315-3.00011-5.

Tucci, V., Saary, J., & Doyle, T. E. (2022). Factors influencing trust in medical artificial intelligence for healthcare professionals: A narrative review. *Journal of Medical Artificial Intelligence*, *5*, 1–13. https://doi.org/10.21037/jmai-21-25

Varshney, P., Singh Rathore, N. P., Sethi, K. K., & Rajput, S. S. (2024). Demystifying XAI: Understanding of applications, limitations, challenges, and future perspectives. *2024 International Conference on Advances in Computing Research on Science Engineering and Technology, ACROSET 2024,* 1–8. https://doi.org/10.1109/ACROSET62108.2024.10743621.

Vogel, P. (2024). Training of AI systems: How to comply with the GDPR?. In M. Ebers & K. Sein (Eds.), *Privacy, Data protection and data-driven technologies* (pp. 143–168). Routledge. https://doi.org/10.4324/9781003502791-9.

Wang, Y. (2023). Balancing trustworthiness and efficiency in artificial intelligence systems: An analysis of tradeoffs and strategies. *IEEE Internet Computing*, *27*(6), 8–12. https://doi.org/10.1109/MIC.2023.3303031.

Wang, Z., Long, C., Xu, S., Gan, B., Shi, W., Cao, Z., & Chua, T.-S. (2023). LGM³A '23: 1st Workshop on large generative models meet multimodal applications. *MM 2023- Proceedings of the 31st ACM International Conference on Multimedia*, 9744–9745. https://doi.org/10.1145/3581783.3610953.

Wischnewski, M., Krämer, N., Janiesch, C., Müller, E., Schnitzler, T., & Newen, C. (2024). Seal we trust? Investigating the effect of certifications on perceived trustworthiness of AI systems. In *Human-machine communication*, *8*, 141–161. https://doi.org/10.30658/hmc.8.7.

Wulf, A. J., & Seizov, O. (2024). "Please understand we cannot provide further information": Evaluating content and transparency of GDPR-mandated AI disclosures. *AI and Society*, *39*(1), 235–256. https://doi.org/10.1007/s00146-022-01424-z.

Xu, S., Luo, Y., Dauwels, J., Khong, A., Wang, Z., Chen, Q., Cai, C., Shi, W., & Chua, T.-S. (2024). LGM³A 2024: The 2nd Workshop on Large Generative Models Meet Multimodal Applications. *LGM3A 2024- Proceedings of the 2nd Workshop on Large Generative Models Meet Multimodal Applications*, 1–3. https://doi.org/10.1145/3688866.3696056.

Yandrapalli, V., & Sharma, S. (2024). Data Governance in the Age of AI, Cybersecurity, Ethics, Sustainability, and Globalization: Challenges and implications. *15th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2024, 2*, 3763–3771.

Zerka, F., Urovi, V., Vaidyanathan, A., Barakat, S., Leijenaar, R. T. H., Walsh, S., Gabrani-Juma, H., Miraglio, B., Woodruff, H. C., Dumontier, M., Dumontier, M., & Lambin, P. (2020). Blockchain for privacy preserving and trustworthy distributed machine learning in multicentric medical imaging (C-DistriM). *IEEE Access*, *8*, 183939–183951. https://doi.org/10.1109/ACCESS.2020.3029445.

Žlahtič, B., Završnik, J., Blažun Vošner, H., & Kokol, P. (2024). Transferring black-box decision making to a white-box model. *Electronics (Switzerland)*, *13*(10), 1895. https://doi.org/10.3390/electronics13101895.

2 Technological and Digital Challenges for Efficient Substitution of Creative Environment in the Era of Generative AI

Rafał Kasprzak and Marta Ziółkowska

2.1 Introduction

The idea of creativity as a hallmark of the human mind has long intrigued scholars. Since the emergence of the "creative class" theory (Florida, 2002), the significant contribution of this social group and trait to economic progress has been progressively acknowledged. It is accepted that aggregation of such individuals in urban areas positively drives innovation potential and the growth of advanced technology industries (Hansen et al., 2009; Hoyman & Faricy, 2009; Sleuwaegen & Boiardi, 2014), displaying unique attributes divergent from traditional industrial era workers (Lisakova, 2020).

Creativity's increasing importance necessitates exploring factors that enhance its potential. Current research extends in intriguing cognitive directions, striving to answer queries such as: how can we stimulate individual creativity, foster the creative process, and modify environmental factors to encourage creativity? Studies seeking factors fostering individual creativity highlight the influence of familial resources (Rusu, 2019; Wolska-Dlugosz, 2015), an educational system focused on skill development (Masadeh, 2021; Rotaru, 2020; Vejian et al., 2016), and personality traits (Afshari et al., 2013).

The emergence of generative AI has significantly influenced the dynamics of hybrid work environments, where employees split their time between remote and in-office work. This blend of work settings presents unique opportunities and challenges for creativity. Below, we explore how generative AI impacts creativity in these hybrid environments. Generative AI tools facilitate collaboration among team members who may be geographically dispersed. By providing platforms for real-time brainstorming, idea generation, and content creation, AI helps maintain creative momentum, regardless of physical location. AI can aggregate insights and suggestions from multiple sources, enriching the creative process. In hybrid teams, this diversity of input can lead to more innovative solutions as team members contribute unique perspectives from different locations (Richter & Richter, 2024).

DOI: 10.4324/9781003586944-3

2.2 Review of Literature – Creative Work Methodologies

Understanding the dynamics of the creative process aids in the identification of ideal work methodologies, which, when consistently applied, yield optimum outcomes. Methods such as the Creative Problem-Solving technique (Lumsdaine & Lumsdaine 1995) are beneficial not just in innovative academic programs (Gerhart & Carpenter 2008) but also in the economic realm (Fernald & Nickolenko 1993; Finke et al., 1992). Further research on creativity has generated numerous techniques to stimulate creativity and apply it to solving socio-economic issues (Kasprzak, 2022). These techniques can effectively shape modern business models (Kasprzak et al., 2020).

Conversely, studies aimed at environments stimulating creativity strive for work methodologies that bypass traditional creativity inhibitors such as structured frameworks, time restrictions, stringent regulations, monotonous tasks, and conventional workplaces (Dul & Ceylan, 2011).

Undeniably, the workplace ambiance created within an organization plays a substantial role in modern entities (Madjar et al., 2002; Samani et al., 2014). The correct implementation is being analyzed ergonomically (Haner, 2005; Oldham & Cummings, 1996). Indeed, creating a work milieu that enhances creativity is a crucial aspect of space planning (Costa et al., 2009; de Croon et al., 2005).

In hybrid environments, team members from diverse disciplines can collaborate more easily with the help of AI. Generative AI can facilitate discussions between designers, marketers, and engineers, promoting interdisciplinary innovation that can lead to groundbreaking ideas. Generative AI complements, rather than replaces, human creativity. By automating mundane tasks and providing suggestions, AI allows human creators to focus on their work's emotional and conceptual aspects, which is crucial for genuine creativity (Ivcevic & Grandinetti, 2024). As teams increasingly rely on AI-generated content, it is essential to maintain a balance between AI assistance and human input. Ethical considerations regarding originality, ownership, and bias must be addressed to ensure creativity remains authentic and inclusive (Dwivedi et al., 2021).

Studies and analyses have identified comprehensive models aiming to stimulate creativity holistically. For instance, the "Four P's" model (Rhodes, 1961) encompasses person (creative individuals), product (creative outputs), process (creative procedures), and press (creativity enablers).

This model promotes a holistic system for enhancing creativity organizationally, which can encapsulate:

- Person: Recruiting individuals with creative potential, distinct cognitive properties, experiences, and skills.
- Product: Making the organization conducive to developing and introducing novel and significantly enhanced products.

- Process: Creating an environment that favors creative work methodologies.
- Press: Identifying environmental elements that impede creativity – both socially and physically.

Moreover, the "Five A's" model (Glăveanu, 2013) extends the above conceptualization by incorporating Actor, Action, Artifact, Audience, and Affordances. In the model, the following changes are proposed:

- Person (with the focus on internal attributes of the person) modified into Actor (with the focus on personal attributes in relation to a societal context).
- Product (with the focus on features of products or consensus around them) modified into Artifacts (with the focus on cultural context of artifact production and evaluation).
- Process (with the focus on primarily cognitive mechanisms) modified into Action (with the focus on coordinated psychological and behavioral manifestation).
- Press (with the focus on the social as an external set of variables conditioning creativity) was divided into Audience and Affordances (with the focus on the interdependence between creators and a social and material world).

The 5xA concept is grounded in contemporary scholarly literature spanning socio-cultural and ecological psychology and the distributed theory of mind. The primary objective of this model is to establish a holistic and consolidated view of the creative process. It encompasses five integral elements that drive creative thought and action:

- Actors – Include individuals and social and cultural factors involved in the creative process. They may be people, groups, or organizations contributing their unique viewpoints and experiences.
- Action – These comprise the tasks performed by the actors, their problem-solving methods, and the steps they undertake throughout the creative process.
- Artifact – Refers to tools, materials, and objects leveraged in the creative process, such as technologies, sketches, documents, or any other aids supporting creative endeavors.
- Audience – These are the beneficiaries of the creative outcomes, such as individuals, groups, or entities that reap the advantages of the work or idea.
- Affordances – Pertain to the conditions, context, and environment that can steer creativity and the constraints in the creative process.

Numerous studies have established that the physical environment significantly influences creativity (Amabile, 1996; Shalley & Gilson, 2004; Woodman et al., 1993). They have underscored certain factors within the physical environment that invigorate creativity, including visual detail complexity, natural surroundings, natural material utilization, restrained use of fabulous shades, and limited usage of manufactured or composite surface materials (McCoy & Evans, 2002).

Environmental aspects such as potted plants (Shibata & Suzuki, 2004; Studente et al., 2016) or accessibility to natural environments (Plambech & Bosch, 2015), colors (Lan et al., 2021; Rook, 2014), and raw materials such as wood (Ridoutt et al., 2002) have been noted as significant stimulators of creativity in empirical studies.

Beyond physical attributes, organizational culture is a pivotal element influencing creativity, potentially enhancing employee creativity. Extensive literature (Hermida et al., 2019; Kiumarsi et al., 2015; Taha et al., 2016) has highlighted certain traits of the organizational culture that promote positive shifts, including psychological safety, healthy interpersonal relationships, effective communication, and efficient employee motivation.

2.3 Methodology

The identified factors positively impacting employee creativity provide the foundation for developing a research model that underpins the empirical research conducted (Figure 2.1).

The proposed research model hypothesizes that an ideal mix of social environment factors (such as an organizational culture that values creativity) and physical environment factors (for instance, the design of employee workspaces) can enhance employee creativity in the existing stationary work model. Furthermore, it is postulated that specific factors could stimulate an employee's creativity in the hybrid work model and effectively replace the current creative stimuli in the stationary model.

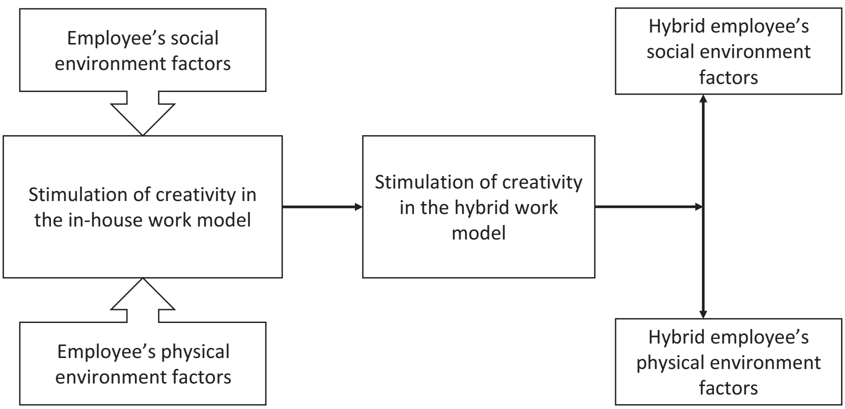

Figure 2.1 Research model

Source: Author's research

The primary research questions stemming from this model cover the following areas of inquiry:

- What are the most effective social environment factors for fostering employee creativity in the stationary work model?
- What are the most effective physical environment factors for fostering employee creativity in the stationary work model?
- What are the most effective social environment factors for fostering employee creativity using Information and Communication Technology (ICT) tools in the hybrid work model?
- What are the most effective physical environment factors for fostering employee creativity using ICT tools in the hybrid work model?
- Using ICT tools, which social environment factors from the stationary work model can be replaced by elements from the hybrid work model?
- Which physical environment factors from the stationary work model can be replaced by elements from the hybrid work model using ICT tools?

The initial empirical research was qualitative, and in-depth interviews were employed to identify critical areas for quantitative analysis in the subsequent phase.

The research project deployed a qualitative approach for preliminary investigation and data analysis, referencing the symbolic-interactionist paradigm. Qualitative research involves gathering and analyzing non-quantitative data (such as text) to comprehend concepts, opinions, or experiences. It is primarily a research methodology in which the researcher progressively understands a social phenomenon by comparing, replicating, cataloging, and classifying the study object (Miles & Huberman, 1994). The results obtained from qualitative research are descriptive, typically presented in words (often using participants' terminology) or images rather than numerically (Fraenkel & Wallen, 1990; Marshall & Rossman, 2022). It emphasizes understanding the perceptions and experiences of participants (Fraenkel and Waller, 1990). This technique allows for exploring multiple realities (Lincoln & Guba, 1985), aiding researchers in understanding how events occur (Fraenkel & Wallen, 1990).

In the described study, a semi-structured, in-depth interview technique was implemented based on a predetermined interview process with established protocols for the interviewer and response collection. This approach benefits the research process, enabling the interviewer to maintain consistency while retaining flexibility (Gaber, 2020).

The data collected were analyzed using MAXQDA 2020, a computer-aided qualitative data analysis software. Themes or patterns in the data can be detected via two primary methods of thematic analysis: "bottom-up" or inductive and "top-down" or deductive, also known as theoretically. Inductive analysis is a data-driven method where data are coded without attempting to fit it into an existing coding frame or the researcher's analytical presupposition (Braun & Clarke, 2006). A combination of inductive and deductive coding provides a more thorough data analysis and validates the research process.

2.4 Results

2.4.1 Features of a Creative Space

The initial area of analytical inquiry pertained to the attributes required in a creative space. Participants associated "creative space" with vivid, appealing colors and comfortable regions encouraging relaxation, where small and large groups can work together. Primary attributes linked with creative environments include vegetation, recreational games or activities, technological and multimedia elements, and the absence of a typical office feel. According to the survey participants, a creative space should foster idea exchanges and brainstorming, featuring ingredients such as flipcharts and writable surfaces for notetaking and spaces for informal conversations. It should also have designated areas for different types of tasks.

Beyond color schemes, respondents also associated creative environments with other sensory elements like light, fragrances, and calming music. Comfort and ergonomic features like a meal area, a fridge with beverages, a café, a library, and even a gym were also considered essential.

However, it is essential to note that, for some respondents, a "creative space" was not a physical entity but rather an attitude – being part of an organization that encourages proactive behavior and sharing unique ideas and solutions. Some viewed creative space as a virtual concept, a shared online space for exchanging ideas, signifying adaptive and varying qualities, as creative work can be undertaken in myriad locations and circumstances. "In my interpretation of the term 'creative space' in our organization, I think it mainly refers to employee rowth. We aim to develop our employees, not restrict them. Although everyone associates the concept of creative space with artistic activities such as painting or inventing, in our organization, it signifies an environment where an individual can freely express new, improved, and interesting ideas" (Respondent 5).

Other participants proposed a varied definition for creative space: "From my experience, I tend to engage in creative work across diverse settings, each with its unique advantages, disadvantages, and requirements... Each task has different demands. For instance, brainstorming about the budget differs from crafting a persuasive post or pondering organizational issues. Hence, it is not feasible to relate one task to a miniature golf setting, another to a large balcony, and a third to a plush sofa or spaces like those in Google" (Respondent 3).

2.4.2 Measuring the Effectiveness of Creative Space

Another research dimension explored was the evaluation of creative spaces' efficacy. According to participants, proposing methods to quantify the effectiveness of creative spaces posed a significant challenge. Many suggested that measuring the effectiveness of creative environments is virtually impossible, primarily due to the need for more suitable data for such assessment or owing to the specificity of individual situations and people.

However, other participants noted that since new ideas inherently involve risks, organizations should embrace these risks, recognizing that not all fresh ideas – including those generated within creative spaces – will yield success. Interestingly, participants suggested several ways to evaluate a creative space's productivity. One proposed method involved contrasting the quality or effectiveness of work produced in a creative area with work conducted in a different environment. "If we were granted the opportunity to examine work produced in distinct spaces, we could indeed measure the promptness and quality of project delivery" (Respondent 1).

Participants further proposed that an effective strategy to gauge the effectiveness of a creative space might encompass an evaluation of the outcomes of the work completed within such a space, such as the idea or product produced. To measure the creative space's efficiency, participants also suggested conducting employee interviews, monitoring the frequency of the creative space's usage, observing whether employee integration and communication quality improve because of the creative freedom, and evaluating the individual elements of the creative space.

2.4.3 The Use of Technology in Creative Spaces

Subsequently, the research attention was steered toward the role of technology within creative environments. Technology presents a vital facet in both tangible and virtual creative areas. Per participants' responses, display screens are fundamental technological equipment within creative rooms. Technical elements facilitating interactive and collaborative work on specific tasks were deemed paramount. Such components include interactive whiteboards, games, rapid data acquisition tools, software supporting idea exchange, a shared drive or email for disseminating ideas, and devices enabling ergonomic and creative remote work.

Participants also highlighted the potential advantages of integrating emerging technologies – like artificial intelligence and virtual reality – into creative spaces. Other technological elements suggested included computers, tablets, satellite TVs, music-triggering sensors, and other equipment associated with creative spaces.

However, it is worth noting that some participants suggested maintaining a separation between the technological zone and the rest of the creative space or even limiting the number of electronic devices, arguing that their presence might hinder creativity. "Our experience over the years has demonstrated that technology may not invariably foster creativity; the opposite may hold. Therefore, perhaps delineating technology-free areas within creative spaces could also be beneficial, positively impacting employees' creativity within such a workspace" (Respondent 1).

2.5 Discussion

Academics and industry practitioners have steadfastly attempted to understand how organizations can boost innovation capabilities by cultivating organizational creativity. In this regard, creating innovative workplaces is progressively considered a significant business practice for building inventive organizations. While fostering creativity is worthwhile, organizations must direct their creative efforts toward fruitful innovation to maintain competitiveness.

Technological advancements in information and communication in the last few decades have led to a transformative process, making individuals and society more interconnected. The widespread adoption and democratization of digital technologies have spurred the general practice of creative activity, spawning an unparalleled array of elements in the digital domain available for creative input and response (Literat, 2018).

It is also noteworthy that emerging active digital technologies, like generative design, assist humans in spawning a myriad of unforeseen solutions, necessitating new thought processes and skills. In this context, the primary responsibility of the designer is to foresee potential outcomes and devise the approach to generate them using these technologies, collaborating with different stakeholders possessing diverse technical abilities.

Generative AI has the potential to significantly enhance creativity in hybrid work environments by facilitating collaboration, increasing efficiency, and fostering innovation. However, organizations must balance leveraging AI capabilities and nurturing human creativity. By addressing the challenges associated with AI integration, teams can create a dynamic and inspiring creative culture that thrives in a hybrid setting.

Companies are compelled to tackle the challenge of crafting innovative workspaces that attract proficient employees and amplify their creativity. Well-known companies like Google, Apple, and Facebook (Nowadays: Meta), renowned for their novel workspace strategies, offer the most conspicuous examples. While numerous other organizations are investing in such workspace designs, a surefire formula for designing workspaces that optimally bolsters employee creativity needs to be built.

Certain workspace features have consistently been demonstrated to enhance creativity. These include plant life, pleasant sounds and smells, window views, adaptable furniture, a suitable office size, an open office layout complemented with private spaces, and relaxation areas.

2.6 Conclusion

Organizations should aim to design workplaces that accommodate various modes of creative work. This implies that workplaces should offer open spaces for teamwork and communication and private rooms for concentrated work. This equilibrium can be maintained by incorporating elements that allow space modifications per situational requirements, like sliding walls or adaptable furniture (such as lightweight chairs, tables, whiteboards, or wheeled items). Personal workspaces should be spacious enough to offer privacy for employees. Moreover, organizations should provide relaxation zones, break areas, and doodling spaces. While these spaces might initially seem to distract employees from their actual work, they provide necessary incubation time for issues that pose challenges to employees. Further, these spaces systematically promote interaction among employees (even across different teams), leading to situations that facilitate the exchange of valuable information and experiences.

References

Afshari, M., Ghavifekr, S., Siraj, S., & Razak, A. Z. (2013). Developing students' creativity. *Life Science Journal, 10*(3), 485–494.

Amabile, T. M. (1996). *Creativity in context: Update to the social psychology of creativity*. Routledge

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa.

Costa, P., Seixas, J., & Roldão, A. (2009). From 'creative cities' to 'urban creativity'? Space, creativity, and governance in the contemporary city. *Working paper*. 2009/80, 1–32. https://repositorio.iscte-iul.pt/bitstream/10071/3231/1/DINAMIA_WP_2009-80.pdf

De Croon, E., Sluiter, J., Kuijer, P. P., & Frings-Dresen, M. (2005). The effect of office concepts on worker health and performance: A systematic review of the literature. *Ergonomics*, *48*(2), 119–134. https://doi.org/10.1080/00140130512331319409.

Dul, J., & Ceylan, C. (2011). Work environments for employee creativity. *Ergonomics*, *54*, 12–20. https://doi.org/10.1080/00140139.2010.542833.

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. https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

Fernald, L. W., & Nickolenko, P. (1993). The creative process: Its use and extent of formalization by corporations. *Journal of Creative Behaviour, 27*(3), 214–220.

Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: Theory, research and application*. The MIT Press.

Florida, R. (2002). *The rise of the creative class (Vol. 9).* Basic books.

Fraenkel, J. R., & Wallen, N. E. (1990). *How to design and evaluate research in education.* McGraw-Hill Education.

Gaber, J. (2020). *Qualitative analysis for planning & policy: Beyond the numbers*. Routledge.

Gerhart, A. L., & Carpenter, D. D. (2008). Creative Problem-Solving Course – Student Perceptions of Creativity and Comparisons of Creative Problem-Solving Methodologies. *Proceedings of the 2008 ASEE Annual Conference and Exposition*.

Glăveanu, V. P. (2013). Rewriting the language of creativity: The five A's framework. *Review of General Psychology*, *17*, 69–81. https://doi.org/10.1037/a0029528.

Haner, U. (2005). Spaces for creativity and innovation in two established organizations. *Creativity and Innovation Management*, *14*(3), 288–298. https://doi.org/10.1111/J.1476-8691.2005.00347.X.

Hansen, H. K., Asheim, B., & Vang-Lauridsen, J. (2009). The european creative class and regional development: How relevant is florida's theory for europe?. In Kong, L. & O'Connor, J. (Eds.), *Creative economies, creative cities. Asian –european perspectives* (pp. 99–120). Springer.

Hermida, Y., Clem, W., & Güss, C. D. (2019). The inseparable three: How organization and culture can foster individual creativity. *Frontiers in Psychology*, *10*, 2133. https://doi.org/10.3389/fpsyg.2019.02133.

Hoyman, M., & Faricy, C. (2009). It takes a village: A test of the creative class, social capital, and human capital theories. *Urban Affairs Review*, *44*(3), 311–333.

Ivcevic, Z., & Grandinetti, M. (2024). Artificial intelligence as a tool for creativity. *Journal of Creativity*, *34*(2), 100079. https://doi.org/10.1016/j.yjoc.2024.100079.

Kasprzak, R. (2022). Design thinking as a creative process of building the value [Design Thinking jako twórczy proces budowania wartości]. In Bombol, M. (Ed.), *About changing world of values in economics, finance and management* [*O zmieniającym się świecie wartości w ekonomii, finansach i zarządzaniu*] (pp. 225–244). Oficyna Wydawnicza SGH.

Kasprzak, R., Tomaszewski, A., & Ziółkowska, M. (2020). Operationalisation of businesss model in the creative industry – initial findinds [Operacjonalizacja modelu biznesu w przemysłach kreatywnych – wstępne wyniki badań]. In Urbanek, G. & Gregorczyk, S. (Eds.), *Strategic management in the age of digital network economy* [*Zarządzanie strategiczne w dobie cyfrowej gospodarki sieciowej*] (pp. 351–364). Wydawnictwo Uniwersytetu Łódzkiego. https://dx. doi.org/10.18778/8220-335-6.21.

Kiumarsi, S., Isa, S. M., & Navi, R. R. (2015). *The Influence of Organizational Culture on Creativity and Innovation: A Review.* https://www.semanticscholar.org/paper/The-Influence-of-Organizational-Culture-on-and-A-Kiumarsi-Isa/8a36aa314ca704854d76492da02a5e92263eb4b9.

Lan, L., Hadji, S., Xia, L., & Lian, Z. (2021). The effects of light illuminance and correlated colour temperature on mood and creativity. *Building Simulation*, *14*, 463–475. https://doi.org/10.1007/s12273-020-0652-z.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.

Lisakova, I. (2020). Формування креативного класу як інтенція розвитку вищої освіти. *Естетика і етика педагогічної дії*. [Formation of the creative class as the intention of the development of higher education]. *Aesthetics and Ethics*, *22*, 42–54. https://doi.org/10.33989/2226-4051.2020.22.221989.

Literat, I., & Glaveanu, V. P. (2018). Distributed creativity on the internet: A theoretical foundation for online creative participation. *International Journal of Communication*, *12*, 893–908.

Lumsdaine, E., & Lumsdaine, M. (1995). Creative problem solving. *IEEE Potentials*, *13*(5), 4–9. https://doi.org/10.1109/45.464655.

Madjar, N. Y., Oldham, G. R., & Pratt, M. G. (2002). There is no place like home? The contributions of work and nonwork creativity support to employees' creative performance. *Academy of Management Journal*, *45*, 757–767.

Marshall, C., Rossman, G. B., & Blanco, G. L. (2022). *Designing Qualitative Research*. SAGE Publications.

Masadeh, T. S. (2021). Teaching practices of EFL Teachers and the enhancement of creative thinking skills among learners. *International Journal of Asian Education, 2*(2), 153–166. https://doi.org/10.46966/IJAE.V2I2.173.

McCoy, J. M., & Evans, G. W. (2002). The potential role of the physical environment in fostering creativity. *Creativity Research Journal*, *14*, 409–426. https://doi.org/10.1207/S15326934CRJ1434_11.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: A sourcebook of new methods*. Sage.

Oldham, G. R., & Cummings, A. (1996). Employee creativity: Personal and contextual factors at work. *Academy of Management Journal*, *39*, 607–634.

Plambech, T., & Bosch, C. C. (2015). The impact of nature on creativity – A study among Danish creative professionals. *Urban Forestry & Urban Greening*, *14*, 255–263. https://doi.org/10.1016/J.UFUG.2015.02.006.

Rhodes, M. (1961). An analysis of creativity. *The Phi delta Kappan*, *42*(7), 305–310.

Richter, A., & Richter, S. (2024). Hybrid work – a reconceptualisation and research agenda. *I-com*, *23*(1), 71–78. https://doi.org/10.1515/icom-2023-0027.

Ridoutt, B. G., Ball, R. D., & Killerby, S. K. (2002). Wood in the interior office environment: Effects on interpersonal perception. *Forest Products Journal*, *52*, 23–30.

Rook, L. (2014). Exposure to the color red enhances creative thinking depending on appetitive-aversive cues. *Creativity Research Journal*, *26*(1), 124–130. https://doi.org/10.1080/10400419.2014.873672.

Rotaru, R. E. (2020). Stimulating primary school children's creativity. *Revista Romaneasca Pentru Educatie Multidimensionala*, *12*(4), 431–439. https://doi.org/10.18662/rrem/12.4/355.

Rusu, M. (2019). Methods of individual creativity stimulation. *Review of Artistic Education*, *18*, 275–286.

Samani, S. A., Rasid, S. Z., & Sofian, S. (2014). A workplace to support creativity. *Industrial Engineering and Management Systems*, *13*, 414–420. https://doi.org/10.7232/IEMS.2014.13.4.414.

Shalley, C. E., & Gilson, L. L. (2004). What leader need to know: A review of social and contextual factors that can foster or hinder creativity. *Leadership Quarterly*, *15*, 33–53. https://doi.org/10.1016/J.LEAQUA.2003.12.004.

Shibata, S., & Suzuki, N. (2004). Effects of an indoor plant on creative task performance and mood. *Scandinavian Journal of Psychology*, *45*(5), 373–381. https://doi.org/10.1111/J.1467-9450.2004.00419.X.

Sleuwaegen, L., & Boiardi, P. (2014). Creativity and regional innovation: Evidence from EU regions. *Research Policy*, *43*, 1508–1522.

Studente, S., Seppala, N., & Sadowska, N. (2016). Facilitating creative thinking in the classroom: Investigating the effects of plants and the color green on visual and verbal creativity. *Thinking Skills and Creativity*, *19*, 1–8. https://doi.org/10.1016/j.tsc.2015.09.001.

Taha, V. A., Šírková, M., & Ferencová, M. (2016). The impact of organisational culture on creativity and innovation. *Polish Journal of Management Studies*, *14*, 7–17. https://doi.org/10.17512/PJMS.2016.14.1.01.

Vejian, G., Kamarudin, N., & Kadir, S. A. (2016). School creative climate: Factors influence fostering creativity school. *International Journal of Education and Training*, *2*(1), 1–5.

Wolska-Długosz, M. (2015). Stimulating the development of creativity and passion in children and teenagers in Family and School Environment – Inhibitors and opportunities to overcome them. *Procedia - Social and Behavioral Sciences*, *174*, 2905–2911. https://doi.org/10.1016/j.sbspro.2015.01.1027.

Woodman, R., Sawyer, J. E., & Griffin, R. W. (1993). Toward a theory of organizational creativity. *Academy of Management Review*, *18*, 293–321. https://doi.org/10.5465/AMR.1993.3997517.

3 Navigating Ethical Dilemmas

Unveiling Greenwashing in the AI Era

Małgorzata Wiktoria Paprocka

3.1 Introduction

In the era of rapid technological advancement, artificial intelligence (AI) is increasingly permeating various aspects of life, including the business sphere. Companies worldwide are leveraging AI to automate and optimize processes, decision-making, and operational efficiency. This leads to competitive advantages; however, as AI becomes more integrated into business operations, new challenges and ethical dilemmas arise. Despite its undeniable innovation, AI can lead to significant issues related to privacy, transparency, and algorithmic biases.

Among these challenges is the issue of greenwashing, where companies mislead consumers about their environmental efforts. In the pursuit of improving their image, firms often use AI to create highly targeted marketing campaigns that do not always reflect reality or manipulate consumers in other ways through specialized algorithms. On the other hand, AI can also support ethical business practices by monitoring eco-friendly initiatives and ensuring greater transparency in reporting.

3.2 Methodology and Foundations

The aim of this chapter is to examine how AI impacts ethical management in business and to explore its role in the phenomenon of greenwashing. Key ethical dilemmas will be addressed, along with recommendations for actions that can enhance transparency and trust in the relationships between companies and their stakeholders. The author presents two research questions: "What are the ethical challenges of using AI in business?" and "How can AI both support and mitigate the phenomenon of greenwashing?" These questions cover two central aspects of the article: an analysis of ethical dilemmas in the context of AI, and an investigation of the role AI plays in the transparency and manipulation of information related to greenwashing activities. The following hypotheses are also proposed: (1) The use of AI in corporate marketing and reporting can contribute to greenwashing practices by enabling more precise manipulation of information

DOI: 10.4324/9781003586944-4

regarding companies' activities; (2) AI, when properly implemented, monitored, and supervised, can effectively enhance transparency and counter greenwashing by analyzing and verifying data related to companies' operations.

A systematic literature review was conducted using an advanced search technique applied to the Scopus scientific database. The author employed the following search code:

TITLE-ABS-KEY ((“*washing” OR “business ethics”) AND (“AI” OR “artificial intelligence”)) AND (LIMIT-TO (SRCTYPE, “j”)) AND (LIMIT-TO (DOCTYPE, “ar”)) AND (LIMIT-TO (SUBJAREA, “BUSI”) OR LIMIT-TO (SUBJAREA, “ECON”)) AND (LIMIT-TO (LANGUAGE, “English”))

The following keywords were used for the literature review: “*washing” [representing any words ending in -washing, for example greenwashing, socialwashing], “business ethics,” “artificial intelligence,” and its abbreviation “AI.” The author limited the search to scientific articles in English, as well as to subareas: Management and Accounting and Economics, Econometrics, and Finance. Only 26 articles on this topic were found. Twenty-five articles obtained from the search were relevant to the research questions under investigation. Given the limited number of sources, only one – completely irrelevant to the subject matter – was excluded from the list. The selected studies primarily focused on ethical considerations in the application of AI, with only five articles directly addressing the issue of greenwashing. From 2019 onward, two to four articles per year were published, with a marked increase in 2024, as illustrated in the chart in Figure 3.1.

Three sources outside the analyzed sample of articles were added to complement the review of ethical dilemmas related to the AI subfield and the subsection on responsible business. The content of the identified articles was critically analyzed by the author.

The foundation part aims to provide the context of the article, focusing on the ethical implications of AI in business development and its role in the phenomenon of greenwashing. AI is becoming an integral part of modern business, bringing

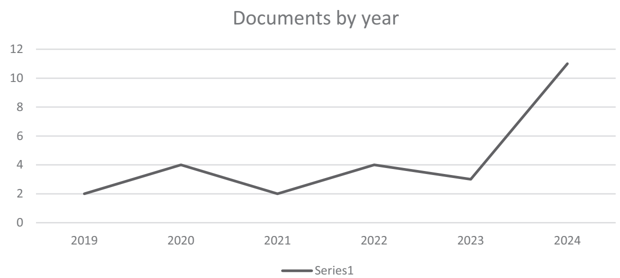

Figure 3.1 Number of documents per year

Source: Scopus database, access: 09.2024.

benefits such as process automation, real-time data analysis, and personalized services. Studies show that companies using AI can improve production efficiency and achieve higher levels of operational efficiency and innovation (Kulkarni et al., 2024; Zhang, 2024). AI in business is currently being applied in logistics, retail, finance, marketing, as well as the health and service sectors. In retail, in particular, AI enables personalization of offers and improved customer service. In the financial sector, AI supports risk management, anomaly identification, and transaction automation, which not only increases efficiency but also minimizes human error. Moreover, in marketing, AI enables advanced recommendation systems that improve the customer experience and increase sales conversions.

Dynamic development of AI is not without ethical challenges. Ethical concerns primarily arise from the fact that AI relies on large datasets, which raises questions about privacy, transparency, and accountability for decisions made by algorithms. Accountability for decisions made based on AI is a key ethical issue. When AI systems make mistakes or produce biased results, it may not be clear who is responsible – whether it is the developers, users, or the AI system itself.

Greenwashing, the practice of misleading consumers about the environmental benefits of a company's product or activities, is becoming a growing problem in the age of data-driven communications and marketing. More and more companies are emphasizing their environmentally friendly actions in order to gain consumer trust. The phenomenon of greenwashing is compounded by AI tools that can be used to manipulate information, making it difficult for consumers to judge whether products actually meet green claims. AI, through advanced data analysis and content personalization techniques, enables the creation of marketing campaigns that more effectively reach specific groups of consumers, often misleading them about the actual environmental impact of products. Nevertheless, AI has the potential to both support and mitigate these practices.

3.3 Ethical Dilemmas in AI – Overview

AI, relying on vast amounts of private data, requires appropriate mechanisms to protect it so that users' privacy is not violated. Research indicates that a significant threat is the collection and storage of data without the consent or full awareness of consumers, which can lead to violations of RODO and other data protection laws (Goncalves et al., 2024). An example of abuse is the collection of biometric data by companies that use AI to analyze customer behavior without their knowledge. In the event of a data breach, the consequences can be severe, not only legally, but also in terms of image, which can lead to a loss of consumer trust (Hu & Min, 2023).

The problem with many AI systems is the so-called black box – the lack of transparency about how the algorithms work and make decisions. Transparency in this area is key to preventing loss of consumer trust. Technologies such as explainable AI, which explains the operation of algorithms in a way that the user can understand, are currently being developed (Akhtar et al., 2024; Martin & Waldman, 2023). Nevertheless, it still remains a significant challenge and risk. Companies implementing AI need to create mechanisms for reporting and explaining how these systems work in order to build trust among stakeholders.

AI, trained on historical data, often inherits biases present in that data, leading to discrimination (Martin & Waldman, 2023). There can also be a deliberate, intentional insertion of biased information into the database on which a given AI model operates – information that is intended to promote, to favor a certain ideology, opinion, approach. Research indicates that one solution is to train algorithms on more diverse datasets and apply techniques to detect and eliminate biases. In addition, companies are increasingly auditing algorithms to detect potential biases in AI systems.

The issue of accountability for decisions made by AI is complex, especially when AI makes decisions autonomously. Research suggests that companies implementing AI should create clear accountability structures for potential algorithmic errors and resulting damages. In the case of algorithmic errors, especially those leading to financial damages or losses to consumers, it is important to determine who should be held accountable – is it the creator of the algorithm, the company that implemented the AI, or perhaps the software itself? It is also important to determine whether algorithmic decision-making systems are considered legitimate at all (Martin & Waldman, 2023). In the health and finance sectors, this liability is crucial, as decisions made by AI can directly affect people's lives (Tóth & Blut, 2024).

The development of AI toward greater technological autonomy raises questions about the limits of control over AI systems. Autonomous cars are an example of technology that requires advanced AI decision-making systems, but at the same time, there is a need to maintain human oversight to prevent AI from making inappropriate decisions without the possibility of human correction. The literature emphasizes the need to develop technologies that ensure AI, even when operating autonomously, remains under human supervision and does not make decisions that may conflict with ethical norms (Ferrell & Ferrell, 2021).

3.4 Responsible Business in the Context of AI

An increasing number of companies are engaging in non-financial reporting, which encompasses activities related to sustainability and corporate social responsibility. In the context of AI, analytical tools enable the automated collection of environmental impact data, allowing firms to monitor CO_2 emissions, water usage, and carbon footprints in real time (Lim, 2024). The use of AI in ESG (Environmental, Social, Governance) reporting leads to more precise and comprehensive environmental reports, which are increasingly being analyzed by institutional investors. AI also assists in identifying areas that require improvement, enabling companies to make more informed strategic decisions regarding sustainability. A notable example is the implementation of AI in supply chains, where analytical systems monitor resource consumption at various stages of production and identify opportunities to reduce pollutant emissions.

The dialogue with stakeholders, including customers, investors, regulators, and local communities, is becoming increasingly complex, and AI tools enable its more effective optimization. Public opinion analysis systems, such as social media monitoring, allow companies to quickly identify stakeholder sentiments and respond to the market's evolving needs. For instance, AI can analyze customer feedback on e-commerce platforms, enabling firms to continuously adapt their products and services to meet customer expectations. Literature on stakeholder dialogue highlights that AI not only enhances external communication but also supports the automation of internal processes. AI systems allow companies to swiftly address concerns raised by stakeholders and to adjust internal policies in response to their needs. Moreover, the ability to analyze large datasets in real-time facilitates proactive stakeholder engagement, ensuring that companies are better equipped to maintain trust and foster long-term relationships and trust (Lai & Lee, 2020).

At the global level, there are a growing number of initiatives aimed at establishing ethical principles for the implementation of AI. One notable example is UNESCO's guidelines on AI ethics, which emphasize the need for transparency, accountability, and adherence to human rights. Similarly, the European AI Alliance promotes legal and technical frameworks designed to ensure that AI is used in a manner consistent with ethical standards and data protection regulations. The integration of such ethical frameworks into corporate governance structures is seen as essential for fostering public trust and ensuring that AI innovations contribute positively to society. All these processes represent not only huge opportunities, but also challenges of the modern business world and humanity in general (Miśkiewicz, 2019).

3.5 AI vs Greenwashing

This section explores the relationship between AI and greenwashing, considering both the negative and positive impacts AI can have in this context. The analysis incorporates findings from scientific research, which illustrate how this relationship manifests in practice. AI can be employed to manipulate information, particularly in product marketing or in non-financial reporting (de Villiers et al., 2024). Examples from the fashion industry show that companies often use AI to create personalized advertising campaigns that mislead consumers about the actual environmental impact of their products. Such marketing techniques may also involve concealing actual production practices. The literature emphasizes that AI, with its ability to analyze vast amounts of data and generate targeted messages, can contribute to the escalation of greenwashing. Consumers are often unaware that marketing messages are so precisely targeted, making it difficult for them to verify the veracity of the information presented. Such behavior even finds its own name in the literature – "machinewashing" (Seele & Schultz, 2022).

On the other hand, AI can also be used to identify instances of greenwashing by analyzing large datasets related to companies' environmental performance and assessing whether their claims align with their actual practices. Technologies such as natural language processing (NLP) enable the scanning of ESG reports to detect inconsistencies between a company's statements and its real-world actions. In a particularly useful way, given the lack of uniform rapport standards, only AI-based models can cope with processing so much heterogeneous data (Antoncic, 2020).

In a study analyzing over 1,000 publicly listed Chinese companies using data from a ten-year period, the authors provide evidence that AI significantly hinders greenwashing practices. However, this effect is more pronounced in companies that are not politically connected, those with lower capital incentives and in companies where there are fewer female directors. With the help of AI, businesses can more efficiently track and validate their eco-friendly initiatives, thereby reducing the risk of misrepresenting their true environmental practices. This process works by mitigating agency issues, alleviating financial constraints, and increasing the oversight from investors and media (Li et al., 2024).

In another study, researchers describe a strong negative relationship between AI capabilities and environmental disclosures characterized by greenwashing. This effect is particularly evident in firms that are more exposed to regulatory climate risk, managed by right-leaning political managers, have stronger governance structures, or exhibit higher CEO pay-for-performance sensitivity (Jiao et al., 2024).

This suggests that the effectiveness of AI in curbing greenwashing is influenced by certain organizational and contextual factors. Therefore, it is essential to establish clear and specific regulations that will prevent, or at least hinder, greenwashing practices, regardless of the industry, company size, or management style.

3.6 Integration of Findings and Recommendations

In practice, there are examples of both companies that use AI for information manipulation (negative) and those that leverage AI to enhance transparency (positive). In the fashion sector, companies have been accused of greenwashing, as their marketing campaigns were found to mislead consumers regarding the true environmental impact of their products. For example, the H&M company is facing a lawsuit for allegedly misleading consumers with false sustainability claims, including deceptive marketing and recycling programs that overstate the environmental benefits of its products. The lawsuit follows a Quartz investigation revealing that many of H&M's sustainability profiles exaggerated the environmental impact of their products (The Sustainable Fashion Forum, 2024).

On the other hand, technology-driven companies such as Patagonia use AI to monitor their supply chain and reduce carbon emissions (Patagonia, 2024).

These contrasting examples illustrate the dual-role AI that can play in either perpetuating or combating greenwashing. While some businesses may exploit AI's ability to create targeted and persuasive marketing messages (Labrecque, 2024), others harness its analytical power to improve sustainability practices, increase accountability, and ensure that their operations align with environmental standards.

The values expressed through principles, norms, and regulatory frameworks should serve as the primary safeguard against greenwashing facilitated by AI (Ferrell et al., 2024). Meanwhile, building stakeholder trust is a key element of sustainable business in the AI era. Companies can leverage AI to foster trust by implementing transparent monitoring and reporting systems. AI can also be integrated into company management processes, particularly in the context of monitoring sustainability efforts. AI can assist in tracking sustainable actions within supply chains by analyzing environmental data in real time, ensuring greater accountability and authenticity in sustainability efforts. Companies should also implement mechanisms to prevent the misuse of AI in the context of greenwashing. This will minimize the risk of information manipulation. The literature also highlights the importance of implementing external audits to verify whether AI systems operate in accordance with principles of ethics and transparency. Recommendations in the literature also emphasize the creation of internal ethical codes regarding AI and the appointment of individuals responsible for overseeing the actions of algorithms.

3.7 Conclusion

This chapter provides a comprehensive analysis of the ethical challenges of AI in the business context, particularly in relation to greenwashing, and proposes concrete recommendations for companies and policymakers. In doing so, the author addressed the research questions and confirmed both the first (1) "The use of artificial intelligence in corporate marketing and reporting can contribute to greenwashing practices by enabling more precise manipulation of information regarding companies' activities" and second (2) "Artificial intelligence when properly implemented, monitored, and supervised, can effectively enhance transparency and counter greenwashing by analyzing and verifying data related to companies' operations' research hypotheses." On one hand, AI can be used to enhance efficiency, personalize services, and improve data analysis, leading to significant business benefits. On the other hand, without appropriate ethical frameworks and transparent processes, AI can support dishonest practices such as greenwashing, misleading consumers, and other stakeholders about the company's actual environmental impact. The analysis shows that companies must place great emphasis on the responsible implementation of AI, particularly in areas related to reporting and marketing of pro-environmental actions. It is confirmed that AI, when properly designed and managed, can contribute to increased transparency and trust in stakeholder relations by helping to identify and combat greenwashing practices. However, a lack of oversight over AI operations can lead to situations where these technologies are used to manipulate information to improve a company's image, regardless of its actual actions. Therefore, AI in business should be implemented with strong ethical principles in mind so that its potential can be fully harnessed in alignment with the values of sustainability and integrity.

A meta-analysis of the Scopus research database reveals gaps in the literature regarding research at the intersection of AI and greenwashing, for which the author encourages further research on the topic.

References

Akhtar, M. A. K., Kumar, M., & Nayyar, A. (2024). Transparency and accountability in explainable AI: Best practices. In M. A. K. Akhtar, M. Kumar, & A. Nayyar (Eds.), *Springer Book Chapter* (Vol. 551, pp. 127–164). Springer. https://doi.org/10.1007/978-3-031-66489-2_5.

Antoncic, M. (2020). Uncovering hidden signals for sustainable investing using big data: Artificial intelligence, machine learning and natural language processing. *Journal of Risk Management in Financial Institutions*, *13*(2), 106–113. https://doi.org/10.69554/CIKJ7477

de Villiers, C., Dimes, R., & Molinari, M. (2024). How will AI text generation and processing impact sustainability reporting? Critical analysis, a conceptual framework and avenues for future research. *Sustainability Accounting, Management and Policy Journal*, *15*(1), 96–118. https://doi.org/10.1108/SAMPJ-02-2023-0097.

Ferrell, O. C., & Ferrell, L. (2021). Applying the Hunt Vitell ethics model to artificial intelligence ethics. *Journal of Global Scholars of Marketing Science: Bridging Asia and the World*, *31*(2), 178–188. https://doi.org/10.1080/21639159.2020.1785918.

Ferrell, O. C., Harrison, D. E., Ferrell, L. K., Ajjan, H., & Hochstein, B. W. (2024). A theoretical framework to guide AI ethical decision making. *AMS Review*, *14*(1–2), 53–67. https://doi.org/10.1007/s13162-024-00275-9.

Goncalves, M., Hu, Y., Aliagas, I., & Cerdá, L. M. (2024). Neuromarketing algorithms' consumer privacy and ethical considerations: Challenges and opportunities. *Cogent Business and Management*, *11*(1), 2333063. https://doi.org/10.1080/23311975.2024.2333063.

Hu, Y., & Min, H. K. (2023). The dark side of artificial intelligence in service: The "watching-eye" effect and privacy concerns. *International Journal of Hospitality Management*, *110*, 103437. https://doi.org/10.1016/j.ijhm.2023.103437.

Jiao, A., Lu, J., Ren, H., & Wei, J. (2024). The role of AI capabilities in environmental management: Evidence from USA firms. *Energy Economics*, *134*, 107653. https://doi.org/10.1016/j.eneco.2024.107653.

Kulkarni, A. V., Joseph, S., & Patil, K. P. (2024). Artificial intelligence technology readiness for social sustainability and business ethics: Evidence from MSMEs in developing nations. *International Journal of Information Management Data Insights*, *4*(2), 100250. https://doi.org/10.1016/j.jjimei.2024.100250.

Labrecque, L. I., Peña, P. Y., Leonard, H., & Leger, R. (2024). Not all sunshine and rainbows: Exploring the dark side of AI in interactive marketing. *Journal of Research in Interactive Marketing*, *18*(5), 970–999. https://doi.org/10.1108/JRIM-02-2024-0073.

Lai, C. H., & Lee, F. (2020). Organizations' dialogic social media use and stakeholder engagement: Stakeholder targeting and message framing. *International Journal of Communication*, *14*, 2246–2271.

Li, D., Zhang, Z., & Gao, X. (2024). Does artificial intelligence deter greenwashing? *Finance Research Letters*, *67*, 105954. https://doi.org/10.1016/j.frl.2024.105954.

Lim, T. (2024). Environmental, social, and governance (ESG) and artificial intelligence in finance: State-of-the-art and research takeaways. *Artificial Intelligence Review*, *57*, 26. https://doi.org/10.1007/s10462-024-10708-3.

Martin, K., & Waldman, A. (2023). Are algorithmic decisions legitimate? The effect of process and outcomes on perceptions of legitimacy of AI decisions. *Journal of Business Ethics*, *183*(3), 653–670. https://doi.org/10.1007/s10551-021-05032-7.

Miśkiewicz, J. (2019). The merger of natural intelligence with artificial intelligence, with a focus on Neuralink company. *Virtual Economics*, *2*(3), 22–29.

Patagonia. (n.d.). (September 17, 2024) Supply chain environmental responsibility program. *Patagonia.* Retrieved, from https://www.patagonia.com/our-footprint/supply-chain-environmental-responsibility-program.html.

Seele, P., & Schultz, M. D. (2022). From greenwashing to machinewashing: A model and future directions derived from reasoning by analogy. *Journal of Business Ethics*, *178*(4), 1063–1089. https://doi.org/10.1007/s10551-022-05054-9.

Sustainable Fashion Forum. (n.d.). H&M is being sued for misleading sustainability marketing—what does this mean for the future of greenwashing?. *The Sustainable Fashion Forum.* Retrieved September 17, 2024, from https://www.thesustainablefashionforum.com/pages/hm-is-being-sued-for-misleading-sustainability-marketing-what-does-this-mean-for-the-future-of-greenwashing.

Tóth, Z., & Blut, M. (2024). Ethical compass: The need for corporate digital responsibility in the use of artificial intelligence in financial services. *Organizational Dynamics, 53*(2), 101041. https://doi.org/10.1016/j.orgdyn.2024.101041.

Zhang, D. (2024). The pathway to curb greenwashing in sustainable growth: The role of artificial intelligence. *Energy Economics, 133*, 107562. https://doi.org/10.1016/j.eneco.2024.107562.

4 Challenges of Trust in Remote Work in the Era of Generative AI

Monika Utzig, Aneta Mikuła, Jozefína Hvastová and Magdaléna Freňáková

4.1 Introduction

In recent years, remote work has become an integral component of numerous industries, with the COVID-19 pandemic accelerating this trend. As technology advances, generative artificial intelligence (AI) is increasingly playing a significant role in the daily tasks of remote workers. However, these changes bring about new challenges, particularly in the context of trust. Trust is a crucial element of effective collaboration, and its absence can lead to numerous issues at both the individual and organizational levels. The study aims to identify challenges to trust in remote work in the era of the development of GenAI.

4.2 The Evolution of Remote Work

Before the COVID-19 pandemic, remote work was prevalent among specific professional groups, such as home-based workers, doctors, lawyers, and increasing knowledge workers (Felstead, 2022; Piro, 2024). The transition of organizations to remote work occurred gradually over several decades due to the widespread adoption of digital technologies (Pushpa et al., 2024). With the increasing availability of digital solutions and internet access, work no longer needs to be confined to a single location – whether at the employer's premises or home – but can be performed in coworking spaces, on trains, at airports, or even on the beach (Donnelly & Johns, 2021).

In the wake of the COVID-19 pandemic, the issue of remote work became highly relevant and attracted the attention of many researchers. In a short period, remote work became necessary for many organizations (Kniffin et al., 2021; Richter, 2020). In Europe, in April 2020, 37% of working individuals reported remote work, reaching nearly 60% in Finland and 50% in several Northern European countries (Eurofound, 2020).

Such a rapid transition to the remote work model revealed its advantages and disadvantages. These can be considered from the perspective of both employees and employers. Several studies have shown an increase in employee productivity

DOI: 10.4324/9781003586944-5

(Arunmozhi et al., 2021; Beňo, 2021; Bloom et al., 2015), attributed to greater work flexibility, reduced commuting time, and fewer distractions (Collings et al., 2021; Thompson, 2019). Other studies have demonstrated a positive relationship between flexible working hours and employee well-being (Arora & Kumari, 2022; Costa et al., 2006; Haddad et al., 2024; Nijp et al., 2012).

On the other hand, challenges associated with remote work include feelings of social isolation due to the lack of physical interaction with colleagues (Fialho, 2022), difficulties in separating work and personal life, which can lead to overwork and family issues (Haddad et al., 2024; Meymand & Bokaie, 2013), technological challenges that may pose barriers for some employees and affect their ability to work effectively (Baumann & Marcum, 2023), and health problems related to musculoskeletal disorders and eye strain from prolonged screen time (Haddad, 2024). From the employer's perspective, the main challenges include organizing remote work, supervising and monitoring task completion, work efficiency, and providing employee support (Raišienė et al., 2020; Wang et al., 2021).

4.3 GenAI in Remote Work

AI is a technology capable of creating seemingly new content based on provided data. This can include generating text, images, and even code (Feuerriegel et al., 2024). In remote work, GenAI can support employees by automating routine tasks, analyzing data, and facilitating communication. As organizations increasingly adopt remote work models, integrating AI technology has become essential for enhancing productivity, communication, and collaboration among remote teams. However, its application also presents specific challenges, particularly in the context of trust.

A crucial effect of utilizing GenAI in work, including remote work, is increased employee efficiency and productivity. AI can automate simple tasks, allowing remote workers to focus on more complex activities (Noy & Zhang, 2023). However, this aspect also raises issues such as skill degradation and lack of employee engagement (Ahmad et al., 2023).

One of the most significant applications of GenAI in remote work is enhancing employee collaboration and communication. The development of realistic avatars and virtual assistants helps bridge the gap created by physical distance in remote collaboration, improving the quality of virtual interactions, which is crucial for maintaining morale and productivity in a remote environment (Christoff et al., 2023). Additionally, AI facilitates idea generation and assists in tasks by broadening perspectives, although concerns exist about the loss of human perspectives and critical thinking (Ahmad et al., 2023).

The development of GenAI, exemplified by tools such as ChatGPT, has impacted the demand for certain remote services. For instance, there has been a decline in demand for human-generated content creation and editing services, particularly in writing, although editing services remain primarily unchanged

(Yuan & Chen, 2023). The skill set sought by employers has also evolved. While AI is applied to tasks related to message drafting and text summarization, there is an increasing emphasis on soft skills and integrity among employees (Cardon et al., 2024).

While GenAI can automate routine tasks, it also creates new opportunities for companies to innovate and offer new services. However, it also brings challenges. These challenges can be divided into technological, ethical, and security domains. The implementation of GenAI systems, especially in remote work, poses technical challenges due to the complexity of integrating these systems with existing organizational processes (Issa et al., 2024). There are concerns about the erosion of technical skills among developers due to over-reliance on GenAI tools, which may impact the technical capabilities of remote work environments (Mbizo et al., 2024).

In the realm of ethical challenges, attention is drawn to security and privacy issues, including potential data breaches and unauthorized access to confidential information (Mbizo et al., 2024). There is also a need for responsible AI design to address ethical challenges such as bias, privacy, misuse, accountability, and responsibility in the context of remote work environments (Mughal, 2018). Concerns are also raised about the potential loss of human perspectives and critical thinking in remote work environments due to GenAI tools. Using AI is essential to maintain human oversight and judgment (Mbizo et al., 2024).

Security threats include concerns about data security and potential threats to organizational security. The rapid development of more advanced GenAI tools has increased security threats related to potential data breaches and other ethical dilemmas in remote work environments. Legal implications and best practices for ensuring cybersecurity in remote work environments have become critical areas of interest for organizations (Wach et al., 2023). As remote work evolves, the demand for robust cybersecurity measures will only grow, requiring continuous research and adaptation. The use of GenAI in remote work environments may also pose threats to organizational security, including the potential for individuals to be deceived by AI-generated content and the compromise of work process integrity (Park et al., 2023).

4.4 Challenges Related to Trust

One of the significant challenges related to trust is ensuring the reliability of GenAI outputs. Users often struggle to trust AI-generated information due to a lack of transparency and verification methods. Moreover, in some activities, the use of AI results in poorer evaluations and reduced trust in the person delegating tasks to AI. Delegating tasks to GenAI is perceived as less morally acceptable than delegating tasks to another person. A scientist choosing such delegation is considered less credible in future projects, and the results of such delegation are rated as less accurate and of lower quality. This perception of AI results as scientifically dubious and significantly impacts researchers' negative evaluation of AI use (Niszczota & Conway, 2023).

Conversely, studies indicate that AI is being utilized to enhance the delivery of public services and internal management (van Noordt & Misuraca, 2022). However, there were concerns about accountability for AI outcomes, which may cause some caution in AI usage by public-sector specialists. A study on public attitudes toward AI in the United Kingdom revealed that society has diverse views on appropriate AI use cases. The same study found that while the general public perceives efficiency and better accessibility as the main advantages of AI, there are concerns that AI will be used to replace professional judgments, for example, in the recruitment process (Modhvadia, 2023).

The widespread use of generative applications, such as ChatGPT, suggests that public-sector employees may have already started relying on these technologies to automate typical administrative tasks. This raises the issue of the business model of these technologies and who will provide them in the future. There remains a potential danger that a few key companies with the resources to create powerful GenAI tools will monopolize the productivity of governments and workers more broadly (Modhvadia, 2023). One of the critical questions for the public sector in the future will be whether it wants to invest in creating its language models supported by open-source technology or focus on procuring from technology companies. Resolving this question will be crucial for the future implementation of GenAI in the public sector and the level of trust in the received results.

The lack of trust in AI results can also be observed among students in higher education. Studies have shown that students are aware of and largely positive about using GenAI to support academic tasks (Bonsu & Baffour-Koduah, 2023). Students recognize potential inaccuracies in GenAI outputs and believe that they need to review the results (Shoufan, 2023). Most students see the necessity to edit GenAI outputs before using them in academic work (Deschenes & McMahon, 2024).

GenAI can significantly enhance communication and creativity within remote work teams. Lack of trust is one of the main obstacles to taking full advantage of the benefits of AI. It can result in reduced cooperation, efficiency, and productivity. Moreover, it can reduce the integration of AI systems and agents into teams (Gillath et al., 2021).

4.5 Methodology

We collected data from random internet users through an online survey to identify the perception of the role and challenges for trust in remote work (Computer-Assisted Web Interview). The survey was entirely anonymous. The questionnaire was distributed among respondents from Poland and Slovakia through an online Google Forms application with 11 questions, two of which were multi-choice. The investigated questions were connected to the perceived trust in remote working and perceived challenges of AI in remote work. We also asked respondents about their experiences with remote work and using AI as well as about trust in results generated by AI.

We collected answers from 343 respondents (136 from Poland and 207 from Slovakia). Among them, 70 (51%) respondents from Poland and 107 (52%) from Slovakia are working or have been working remotely. The survey was conducted from July to September 2024. The survey participants were primarily women (74% of respondents from Poland and 63% of respondents from Slovakia) at the age below 40 (91% of respondents from Poland and 68% of respondents from Slovakia), assessing their financial situation as good (56% of respondents from Poland and 24% of respondents from Slovakia) or average (38% of respondents from Poland and 52% of respondents from Slovakia). The respondents are living in the city. Overall, 46% of Polish respondents live in the biggest cities, with 1000 thousand or more citizens, while 30% of respondents from Slovakia live in the biggest cities. Overall, 25% of Slovak respondents (and 24% of Polish respondents) pointed to a town with under 100 thousand citizens as their living place.

The analyzed sample does not represent the population, but it can help identify the perception and challenges for trust in remote work in the era of AI.

4.6 Results and Discussion

We asked respondents if they perceive remote work as requiring more trust than stationary work. Overall, 67% of Slovak and 62% of Polish respondents strictly agreed, while 18% of respondents from Slovakia and 15% from Poland said it is hard to say. So, the result of our survey shows that in the opinion of respondents, remote trust is perceived as requiring more trust than stationary work.

In the next step, we asked respondents to rate how much they agreed with some statements reflected in remote and stationary work. The results (Table 4.1) show some differences between respondents from Poland and Slovakia.

Both Polish and Slovak respondents recognized that it is more difficult in remote work than in stationary work to maintain high employee motivation, monitor employees, maintain data security, and communicate with other employees.

The respondents' answers from Poland and Slovakia are pretty similar. More often than Polish, Slovak interviewees pointed out that monitoring employees is easier in remote work than in stationary work. More often than Polish, Slovak respondents answered that the communication process is easier during remote work.

We also asked the responders about their experiences and opinions about using AI at work. Only 30% of respondents from Poland and 48% from Slovakia use or have used AI in their work. On the other hand, only 10% of Polish and 24% of Slovak respondents trust AI-generated content, while 38% of respondents from Poland and 29% from Slovakia do not trust it. Our results are consistent with previous research, which shows that trust in AI is definitely low (Stawicka & Anderson, 2023).

In the next step, we asked for an opinion about a list of statements connected to the usage of AI at work (Figure 4.1).

Table 4.1 The share of answers to the question connected to issues is more difficult/easier in remote work than in stationary work

	Share Poland			*Share Slovakia*		
Statement	*More difficult (%)*	*I have no opinion (%)*	*Easier (%)*	*More difficult (%)*	*I have no opinion (%)*	*Easier (%)*
It is more difficult/ easier to maintain high employee motivation in remote work than in stationary work	60	22	18	62	15	23
It is more difficult/ easier to monitor employees in remote work than in stationary work	69	25	6	62	15	23
It is more difficult/ easier to maintain data security in remote work than in stationary work	60	34	6	64	26	10
In remote work, it is more difficult/ easier to communicate with other employees than in stationary work	54	32	14	57	13	30

Source: Authors' elaboration based on questionnaire survey.

Respondents from both countries most frequently agreed that using AI brings some threats. Overall, 75% of Polish and 60% of Slovak respondents agreed with the statement that there is a possibility that the use of AI in work will lead to copyright abuse. Overall, 75% of interviewees from Poland and 48% from Slovakia agreed that employees may pass off AI-generated content as their own. The main difference between Polish and Slovak respondents is that they are concerned about employers using AI to monitor employees working remotely. Overall, 54% of respondents from Slovakia and 18% from Poland agreed with that statement.

4.7 Conclusions and Limitations

The chapter discusses the role of trust in remote work in the era of GenAI. Our research showed that using AI at remote work brings advantages and threats.

An essential limitation of the research is that the study covers only two countries. For further research, more widely distributed questionnaires will make it possible to compare more countries and identify more factors perceived as essential challenges for trust in remote work in the era of AI. The limitation of the study is also the fact that employees completed the survey questionnaire. It will also be interesting to elaborate on a survey among employers.

This study can help managers better understand and manage employees at a time when a significant number of people have become familiar with the form of remote learning and work. Further research in this area is crucial, as organizations need to be flexible in the changing work environment and take actions aimed at workers from different generations.

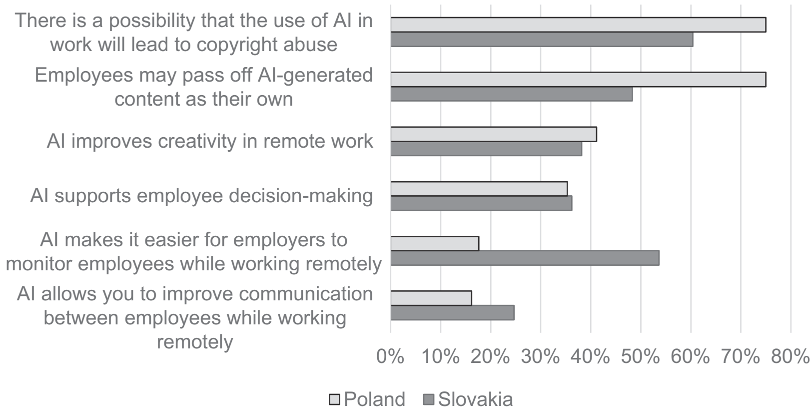

Figure 4.1 The share of respondents agreed with some statements from Polish and Slovak interviewees

Source: Authors' elaboration based on questionnaire survey.

References

Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M., Irshad, M., Arraño-Muñoz, M., & Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, *10*(1), 1–14.

Arora, S., & Kumari, N. (2022). Imagining benefits and challenges for future hybrid workplace to enable reentry for women on career break. *International Journal of Networking and Virtual Organisations*, *26*(3), 231–248.

Arunmozhi, M., Kumar, R. K., & Srinivasa, B. A. (2021). Impact of COVID-19 on global supply chain management. In A. R. Sakthivel, J. Kandasamy, & J. P. Davim (Eds.), *Managing supply chain risk and disruptions: Post COVID-19*, (pp. 1–18). Springer.

Baumann, H. M., & Marcum, T. M. (2023). Human capital and legal perspectives on remote work: Recommendations for organizations. *Management Research Review*, *46*(12), 1711–1726. https://doi.org/10.1108/MRR-06-2022-0412.

Beňo, M. (2021). The Advantages and disadvantages of E-working: An examination using an ALDINE analysis. *Emerging Science Journal*, *5*, 11–20. https://doi.org/10.28991/esj-2021-SPER-02.

Bloom, N., Liang, J., Roberts, J., & Ying, Z. J. (2015). Does working from home work? Evidence from a Chinese experiment. *The Quarterly Journal of Economics*, *130*(1), 165–218.

Bonsu, E. M., & Baffour-Koduah, D. (2023). From the consumers' side: Determining students' perception and intention to use ChatGPT in Ghanaian higher education. *Journal of Education, Society & Multiculturalism, 4*(1), 1–29. https://doi.org/10.2478/jesm-2023-0001.

Cardon, P., Fleischmann, C., Logemann, M., Heidewald, J., Aritz, J., & Swartz, S. (2024). Competencies needed by business professionals in the AI age: Character and communication lead the way. *Business and Professional Communication Quarterly*, *87*(2), 223–246. https://doi.org/10.1177/23294906231208166.

Christoff, N., Neshov, N. N., Tonchev, K., & Manolova, A. (2023). Application of a 3D talking head as part of telecommunication AR, VR, MR system: Systematic review. *Electronics*, *12*(23), 4788. https://doi.org/10.3390/electronics12234788.

Collings, D. G., McMackin, J., Nyberg, A. J., & Wright, P. M. (2021). Strategic human resource management and COVID-19: Emerging challenges and research opportunities. *Journal of Management Studies*, *58*(5), 1378–1382. https://doi.org/10.1111/joms.12695.

Costa, G., Sartori, S., & Åkerstedt, T. (2006). Influence of flexibility and variability of working hours on health and well-being. *Chronobiology International*, *23*(6), 1125–1137. https://doi.org/10.1080/07420520601087491.

Deschenes, A., & McMahon, M. (2024). A survey on student use of generative AI chatbots for academic research. *Evidence Based Library and Information Practice*, *19*(2), 2–22. https://doi.org/10.18438/eblip30512.

Donnelly, R., & Johns, J. (2021). Recontextualising remote working and its HRM in the digital economy: An integrated framework for theory and practice. *The International Journal of Human Resource Management*, *32*(1), 84–105. https://doi.org/10.1080/09585192.2020.1737834.

Eurofound (2020). Living, working and COVID-19. Publications of the European union. Available on https://www.eurofound.europa.eu/en/publications/2020/living-working-and-covid-19 (07.08.2024).

Felstead, A. (2022). *Remote working: A research overview*. Routledge.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative ai. *Business & Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7.

Fialho, J. (2022). Benefits and challenges of remote work. In R. Pereira, I. Bianchi, & Á. Rocha (Eds.), *Digital technologies and transformation in business, industry and organizations* (pp. 1–17). Springer International Publishing.

Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., & Spaulding, R. (2021). Attachment and trust in artificial intelligence. *Computers in Human Behavior*, *115*(106607), 1–10. https://doi.org/10.1016/j.chb.2020.106607.

Haddad, C., Zakhour, M., Sacre, H., Karam, J., El Khatib, S., Akel, M., & Salameh, P. (2024). Perceived advantages and drawbacks of teleworking during the economic crisis and COVID-19: A cross-sectional study among university degree holders. *Archives of Environmental & Occupational Health*, *79*(2), 91–105. https://doi.org/10.1080/19338244.2024.2366815.

Issa, H., Kadian, A., Ahuja, S., & Nishant, R. (2024). When a dream turns into a nightmare: A case study of an education technology startup to uncover the dark side of generative AI. *Communications of the Association for Information Systems*, *54*, 1048–1078. https://doi.org/10.17705/1CAIS.05444.

Kniffin, M., Narayanan, J., Anseel, F., Antonakis, J., Ashford, S. P., Bakker, A. B., … & Vugt, M. V. (2021). COVID-19 and the workplace: Implications, issues, and insights for future research and action. *The American Psychologist, 76*(1), 63–77. https://doi.org/10.1037/amp0000716.

Mbizo, T., Oosterwyk, G., Tsibolane, P., & Kautondokwa, P. (2024, July). Cautious optimism: The influence of generative AI tools in software development projects. In A. Gerber, (Ed.), *Annual conference of south african institute of computer scientists and information technologists* (pp. 361–373). Springer Nature Switzerland. Available on https://link.springer.com/chapter/10.1007/978-3-031-64881-6_21, Accessed 08.09.2024.

Meymand, Z. Z., & Bokaie, M. (2013). Tele-working and its effect on human resource efficiency. *Advances in Environmental Biology*, *7*(10), 3071–3076.

Modhvadia, R. (2023). *How do people feel about AI? A nationally representative survey of public attitudes to artificial intelligence in Britain*. The Ada Lovelace Institute and The Alan Turing Institute. Retrieved July 06, 2024, from https://www.adalovelaceinstitute.org/report/public-attitudes-ai.

Mughal, A. A. (2018). Artificial intelligence in information security: Exploring the advantages, challenges, and future directions. *Journal of Artificial Intelligence and Machine Learning in Management*, *2*(1), 22–34.

Nijp, H. H., Beckers, D. G., Geurts, S. A., Tucker, P., & Kompier, M. A. (2012). Systematic review on the association between employee worktime control and work-non-work balance, health and well-being, and job-related outcomes. *Scandinavian Journal of Work, Environment & Health*, 38(4), 299–313. https://doi.org/10.5271/sjweh.3307.

Niszczota, P., & Conway, P. (2023). Judgements of research co-created by Generative AI: Experimental evidence. *Economics and Business Review*, *9*(2), 101–114. https://doi.org/10.18559/ebr.2023.2.744.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586.

Park, A., Kietzmann, J., Killoran, J., Cui, Y., van Esch, P., & Dabirian, A. (2023). Nothing is harder to resist than the temptation of AI. *IT Professional*, *25*(6), 13–20. https://doi.org/10.1109/mitp.2023.3340529.

Piro, V. (2024). From homework to remote work. Reflecting on the porosity of spaces, times and social relationships at home and in workplaces. *Rassegna Italiana Di Sociologia*, *65*(1), 91–100. https://doi.org/10.1423/112993.

Pushpa, A., Shukla, N., Horal, L., Kivshyk, O., Stepaniuk, O., & Reznik, N. P. (2024). Evolving horizons of work: Unravelling the conceptual and future research dimensions of digital workspaces. In R. Khamis, & A. Buallay (Eds.), *AI in business: Opportunities and limitations* (Vol. 1, pp. 585–598). Springer Nature Switzerland.

Raišienė, A. G., Rapuano, V., Varkulevičiūtė, K., & Stachová, K. (2020). Working from home—Who is happy? A survey of Lithuania's employees during the COVID-19 quarantine period. *Sustainability*, *12*(13), 5332, 1–21. https://doi.org/10.3390/su12135332.

Richter, A. (2020). Locked-down digital work. *International Journal of Information Management, 55*, 102157. https://doi.org/10.1016/j.ijinfomgt.2020.102157.

Shoufan, A. (2023). Exploring students' perceptions of chatGPT: Thematic analysis and follow-up survey. *IEEE Access*, *11*, 38805–38818. https://doi.org/10.1109/ACCESS.2023.3268224.

Stawicka, E., & Anderson, T. (2023). Digital trust in artificial intelligence. In J. Paliszkiewicz, J. L. Gusmano, & J. Gołuchowski (Eds.), *Trust, digital business and technology. Issues and challenges* by, (pp. 144–155). Routledge.

Thompson, B. Y. (2019). The digital nomad lifestyle:(remote) work/leisure balance, privilege, and constructed community. *International Journal of the Sociology of Leisure*, *2*(1), 27–42. https://doi.org/10.1007/s41978-018-00030-y.

Van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: Results of landscaping the use of AI in government across the European Union. *Government Information Quarterly*, *39*(3), 101714.

Wach, K., Duong, C. D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., Paliszkiewicz, J., & Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of chatGPT. *Entrepreneurial Business and Economics Review*, *11*(2), 7–30. https://doi.org/10.15678/EBER.2023.110201.

Wang, B., Liu, Y., Qian, J., & Parker, S. K. (2021). Achieving effective remote working during the COVID-19 pandemic: A work design perspective. *Applied Psychology*, *70*(1), 16–59. https://doi.org/10.1111/apps.12290.

Yuan, Z., & Chen, H. (2023). "The impact of ChatGPT on the demand for human content generating and editing services: Evidence from an online labor market". *ICIS 2023 Proceedings*, *1*, 2669. https://aisel.aisnet.org/icis2023/techandfow/techandfow/1.

5 Generative AI as a Driver of Trust in Innovation

Michał Borowy

5.1 Introduction

The rapid development of generative artificial intelligence (GEN AI) has introduced a new paradigm in technological innovation, impacting numerous sectors and redefining how businesses operate. This chapter explores the intersection of trust and innovation, focusing on how the adoption of GEN AI fosters confidence in technological advancements among companies and individuals. Trust, a multifaceted concept, varies across disciplines. In psychology, it reflects a belief in interpersonal relationships (Rotter, 1967, 1980; Watson, 2005), while in management, it centers around credibility, competence, and goodwill (Mayer et al., 1995; McAllister, 1995; Watson, 2005; Watson, 2005). From an economic perspective, trust is linked to belief in investment profitability (Vizyble, 2024). It, therefore, results from rational decision-making, which plays a pivotal role in reducing costs and increasing the revenue of companies. This economic approach to trust is particularly relevant in evaluating the potential of GEN AI.

As businesses increasingly adopt GEN AI technologies to drive efficiency, streamline processes, and foster innovation, trust in these technologies has become crucial. GEN AI, a subset of AI, enables machines to create new text, images, and audio content that closely resembles human outputs. This capacity has unlocked vast applications across healthcare, marketing, manufacturing, and finance industries. The ability of GEN AI to mimic human creativity while significantly reducing costs and increasing productivity is reshaping how businesses approach innovation.

This chapter assesses the degree of trust placed in GEN AI by analyzing investment trends, adoption rates across industries, and public sentiment toward businesses utilizing AI technologies. By doing so, it sheds light on how GEN AI acts as a catalyst for new technologies and business models, driving economic growth and fostering a culture of innovation.

This chapter consists of five substantive parts: A critical review of the subject matter literature, research methodology, presentation of research results, discussion, together with presentation assumptions to the *Transaction cost*

DOI: 10.4324/9781003586944-6

acceptability level model for GEN AI deployments, and the final conclusions are included in the last section.

5.2 Literature Review

Trust phenomenon can be interpreted in different dimensions (Borowy & Karpio, 2024; Paliszkiewicz, 2018). In psychology, it refers to an individual's belief in interpersonal relationships (Rotter, 1967, 1980; Watson, 2005). In management science, it refers to the credibility, competence, and goodwill in professional relationships (Mayer et al., 1995; McAllister, 1995; Watson, 2005; Watson, 2005). In marketing, it is based on dialogue in business exchange relationships (Delgado et al., 2003; Morgan & Hunt, 1994; Schurr & Ozanne, 1985; Watson, 2005). In economics, on the other hand, trust is considered from the point of view of its ability to reduce transaction costs (Ambroziak et al., 2016; Kramer & Cook, 2004; North, 1990) and profitability (Vizyble, 2024). This chapter focuses mainly on the second aspect, which is related to the pragmatic approach based on business calculations. Indeed, the subject of the analysis in this work is the effects of rational entrepreneurial decisions, as expressed by investments in GEN AI, which confirm confidence in innovation. However, this chapter also addresses the issue of transaction costs by proposing assumptions (variables) for a model describing the benefits and costs resulting from the use of GEN AI.

GEN AI *refers to computational techniques that are capable of generating seemingly new, meaningful content such as text, images, or audio from training data* (Feuerriegel et al., 2024, p. 111). In the advisory environment, this phenomenon is interpreted similarly as a new technology, which, based on the available artifacts, can create new ones that reflect the features of the training data but do not repeat them (Gartner, 2024). Other authors add that inputs can consist of natural language prompts or other non-traditional, non-code forms of input (Forrester, 2023). All this means that GEN AI opens up a wide range of applications for companies and societies, the scope of which we cannot yet determine.

Historically, research on AI was undertaken as early as the 1950s (Turing, 1950). Growth in this area of knowledge accelerated in the 1990s when machine learning, and consequently deep learning, was developed. There was then a significant increase in interest in AI, understood as *intelligence demonstrated by machines*, which is possible to implement in various socio-economic aspects (Delipetrev et al., 2020, p. 5).

A breakthrough in the development of GEN AI came in the last decade of the 21st century with the publication of several key scientific papers that introduced new approaches to generative model building. These include works such as *Generative Adversarial Nets* (Goodfellow et al., 2014) and *Attention is All You Need* (Vaswani et al., 2017). The first publication introduced the concept of Generative Adversarial Networks – one of the most important technologies in the development of GEN AI. It allows the generation of realistic images, sounds, and other data.

The second one introduced the transformer mechanism, which is the basis for large language models – LLMs, such as GPT (Generative Pre-trained Transformers).

Thanks to its ability to generate creative and realistic outputs, GEN AI, in a short period of time, has gained trust and become a focal point for various sectors. The introduction of tools such as ChatGPT, Dall-E, or Midjourney has increased access to LLMs, enabling the creation of human-like content and the significant popularity of GEN AI from 2022 onward (García-Peñalvo & Vázquez-Ingelmo, 2023). Currently, this technology is being used successfully in most spheres of the economy, contributing in particular to increasing productivity, automating manufacturing processes, as well as shortening the development path of new products or services (Chui et al., 2023; Deloitte, 2023; Deloitte, 2024). Some key applications in different sectors are outlined below.

Healthcare and pharmacy. AI supports the drug discovery process, medical data analysis, and diagnostics; AI models help generate potential molecular compounds for drug development by predicting how molecules interact with each other, speeding up drug discovery; generative models are being used to improve medical imaging (like X-rays or MRIs), fill gaps in scans, and generate detailed images to help diagnose process; can support patient care processes by analyzing voluminous patient documentation; and automatically generate summaries of important information; they can also convert doctors' notes (also verbal) into structured electronic medical records, ensuring accurate documentation with minimal effort (Coursera, 2024; Horban, 2024).

Marketing and Advertising. AI is used to automate content creation, personalize advertising campaigns, generate graphics, text, and movies and optimize marketing strategies based on data. Technology also has the tools to translate marketing messages disseminated in new territories. Apart from that, it enables trend prediction and analysis of consumer behavior (Coursera, 2024; Horban, 2024)

Manufacturing. In the industry, AI helps design products, optimize production processes, simulate machine operation, and forecast raw material requirements. GEN AI can support engineers in speeding up design processes by enabling them to generate designs and evaluate projects in terms of assumed constraints. The technology, based on historical data, can also alert on potential maintenance needs of heavy equipment resulting from its operation or generate delivery schedules and recommendations for suppliers (Chui et al., 2023; Coursera, 2024)

Technology and IT: GEN AI is used in software development, automation of programming tasks, data processing, and development of AI-based tools. The technology assists programmers in creating and completing code, can be a tool for programmers to interact with the software without needing to know the programming language, can act as a translator, and can assist in testing different software use scenarios (Boza, 2021; Chui et al., 2023; Coursera, 2024).

Finance. The financial sector is using GEN AI for market forecasting, risk analysis, service personalization, and automation of wealth management and customer service processes. Technology can be used to create investment strategies by recommending the best investments that take into account the user's or client's assumed goals. It can find and complete transactions much more quickly, maintaining certain parameters and providing personalized customer service. It can also monitor regulations, inform about changes, and prepare the necessary documentation (Chui et al., 2023; Coursera, 2024).

Media. GEN AI can help create and edit visual content, create short videos, and make it easier to work with content management systems. It can also create completely new video content or generate visual and graphic effects and allows viewers to create their own custom realizations. It also indexes large media libraries, making it easy to find relevant content in any set (Coursera, 2024).

Education. GEN AI provides the tools to deliver learning activities in a more dynamic, inclusive, and, at the same time, personalized way. It enables the creation of materials tailored to the needs of different subjects and learning styles, generates quizzes and educational games, creates language exercises that mimic real-life conversations that facilitate language acquisition in language classes, or enables virtual experiments, often enabling them to be carried out at all. On the other hand, GEN AI tools also allow administrative tasks to be automated, streamlining content creation and assessment processes and allowing trainers to devote more attention to classroom teaching and the development of innovative training modules (Horban, 2024; Kulik, 2024).

All the applications described above show how GEN AI is widely used, contributing to increasing efficiency and productivity. It undoubtedly provides an incentive for enterprises to implement such technologies further. This applies to both large companies (Gursharan, 2024) and the small- and medium-sized enterprises (SME) sector (Karkhanis, 2023; Slack, 2023), especially startups (Horban, 2024). Thus, GEN AI becomes a catalyst for new technologies (Chui et al., 2023) and an incentive for companies to develop business models based on innovation (Gidwani & Bello, 2024; Kanbach et al., 2024; Norbäck & Persson, 2024).

5.3 Methodology

According to the literature review, GEN AI is a catalyst for new technologies and an incentive for companies to develop business models based on innovation (Chui et al., 2023; Gidwani & Bello, 2024; Kanbach et al., 2024; Norbäck & Persson, 2024). So, considering GEN AI as a driver of trust in innovation in this work, we analyze the level of trust in such technology de facto. This is because increased trust in them implies increased trust in innovation. And if so, GEN AI can be seen as a driver of trust in innovation.

From an economic perspective, trust is linked to belief in investment profitability (Vizyble, 2024). In other words, it results from pragmatic management decisions based on calculations. In the case of this study, the pragmatic decisions relate to investment in GEN AI, which is a source of innovation that enables significant cost reduction, productivity improvement, and higher income. Therefore, the main purpose of this chapter is to assess the degree of trust in innovation expressed by business investments in GEN AI solutions implementation.

In order to achieve the research objective a statistical analysis of data on (1) private investment in GEN AI, (2) level of adoption AI/GEN AI solutions by companies, (3) share of companies using GEN AI by industry, (4) use of GEN AI in organizations, by function, (5) cost decrease and revenue increase from GEN AI adoption, and (6) social trust in business using AI. The research material was collected from the latest reports of Stanford University, McKinsey & Company, AIPRM, Statista, and Forbes Advisor from 2023 to 2024.

5.4 Findings

In the case of private investment in GEN AI, data published by Stanford University in *Artificial Intelligence Index Report 2024* (Stanford University, 2024) allow us to confirm a definite upward trend in their value in the period 2019–2023. The largest, almost nine-fold increase occurred in the period 2022–2023, when the technology became significantly more recognizable, and entrepreneurs became much more familiar with the possibilities of its use (García-Peñalvo & Vázquez-Ingelmo, 2023). In the international capital market, the leader in private investment in GEN AI is the United States, with € 22.5 billion. The countries of the European Union, together with the United Kingdom, have reached € 0.74 billion in such investments, and China € 0.65 billion (European Parliament, 2024; Stanford University, 2024). All of these research results give a clear indication that businesses on a global scale have come to trust the possibilities that GEN AI offers.

According to data published in the McKinsey report – *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value*, global adoption of GEN AI doubled in just one year from 2023 to 2024, to a level of 65% and analysts' predictions point to further strong growth (Singla et al., 2024). This trend is confirmed in the results of another research. As reported by Deloitte, 67% of those organizations they surveyed are increasing investment in GEN AI, seeing great value observed to date (Deloitte, 2024). Also, GOOGLE's analysis shows that 64% of executives want to adopt GEN AI; however, most recognize that their organization lacks the most critical skills (Moyer, 2023). Similar conclusions are presented by Lucidworks which shows that 63% of companies plan to increase investment in this area over the next 12 months (Lucidworks, 2024). Thus, the above data also confirm that businesses, on a wide scale, have come to trust the possibilities that GEN AI offers.

Given the impressive growth in the application of GEN AI, it is also important to analyze structurally the level of use investigated technologies in the various industries as well as at which functional area they are applied mostly.

Available data (AIPRM, 2024; Statista, 2024) shows that in the 2023 US market, GEN AI technologies are most commonly used in industries such as marketing and advertising -37%, technology -35%, and consulting -30%. Much less frequently in teaching -19%, accounting -16%, or healthcare -15%. In this case, the example of one country is relatively meaningful, as the United States is currently the world leader in terms of the level of private investment in GEN AI (European Parliament, 2024; Stanford University, 2024).

An analysis of the use of GEN AI in organizations, by function, in 2024, was possible thanks to the results of a survey by McKinsey (Singla et al., 2024). It shows that organizations were most likely to use GEN AI for tasks related to marketing and sales -34%, product and/or service development -23%, and IT -17%. So far, the technologies were marginally used in such areas as manufacturing, supply chain and inventory management, strategy and corporate finance – respectively: 4, 6 and 7%. However, given the rapid progress of GEN AI, their importance may also change rapidly in these areas.

Comparing the results on: (1) the percentage of companies using GEN AI by industry and (2) the use of GEN AI in organizations, by function, we note that the technologies in question are most commonly used as a tool for marketing and sales and product and/or service development, in such industries as: marketing and advertising, technology and consulting. This shows that GEN AI plays an important role in creating product innovations. Moreover, given the data on the IT and marketing spheres, it can also be assumed that, according to the typology of the Oslo Manual (OECD & Eurostat, 2018), they are also important in creating business process innovations. So, GEN AI also has an impact on increasing trust in innovation.

The last part of the economic analysis refers directly to the profitability of investments in GEN AI. The available data from the McKinsey report (Singla et al., 2024) confirm that the use of the technologies in question brought tangible benefits in the form of reduced costs and increased revenue to all the organizations surveyed, regardless of their business profile. Especially noteworthy in the case of this analysis is the observation regarding profitability in the area of product and/or service development. Because it is already at a noticeable level to the entrepreneur, -37% survey confirmed a reduction in costs and 35% an improvement in revenue. In this area of the organization's activity, according to the Oslo Manual, product innovation and business process innovation can be included. This means that GEN AI is important for companies' innovation performance. This observation is in line with the other author's findings, who consider GEN AI as a catalyst for new technologies (Chui et al., 2023) or an incentive for companies to develop business model innovation (Gidwani & Bello, 2024; Kanbach et al., 2024; Norbäck & Persson, 2024). The arguments presented also confirm that GEN AI is a driver of trust in innovation.

A complementary element to the analysis carried out was an attempt to understand the public reaction, the customers of the companies where the changes described are dynamically taking place. Some response in this area was found in the results of a study published by Forbes Advisor (Haan, 2024). They show that

65% of consumers still trust businesses that use AI. For businesses, it is particularly important information from the point of view of corporate social responsibility and the purchasing propensity of potential customers. All the more importantly, consumer awareness is now at a high level, and there is increasing online awareness of the risks of using GEN AI (Chui et al., 2023; Forrester, 2024; Singla et al., 2024).

5.5 Discussion

Trust is mainly associated with mental determinants of interpersonal relationships (Delgado et al., 2003; Mayer et al., 1995; McAllister, 1995; Morgan & Hunt, 1994; Rotter, 1967, 1980; Schurr & Ozanne, 1985; Watson, 2005). The approach in this study is closer to economic science (Vizyble, 2024). It has been interpreted here through the prism of entrepreneurs' pragmatic decisions that lead to reduced costs and generate profitability for the company. It is expressed through the level, scale, and structure of private investment in GEN AI. In general, as can be seen from the findings presented, companies trust GEN AI. This is because they see it as a powerful tool that improves efficiency, reduces costs, enables scaling of operations, and provides new opportunities for personalization and innovation. AI is becoming a key component of business strategies, offering companies significant advantages in a competitive environment.

However, as the benefits of GEN AI implementation increase, companies (and consumers) also experience increasing relevant risks. Threats may concern mostly inaccuracy, intellectual property rights, cybersecurity, personal individual privacy, or regulatory compliance (Forrester, 2024; Singla et al., 2024). It can lead to increased transaction costs (Ambroziak et al., 2016; North, 1990). These are related to technology infrastructure, regulation, data and intellectual property protection, risk management, as well as adapting the technology to the company's specific needs. These costs can be high, but companies choosing to invest in GEN AI often see them as necessary to gain long-term benefits and competitive advantage.

An interesting addition to the analysis would be the creation of a *Transaction cost acceptability level model for GEN AI deployments*. The model could be used to identify the boundary up to which companies are willing to cover the transaction costs associated with the implementation of such technologies. It could also clarify which transaction costs are of greater or equal importance compared to the potential benefits. Also, which determinants of transaction costs are most important? Proposed variables for the model are presented below, split into two groups: (1) GEN AI transaction costs and (2) benefits of implementing GEN AI, as well as a scheme of the model proposal in Figure 5.1.

1 GEN AI transaction costs variables:

- **Technology** – infrastructure deployment cost (the cost of purchasing, integrating, and maintaining the hardware and necessary software to get GEN AI up and running; cost of adapting to existing systems; costs associated with upgrading servers, purchasing GPUs, or accessing cloud computing; operational support costs.

- **Standards and regulations** – cost of legal advisory on GEN AI responsibility, intellectual property protection, insurance, cost of adapting systems, and procedures.
- **Training and staff development** – technical training (operating AI-based tools, understanding their functions); changes in organizational structure (training and recruitment of new staff with the right skills).
- **Licenses and software –** for the use of AI-based tools such as GPT models; subscription costs.
- **Data security and privacy protection** – costs associated with ensuring compliance with data protection legislation (e.g. RODO); safeguards against cyber threats (IT and monitoring systems).
- **Risks associated with errors and unpredictability** – monitoring and correction costs; costs associated with repairing reputation or handling legal consequences.
- **Supplier negotiation and contracting** – the cost of finding the right AI supplier to meet all the company's needs; the cost of legal advisor involved in negotiating complex contracts for AI technology implementation, technical support, licensing, or data protection.
- **Industry-specific adaptation** – cost of testing (tailoring to industry-specific requirements) accuracy, and functionality; testing and iteration.
- **Ethical issues and social responsibility – the cost of ethical audits to ensure that their AI operates in** accordance with social and legal standards; the cost of risk management.

2 GEN AI implementation benefit variables

- **Automation and savings** – automatic generation of reports, analyses, marketing content, programming codes, or graphic designs, which significantly speeds up work; reduction of routine tasks (repetitive, time-consuming activities) allowing employees to focus on more creative and valuable tasks.
- **Scalability and flexibility** – the ability to easily scale operations without the need for a proportional increase in staff, e.g. serving more customers: Chatbots and virtual assistants can serve customers 24/7, offering personalized responses without the need for human employees; adaptability to changing market conditions, e.g. increasing or decreasing resources as required.
- **Productivity improvement** – productivity growth in both operational and creative areas, e.g. content creation (GEN AI can automatically generate high-quality content such as articles, blog posts, advertising campaigns, or marketing materials, increasing the productivity of marketing and creative teams; programming (AI tools can support developers by automatically generating code snippets, suggesting optimizations and reducing software bugs.

- **Better personalization** – GEN AI enables companies to deliver more personalized products and services, e.g. it can analyze customer data and, based on this, create personalized marketing campaigns, product offers, or recommendations, leading to higher customer satisfaction and increased conversions; customer service (virtual assistants and chatbots can tailor responses to individual customers, improving the service experience).
- **Innovation and new product development** – GEN AI can be a source of innovation, supporting the development of new products and services, e.g. design and prototyping (AI can generate new product concepts, designs, or models, reducing the time needed for prototyping and innovation; artistic creation (in fields such as fashion, graphic design, music, art, and others), AI supports the creative process, generating inspiration and new ideas, that can be turned into viable products or services, opening up new business opportunities.
- **Cost reduction** – automating tasks with GEN AI allows companies to reduce operational costs, such as less staff required (automating routine tasks reduces costs associated with hiring and training new staff); savings through optimization (AI can optimize production, logistics, and sales processes, leading to reduced waste and lower operational costs).
- **Increased competitiveness** – the use of GEN AI gives companies a competitive advantage, for example by faster time-to-market (AI allows for faster testing and introduction of new products, reducing the time to market advantage); better data analysis (GEN AI helps companies analyze huge amounts of data, leading to better business decisions and a more effective market strategy).
- **Improved customer service** – 24/7 availability: chatbots and virtual assistants running on AI are available 24/7, increasing customer satisfaction and providing immediate service; rapid response to customer needs: AI can automatically resolve common customer issues, minimizing support wait times.
- **Improved decision-making processes** – predictive analytics: AI can analyze historical data and predict future market trends, which supports strategic planning and reduces risk; automated reports and recommendations: AI can generate data-driven reports and recommendations, enabling managers to make more informed decisions.
- **Risk management and compliance** – risk monitoring, e.g. AI can automatically analyze data and identify potential risks (e.g. in finance or manufacturing), which helps companies respond to problems more quickly; regulatory compliance, e.g. AI can support companies in monitoring and ensuring compliance with regulations, which reduces the risks associated with audits and sanctions.

The model proposal is only a starting point for an in-depth empirical analysis, the results of which may help to determine further research directions. They may also provide hints for the business environment, especially the SME's sector, for which investment in innovation is a greater financial challenge than for large companies.

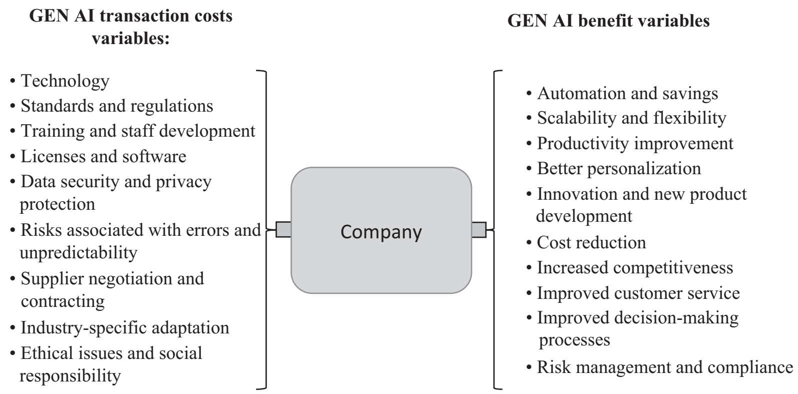

Figure 5.1 Transaction cost acceptability level model for GEN AI deployments

Source: Own elaboration

5.6 Conclusion

This study highlights the critical role of GEN AI as both a driver of innovation and a catalyst for building trust in emerging technologies. Through an analysis of investment trends, adoption rates, and industry applications, the research demonstrates that trust in GEN AI is closely tied to its potential to enhance productivity, optimize processes, and reduce costs. As businesses increasingly rely on GEN AI to support decision-making and drive innovation, the technology's integration is reshaping economic models and transforming various sectors, including healthcare, marketing, manufacturing, and finance.

From a theoretical standpoint, this study contributes to the understanding of trust in innovation by integrating economic perspectives with the technological potential of GEN AI. Traditionally, trust has been studied through psychological and interpersonal lenses, but this research positions trust as a pragmatic economic decision that aligns with profit-driven motivations. The findings suggest that trust in GEN AI technologies is not only about reducing costs but also about fostering long-term innovation strategies. This extension of trust theory opens new avenues for understanding how businesses decide to adopt and invest in disruptive technologies.

In practice, the study underscores the transformative potential of GEN AI in driving productivity and innovation across industries. For practitioners, the results point to the importance of integrating GEN AI into business models to remain competitive in a rapidly evolving technological landscape. The findings also highlight the importance of fostering consumer trust, as public acceptance of AI technologies remains a critical factor for widespread adoption. Companies that can effectively communicate the benefits and ethical considerations of GEN AI are likely to build stronger relationships with stakeholders, including customers and investors.

Despite the valuable insights provided, this study faces some limitations. These are mainly due to the fact that the study focuses predominantly on the economic aspects of trust, which may overlook other dimensions, such as ethical concerns and regulatory frameworks that are increasingly shaping AI adoption.

Future research should investigate the ethical and legal implications of GEN AI in more depth, particularly regarding data protection, transparency, and regulatory frameworks. Moreover, it would be beneficial to explore how the transaction costs associated with GEN AI implementation, such as infrastructure investments, legal compliance, and risk management, influence the adoption decisions of SMEs. Such future inquiries would provide a more holistic understanding of the challenges and opportunities associated with GEN AI and its role in fostering trust and innovation.

References

AIPRM. (2024). *50+ Generative AI Statistics 2024*. Retrieved from https://www.aiprm.com/generative-ai-statistics/.

Ambroziak, E., Starosta, P., & Sztaudynger, J. J. (2016). Trust, willingness to help and fairness, and economic growth in Europe [Zaufanie, skłonność do pomocy i uczciwość a wzrost gospodarczy w Europie]. *Ekonomista, 5,* 647–671.

Borowy M., & Karpio K. (2024). Trust in modern digital technologies - an example of artificial intelligence. In J. Paliszkiewicz & G. Jerzy (Eds.), *Trust and artificial intelligence development and application of AI technology* (pp. 49–60). Taylor & Francis Group.

Boza, P., & Evgeniou, T. (2021). *Implementing Ai Principles: Frameworks, Processes, and Tools.* INSEAD Working Paper No. 2021/04/DSC/TOM. Retrieved from SSRN: https://ssrn.com/abstract=3783124.

Chui, M., Hazan, E., Roberts, R., Singla, A., & Smaje, K. (2023). The economic potential of generative AI. *McKinsey.* Retrieved from https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/the%20economic%20potential%20of%20generative%20ai%20the%20next%20productivity%20frontier/the-economic-potential-of-generative-ai-the-next-productivity-frontier.pdf.

Chui, M., Yee, L., Hall, B., & Singla, A. (2023). The state of AI in 2023: Generative AI's breakout year. *McKinsey & Company.* Retrieved from https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year.

Coursera. (2024, July 24). 20 *Examples of Generative AI Applications Across Industries.* Retrieved from https://www.coursera.org/articles/generative-ai-applications.

Delgado-Ballester, E., Munuera-Aleman, J. L., & Yagoe-Guillin, M. J. (2003). Development and validation of a brand trust scale. *International Journal of Market Research, 45(1)*, 335–353.

Delipetrev, B., Tsinarakii, C., & Kostić, U. (2020). Historical evolution of artificial intelligence. Publications Office of the European Union. Retrieved from https://publications.jrc.ec.europa.eu/repository/handle/JRC120469.

Deloitte. (2023). *The Generative AI Dossier. A selection of high-impact use cases across six major industries.* Deloitte AI Institute. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/us/Documents/consulting/us-ai-institute-gen-ai-use-cases.pdf.

Deloitte. (2024). Now, we have decided on the next step: Moving from potential to performance, Deloitte's State of Generative AI in the Enterprise Quarter three report. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/us/Documents/consulting/us-state-of-gen-ai-q3.pdf.

European Parliament. (2024). *AI investment: EU and global indicators.* Retrieved from https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/760392/EPRS_ATA(2024)760392_EN.pdf.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative ai. *Business & Information Systems Engineering, 66(*1), 111–126.

Forrester. (2023, February 14). *Forrester: Ignoring Generative AI Will Be a Costly Mistake for Enterprises*. Retrieved from https://www.forrester.com/press-newsroom/forrester-generative-ai-research/.

Forrester. (2024). *Generative AI Trends for Business: Why, When, and Where to Begin.* Retrieved from https://www.forrester.com/technology/generative-ai/.

García-Peñalvo, F., & Vázquez-Ingelmo, A. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in generative AI. *International Journal of Interactive Multimedia and Artificial Intelligence, 8(4*), 7–16.

Gartner. (2024). *Gartner Experts Answer the Top Generative AI Questions for Your Enterprise* Retrieved from https://www.gartner.com/en/topics/generative-ai.

Gidwani, S., & Bello, A. (2024, April 1). *Using generative AI to accelerate product innovation*. Retrieved from https://www.ibm.com/think/topics/generative-ai-product-development.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014).* pp. 2672–2680. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf.

Gursharan, S. (2024, April 4). *How Big Companies Use Generative AI to Grow in 2024.* Retrieved from https://www.webspero.com/blog/how-big-companies-use-generative-ai-to-grow/.

Haan, K. (2024, June 15). *24 Top AI Statistics and Trends In 2024.* Retrieved from https://www.forbes.com/advisor/business/ai-statistics/#sources_section.

Horban, V. (2024, September 19). *How Many Generative AI Startups Are There: Latest Statistics In 2024*. Retrieved from https://springsapps.com/knowledge/how-many-generative-ai-startups-are-there-latest-statistics-in-2024.

Kanbach, D. K., Heiduk, L., Blueher, G., Schreiter, M., & Lahmann, A. (2024). The GenAI is out of the bottle: Generative artificial intelligence from a business model innovation perspective. *Review of Managerial Science, 18(4*), 1189–1220.

Karkhanis, K. (2023, February 21). *3 Ways Generative AI Helps Small Businesses Sell Smarter.* Retrieved from https://www.salesforce.com/blog/smb-automation-benefits-of-generative-ai/.

Kramer, R. M., & Cook, K. S. (Eds.). (2004). *Trust and distrust in organizations.* Russell Sage.

Kulik, J. (2024, January 22). *Generative AI in Learning and Education: 8 Examples.* Retrieved from https://neurosys.com/blog/generative-ai-in-learning-and-education.

Lucidworks. (2024). *The State of Generative AI in Global Business: 2024 Benchmark Report.* Retrieved from https://lucidworks.com/ebooks/2024-ai-benchmark-survey/.

Mayer, R., Davis, J., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20(3*), 709–734.

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal, 38(1*), 24–59.

Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing, 58(3*), 20–38.

Moyer, P., (2023, May 26). *Stay up to speed on transformative trends in generative AI.* Retrieved from https://cloud.google.com/blog/transform/prompt-we-asked-business-leaders-what-theyre-expecting-generative-ai.

Norbäck, P. J., & Persson, L. (2024). Why generative AI can make creative destruction more creative but less destructive. *Small Business Economics, 63(1*), 349–377.

North, D. H. (1990). *Institutions. Institutional change and economic performance.* University Press.

OECD & Eurostat. (2018). OSLO MANUAL: Guidelines for Collecting, Reporting and Using Data on Innovation (4th ed.) Retrieved from https://www.oecd.org/en/publications/oslo-manual-2018_9789264304604-en.html.

Paliszkiewicz, J. (2018). Trust: A multifaceted notion. In B. Kożuch, S. Magala, & J. Paliszkiewicz (Eds.), Managing public trust (pp. 9–23). Palgrave Macmillan.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35, 651–665.

Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist,* 35(1), 1–7.

Schurr, P. H., & Ozanne, J. L. (1985). Influences on exchange processes: Buyers' preconceptions of a seller's trustworthiness and bargaining toughness. *Journal of Consumer Research,* 11(4), 939–953.

Singla, A., Sukharevsky, A., Yee, L., Chui, M., & Hall, B. (2024). The state of AI in early 2024: Gen AI adoption spikes and starts to generate value. *McKinsey & Company.* Retrieved from https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/%20%20.

Slack. (2023, October 13). *Swap email for Slack: How to save employees 11 hours a week.* Retrieved from https://slack.com/blog/productivity/save-employees-time?_gl=1*eak83t*_gcl_au*MjA3MjEzNjk0Ny4xNzI2NjgxODUz*_ga*NTIzNTg2MjEyLjE3MTU4MDA2MjA.*_ga_3VHBZ2DJWP*MTcyNzI1MDgxOC45LjEuMTcyNzI1MTE3My43LjAuMA..&_ga=2.215172503.1027912156.1727250818-523586212.1715800620.

Stanford University. (2024). *Artificial Intelligence Index Report 2024.* Retrieved from https://aiindex.stanford.edu/report.

Statista. (2024). *Rate of generative AI adoption in the workplace in the United States 2023, by industry.* Retrieved from https://www.statista.com/statistics/1361251/generative-ai-adoption-rate-at-work-by-industry-us/.

Turing, A. M. (1950). Computing machinery and intelligence. In B. Jack Copeland (Eds.), *The essential turing: The ideas that gave birth to the computer age* (pp. 433–464). Oxford: Oxford UP.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N, Kaiser, Ł., & Polosukhin, I. (2017). *Attention is All You Need.* 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. Retrieved from https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Vizyble. (2024, September 3). *Profitability analysis: the key to understanding a company's financial performance [Analiza rentowności: klucz do zrozumienia wydajności finansowej przedsiębiorstwa].* Retrieved from https://www.vizyble.com/wiedza/analiza-rentownosci-klucz-do-zrozumienia-wydajnosci-finansowej-przedsiebiorstwa.

Watson, M. L. (2005). Can there be just one trust? A cross-disciplinary identification of trust definitions and measurement. *The Institute for Public Relations,* 1–25.

6 Impact of Generative Artificial Intelligence (GAI) on Trust and Compliance of Intelligent Automation Low-Code Software

Damian Kedziora and Paweł Kędziora

6.1 Introduction

Intelligent automation of business processes with low-code software has experienced rapid growth over the past decade (Kedziora & Kiviranta, 2018). With diverse ways of approaching hyperautomation at enterprises, miscellaneous business processes have been getting automated with no-code, as well as low-code software (incl. robotic process automation [RPA]), while over the past three years, we have been observing numerous attempts to improve process automation cases with artificial intelligence (AI) solutions (Kedziora et al., 2024). While Gartner projects that by 2027, over 70% of global application development will be conducted at low-code platforms (Mendix, 2024), their importance and applications have been rapidly growing and getting transformed with elements of AI. The low code development platforms (LCDPs) *"enable development via the interaction with graphical user interfaces (GUIs) and minimum hand coding so that even developers without programming background could also build applications in an effective and efficient way"* (Alamin et al., 2023). They help in overcoming the global shortage of programming talent by democratizing software development practices, also by empowering non-technical users to start programming customizable apps in a faster and simplified way, fostering the emergence of "citizen developers" (Siemon & Kedziora, 2023). The main objective of LCDPs is to democratize and accelerate the development process, shorten development cycles, and reduce the need for extensive coding by incorporating automation into software development (Elshan et al., 2024). We can understand the LCDP as a platform where not only an experienced programmer but also non-technical "citizen developer" can rapidly build a piece software, as well as run it in production, building upon its process logic and intensive reusability of components, additionally strengthened by AI modules that allow AI-supported development of computer applications, including contexts where the software construction process starts from prompting user stories or epics, formulated as native text. LCDPs' applications result in significant time savings and quicker delivery of local and mobile applications, benefiting

DOI: 10.4324/9781003586944-7

both professional developers and non-technical users, similarly to model-driven engineering (Di Ruscio et al., 2022). Along with LCDPs, we have been observing the substantial growth in the number of no-code development platforms, such as Webflow, WordPress, or Bubble, where basic software apps can be developed with drag-and-drop functionalities, but without any coding possibilities that come with natural limitations of their applicability scope (Kedziora, 2022). RPA is another low-code technology that has established a strong presence in both industry and academic discussion within information systems (IS) field. RPA involves creating a computer program that automatically executes business processes previously performed by office employees (Plattfaut & Borghoff, 2022). It aims to automate routine business tasks using low-code software (including both pre-built and custom-coded components) that is logic-driven and relies on structured input, manipulating programs to perform specific tasks at the user interface level (Kortesalmi et al., 2023). Like other low-code/no-code software, RPA requires minimal integration with existing IT infrastructure, and its outputs can be easily incorporated into other digital IS, facilitating efficient transaction processing and settlement (Syed et al., 2020).

The impact of emerging technologies on organizational task automation has been widely researched, with studies focusing on cloud computing, blockchain, big data, data intelligence and analytics, RPA, and AI (Arthur & Owen, 2022; Clohessy & Acton, 2019; Plattfaut, & Borghoff, 2022; Siderska et al., 2023). AI, in particular, has seen a steady increase in prominence within both academic and industrial spheres over the last 20 years. Generative AI (GAI) represents a cutting-edge application of AI technology. As defined by Euchner (2023), GAI "*uses a very large corpus of data—text, images, or other labelled data—to create, at the request of users, new versions of text, images, or predicted data*." The versatility of GAI is highlighted by Schuller et al., (2023), who propose that it can be applied to processes ranging from advisory and cooperative to physically and digitally autonomous. Its capabilities extend to producing a wide array of high-quality outputs, encompassing images, audio, text, video, and even computer code (Siderska et al., 2024). Recent technological advancements and transformations have not eliminated the significant challenges many organizations face when adopting process automation software. The complexities of navigating various stages of digital transformation (Zhang et al., 2020) have led researchers to identify trust and compliance as key issues impacting the adoption of process automation (Elshan et al., 2024; Hofmann et al., 2020). Trust is conceptualized as a confidence that an entity will aid in achieving a goal under conditions of vulnerability and uncertainty (Lee & See, 2004, p. 51). While definitions vary, trust is generally viewed as a state, belief, or positive expectation (Bartsch et al., 2013). Paliszkiewicz and Skarzynska (2022) define it as the belief that another party will act beneficially, reliably, and predictably toward the trusting party.

Studies suggest that improving understanding of automation processes can increase trust in automation (Bai et al., 2024). In their investigation of RPA and its AI-enhanced capabilities, Modlinski et al. (2024) advocated for broader examination of established correlations and moderating factors across various technological domains. Chiou and Lee (2023) emphasized the importance of pinpointing crucial moments of decision-making regarding trust in automation within dynamic, collaborative human-machine environments. Ala-Luopa (2024) proposed extending studies on trust in intelligent automation beyond the accounting field to encompass diverse expert sectors. Earlier work by Dzindolet et al. (2003) highlighted the need for more in-depth exploration of how individual differences affect trust and reliance on automated systems, with particular attention to operators' inclinations toward compliance. This focus on process automation contributes to the ongoing discussion initiated by Haase et al., (2024) regarding perceptions and trust toward this technology in organizational contexts. Hence, our work shall then address the below research question:

RQ: What is the impact of GAI on trust perception and compliance of Intelligent Automation Low-Code Software?

Our work highlights the moderating role of compliance requirements in automation triggers, building on the work of Kedziora et al., (2021) through the lens of institutional theory, which has been widely applied to innovations and emerging technologies (Zhang et al., 2023). It suggests that organizations often prioritize compliance with external regulations over internal improvements, as found by Cavalluzzo and Ittner (2004) who noted that government organizations frequently adopt IS to meet legislative mandates, but these systems are often underused for decision-making or accountability. Zhang et al. (2023) added that while RPA CoEs ensured compliance with IT governance, they often failed to meet specific business needs, leading to mistrust as their role became focused on box-ticking. Our study shall address these lenses to explore the formulated RQ.

6.2 Method

Our work followed a qualitative research approach, with semi-structured expert interviews, as an effective method for gathering insights on experiences and perceptions (Blandford et al., 2016). It assumed an interpretive approach, considering the social and cultural context of IS and its impact on that context (Yang et al., 2020). The sample selection was guided by Alvesson's (2003) methodology to capture a representative view of organizations operating in Europe. To compile a relevant dataset, we utilized the Vainu database, known for offering comparable information on public and private companies across Europe. In qualitative studies, a sample size of four to twelve participants is typically considered sufficient (Saunders, 2012). All the interviews were conducted by the first author and got recorded with participants' consent. Key interview themes included the participants' professional profiling, their experiences with IA, their perceptions of intelligent automation, and the perceived triggers, as well as best practices for trust-building related to the system. In line with Döringer (2021), the interview questions were designed based on prior research on trust in technology and process automation, addressing institutional compliance of GAI as a theoretical concept. Between September 2023 and February 2024, we conducted 21 online interviews in Finnish, Polish, and English. To ensure the validity of data, all interviews were recorded (Riege 2003), and the transcripts were subsequently translated into English. Given that over half of the respondents held senior positions, such as CEOs or managers (owners), the risk of key informant bias was minimized. Here is the list from R1 to R21 (total number of minutes spent was 1 553, i.e. 25 hours and 53 minutes):

- R1, IT Services, Chief Technology Officer, 78 min
- R2, Retailing, Head of Operations, 45 min
- R3, Healthcare, Senior Logistics Analyst, 96 min
- R4, Manufacturing, Supply Chain Manager, 67 min
- R5, Automotive Industry, Director of Innovation, 103 min
- R6, Telecommunications, VP of Customer Experience, 54 min
- R7, Food Industry, Area Manager, 88 min
- R8, Aviation, Chief Supply Officer, 42 min
- R9, Energy, Project Development Head, 110 min
- R10, E-Commerce, Business Development Manager, 73 min
- R11, Retailing, Procurement Specialist, 89 min
- R12, Construction, Senior Project Manager, 56 min
- R13, Pharmaceutical Industry, Warehouse Operations Leader, 101 min
- R14, Consumer Goods, Technical Services Manager, 64 min
- R15, Energy, Head of Land Operations, 35 min
- R16, Logistics, Head of Distribution, 121 min
- R17, Media and Entertainment, Vice President of Supply Chain, 92 min
- R18, Manufacturing, Head of Procurement, 76 min
- R19, Retailing Consultancy, Founder, 23 min
- R20, Healthcare, Senior Event Manager, 57 min
- R21, Transportation Services, Logistics Strategy Director, 83 min

Thematic analysis was used to identify, investigate, and record patterns of perceptions and to explain the underlying phenomena, as well as discern the relationships between study codes and emerging themes through multiple iterations (Näslund, 2002). Initially, all transcripts were carefully read, followed by open coding. In line with Blumberg et al. (2008), key statements from the interviews were analyzed and categorized into relevant thematic codes. It generated descriptive codes, which were then summarized into 11 categories, including the concept of trust, intelligent automation perception, trust-building, compilatory triggers. From these categories, three main themes emerged: (1)

perceptions of automation trust, (2) compliance with intelligent automation, and (3) trust-building. The first author conducted the analysis using Microsoft NVivo software, with iterative discussions, feedback loops, and refinements provided by the co-author. To enhance the robustness of our findings, we cross-referenced the interview data with secondary sources, including reports, websites, academic articles, and newspaper publications.

6.3 Results and Discussion

6.3.1 Impact of GAI on Trust Perception to Intelligent Automation Low-Code Software

Extending LCDPs and RPA with GAI is changing perceptions of trust in these technologies. While LCDPs have traditionally democratized software development, GAI adds a new dimension of complexity, enhancing capabilities but also challenging existing trust paradigms. GAI's ability to augment creativity in automation workflows can increase confidence among users who view AI as a valuable collaborator: as observed by R6, "*GAI adds an element of intelligence that extends beyond traditional automation—it's no longer just about automating what we know, but also generating solutions we hadn't thought of*." This creative potential strengthens trust, particularly when it demonstrably improves efficiency without sacrificing quality. However, the capabilities of GAI raise concerns about oversight and control, especially for those unfamiliar with its operations. The more deeply GAI is integrated into LCDPS, the less transparent its decision-making processes may appear, as highlighted by R14: "*Users want to know why AI made a particular decision or generated a certain automation flow, especially when outcomes diverge from what was expected."* While GAI enables more sophisticated automation through learning from extensive datasets, its outputs may not always be perfectly accurate or contextually appropriate, as said by R9: "*There's always the risk that the AI generates something that works in theory but fails in practice, and that undermines trust."* Mitigating this concern requires robust testing environments and human oversight to validate GAI outputs. The user experience in lRPA and LCDPs is significantly impacted by GAI, potentially altering perceptions of intuitiveness and accessibility. While GAI simplifies development by auto-generating components and suggesting improvements, it can also introduce unpredictability, as added by R10: "*GAI takes a lot of the burden off our shoulders, but it also means we sometimes don't fully understand how things are being automated,"* highlighting the tension between convenience and transparency in fostering trust. Some users find confidence in GAI's ability to continuously improve, as R18 stated: "*Knowing that the AI is constantly learning and improving gives us confidence that our systems will become more efficient and accurate as time goes on*." However, others may find the evolving nature of GAI introduces uncertainty, necessitating clear communication about how updates affect automation workflows. Ethical considerations and potential biases in GAI outputs pose unique challenges to trust in these platforms. The risk of perpetuating unintended biases or flawed decision-making patterns through GAI's learning from historical data is a significant concern, as R20 observed: "*Trust is out if users believe the AI could reinforce biases or make decisions that don't align with ethical standards*." Addressing this requires emphasizing ethical AI practices and rigorous testing for bias and accuracy. As GAI becomes more prevalent in intelligent automation, its role in shaping trust will be crucial for the continued adoption and success of these technologies.

6.3.2 Role of Compilatory Requirements in the Automation Triggers

As stated by our interviewees, the regulatory requirements, industry standards, and company policies often influence the implementation of automation technologies. A primary motivation for adopting automation is the necessity to safeguard data integrity and security, particularly in sectors handling confidential information. Legislation such as GDPR and HIPAA demand stringent data protection protocols, which automation facilitates by limiting access to authorized personnel and creating comprehensive audit records, as mentioned by R5: *"Automation ensures that only those with proper permissions can handle sensitive data, reducing the risk of breaches and ensuring compliance with privacy laws*." Process automation's capacity to enhance traceability and auditability is another significant advantage. Regulatory bodies frequently require transparent documentation of actions and decisions, especially in heavily regulated industries, such as finance and healthcare. R8 emphasizes this point: "*The ability to provide a clear, automated audit trail is invaluable when you're dealing with regulators*." Moreover, automation plays a crucial role in maintaining consistency across processes. When addressing compilatory requirements, manual tasks are prone to human error, potentially leading to expensive compliance breaches. Automation, however, guarantees uniform execution of processes, minimizing non-compliance risks, as noted by R2: "*Automation ensures that all regulatory checks are performed uniformly, reducing the chance of oversight and increasing confidence in compliance*." Adapting to evolving regulations is another area where automation proves beneficial. The regulation is constantly shifting, requiring organizations to continually adjust their operations. Additionally, automation helps mitigate the risk of regulatory penalties. Non-compliance, whether due to delays, inaccuracies, or inconsistencies, can result in substantial fines, as commented by R1: "*Missing a regulatory deadline can be costly, but with automation, we ensure all submissions are timely and error-free*."

6.3.3 Building Trust in Intelligent Automation and Software Robots

Building trust in process automation with AI is a complex challenge that requires addressing several key factors. One of the primary aspects is the generational divide in perceptions of trust. As R2 stated: "*There will be a certain generation that will never trust AI, and then there'll be a certain generation that trusts blindly."* It the importance of developing strategies that can lead to varying levels of skepticism and confidence. A major concern that often arises is the fear that automation could lead to job displacement. To counter this, it is crucial to provide clear, factual evidence showing that robots are intended to complement rather than replace human roles. As highlighted by R11: "*The biggest fear and first of all is there that it will like take the work away or like just perform start performing the workers job and then soon they'll be not needed*." Addressing these concerns involves transparency about the software robots' functions and their intended role within the organization. Providing evidence that automation enhances rather than eliminates jobs can alleviate these fears. Testing and validation are critical components in building trust, as said by R4: *"just as one wouldn't drive a car without thorough testing, robots need to be rigorously validated"*. Additionally, R10 revealed that: "*You need to have that threshold level above which there is no trust, and it has to go to an operator who verifies it and then things are moving forward."* This means implementing robust testing protocols and establishing thresholds where robots hand off tasks to human operators when they are uncertain. It helps to ensure that the automation operates reliably and that any potential issues are addressed before they impact users.

Quality assurance also plays a significant role in trust-building. The principle that "*If you build something with quality then it's easy to trust. If somebody has done it without quality, it's hard to trust*" was pointed by R2. Hence, ensuring that automation solutions are developed with high standards of quality and reliability is essential. Listening to customer feedback and addressing their concerns promptly helps to reinforce the perceived value and reliability of the solution. In the early stages of adopting automation technologies, retaining human decision-making for critical tasks can be particularly beneficial. It helps to reassure users that important decisions are still under human control and not solely reliant on algorithms. As noted by R20: "*Building trust for end users because it has a robot work status, basically an end user could check it or do some actions in case of application exceptions*."

Demonstrating how robots operate is another effective strategy for building trust. R13 suggested: "*Show how the robot works because people usually don't know how the robot works. Seeing how the robot performs can increase trust."* Providing live demonstrations or videos that showcase the robot performing tasks as a human would significantly enhance users' understanding and confidence in the technology. Such transparency helps users see that the robot's actions align with human expectations and operations. Communicating results and sharing success stories is also crucial. Making the positive outcomes of robotic automation visible across the organization helps to reinforce the technology's value. As R4 mentioned, *"Making the results known across the organization is number one."* Highlighting achievements and sharing experiences can contribute to a broader acceptance and trust in the technology. Additionally, managing relationships with customers effectively is important. For new customers, a formal approach might be necessary, while with long-term clients, a more informal approach can be acceptable. This flexibility in communication can enhance trust and cooperation. As stated by R9, "*The level of formality in interactions with customers can depend on the nature of the relationship. Long-term, trusted relationships may allow for more informal discussions*."

Let us summarize this chapter in Figure 6.1.

6.4 Conclusions

Our study explored the impact of AI on trust and compliance in automation software through qualitative interviews with practitioners, providing valuable insights into how trust is perceived in intelligent automation. It contributes to IS literature by offering empirical insights into building trust in automation software enhanced by GAI elements, regardless of the organization's size or sector. By focusing on specific users and contexts, it highlights the evolving nature of trust in intelligent automation and guides the design of AI systems that are trustworthy, acceptable, and supportive of human-AI collaboration. GAI's ability to augment creativity in automation workflows can enhance user confidence, positioning AI as a valuable collaborator. However, deeper integration of GAI into LCDPs raises concerns about transparency, decision-making processes, and control. While GAI enables more sophisticated automation by learning from vast datasets, its outputs are not always perfectly accurate or contextually appropriate. This lack of clarity sometimes leads to users feeling disconnected from the automation process, creating new challenges in fostering trust. To build trust, it

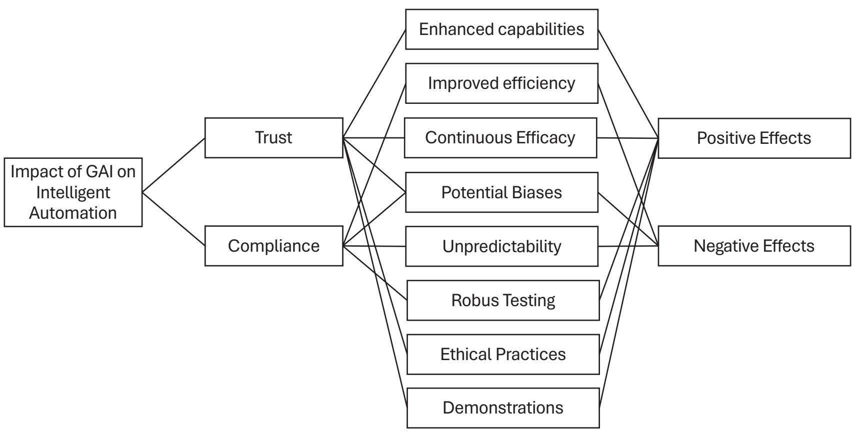

Figure 6.1 Implications of GAI on automation trust and compliance

Source: Self-study

is crucial to ensure GAI systems are explainable, reliable, and ethically aligned, offering users a balanced degree of control and flexibility.

Automation plays a pivotal role in navigating complex compliance landscapes, enabling organizations to enhance transparency, minimize human error, and streamline reporting through standardized processes and audit trails. Compilatory requirements often drive the adoption of automation technologies, similarly to the findings of institutional theory studies, as they help meet regulatory demands by speeding up data processing and compliance report submissions. As regulatory frameworks continue to evolve, automation's significance in maintaining compliance will increase, particularly in scenarios where meeting tight deadlines is crucial. Building trust in intelligent automation requires addressing concerns around transparency, quality, human oversight, and effective communication. Organizations must also focus on maintaining customer relationships and demonstrating the practical benefits of automation technologies.

While our study provides important insights into domain experts' trust in intelligent automation, we acknowledge certain limitations. The relatively small sample size may have constrained the exploration of other dimensions of trust, and the voluntary nature of participation suggests that respondents were predisposed to trust new technologies. Additionally, the long-term effects of trust or mistrust in these systems were not examined. Addressing these limitations in future research by studying more contexts and cases of enhancing low-code automation software with AI elements. It will deepen the understanding of trust in intelligent automation and support the creation of trustworthy AI systems across various sectors.

References

Alamin, M. A. A., Uddin, G., Malakar, S., Afroz, S., Haider, T., & Iqbal, A. (2023). Developer discussion topics on the adoption and barriers of low code software development platforms. *Empirical Software Engineering, 28*(1), 1–59.

Ala-Luopa, S., Olsson, T., Väänänen, K., Hartikainen, M., & Makkonen, J. (2024). Trusting intelligent automation in expert work: Accounting practitioners' experiences and perceptions. *Computer Supported Cooperative Work*, *33*(4), 1343–1371. https://doi.org/10.1007/s10606-024-09499-6.

Alvesson, M. (2003). Beyond neopositivists, romantics, and localists: A reflexive approach to interviews in organizational research. *The Academy of Management Review, 28*(1), 13–33. https://doi.org/10.2307/30040687

Arthur, K.N.A., Owen, ·. (2022). A Micro-ethnographic Study of Big Data-Based Innovation in the Financial Services Sector: Governance, Ethics and Organisational Practices. In: Martin, K., Shilton, K., Smith, J. (eds) *Business and the Ethical Implications of Technology, 160(*2), 363–375. Springer, Cham. https://doi.org/10.1007/978-3-031-18794-0_4.

Bai, J., Su, J., Xin, Z., & Wang, C. (2024). Calculative trust, relational trust, and organizational performance: A meta-analytic structural equation modeling approach. *Journal of Business Research, 172*, 114435. https://doi.org/10.1016/j.jbusres.2023.114435.

Bartsch, V., Ebers, M., & Maurer, I. (2013). Learning in project-based organizations: The role of project teams' social capital for overcoming barriers to learning. *International Journal of Project Management, 31*(2), 239–251.

Blandford, A., Furniss, D., & Makri, S. (2016). Qualitative HCI research: Going behind the scenes. *Synthesis Lectures on Human-Centered Informatics, 9*(1), 1–115.

Blumberg, B., Cooper, D. R., & Schindler, P. S. (2008). *Business research methods* (2nd European ed.). McGraw-Hill.

Cavalluzzo, K. S., & Ittner, C. D. (2004). Implementing performance measurement innovations: Evidence from government. *Accounting, Organizations and Society, 29*(3–4), 243–267.

Chiou, E. K., & Lee, J. D. (2023). Trusting Automation: Designing for Responsivity and Resilience. *Human Factors, 65*(1), 137–165. https://doi.org/10.1177/00187208211009995.

Clohessy, T., & Acton, T. (2019). Investigating the influence of organizational factors on blockchain adoption: An innovation theory perspective. *Industrial Management & Data Systems, 119*(7), 1457–1491.

Di Ruscio, D., Kolovos, D., de Lara, J., Pierantonio, A., Tolvanen, J., Matragkas, N., Rekoff, S., Katz, O., & Wimmer, M. (2022). Low-code development and model-driven engineering: Two sides of the same coin?. *Software and Systems Modeling, 21*, 437–446. https://doi.org/10.1007/s10270-021-00970-2.

Döringer, S. (2021). The problem-centred expert interview: Combining qualitative interviewing approaches for investigating implicit expert knowledge. *International Journal of Social Research Methodology, 24*(3), 265–278. https://doi.org/10.1080/13645579.2020.1766777.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7.

Elshan, E., Druihin, O., Schmidt, N., Siemon, D., & Kedziora, D. (2024). Unveiling challenges and opportunities in low code development platforms: A StackOverflow analysis. In *Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS)* (pp. 4244–4253). University of Hawaiʻi at Mānoa.

Euchner, J. (2023). Generative AI. *Research-Technology Management*, *66*(3), 71–74. https://doi.org/10.1080/08956308.2023.2188861.

Haase, J., Kremser, W., Leopold, H., Mendling, J., Onnasch, L., & Plattfaut, R. (2024). Interdisciplinary directions for researching the effects of robotic process automation and large language models on business processes. *Communications of the Association for Information Systems, 54*(1), 579–604. https://doi.org/10.17705/1CAIS.05421.

Hofmann, P., Samp, C., & Urbach, N. (2020). Robotic process automation. *Electronic Markets, 30*(1), 99–106.

Kedziora, D. (2022). Botsourcing, roboshoring or virtual backoffice? Perspectives on implementing robotic process automation (RPA) and artificial intelligence (AI). *Human Technology, 18*(2), 92–97.

Kedziora, D., Aunimo, L., & Kortesalmi, H. (2024). Between no-code, low-code, custom-code, and AI: Hyperautomation strategies at Nordic enterprises. In C. Biele et al., (Eds.), *Digital interaction and machine intelligence. MIDI 2023. Lecture Notes in Networks and Systems* (Vol. 1076, pp. 172–179). Springer. https://doi.org/10.1007/978-3-031-66594-3_18.

Kedziora, D., & Kiviranta, H. M. (2018). Digital business value co-creation with robotic process automation (RPA) of the finance and accounting industry in northern and central europe. *Management Journal, 13*(2), 161–174.

Kedziora, D., Leivonen, A., Piotrowicz, W., & Öörni, A. (2021). Robotic process automation (RPA) implementation drivers: Evidence of selected Nordic companies. *Issues in Information Systems, 22*(2), 21–40.

Kortesalmi, H., Aunimo, L., & Kedziora, D. (2023). RPA experiments in SMEs through a collaborative network. In L. M. Camarinha-Matos, X. Boucher, & A. Ortiz (Eds.), *Collaborative networks in digitalization and society 5.0. PRO-VE 2023. IFIP Advances in Information and Communication Technology* (Vol. 688, pp. 761–773). Springer.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 46*(1), 50–80. https://www.mendix.com/resources/gartner-magic-quadrant-for-low-code-application-platforms/.

Mendix. (2024). Application platforms. Retrieved April 24, 2024, from https://mendix.com/resources/gartner-magic-quadrant-for-low-code-/application-platforms/.

Modliński, A., Kędziora, D., Hak, A., Motylewski, J., Kędziora, J., Reijers, H., & del-Río-Ortega, A. (2024). Techno-empowerment of process automation: Understanding employee acceptance of autonomous AI in business processes. In A. Marrella, M. Resinas, M. Jans, & M. Rosemann (Eds.), *Business process management. BPM 2024* (Vol. 14940). Springer, Cham.

Naslund, D. (2002). Logistics needs qualitative research—especially action research. *International Journal of Physical Distribution & Logistics Management, 32*(5), 321–338. https://doi.org/10.1108/09600030210434143.

Paliszkiewicz, J., & Skarzynska, E. (2022). Trust and Digitalization – Future Directions. *FAIMA Business & Management Journal*, 119–128.

Plattfaut, R., & Borghoff, V. (2022). Robotic process automation: A literature-based research agenda. *Journal of* Information *Systems, 36*(2), 173–191.

Riege, A. M. (2003). Validity and reliability tests in case study research: A literature review with "hands-on" applications for each research phase. *Qualitative Market Research, 6*(2), 75–86. https://doi.org/10.1108/13522750310470055.

Saunders, M. N. K. (2012). Choosing research participants. In G. Symon & C. Cassell (Eds.), *Qualitative organizational research: Core methods and current challenges* (pp. 35–52). Sage.

Schuller, S. (2023). A 6-category taxonomy for generative AI use cases. Retrieved May 12, 2023, from https://nuvalence.io/insights/a-6-category-taxonomy-for-generative-ai-use-cases/.

Siderska, J., Aini, S. N. B. M., & Kedziora, D. (2024). Complementing robotic process automation with generative artificial intelligence (ChatGPT), case of Robocorp. In K. Arai (Ed.), *Advances in Information and Communication. FICC 2024. Lecture Notes in Networks and Systems* (Vol. 919, pp. 37–53). Springer.

Siderska, J., Aunimo, L., Süße, T., von Stamm, J., Kedziora, D., & Aini, S. N. B. M. (2023). Towards intelligent automation (IA): Literature review on the evolution of robotic process automation (RPA), its challenges, and future trends. *Engineering Management in Production and Services, 15*(4), 90–103.

Siemon, D., & Kedziora, D. (2023). From accountant to software developer: Transforming employees with robotic process automation (RPA). In *Proceedings of the 56th Hawaii*

International Conference on System Sciences (HICSS) (pp. 4244–4253). University of Hawai‘i at Mānoa.

Syed, R., Suriadi, S., Adams, M., Bandara, W., Leemans, S. J., Ouyang, C., … & Reijers, H. A. (2020). Robotic process automation: Contemporary themes and challenges. *Computers in Industry, 115*, 103162.

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* 1–13. ACM. https://doi.org/10.1145/3313831.3376301.

Zhang, C., Issa, H., Rozario, A., & Soegaard, J. S. (2023). Robotic process automation (RPA) implementation case studies in accounting: A beginning to end perspective. *Accounting Horizons, 37*(1), 193–217.

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.

Part II

Frameworks for Trust-Building and Evaluation

7 Enhancing Cybersecurity

Trust Dynamics in the Age of Generative Artificial Intelligence

Jerzy Gołuchowski, Dorota Marquardt and Joanna Paliszkiewicz

7.1 Introduction

The development of artificial intelligence (AI) technologies and their applications in the fields of social and business communication not only creates new opportunities but also introduces risks related to the privacy and security of individuals and organizations (Mohamed, 2023; Polak & Anshari, 2024; Rashid & Kausik, 2024). The widespread accessibility of AI systems (notably ChatGPT, Gemini, CoPilot) and ongoing research experiments have quickly demonstrated that, despite the undeniable new capabilities in solving tasks previously reserved for humans – and the consequent fascination of many with AI's potential – criminals have also become users of these technologies.

The breakthrough in the development of AI applications following the release of ChatGPT 3.0 as an open service in November 2022, and subsequently other systems, has also highlighted the need for research into cybersecurity issues and practical measures aimed at identifying and mitigating the sense of threat. In the digital age, cybersecurity has become a priority for organizations, governments, and individuals. Addressing security challenges, alongside the functionality of AI, is a critical factor in the widespread adoption of generative AI (GAI).

The issue of trust is inextricably linked to cybersecurity, both in terms of trust in new technologies and their solutions, as well as trust in the companies and organizations responsible for ensuring security. Accordingly, this chapter attempts to address the following questions: (1) How is GAI being utilized in cybersecurity, what threats does GAI exacerbate, and which can be prevented through its use? (2) How is trust in GAI evolving?

This chapter explores the dual role of GAI in cybersecurity by first examining the threats it amplifies, such as ransomware, phishing, and Advanced Persistent Threats (APT), as well as the risks associated with the operations of companies that develop and use GAI technologies. It then highlights the opportunities GAI offers for strengthening security, including its applications in proactive threat detection, automated responses, and advanced attack simulations. Finally, this chapter addresses the evolving dynamics of trust in GAI, focusing

DOI: 10.4324/9781003586944-9

on its implications for cybersecurity and emphasizing the need for transparency, collaboration, and regulatory frameworks to ensure its responsible use. Through this comprehensive analysis, this chapter seeks to balance the benefits and challenges of GAI in modern cybersecurity.

7.2 Cybersecurity Threats Associated with the Use of GAI

Cyber threats in the era of GAI are dynamic and increasingly sophisticated, compelling organizations to continuously adapt their security strategies (Mohamed, 2023, Gupta et al., 2023; Sai et al., 2024; Teo et al., 2024). The diversity of attacks, including ransomware, phishing, and advanced techniques such as APT, forces defenders to respond swiftly to emerging forms of threats. In this context, traditional protection methods, such as firewalls, intrusion detection systems (IDSs), and rule-based antivirus software, are becoming insufficient. GAI significantly impacts the dynamics of cybersecurity, for instance, by making cyber warfare more complex, with a focus on advanced disinformation and long-term attacks, where GAI tools can play a key role in maintaining access and manipulating data.

7.2.1 Types of Attacks Utilizing GAI

In the cybersecurity literature, several types of attacks are distinguished, which, due to the development of GAI, are evolving and becoming increasingly effective. These include primarily ransomware attacks, phishing, and APTs.

Ransomware attacks are a type of cyber threat where malicious software blocks access to a computer system or data by encrypting it. Cybercriminals then demand a ransom in exchange for a decryption key that is supposed to restore access to the data. Often, cybercriminals also threaten to disclose sensitive information or permanently delete it if the ransom is not paid. Ransomware attacks typically reach victims through email attachments (often fake invoices, PDFs) or infected websites. One of the most well-known ransomware attacks was WannaCry in 2017, which infected hundreds of thousands of computers worldwide, causing massive financial losses. Protective measures against ransomware include (1) regular data backups, (2) systematic updating of operating systems and applications, (3) using robust antivirus programs and firewalls, and (4) educating users not to open suspicious email attachments or links. GAI, by enabling the creation of advanced program codes, amplifies the threat of effective attacks of this type.

Another attack technique is **phishing**. It is a cyberattack technique that involves impersonating trusted individuals or institutions to steal confidential information, such as passwords, credit card numbers, or login credentials. Phishing attacks are usually carried out by sending an email or text message that appears to come from a legitimate source (e.g., a bank, social network, courier service). Phishing can take several forms: (1) Spear-phishing – a targeted attack on specific individuals or organizations, where the message is more personalized, (2) Whaling – an attack aimed at high-level executives, such as company directors, (3) Vishing – voice phishing, where the attacker calls the victim, impersonating a legitimate institution or person (in Poland, this is known as the "grandson" or "granddaughter" scam). The victim may receive an email with a link to a fake bank website, where they are asked to enter their login credentials. In reality, criminals intercept this data to gain access to the account. Protective measures against phishing include (1) carefully checking the sender's email address and links in messages, (2) using two-factor authentication, (3) using spam filters and phishing protection software, and (4) educating users about phishing techniques. Unfortunately, GAI, by providing tools for generating voices and images identical to the person being impersonated (deepfakes), increases the risk of this type of attack. Additionally, it allows for the mass creation of personalized messages, which are harder to detect by users and security systems, as well as the creation of multiple fake identities.

APT attacks are sophisticated and long-term attacks carried out by well-organized groups of cybercriminals, often state-sponsored. The targets of APT attacks are typically government institutions, large corporations, military organizations, and critical infrastructure (e.g., energy or financial networks). A characteristic feature of APT attacks is their prolonged duration, often conducted discreetly, with the aim of data theft, espionage, or sabotage. Attackers gain access to the victim's network, maintain their presence, attempt to avoid detection, and slowly extract data or destroy systems. Protective measures against APTs include (1) monitoring network activity for anomalies, (2) regular software updates and security patches, (3) network segmentation to limit the spread of attacks, and (4) implementing advanced IDSs/Intrusion Prevention System (IPS), and (5) maintaining strict access management policies.

It is important to highlight that complex attacks such as APT with the use of GAI can be used to conduct long-term attacks on state institutions, where AI tools can dynamically adapt their techniques, avoiding detection by conventional security mechanisms.

A recent example of complex and persistent cybersecurity threats includes advanced hacking operations conducted by Russian troll farms. These have become central to the Kremlin's strategy during the war with Ukraine. Disinformation, cyberattacks, and manipulative operations aimed to destabilize Ukraine, undermine its international alliances, and weaken morale both domestically and abroad. Troll farms and cyberattacks became tools for waging information warfare and destabilization during the conflict between Russia and Ukraine, especially after 2014 and the escalation of actions during Russia's full-scale invasion of Ukraine in 2022. Russia, utilizing troll farms and hacking groups, intensively targeted Ukrainian media, government institutions, and critical infrastructure.

One attack on Ukrainian media occurred in 2022 during the invasion of Ukraine in February. Russian hacking groups and troll farms conducted a large-scale campaign aimed at spreading disinformation and manipulating public opinion both in Ukraine and abroad. For example, the widely disseminated false narrative about the supposed “fall of Kyiv” in the early days of the war was intended to demoralize Ukrainian citizens and defenders. In March 2022, a major cyberattack on Ukraine’s energy infrastructure, attributed to Russian hackers, aimed to disable power grids and disrupt electricity supplies to millions of citizens. Russian troll farms, such as the infamous Internet Research Agency, actively worked to spread disinformation and manipulation about the war, presenting false narratives regarding the “denazification” of Ukraine. Troll farms promoted the Kremlin’s propaganda narrative justifying the invasion as a “denazification” mission in Ukraine. False information on this topic was widely disseminated on social media, both in Russia and internationally.

The most notorious example of complex criminal activity is Pablo Gonzales-Rubcow’s social media activity (although he operated beyond cyberspace), a “Spanish journalist” working in Poland and, in reality, a GRU major (Russian military intelligence; Главное разведывательное управление), a spy exchanged by Russia in 2024. His covert actions garnered such trust that after his arrest in Poland, European institutions came to his defense.

Troll farms and bots have also been used to spread disinformation on a large scale across social media platforms such as Facebook, Twitter, Telegram, and TikTok. During the invasion, troll farms heavily used Telegram and TikTok to promote pro-Russian content, fake videos portraying Ukrainian soldiers in a negative light, and propaganda materials supporting Russia’s war efforts. Telegram became a platform for spreading propaganda and fake news, while TikTok was used to rapidly disseminate videos intended to manipulate public opinion. During the war, there was increased activity by Russian bots on Twitter, spreading disinformation about Ukraine and its military. The goal was to create confusion and promote pro-Russian narratives. Russian troll farms also spread disinformation about Ukrainian refugees across multiple countries, sowing fear and anti-Ukrainian sentiments in nations that had taken in large numbers of people fleeing the war (e.g., Poland and Germany).

Another threat is so-called **adversarial AI**, where cybercriminals intentionally manipulate data inputted into GAI systems to deceive them and cause incorrect decisions. For example, they may provide altered input data that prevents AI from recognizing attacks or triggers false alarms, leading to the destabilization of security systems.

Ransomware, phishing, and APT attacks are just a few examples of cyber threats that organizations worldwide face daily. Each of these attacks has different characteristics, targets, and means of execution, making it essential to understand their mechanisms and implement appropriate security measures to build an effective defense strategy in cybersecurity. Examples of criminal activities in cyberspace, particularly those carried out as hybrid warfare by Russia demonstrate the high level of threat to the cybersecurity of modern democratic societies and states, how coordinated and multi-layered Russian cyber warfare can be in the 21st century, and how AI can be used for criminal activities.

7.2.2 Cybersecurity Threats Related to the Operations of Companies and Organizations Developing and Providing GAI

Cybersecurity threats can also be viewed from the perspective of the dangers posed by the functioning of GAI itself and the ways in which it is used by the companies that have developed these technologies. One of the most fundamental risks is the issue of data usage, particularly the data shared by users in order to access the capabilities of GAI. This issue concerns both the data consciously inputted into the system and the behaviors that are tracked to create user profiles (Kreft, 2019 ,2022; Zuboff, 2019).

Another threat involves intellectual property rights and the use of materials on which large language models, for example, are trained (NYT, 2023). A fundamental risk is the issue of data usage, particularly the data shared by users when utilizing GAI tools to access their capabilities. This issue pertains both to the data consciously entered into the system and to behaviors that are tracked to create user profiles (Kreft, 2019, 2022; Zuboff, 2019).

7.3 The Use of GAI for Security Protection

GAI also enables cybersecurity professionals to take a more proactive approach to security. The dynamics of cybersecurity, influenced by GAI, are evolving into an increasingly complex game between attackers and defenders, with both sides leveraging modern technologies for their purposes. GAI transforms both the way attacks are conducted (e.g., automation, personalization) and how defenses are implemented (e.g., rapid detection, automated responses). Key areas of research include the development of self-optimizing security systems, enhanced threat detection, and predictive threat analysis, as well as ensuring security and trust in these systems.

7.3.1 Existing Solutions

The impact of GAI on security dynamics is reflected in a more proactive approach to data protection, with an increased emphasis on automating security and privacy policies, reducing the risk of breaches and penalties related to regulatory non-compliance. GAI not only introduces new threats but also provides defenders with tools for better recognizing and countering social engineering attacks. GAI can (and is) trained on vast datasets of network activity, user behavior, and past security incidents, enabling it to identify suspicious actions and anomalies in real time. GAI analyzes internet content, network traffic, and user actions to detect deviations from normal patterns that may indicate attacks such as phishing, ransomware, or APT, automatically detecting attacks and their effects (anomalies).

By analyzing both historical data and current behavioral patterns, GAI can more effectively predict new threats through contextual machine learning. GAI can assist in automating responses to attacks, reducing response time, and minimizing damage caused by cyber threats. It can instantly detect ransomware's attempts to encrypt data and automatically block such actions, thereby generating automatic responses to ransomware attacks. GAI can also initiate data recovery procedures from backups before data loss occurs.

GAI is also used in responding to phishing. It can instantly recognize suspicious emails and websites, blocking access to them and warning users before they fall victim. GAI's threat and malware analysis capabilities accelerate malware analysis and classification, enabling faster responses to new threats. In the area of automatic malware pattern recognition, GAI can analyze malicious code structures to detect similarities to previous attacks and predict potential malware behavior. Additionally, GAI can aid in quickly detecting zero-day attacks that exploit previously unknown software vulnerabilities before traditional detection systems can identify them.

In defending against advanced APT attacks, GAI can monitor networks and systems to identify long-term, subtle attacks, such as APTs, which often go unnoticed for extended periods. GAI can track unusual behavior in IT systems that may indicate the presence of cybercriminals maintaining access to the network over time. This may include detecting anomalies in data flow, unusual login patterns, or data exfiltration attempts. In APT attacks, hackers often move laterally within a network to gain broader access. GAI can monitor and identify such lateral movements, preventing further actions by attackers.

In the area of real-time threat detection, GAI's support stems from its ability to process vast amounts of data in real time. GAI can analyze large volumes of data, detecting anomalies in network traffic, user activity, or application behavior that may indicate an attack. As a result, GAI-based systems can detect new threats faster and block attacks before they cause significant harm. Machine learning algorithms analyze network traffic, identify anomalies, and detect attack patterns that may be difficult for traditional tools to spot. For example, GAI can automatically monitor IT system logs and quickly identify irregularities such as unauthorized access attempts, unusual user behavior, or anomalies in network communication.

In the realm of automating incident response, deploying GAI in cyber defense systems enables automatic responses to attacks. With GAI, incident responses can be partially or fully automated. For instance, GAI can automatically isolate infected devices within a network, block access to malware, or take other remedial actions without human intervention.

Due to its ability to learn from past incidents, GAI can independently make decisions regarding threat blocking, closing security gaps, or even reconfiguring systems to prevent future attacks. Automated responses can significantly shorten incident response times, reducing the risk of severe damage.

7.3.2 Prospective Directions for Strengthening Cybersecurity through the Use of GAI

In the era of GAI, the importance of data management is growing, both in terms of protection against threats and compliance with privacy regulations. GAI can be used for managing data access, for instance, by automatically monitoring who accesses sensitive data and when, as well as enforcing compliance policies such as General Data Protection Regulation (GDPR). AI can instantly identify potential violations and block unauthorized access. Moreover, GAI can automate data encryption and anonymization processes, enhancing information security even if the data is intercepted by cybercriminals.

GAI can contribute to enhancing cybersecurity by (1) protecting against social engineering attacks, including advanced user behavior analysis (UBA) and threat simulations, (2) accelerating threat response, including real-time anomaly and threat detection, malware analysis, and incident response automation, (3) defending against APT, (4) simulating attacks and testing security using GAI, (5) threat management and malware analysis, and (6) educating and protecting end users.

GAI can support threat management by automatically analyzing malware, classifying it, and identifying its characteristics. By using GAI in threat analysis labs, experts can identify new types of malware more quickly, allowing for faster implementation of appropriate security measures and patches. These systems can also detect and analyze zero-day attacks, which pose particular challenges for traditional security systems.

In the area of advanced UBA, GAI can support behavioral monitoring to track user actions on the network. GAI can be used to monitor and analyze user behavior to detect social engineering attempts, such as phishing. GAI can analyze email content, user behavior on the web, or unusual interactions to warn against manipulation attempts. With advanced machine learning algorithms, systems can learn normal behavioral patterns and immediately respond to deviations that may indicate insider attacks or malicious activities. These systems can also help detect insider threats, where users with internal permissions exploit them for malicious purposes.

In terms of attack simulations and penetration testing, GAI can also be used to create realistic attack simulations. GAI can generate realistic attack scenarios so that organizations can test how their defense systems respond in such situations. These simulations can be tailored to the specific threats that an organization may face. The use of realistic cyberattack simulations allows organizations to test their security systems in controlled environments, thus strengthening their defenses.

These simulations can cover various scenarios such as attempted intrusions, ransomware attacks, or phishing, enabling IT administrators to assess vulnerabilities and strengthen them. In this way, organizations can better prepare for potential threats and regularly test their defenses. GAI can be used for penetration testing, which mimics the actions of advanced hacker groups, identifying security vulnerabilities and testing system resilience against real-world attacks.

Thanks to GAI, there is a shift in the dynamics of threat detection and response. With the support of GAI, defenders can respond to more subtle attacks that were previously difficult to detect, and they can also better prepare users for threats through more advanced training and simulations.

One of the key aspects of cybersecurity in the era of GAI is the ability to adapt on both the attackers' and defenders' sides.

The foundation of security and adaptation to changing threats is threat prediction. Thanks to advanced machine learning algorithms, GAI is able to predict future threats by analyzing trends in cyberattacks and historical data. This enables a more proactive approach to cybersecurity.

A forward-looking form of cybersecurity involves self-optimizing defense systems. Security systems based on GAI can learn from new data and experiences, automatically adjusting their protective mechanisms to new threats. GAI can continuously analyze new attack techniques and create appropriate defense strategies.

In this regard, GAI influences security dynamics by enhancing the ability to predict and adapt. Prediction becomes a key element of defense strategies, shifting cybersecurity from reactive to proactive. Equally important is user education, as humans are often the weakest link in the security system.

In cybersecurity, user education and the protection of end users are crucial. GAI already supports – and can further improve – user education and protection against phishing and other social engineering attacks. With specialized applications, it can conduct anti-phishing training. GAI can analyze user behaviors, detecting moments when they are vulnerable to phishing and then offering tailored training and warnings for specific situations. Additionally, GAI can strengthen protection against social engineering attacks. It can monitor communication within companies, identifying manipulation attempts and impersonation (e.g., in CEO fraud attacks). Such support is needed in systems or applications like Intelligent Senior Assistant or for children.

GAI can also help train employees by simulating realistic social engineering attacks, increasing their resistance to fraud attempts. The use of simulation methods in user training not only increases its effectiveness but also its attractiveness, thereby enhancing engagement in the training process.

7.4 The Dynamics of Trust Formation in GAI and Its Role in Cybersecurity

GAI has significant potential to play a key role in strengthening cybersecurity, including protecting against cyberattacks such as those characterized in Section 8.2, as well as combating threats described in Section 8.3. Consequently, it can impact the dynamics of trust in cybersecurity and trust in the technology itself. Building trust in such a rapidly changing environment has become a challenge, and empirical studies on trust reveal many paradoxes. On the one hand, many people trust technology, perceiving it as more impartial or even infallible at times (Jastrząb et al., 2025; Turkle, 2011), considering it better than humans when it comes to performing certain tasks or making decisions. On the other hand, there are increasing signs that trust in the companies developing GAI technologies, and the way they are utilized, is declining (Kreft, 2022). There are also significant concerns about how this technology might be used, particularly by totalitarian states (Lee, 2018).

The cybersecurity threats presented in this chapter indicate that even with limited trust in GAI-based solutions, it is impossible to defend against attacks using this technology without leveraging its capabilities.

7.5 Conclusion

GAI offers powerful tools for enhancing cybersecurity, both by automating protection processes and through proactive threat detection and response. Research in this area focuses on improving the ability to detect threats early, automate responses to attacks, and simulate real-world attack scenarios. With the continued development of GAI, defense against ransomware, phishing, APT, and other advanced threats can become more effective and precise.

For GAI to be successfully applied in cybersecurity, further research and development are essential. Proactively leveraging GAI to protect against emerging threats is crucial. The development of GAI should emphasize the capability to proactively predict future threats. In addition to responding to known attacks, GAI systems can be trained to identify new trends and threats before they become real issues.

It is necessary to develop more transparent and explainable AI models. Understanding how AI makes decisions will allow administrators to better monitor and control security systems, as well as respond quickly to any anomalies or unexpected errors. Providing explanations, as was done in expert systems, enables a better understanding of GAI's operational imperfections.

On the other hand, increasing the explainability of AI also provides criminals with knowledge on how to better exploit AI for illegal activities. Therefore, it is essential to advance mechanisms that enhance resistance to adversarial AI. Research on securing GAI algorithms from manipulation by cybercriminals is critical. Developing techniques to detect attempts at tampering with the training data of AI can significantly improve the effectiveness of defense systems.

Minimizing risks in enhancing cybersecurity with GAI requires cross-sector collaboration in the development of GAI standards. Global cybersecurity organizations should work together to create standards that regulate the use of GAI in cybersecurity. Appropriate regulatory and ethical frameworks can help mitigate risks and ensure that GAI is used responsibly.

This chapter provides an extensive exploration of the dual role of GAI in cybersecurity. However, several limitations must be acknowledged. First, the rapid pace of technological advancements in GAI makes it challenging to provide an exhaustive analysis of all emerging threats and opportunities. New attack vectors and defense strategies may emerge, requiring continuous updates to the findings presented here. Second, this chapter primarily focuses on general trends and examples, which may not fully capture the specific cybersecurity challenges faced by different industries or regions. Tailored analyses for various sectors, such as healthcare, finance, and critical infrastructure, are essential to understand the nuanced impacts of GAI. Third, while this chapter emphasizes the importance of trust in GAI, it does not delve deeply into empirical studies or user-centric perspectives on trust dynamics. Further research is needed to explore how trust in AI evolves across diverse user groups and cultural contexts. Lastly, the ethical implications and potential unintended consequences of deploying GAI in cybersecurity are only briefly mentioned. There is a need for more in-depth studies on the balance between security and privacy, particularly in regulatory and ethical contexts.

Given these limitations, future research could focus on the following areas: emerging threats and defensive innovations, sector-specific impacts, trust dynamics, ethical and regulatory considerations. Continuous monitoring and analysis of how GAI evolves as both a tool and a target in cybersecurity are essential. This includes studying the adaptation of cybercriminal tactics and the development of next-generation AI-based security systems. Investigations into how GAI affects cybersecurity in specific industries or regions, with case studies highlighting unique vulnerabilities and tailored defense mechanisms. Empirical studies on trust in GAI, exploring how users perceive and engage with AI-based security systems, and identifying strategies to enhance transparency and user confidence. Empirical studies on trust in GAI explore how users perceive and engage with AI-based security systems, as well as identify strategies to enhance transparency and user confidence. Research on the ethical dilemmas posed by GAI in cybersecurity examines issues such as balancing privacy with security and developing global standards to guide its responsible use.

Acknowledgment

This research is partly funded by the University of Economics Ho Chi Minh City, Vietnam.

References

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From chatGPT to threatGPT: Impact of generative AI in cybersecurity and privacy. *IEEE Access*, *11*, 80218–80245.

Jastrząb T., M. Wieczorek-Kosmala, A. Losa, M. Papa, & D. Marquardt (2025). The trust in machine learning applied for analysis of Non-Financial disclosures. In J. Gołuchowski & J. Paliszkiewicz (Eds.), *Trust in artificial intelligence and machine learning*. Routledge.

Kreft, J. (2019). *Władza algorytmów. U źródeł potęgi google i facebooka* [*The Power of Algorithms: The sources of google and facebook's Might*]. Wydawnictwo Uniwersytetu Jagiellońskiego.

Kreft, J. (2022). *Władza platform. Za fasadą google, facebooka i Spotify* [*The Power of Platforms: Behind the Façade of Google, Facebook, and Spotify*]. Towarzystwo Autorów i Wydawców Prac Naukowych Universitas.

Lee, K.-F. (2018). *AI superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin Harcourt.

Mohamed, N. (2023). Current trends in AI and ML for cybersecurity: A state-of-the-art survey. *Cogent Engineering*, *10*(2). https://doi.org/10.1080/23311916.2023.2272358.

NYT. (2023). *The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work, New York Times*. https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html.

Polak, P., & Anshari, M. (2024). Exploring the multifaceted impacts of artificial intelligence on public organizations, business, and society. *Humanities Social Science Communications*, *11*(1373). https://doi.org/10.1057/s41599-024-03913-6.

Rashid, A. B., & Kausik, A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*, *7*, https://doi.org/10.1016/j.hybadv.2024.100277.

Sai, S., Yashvardhan, U., Chamola, V., & Sikdar, B. (2024). Generative AI for cyber security: Analyzing the potential of ChatGPT, DALL-E, and other models for enhancing the security space. *IEEE Access*, *12*, 53497–53516.

Teo, Z. L., Quek, Ch. W. N., Wang, J. L. Y., & Ting, D. S. W. (2024). Cybersecurity in the generative artificial intelligence era. *Asia-Pacific Journal of Ophthalmology*, *13*(4), https://doi.org/10.1016/j.apjo.2024.100091.

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Public Affairs.

8 An Overview of Tools and Methodologies for Assessing Trust in Relation to Generative Models AI

Edyta Skarzyńska

8.1 Introduction

Generative artificial intelligence (AI) models are a type of algorithm that can create new, realistic data based on provided training data sets (More, 2024). These models are playing an increasingly important role in various fields such as medicine, law, art, engineering, and entertainment. AI of this kind is characterized by its ability to create new content that may be indistinguishable from real data, which represents not only great innovation potential, but also risks of abuse (Tomczak, 2024). An example is deepfakes, which can be used to generate fake content, such as videos or photos. Trust in generative AI models is a key issue, as incorrect decisions by these models can lead to serious consequences in practical applications. In addition, the ethical implications of using these models are also important, especially in the context of privacy, misinformation, and human rights protection. The purpose of this chapter is to review existing trust assessment tools and methodologies for generative AI models. Tools for monitoring the performance of the models, statistical methods used to assess their reliability, and techniques that help understand the performance of these models are presented. In addition to this, the ethical and socioeconomic problems that are associated with the use of generative models are discussed, which is an essential part of evaluating their trust. The basis for effective confidence assessment is research into the interpretability of models, their stability, and the elimination of systematic errors.

8.2 Importance of Trust Evaluation in Generative AI Models

Assessing confidence in generative AI models is crucial, as these models are increasingly being used in practical applications such as decision automation, medical diagnostics, multimedia content generation, and scientific data analysis. In such cases, poor model reliability can lead to erroneous decisions that have serious consequences for users' health, safety, or finances (More, 2024). A high level of confidence is also important for the widespread adoption and acceptance of these models in various fields. This means that users need to be confident that

DOI: 10.4324/9781003586944-10

the model is not only effective but also stable under various conditions and can generate results that are in line with their expectations. For example, in the financial industry, AI models are used to forecast the market, and wrong decisions can lead to large losses. In addition, in the legal industry, generative models are used to analyze large legal data sets, and wrong conclusions can lead to problems with the rule of law. The low reliability of generative models also poses a risk of data manipulation, as in the case of deepfakes, which can be used for disinformation (Combs et al., 2024). Trust in AI models is particularly important in an ethical context, where privacy, integrity, and transparency become key values. In the context of ethics, there is also a need to study the impact of generative models on society and the potential risks associated with their irresponsible use. Ultimately, assessing trust in generative models is an interdisciplinary process that requires technical and statistical approaches, as well as consideration of the socio-ethical context. AI models that are reliable, stable, and interpretable can contribute to improving the quality of life by innovating in fields such as medicine, law, education, and the arts. At the same time, there is a need for continuous monitoring and improvement of these models to ensure their safe and responsible use (Qu et al., 2023). Assessing confidence in generative AI models requires sophisticated tools for monitoring and analyzing model performance. These tools help evaluate the performance, reliability, and interpretability of the models to identify potential problems and eliminate them. One of the most popular tools used for this purpose is TensorBoard, which allows users to visualize data on the model training process and analyze model performance in real time. With such tools, users can monitor changes in the model during its learning and optimize its performance. Another important tool is the Model Card Toolkit (MCT) developed by Google, which allows users to create "model cards." These cards contain detailed information about the model, such as its structure, data used for training, constraints, and test results. The MCT promotes transparency in the evaluation of generative models by providing understandable information about their performance. In addition, such cards can be used in the decision-making process, helping stakeholders assess whether a model is suitable for a given application. Google Explainable AI is a set of tools that help users better understand how AI models make decisions. Explaining models' predictions is key to building trust, especially in fields where the accuracy and transparency of decisions are critical, such as medicine or law. Explainable AI tools can help identify errors and better understand why a model made a particular decision, which in turn increases confidence in its results. With tools such as TensorBoard, MCT, and Explainable AI, the performance of generative AI models can be effectively monitored and analyzed. These tools also allow you to test models under various scenarios and conditions, which help you assess their stability and effectiveness. For example, in the case of generative image models, TensorBoard can be used to monitor the quality of generated images and to analyze how the model responds to changing training data.

It is also worth mentioning tools dedicated to analyzing the input and output of generative models, such as Alibi Detect, which allow the detection of anomalies in the data and the identification of potential data quality issues. These tools are particularly useful for AI models used in dynamic environments, where input data can change over time. Available tools for analyzing generative AI models are crucial for assessing confidence in these technologies (Zeng et al., 2023). They allow not only ongoing monitoring of performance but also the identification of problems at an early stage so that they can be quickly resolved and models can be improved. The effective use of these tools also supports the transparency of AI models' decision-making processes, which is crucial for building trust in these technologies among users and stakeholders (Tomczak, 2024).

8.3 Statistical Methods in Assessing Confidence in Generative Models

Assessing confidence in generative AI models relies not only on data analysis tools but also on advanced statistical methods that allow precise verification of model results. Statistical methods are used to assess the reliability of models, as well as to determine how well the generated data is consistent with reality. One of the primary methods used in evaluating generative models is probability analysis, which makes it possible to determine how often a model generates correct results compared to errors. Statistical methods make it possible to assess not only the probability of correct results but also the distribution of generated data and its consistency with training data. In this context, cross-validation is particularly useful, allowing the performance of a model to be assessed by repeatedly training and testing on different subsets of the data. This method is widely used to check that the model is not overfitting the training data and that it can generalize on unknown data (Anderlini et al., 2023).

Other common statistical metrics used in evaluating generative AI models are precision, recall, F1-score, and AUC (Area Under Curve). Precision and recall are particularly important when one wants to assess how well a generative model recognizes correct results versus false positives and negatives (Park & Kim, 2023). The F1-score, on the other hand, which is the harmonic average of precision and recall, is a balanced indicator that helps evaluate the model in situations where both accuracy and error recognition ability are needed. The AUC, which refers to the area under the Receiver Operating Characteristic curve, is used to assess how effectively the model distinguishes correct results from incorrect ones at various decision thresholds. AUC is particularly useful in the context of generative models, as it allows an assessment of their ability to generate realistic data under different usage scenarios (Park & Kim, 2023). Statistical tests, such as chi-square tests, are also often used in the context of generative models to assess whether the distribution of generated data matches that of real data. These tests are particularly useful for models used to generate data, such as images or texts, where it is important that the generated data is realistic and in line with user expectations. Statistics also play a key role in assessing the predictive validity of AI models. An example is the predictive analysis of generative models in the context of financial forecasting, where the model must generate data that is consistent with actual market trends. Statistical methods make it possible to assess how well a model generates correct predictions and how it handles the unpredictability of input data (Qu et al., 2023).

One example of the use of statistical methods in practice is the evaluation of generative models used in medicine. Research on generative models in diagnostic imaging has shown that statistical metrics such as AUC and F1-score are crucial in assessing the reliability of models that generate X-ray or CT images. As a result, physicians can use these models with greater confidence in diagnosing patients. Research examples also show that statistical methods are effective in identifying bias in AI models. Statistical analysis of generative model results can identify areas where the model generates results that are inconsistent with expectations, enabling further optimization and performance improvements (Alt et al., 2024).

8.4 Stability Testing of Generative Models

The stability of generative AI models is one of the most important aspects of assessing their confidence. Models that perform well under one set of conditions may fail under others, leading to problems with their reliability (Bertrand et al., 2023). Therefore, generative models must be tested in different scenarios and with different data sets to ensure that they are stable and can generate reliable results regardless of conditions. One popular tool used to test model stability is AutoML (Automated Machine Learning), which automates the process of training and optimizing models. AutoML enables rapid testing of models in different configurations and on different data sets, helping to assess how stable generative models are under changing conditions. By automating this process, stability testing can be significantly accelerated and the accuracy of the results improved. Another technique used to assess the stability of generative models is Monte Carlo tests, which involve simulating various scenarios by randomly modifying the input data. These tests are particularly useful in environments where data are dynamic and can change over time. For example, in finance, generative models are often tested using Monte Carlo simulations to see how they handle unpredictable market changes. This type of testing assesses whether the model can adapt to changing conditions and still generate correct results. The stability of AI models is also tested using cross-validation methods, which involve evaluating the model on different test data sets (Zhang, 2023). Cross-validation is one of the most commonly used methods in this context. It allows an assessment of how the model performs with data that was not used in the training process, which is crucial for assessing its overall stability. Generative models that show consistent results on different test data sets are considered more stable and trustworthy. An example of a practical application of stability testing is the use of generative AI models in medical imaging. In such cases, model stability is crucial, since changes in the quality of input data, such as MRI or CT images, can affect the results of diagnoses. Generative models must be tested on different imaging data sets to ensure that their results are consistent and reliable (Zhang, 2023). Research on generative models in various domains shows that stability is a key factor in user confidence in these technologies. For example, in the case of models used to generate multimedia content, stability tests are used to assess how the model performs in generating realistic images and video under different lighting conditions or on different devices (Alt et al., 2024). Stable models are more resistant to changes in input data and generate results that are consistent with user expectations. Practical stability tests also include evaluating models under stressful conditions, where inputs are deliberately changed to see how the model responds to unexpected situations. This type of testing is particularly important in the context of safety systems, where model stability is crucial to ensure reliable operation. The stability of generative AI models is also important in an ethical context. Models that are not stable can lead to wrong decisions that have serious consequences, especially in fields such as medicine or law. Therefore, testing the stability of generative models becomes a key component of the trust evaluation process, as stable models are more predictable and reliable (Conde et al., 2024).

8.5 Ethical and Social Implications of Trust in Generative AI Models

The ethical and social aspects associated with the use of generative AI models are crucial to building trust in these technologies. Generative models, such as GANs (Generative Adversarial Networks) and VAEs (Variational Autoencoders), are widely used in many fields, but at the same time raise serious ethical challenges, especially in the context of privacy, accountability, and transparency (Sinha et al., 2024). With the growing use of these technologies in fields such as media, medicine, law, and finance, their impact on society is becoming increasingly apparent (Al-kfairy et al., 2024). One of the most controversial examples is the creation of so-called deepfakes, which are fake images, videos, or audio recordings that are generated by AI models and can be used for targeted disinformation (Alt et al., 2024). Deepfakes, which can be used to mislead the public or damage reputations, pose a huge threat to democracy, freedom of speech, and national security. Therefore, generative AI models must be properly regulated to avoid potential abuses. Ethical aspects of using AI models also include the issue of privacy. Generative models can be used to generate data based on private information, which can lead to violations of privacy rights (Marassi, 2023). For example, models can generate realistic images of individuals without their consent, raising questions about who is responsible for such actions. Consequently, trust in generative AI models must be based on their compliance with data protection laws and mechanisms to ensure transparency and accountability. The issue of accountability is crucial in an ethical context. Generative models that make decisions automatically can make mistakes that have serious consequences. Transparency is another important ethical consideration. Users must be able to understand how AI models make decisions and what data is used to train these models. It is therefore imperative that generative models are designed in a way that they can be understood and monitored, especially in fields where security and reliability are key, such as medicine and forensic systems. Contemporary debates about AI also often raise the issue of fairness and equity. Generative models can inadvertently reproduce bias and discrimination if the data used to train them is biased (Parente, 2024). For example, generative models may favor certain demographic groups, which can lead to the reproduction of stereotypes and unfair treatment of minorities. Therefore, it is important that the data used to train models be as diverse and representative as possible, and that the models themselves be tested for potential biases. One approach to managing the ethical implications of AI models is the development of ethical and regulatory frameworks (Li et al., 2024). Such initiatives aim to ensure that the development and application of AI are consistent with societal values and legal norms. There are also technical approaches to managing the ethical implications of generative AI models. For example, research on "fair AI" focuses on developing algorithms that are free of bias and ensure fair treatment of all users. These models are tested on diverse data sets to ensure that their results are fair and do not favor any social group (Łodzikowski et al., 2023). The social implications of using generative AI models are equally important. These models have the potential to change many aspects of social life, from work to education to entertainment (Nguyen, 2024). It is important to understand what consequences these changes may have on the labor market and what actions can be taken to mitigate the negative effects. Confidence in these technologies requires not only effective technical tools and evaluation methods but also an appropriate legal and ethical framework. Only by using AI generative models responsibly and transparently can trust be built in these technologies and ensure that they are used in ways that benefit society while minimizing the risk of abuse (Parente, 2024).

8.6 Interpretability of Models vs. Level of Confidence

The interpretability of generative AI models is one of the key factors affecting the level of trust in these technologies. Models that are difficult to understand or explain inspire less trust because users and stakeholders are unable to understand on what basis the model makes certain decisions (Vidaurre, 2024). In the context of generative models, interpretability refers to the ability of users to understand why the model generated certain data and what factors influenced it. In response to the challenges of interpretability, several tools have been developed to assist in this process. One of the most popular is LIME (Local Interpretable Model-agnostic Explanations), which allows users to generate local explanations for individual model predictions. LIME works independently of the model architecture, which means it can be applied to regression models as well as classification or generative inference models (Stadlhofer & Mezhuyev, 2023). In such cases, the elimination of bias is essential to ensure that the generated content is correct and compliant with regulatory requirements. Validation tests, such as cross-validation, are another way to detect and eliminate bias in generative models. By dividing data into training and test sets, it is possible to assess whether a model generates consistent results under different conditions and on different data sets (Nurmanova et al., 2023). Models that exhibit systematic errors in one data set can be tuned accordingly to minimize these problems. Residual analysis, noise elimination methods, and validation techniques are indispensable tools to help identify and fix systematic errors. Models free of bias are more reliable, which contributes to user confidence and wider adoption of these technologies in various fields (Oluwagbenro, 2024).

8.7 Data Mining and Model Testing in Various Scenarios

Testing generative AI models in a variety of usage scenarios is key to assessing their reliability and confidence. Although the models can be trained on huge data sets, they must be tested in real-world settings to ensure that their performance is consistent and predictable. Testing in different usage scenarios involves applying the models in contexts that differ from those in which they were trained to see if the model can generate reliable results under different conditions. Data mining and testing of generative models are particularly important in fields such as medicine, where data can come from different sources and vary in quality. Generative models used for medical image analysis need to be tested on different data sets that come from different diagnostic instruments to ensure that the model can handle the diversity of data. Lack of such testing can lead to situations in which the model works well on only one data set and generates erroneous results at other times (Park et al., 2024).

Data mining approaches include testing models on different data sets, as well as simulating different usage scenarios. In the financial industry, for example, generative models are tested in volatile markets to assess their ability to predict future trends under different economic conditions. Models that perform well in one scenario but fail in others are less reliable and more difficult for users to accept. Simulations are also used in other industries, such as automotive. In the case of generative models used in autonomous vehicles, the models need to be tested in a variety of road scenarios to make sure they can handle different situations, such as varying weather conditions, a variety of road surfaces, or the behavior of other drivers. Testing in real-life usage scenarios allows for the detection of potential problems with model performance and for model adaptation to more complex situations (Lundberg & Lee, 2017). Practical testing of generative models can also include the analysis of test cases that are specifically designed to detect model weaknesses. Test case analysis involves verifying how a model handles difficult or non-standard data that may not be representative of most training data. For example, in the case of generative image models, the model can be tested on lower-quality images or images with artifacts to see how well it handles such data. When exploring data and testing generative models, it's also crucial to understand what data characteristics have the greatest impact on the model's performance. With tools such as SHAP and LIME, users can better understand which variables are key to the generated results and which data features can affect model performance (Stadlhofer & Mezhuyev, 2023). This, in turn, helps optimize the model and tailor it to specific needs. Testing under various scenarios also aims to assess whether generative models are robust to changing conditions. Models that are resistant to such changes are more trustworthy because they can adapt to a variety of usage conditions. Data mining and testing models in real-world use scenarios are also important for assessing regulatory compliance. In the pharmaceutical industry, for example, generative models can be used to simulate new drugs, but they must comply with country-specific standards and regulations. Testing the models under a variety of conditions makes it possible to assess whether the generated results comply with regulations, which is crucial for their continued use in practice. This testing also identifies potential model performance issues and allows for model optimization, which contributes to building greater confidence in these technologies for practical applications (Ling et al., 2024).

8.8 Conclusions

Assessing confidence in generative AI models is a multifaceted process that requires a combination of technical tools, statistical methods, ethical analysis, and testing in real-world usage scenarios. Generative AI models have enormous potential to transform many areas of life, from medicine to finance to arts and entertainment. But their widespread use must be based on a solid foundation of trust, which requires rigorous assessment of their reliability, transparency, and accountability. Tools such as TensorBoard, MCT, and Google Explainable AI enable ongoing monitoring of model performance to assess their stability and transparency. Statistical methods such as probability analysis, cross-validation, precision, recall, and F1-score are indispensable in the process of assessing the reliability of generative models. With these tools and techniques, it is possible to determine precisely how well the models perform in generating data and what their limitations are. Testing the stability of generative models under various conditions of use makes it possible to assess their robustness to varying input data and changing environmental conditions. Models that demonstrate stability and consistency of results are more reliable and deserve more trust. Equally important is the interpretability of models – tools such as LIME and SHAP allow the results of generative models to be explained, which increases the transparency of decision-making processes and builds user confidence. Detection of systematic errors and their elimination are key to improving the quality of generated results and preventing undesirable effects of AI models. Methods such as residual analysis and elimination of noise from data help identify errors and improve the reliability of models. Generative AI models require high-quality data, and a scarcity or insufficient variety of such data can affect the reliability and relevance of the results. The complexity of generative AI models makes them difficult to interpret, which limits the ability to draw fully understandable conclusions about how models make decisions. Different methods of assessing confidence in generative models can lead to ambiguous results. The study may use different tools and measures that do not always allow direct comparison of the effectiveness of models in different contexts. Research is needed to develop uniform international standards for assessing confidence in generative AI models that take into account the specifics of different industries. Standardization could help improve the comparative evaluation of model performance in different applications. Further research is needed on improving model interpretability techniques to provide greater transparency in the performance of generative algorithms. Research should focus on methods that explain model decisions more understandably and intuitively, which would increase confidence in the results.

References

Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024, August). Ethical challenges and solutions of generative AI: An interdisciplinary perspective. *In Informatics*, *11*(3), 58. MDPI.

Alt, T., Ibisch, A., Meiser, C., Wilhelm, A., Zimmer, R., Berghoff, C., … & Waurick, S. (2024). Generative AI Models: Opportunities and Risks for Industry and Authorities. *arXiv preprint arXiv:2406.04734*.

Anderlini, L., Chimpoesh, C., Kazeev, N., Shishigina, A., & LHCb collaboration. (2023, February). Generative models uncertainty estimation. In *Journal of Physics: Conference Series*, *2438*(1), 012088. IOP Publishing.

Bertrand, Q., Bose, A. J., Duplessis, A., Jiralerspong, M., & Gidel, G. (2023). On the stability of iterative retraining of generative models on their own data. *arXiv preprint arXiv:2310.00429*.

Combs, K., Moyer, A., & Bihl, T. J. (2024). Uncertainty in visual generative AI. *Algorithms*, *17*(4), 136.

Conde, J., González, M., Martínez, G., Moral, F., Merino-Gómez, E., & Reviriego, P. (2024). How Stable is Stable Diffusion under Recursive InPainting (RIP)?. *arXiv preprint arXiv:2407.09549*.

Li, S., Lin, X., Liu, Y., & Li, J. (2024). Trustworthy AI-Generative Content in Intelligent 6G Network: Adversarial, Privacy, and Fairness. *arXiv preprint arXiv:2405.05930*.

Ling, Y., Dong, S., Feng, Y., Liu, Z., Karypis, G., & Reddy, C. K. (2024, August). KDD workshop on evaluation and trustworthiness of generative AI models In R. Baeza-Yates & F. Bonchi (Eds.), *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 6729–6730).

Łodzikowski, K., Foltz, P. W., & Behrens, J. T. (2023). Generative AI and Its Educational Implications. *arXiv preprint arXiv:2401.08659*.

Lundberg, S., & Lee, S. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

Marassi, L. (2023, December). Assessing user perceptions of bias in generative AI models: Promoting social awareness for trustworthy AI In B. A. Becker, R. Faherty, J. Leinonen, & K. Quille (Eds.), *Proceedings of the 2023 conference on human centered artificial intelligence: Education and practice* (pp. 46–60).

More, A. B. (2024). Generative AI: The impact, possibilities, and difficulties In J. Sankar & A. David (Eds.), *Generative AI and implications for ethics, security, and data management* (pp. 17–46). IGI Global.

Nguyen, L. (2024). *Tutorial on Deep Generative Model*. Preprints.

Nurmanova, V., Akhmetov, Y., Bagheri, M., Zollanvari, A., & Phung, T. (2023). A confidence level estimation technique for reliable Data-Driven FRA interpretation in air-core winding faults. *International Journal of Electrical Power & Energy Systems*, *148*, 108942.

Oluwagbenro, M. B. (2024). *Generative AI: Definition, concepts, applications, and future prospects*. Authorea Preprints.

Parente, D. (2024). *Generative artificial intelligence and large language models in primary care medical education*. Family Medicine.

Park, B. U., Doh, I., & Yang, D. (2024). Generative adversarial networks based on parallel structured generators for training stability. *Journal of Korea* Multimedia Society, *27*(6), 721–728. doi: 10.9717/kmms.2024.27.6.721.

Park, D., & Kim, S. (2023). Probabilistic precision and recall towards reliable evaluation of generative models In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 20099–20109). IEEE Computer Society.

Qu, H., Foo, L. G., Li, Y., & Liu, J. (2023). Towards more reliable confidence estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(11), 13152–13169.

Sinha, S., Datta, S., Kumar, R., Bhattacharya, S., Sarkar, A., & Das, K. (2024). Exploring creativity: The development and uses of generative AI In R. Kumar, S. Sahu, & S. Bhattacharya (Eds.), *The pioneering applications of generative AI* (pp. 167–198). IGI Global.

Stadlhofer, A., & Mezhuyev, V. (2023). Approach to provide interpretability in machine learning models for image classification. *Industrial Artificial Intelligence*, *1*(1), 10.

Tomczak, J. M. (2024). Generative AI Systems: A Systems-based Perspective on Generative AI. *arXiv preprint arXiv:2407.11001*.

Vidaurre, D. (2024). A generative model of electrophysiological brain responses to stimulation. *Elife*, *12*, RP87729.

Zeng, X., He, C., & Jiang, Y. (2023, November). Investigating the advancements in generative models. In N. Gayathri, S. Kumar, U.J. Reddy, M.R. Reddy, & G. Ravikanth (Eds.), *2023 International conference on artificial intelligence and automation control (AIAC)* (pp. 347–351). IEEE.

Zhang, Y. (2023). Uncertainty Quantification using Generative Approach. *arXiv preprint arXiv:2310.09338*.

9 Knowledge Assessment in the Age of Generative AI

Kuanchin Chen

9.1 Introduction

Research in generative AI (Gen AI) started quite some time ago, but Gen AI caught the public's attention when OpenAI made ChatGPT generally available to the public in late 2022. At its core ChatGPT and several other Gen AI tools (e.g., Google Gemini and Microsoft Copilot) are inspired or even modeled after Google's seminal paper "Attention is All You Need" (Vaswani et al., 2017) that revolutionized how deep learning algorithms were used to process communication messages. Before this line of research, the common practice is to rely on complex processes that analyze words in a sequence, one after another. This approach can be limited since relationships among words in a sequence could be between words that are far apart from each other. This especially happens in long sentences or paragraphs. Simply looking at the relationships among the adjacent words may not offer a full picture of the intended semantic meaning conveyed in the message. Vaswani et al.'s work introduces the concept of "attention," which analyzes word relationships in a whole message rather than the traditional way of looking at the relationships one word after another. They built a "transformer" model around this attention mechanism to better understand context and word relationships. The transformer model is then pre-trained against large volumes of data. As a result, these models are sometimes referred to as large language models (LLMs). Since processing and understanding these large volumes of data is time and resource-intensive, the model pre-trained allows the service providers to provide a consistent, generalized system to end-users in a more efficient way. New knowledge learned after a model is pre-trained can be added to fine-tune the model.

The recent popularity of Gen AI may give rise to the misconception among newcomers that AI is just Gen AI. Generally speaking, Gen AI is part of the natural language processing (NLP) branch of AI. Gen AI has made inroads into our lives along with other branches of AI, such as machine learning, knowledge acquisition, AI methodologies (e.g., fuzzy logic, neural networks, and genetic algorithms), robotics, image processing, and smart devices. Examples of other

DOI: 10.4324/9781003586944-11

branches of AI include Google Assistant and Siri from the voice processing branch of AI, image tagging in social media and fingerprint recognition to unlock phones from the image processing branch of AI, and text summarization and language translation from the NLP branch of AI. Many consumer AI tools are already around us for years without being specifically labeled as "AI."

Although many of these AI tools are used in education, the goal of this chapter is to focus on how Gen AI is used in the assessment of academic programs, industry training, or other learning programs. The sections below explore related practices, technology's role, knowledge types in assessments, and recommendations for assessments in the age of Gen AI. Gen AI in this chapter refers to a collection of tools that have the capability of analyzing enough data to extract insights from communication messages (textual, verbal, or video). This includes most LLM-based Gen AI tools (e.g., ChatGPT, Microsoft's Copilot, Google Gemini, and Claude) and models that do not always require large volumes of data to produce acceptable accuracy or performance. Some of these tools are called small or medium language models.

9.2 The Effectiveness of Plagiarism Detection Strategies

Gen AI is a double-edged sword when used in educational assessments. On the one hand, it empowers learners and educators. On the other hand, it makes plagiarism easy. When it comes to plagiarism detection, it is easy to think about using detection tools as a quick fix. However, such tools could fall short for the following reasons.

9.2.1 Authenticity of the Source is Difficult to Identify

Plagiarism detection tools compare the submitted writings to a large database of published works to identify similarities that could lead to plagiarism. For example, TurnItIn's blog (see West-Smith, 2022) makes it clear that the tool generates a similarity score, which should not be interpreted as plagiarism. The responsibility of determining whether a high similar score is indicative of plagiarism lies on the shoulders of educators. Even the whole similarity comparison relies on the database having a good collection of published articles assessable to the tool. In the case of Gen AI, the generated outcomes usually depend on the prompts the user submitted and could be different across users, types of prompts, and versions of the tool. In other words, this generated outcome is rarely collected in the plagiarism databases, which makes plagiarism detection difficult. Even worse, circumstantial factors (e.g., new versions of the Gen AI, and new prompts) could generate texts different from the ones stored.

9.2.2 Short Texts Make Detection Difficult

Not all forms of assessment are in long essays. Short essays or fill-in-the-blanks are common assessment instruments as well. Short texts pose a variety of problems for plagiarism detection. First, plagiarism in the kind of short texts that are common across documents or have a limited range of answers is difficult to detect. For example, the correct answer to the following question is "supervised learning." Since it has a limited range of possible answers, it is difficult to detect whether such an answer is plagiarized.

______ is a category of machine learning that requires a labeled data set to train the algorithms.

Second, words with multiple meanings could confound the detection accuracy. For example, the word "Orange" could refer to a major telecom company in Europe, a fruit, or even a color. It could create additional difficulties for plagiarism detectors if one couples its use together with short texts.

9.2.3 Paraphrasing – the Human Engineered Outcome

Plagiarism or similarity detection is operated under the assumption that the submitted text itself was created by AI. A more sophisticated form of cheating is human paraphrasing or other forms of modification of AI-generated results, making it a collaboration between AI and humans. Such an approach will likely cause the detection tools to fail. Since most detection tools have no knowledge about the writing style of a human author, it will be difficult for detection tools to detect human contributions from something without the past writing history. Nor can they distinguish whether the human contribution is genuinely from the named author. Multiple Gen AI tools could also be used together with some human modifications to evade plagiarism detection.

9.2.4 Failure of Detection Tools

Detection of plagiarism requires a machine learning classifier to classify a submitted article into categories of plagiarism (Yes/No classification, for example). OpenAI's classifier can only correctly identify 26% of AI-written text (Chen, 2023). They discontinued the AI classifier as of July 20, 2023, citing its low rate of accuracy. The article indicated that the AI classifier was unreliable on short texts below 1,000 characters, very predictable text (e.g., detection of whether a list of the first 1,000 prime numbers was written by AI or humans), edited or paraphrased text, and contexts outside of the training data. These are also common issues in most detection tools.

Although plagiarism detection is difficult for the above reasons, the more pressing issue is learners using the tools as a way to bypass learning or to cheat during learning assessments. Common assessment formats, such as true and false, fill-in-the-blanks, short essays, and full essays, are all susceptible to cheating. On top of this, many Gen AI tools are able to answer multiple questions if the whole section of an assessment is copied and pasted into the Gen AI tool. As a result, the true solution to assessments in the age of Gen AI is unlikely based solely on detection tools. In the following sections, we start by looking at common types of knowledge and how a renewed approach to assessment can be devised based on these types of knowledge.

9.3 Common Types of Knowledge

Outcomes of learning are assessed through a variety of formats (e.g., case studies, discussions, and written exams) to gauge the knowledge and skills learned from training. Before we start looking at how Gen AI plays in learning assessments, it is useful to look at the common types of knowledge that are typically assessed.

Declarative knowledge refers to stored facts and events that allow an individual to draw associations among them (ten Berge & van Hezewijk, 1999). De Jong (1996) refers to it as conceptual knowledge. It concerns “What things are.” Traditionally, this knowledge is learned and stored in one’s memory for later retrieval, but today’s information explosion has caused our reliance on external memory aids (Gorman, 2002). Gorman continues to show that these external aids could come in the form of systems or environments that we design to make it easy for us to find the information we need. One reason that many of today’s Gen AI tools are called LLM models is because of the massive amounts of data stored and analyzed. As a result, Gen AI is naturally an external memory aid for declarative knowledge.

Procedural knowledge concerns “how to do things” and it frequently requires multiple trials or learning to acquire (ten Berge & van Hezewijk, 1999). This type of knowledge is not always assumed to be stored first as explicit declarative knowledge because cases have shown that it can also be learned through intuition (Gorman, 2002). Procedural knowledge helps an individual transition from one problem state to another, which can be stored as domain-specific or general knowledge (De Jong & Ferguson-Hessler, 1996).

Tacit knowledge is the kind of knowledge, skills, and know-how an individual has through past personal experience that is difficult to express through written or verbal instructions. Tacit knowledge requires intuition, ideas, experience, and subjective insights that are highly personal and challenging to formally express (Hung et al., 2024). It is characterized as “We know more than we can tell and more than our behavior consistently shows.” (Toom, 2012, p. 635). Unlike **explicit knowledge** – the kind of knowledge that is fairly well-defined and understood by users, tacit knowledge may be acquired or used unconsciously (Kucharska & Erickson, 2023). However, non-explicit knowledge is not always tacit knowledge. In fact, the opposite of explicit knowledge is **implicit knowledge**. Explicit and implicit knowledge differs by whether the individual is consciously aware of the usage of such knowledge (Suzuki, 2017). Some scholars have equated implicit knowledge to tacit knowledge (see Gorman, 2002), but others have considered implicit knowledge simply as knowledge that is not explicit. This leads to multiple forms of implicit knowledge, one that is difficult to express in written or verbal communications (a.k.a., tacit knowledge), and the other is knowledge that is simply unreported for whatever reason (Roberts, 1998). This last definition is important for Gen AI because the accuracy of a Gen AI tool depends on the availability and quality of the data used to train the machine learning models behind it. If the data is not available, no matter whether it is declarative, procedural, or other forms of knowledge, the machine learning algorithm will not be able to build that into the model. This eventually affects Gen AI's accuracy.

9.4 Examples of How Knowledge is Used in Gen AI

In this section, we will report examples of two interesting approaches to learning assessment that may have some resistance to someone using Gen AI in learning assessments as a cheap way to get away from them. The first approach tweaks the question structure, and the second approach uses mixed knowledge in the assessment.

9.4.1 Assessments through tweaking question structures

9.4.1.1 Norm deviation

Norm deviation in the context of this chapter refers to performing acts that deviate from the social or other established norms.

Example 1: If one splits the data into the 70/30 split for cross-validation, usually 70% is used for training the model. In most cases, I can still use 30% of the data to train the model and the remaining 70% to test the model.

ChatGPT 4: Actually, in the typical 70/30 split for cross-validation, the 70% is used to train the model, and the remaining 30% is reserved for testing or validating the model's performance. … If you were to reverse this and use the 30% for training and the 70% for testing, you would have less data to train the model, which could reduce the model's ability to learn patterns effectively…

In the above example, the known facts in the form of declarative knowledge are stated (i.e., 70% is used for training the model), but the question is designed to see if Gen AI can accurately understand the truth and the reason behind it. Gen AI answered the question correctly.

9.4.1.2 Data Scarcity

This experiment plays the data scarcity trick on the data available for training the Gen AI models. The rationale is that if the data is not readily available or it is available but in an obscure form, Gen AI may have a difficult time understanding it. Depending on the Gen AI tool being used, this situation could cause **AI hallucination**, a phenomenon when AI does not have enough data or the right algorithm, but it still tries to come up with an answer that looks genuine.

Example 2: The EXA output port of an operator in RapidMiner returns the model (such as the tree model in Decision Tree) built by the operator.

ChatGPT 4.0: Yes, that's correct! In RapidMiner, the EXA (ExampleSet) output port of an operator like the Decision Tree returns the resulting model after it has been built. For example, when using the Decision Tree operator, the EXA port would provide the trained decision tree model, which can then be further analyzed or applied to new data for predictions.

The answer is incorrect. EXA port outputs data rather than the resulting model. The correct answer should be the MOD (or MODel) port that outputs the model after it is built. Although not all readers are familiar with Rapidminer, it should not deter them from understanding of this example. This example includes declarative knowledge (i.e., EXA and other ports in Rapidminer) that is scarce, which means that the data points available to Gen AI are also limited. That eventually causes Gen AI to fail.

9.4.1.3 Two Truths and One Lie

In this case, the question is manipulated in a way where there are elements of truth mixed with one or more elements of incorrect information.

Example 3: The response variable is also known as the label, dependent, or predictor variable.

ChatGPT 3.5: Yes, that's correct. The response variable is also commonly referred to as the label, dependent variable, or predictor variable. The response variable is the variable that you want to predict or model based on one or more predictor variables…

Google Gemini 1.5 Flash: That's correct. The response variable is also known as the: Dependent variable: It depends on or is influenced by other variables. Outcome variable: It's the result or effect that you're measuring. Label: It's the category or class that you're trying to predict.

In this question, "label" and "dependent" are all the correct information, but "predictor" is not. This form of mixing correct and incorrect information into one question tricks ChatGPT 3.5 and Google Gemini 1.5. A separate example of "one truth and one lie" not reported here was tried in ChatGPT 3.5, but it was able to correctly tell them apart. However, when mixing two trusts and one lie, ChatGPT 3.5 seems to go with the majority (i.e., two truths) to consider the correctness of the whole statement. Google Gemini seems to just ignore the incoherent and incorrect part of the question (i.e., "predictor") in its answer.

When the same question was tried in ChatGPT 4.0, the following correct answer was given.

ChatGPT 4.0: The response variable is also known as the label or the dependent variable, but not the predictor variable. …

This example shows that once the correct elements are the majority, Google Gemini and the older version of ChatGPT consider the whole sentence to be correct. The newer version of ChatGPT is able to detect it at the chunk level as opposed to only at the sentence level. Strategies like this can only serve as a temporary way to deter cheating. AI evolves all the time. Relying on this temporary fix requires constantly checking one's questions against the new version of the software.

9.4.1.4 Component relationships in procedure knowledge

Similar to Gen AI's capabilities in declarative knowledge, its ability to represent procedure knowledge is also quite well. However, showing the correct procedure only demonstrates the tool's ability to correctly order the pieces of knowledge but it does not necessarily represent its ability in the associations and order of these knowledge pieces.

<u>Example 4a</u>: Show me the steps to perform a tire change.
ChatGPT 4.0:
Here are the basic steps to change a tire on your vehicle:

1 Find a safe location: Look for a level and sturdy surface to change your tire on, such as a paved shoulder or a flat parking lot. Turn on your hazard lights to alert other drivers of your presence.
2 Gather your tools: You will need a jack, a lug wrench, and a spare tire. Make sure you have these tools in your vehicle before you have a flat tire.
…
5 Remove the flat tire: Remove the lug nuts and take off the flat tire. Place it next to the vehicle.
…

<u>Example 4b</u>: Can I do step 5 before step 1?
ChatGPT 3.5: No, it's important to follow the steps in the proper order. …

<u>Example 4c:</u> Can I use tree branches in place of a jack?
ChatGPT 3.5: Using tree branches or other improvised objects in place of a proper jack is not recommended and can be very dangerous. …

Example 4d: Can I just lift the car without using a jack?
ChatGPT 3.5: Lifting a car without using a jack is not safe and is not recommended. …

Example 4e: But, I am a superman. I have the power to lift the car without a jack.
ChatGPT 3.5: While I understand that you may feel confident in your abilities, lifting a car without a jack is still not safe and is not recommended, even for a "superman." …

As one can see, even ChatGPT 3.5 has the capability to not just correctly show procedural knowledge but also understand the order of the steps and the implications of changes to the recommended steps.

9.4.2 Assessments through mixed knowledge

Tweaking question structures only works so far. Since the underlying Gen AI algorithms are evolving, there is no guarantee that what does not work today will continue to be an issue for Gen AI in the future. This is exemplified in the "two truths and one lie" example in the previous section, where the older version of ChatGPT considered the correctness of a statement by going with the majority of individual pieces of information. The later version of ChatGPT was able to consider the correctness of all pieces of information. What this section presents are the approaches that tap into implicit and tacit knowledge. As we recall, the literature shows two categories of implicit knowledge: tacit knowledge versus knowledge that is not explicitly expressed or reported. The former refers to difficulties in expressing the knowledge or unconsciously using the knowledge, while the latter refers to knowledge that is not expressed. Both will pose difficulties for Gen AI.

9.4.2.1 Expressions containing implicit knowledge

Example 5: It is not possible to upload RMP files to our e-learning systems.
ChatGPT 4.0: It sounds like the e-learning system you are using might not support RMP file formats. RMP files are usually associated with project management or specialized software, so they might not be recognized by standard e-learning platforms.

Gen AI's response is incorrect. Tacit knowledge appears in this part of the sentence: "our e-learning systems," which belongs to the second category of implicit knowledge. In this specific case, Gen AI has no knowledge about the e-learning system being in use. Not only does it know nothing about the underlying tacit knowledge, but it also incorrectly associates RMP files with project management software.

9.4.2.2 Drawing an underlying inference from known facts

The following example shows a way to ask Gen AI to draw an inference from known facts.

Example 6: Does XXX University always charge late fees for adding a course about one week after classes start?
ChatGPT 4.0: Western Michigan University does not always charge late fees specifically for adding a course about one week after classes start …

XXX in the above example refers to the author's university. The answer provided by Gen AI is not entirely true. The university charges a late add fee approximately one week after classes begin. This is a known fact, but it is not explicitly published. Although one will be able to learn about this fact by analyzing calendars from multiple semesters, it is this implicit knowledge that tricks Gen AI.

9.4.2.3 Lack of specifics from tacit knowledge

As shown in the previous sections, tacit knowledge relies on one's own experience and may be specific to a context.

Example 7: When is the best time to switch from lengthy lectures to visual examples in classroom teaching?
ChatGPT 4.0: The best time to switch from lengthy lectures to visual examples in classroom teaching is when students start to show signs of disengagement, confusion, or information overload. Here are key moments when a shift to visual examples can be most effective … [Generic recommendations such as "After covering complex or abstract concepts," "when students' attention starts to wane," "to illustrate real-world applications," "During interviews or summaries" and "When addressing diverse learning styles" are given.]

Here in this following example, we see that Gen AI could fall short on the specifics and contextual possibilities. For example, visual learners tend to understand abstract concepts better than auditory learners. Not only do Gen AI tools fall short on the specifics, but Gen AI tools do not act the same way. The following is the response from Google Gemini, which shows a similar result, but it does provide certain details, such as "to break up long lectures" and "to accommodate different learning styles."

Google Gemini: The optimal time to switch from lengthy lectures to visual examples in classroom teaching can vary depending on the subject matter, student engagement, and individual learning styles. However, here are some general guidelines:

- **When student engagement starts to wane** …
- **When a concept is complex or abstract:** …

- **To reinforce key points:** …
- **To break up long lectures:** …
- **To accommodate different learning styles:** …

9.5 Conclusion

Detection approach to identify plagiarism or efforts to fool learning assessments may easily fall short. This is because those tools are limited by their algorithms and also by the kind of data used for detection. It is generally difficult to tell if some data (e.g., such as short statements) are from humans or AI. Even with longer statements or essays, detection tools use the assumption that the data fed into them are either from humans or from AI. Collaboration between the two, between multiple Gen AI tools, and between multiple people could evade setting off the alarm. This is especially worse when the assessment formats chosen lean toward short answers (e.g., true and false, multiple choice, matching, and short essays).

In this chapter, two approaches are proposed that rely on the type of knowledge being assessed. The first approach relies on tweaking the structure of assessment questions so that Gen AI has trouble answering them correctly. Despite some success, the underlying assumption of this approach is that what does not work today for Gen AI will continue to do so in the future. Unfortunately, this is not a valid assumption since technology advances constantly. Therefore, this approach is only a temporary solution, whose viability relies on something outside of one's control (i.e., whether the technology has evolved to a point that stops the strategy from working). Because of this dependency, it also requires one to regularly check their approach against new versions of the software.

The second approach proposed in this chapter relies on the types of knowledge being assessed. More specifically, it relies on two categories of implicit knowledge: one that is difficult to express (i.e., tacit knowledge) and the other one that is simply not reported or made available. The former poses two issues to most Gen AI tools (i.e., data unavailability and data quality issues), while the latter causes most Gen AI to either hallucinate or present an answer generically with little contextual relevance. Examples are provided to illustrate how these approaches work, but they are not an exhaustive list of possibilities. Other possibilities could be tweaking assessment formats into something situation or context-aware (e.g., free-form discussions or debates) and referencing an event, data point, or data format (e.g., images and videos) that are either difficult to obtain or unable for Gen AI to analyze.

It is worth noting that it is not recommended to always hold a static view of technology. Whatever assessment techniques proposed could still fail when new technology or capabilities are invented. For example, when an individual is aware of using tacit knowledge, willing to formulate it in some form of expression, and share such knowledge, tacit knowledge then becomes explicit knowledge (Kucharska & Erickson, 2023). Although this form of transformation of tacit knowledge allows Gen AI to analyze, it could provide only a limited number of data points viable for the specific situation or context.

Gen AI is a double-edged sword. It helps and could also hurt. Gen AI has the potential to disrupt the existing norms or to empower the underrepresented population. It will be interesting to see how AI collaborates as a partner, team member, or personal coach. For example, will gender dyads affect team learning (Chen & Rea, 2018) the same way in the age of Gen AI, or is it time to start thinking about triads with the new addition being the AI itself? Similarly, do the conditions that enable technology trust work the same way for Gen AI? Does Gen AI produce the same effects compared to other types of technology? This is the reason that Chen and Paliszkiewica (2024) cautioned that the double-edge nature of Gen AI could come in the form of task or even job replacement if one does not plan well ahead.

References

Chen, R. (2023). New AI classifier for indicating AI-written text, Retrieved from https://openai.com/index/new-ai-classifier-for-indicating-ai-written-text/.

Chen, K., & Paliszkiewicz, J. (2024). Swift trust in the age of artificial intelligence. In J. Paliszkiewicz, K. Chen, & J. Gołuchowski (Eds.), *Privacy, trust and social media* (pp. 98–107). Routledge.

Chen, K., & Rea, A. (2018). Do pair programming approaches transcend coding? Measuring agile attitudes in diverse information systems courses. *Journal of Information Systems Education*, *29*(2), 53–64.

De Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist*, *31*(2), 105–113. https://doi.org/10.1207/s15326985ep3102_2.

Gorman, M. E. (2002). Types of knowledge and their roles in technology transfer. *Journal of Technology Transfer*, *27*, 219–231.

Hung, S.-Y., Tsai, J., Chen, K., Chen, C., & Yeh, T.-T. (2024). Tacit knowledge sharing in information systems development projects: Social interdependence and regulatory focus perspectives. *Information Technology & People*, *37*(4), 1449–1477. https://doi.org/10.1108/ITP-08-2022-0587.

Kucharska, W., & Erickson, G. S. (2023). Tacit knowledge acquisition & sharing, and its influence on innovations: A Polish/US cross-country study. *International Journal of Information Management*, *71*, 102647. https://doi.org/10.1016/j.ijinfomgt.2023.102647.

Roberts, P. (1998). Implicit knowledge and connectionism what is the connection? In *Implicit and explicit mental processes* (1st ed., pp. 118–132). Taylor & Francis.

Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics*, *38*(5), 1229–1261. https://doi.org/10.1017/S014271641700011X.

Ten Berge, T., & Van Hezewijk, R. (1999). Procedural and declarative knowledge: An evolutionary perspective. *Theory & Psychology*, *9*(5), 605–624. https://doi.org/10.1177/0959354399095002.

Toom, A. (2012). Considering the artistry and epistemology of tacit knowledge and knowing. *Educational Theory*, *62*(6), 621–640. https://doi.org/10.1111/edth.12001.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. *Profeedings of Advances in Neural Information Processing Systems*. Advances in neural information processing systems. https://arxiv.org/pdf/1706.03762.pdf.

West-Smith, P. (2022). Does turnitin detect plagiarism? Retrieved from https://www.turnitin.com/blog/does-turnitin-detect-plagiarism.

10 Opening the Black Box

Achieving Trust through Transparency and Explainability in Generative AI

Casey Phillips

10.1 Introduction

As we stand on the verge of a new era in artificial intelligence (AI), one dominated by generative AI systems capable of producing human-like text, images, and even code, a critical question looms large: How can we trust systems we don't fully understand? This chapter delves deep into the crucial concepts of transparency and explainability in generative AI systems, exploring why these principles are not just technical considerations, but fundamental requirements for building trust in the AI era.

The rapid advancement of AI technologies, exemplified by systems like ChatGPT and DALL-E, has brought unprecedented capabilities to our fingertips. These systems can engage in human-like conversations, create art, write essays, and even assist in complex problem-solving tasks. However, with this, power comes a responsibility to ensure that these systems are not black boxes, making decisions that impact human lives without any accountability or understanding.

Our journey through this chapter will take us across the landscape of AI transparency, examining its multifaceted role in mitigating bias, promoting consumer confidence, and ensuring ethical AI development. We'll look at real-world examples of how companies are implementing transparency in their AI systems, from music recommendations to e-commerce platforms, and even in critical areas like healthcare and finance.

Along the way, we'll consider the impact of recent policy initiatives, such as the Biden-Harris Administration's Executive Order on AI, and their implications for the future of AI development. We'll dissect the challenges faced by developers, policymakers, and users in achieving meaningful transparency in increasingly complex AI systems.

Importantly, we'll dive into practical approaches to implementing transparency, examining concepts like Explainable AI (XAI) and discussing how they can be applied in real-world scenarios. Through case studies and expert insights, we'll illuminate both the successes and the pitfalls in current efforts to make AI more transparent and accountable.

DOI: 10.4324/9781003586944-12

As we look to the future, we'll explore emerging technologies and methodologies that promise to enhance AI transparency, from advanced visualization techniques to novel algorithmic approaches. We'll also consider the broader implications of AI transparency for society, discussing how it intersects with issues of privacy, innovation, and human-AI collaboration.

By the end of this chapter, you'll have a comprehensive understanding of why opening the "black box" of AI is essential, and how transparency and explainability can lead us toward a future where AI is not just powerful, but also trustworthy and accountable. Whether you're a developer, policymaker, business leader, or simply someone interested in the future of technology, this chapter will equip you with the knowledge and insights needed to navigate the complex landscape of AI transparency in the generative era.

10.2 The Recent Evolution and Impact of AI

10.2.1 The Generative AI Revolution

The latest phase in AI's evolution, and the one most relevant to our discussion of transparency, is the rise of generative AI. This began in earnest with the development of Generative Adversarial Networks (GANs) by Ian Goodfellow and his colleagues in 2014 (Goodfellow et al., 2014). GANs enabled AI systems to generate new, original content, from images to music.

However, it was the development of large language models (LLMs) that truly ushered in the era of generative AI as we know it today. The release of GPT-3 by OpenAI in 2020 marked a significant milestone, demonstrating unprecedented capabilities in natural language understanding and generation (Brown et al., 2020).

The subsequent release of ChatGPT in late 2022 brought these capabilities to the mainstream, allowing millions of users to interact directly with an LLM. As reported by OpenAI (2022), ChatGPT demonstrated the ability to engage in human-like conversations on a wide range of topics, create various forms of content, and even assist with complex tasks like coding and analysis.

The impact of these systems has been profound. They have opened up new possibilities for automation, creativity, and problem-solving across various industries. For instance, GitHub's Copilot, based on OpenAI's Codex model, can generate entire functions of code from natural language descriptions, potentially revolutionizing software development (GitHub, 2021).

However, this rapid advancement has also raised significant concerns. Questions about the authenticity of AI-generated content, the potential for these systems to perpetuate or amplify biases, and the implications for jobs and society at large have come to the forefront of public discourse.

10.2.2 The Democratization of AI

One of the most significant impacts of recent developments in AI, particularly generative AI, has been its democratization. As Jason Stanley, head of insight at Local

Logic, points out in a recent CMSWire article, "ChatGPT's easy-to-use interface and integration capabilities have led to the democratization of generative AI, enabling even small businesses to implement sophisticated AI solutions, thereby causing disruption" (Clark, 2023).

This democratization has accelerated the integration of AI into various aspects of our lives. AI-powered tools are now accessible to individuals and small businesses, not just large corporations with substantial R&D budgets. This has led to a proliferation of AI applications across industries, from customer service chatbots to personalized content recommendations, AI-assisted writing tools, and much more.

For example, platforms like Jasper and Copy.ai use generative AI to help content creators and marketers generate written content, from blog posts to social media updates. In the visual arts, tools like Midjourney and DALL-E allow users to generate images from text descriptions, opening up new possibilities for designers and artists.

However, this democratization has also amplified concerns about privacy, data security, and the ethical use of AI. As AI becomes more pervasive and accessible, questions about how these systems make decisions, what data they use, and how they can be held accountable have become increasingly pressing.

Moreover, the ease with which convincing text, images, and even videos can be generated by AI has raised concerns about misinformation and the potential for misuse. The ability to create "deepfakes" – highly realistic but fake videos or audio recordings – has particularly alarmed many observers.

This context of rapid advancement, widespread adoption, and growing concerns sets the stage for our discussion of transparency in AI systems. As these technologies become more powerful and more integrated into our daily lives, the need for understanding how they work, what they're capable of, and what limitations they have becomes increasingly critical.

In the following sections, we'll explore why transparency is so crucial in the age of generative AI, how it can be implemented in practice, and what challenges and opportunities lie ahead as we work toward creating AI systems that are not just powerful, but also trustworthy and accountable.

10.3 The Importance of Transparency in AI Systems

10.3.1 Understanding AI Transparency

Before delving into why transparency is crucial in AI systems, it's important to define what we mean by "transparency" in this context. AI transparency refers to the degree to which the decisions or outputs of an AI system can be understood, interpreted, or explained by humans. This includes understanding the data used to train the system, the algorithms employed, the decision-making process, and the limitations and potential biases of the system.

Transparency in AI is not just about making the code open-source or providing technical documentation. It's about making AI systems understandable and interpretable to various stakeholders, including developers, users, policymakers, and the general public. This often involves providing explanations in language and formats that non-experts can understand.

10.3.2 Alleviating Bias and Preventing Harm

One of the most critical reasons for transparency in AI systems is to identify and mitigate bias, thereby preventing potential harm. AI systems, despite their impressive capabilities, are not immune to bias. In fact, they can sometimes amplify existing societal biases, leading to decisions that can significantly impact people's lives.

Consider the following examples:

1. **COMPAS Recidivism Algorithm**: In 2016, an investigation by ProPublica found that the COMPAS algorithm, used in the US criminal justice system to predict the likelihood of a defendant becoming a recidivist, was biased against Black defendants. The algorithm was more likely to falsely label Black defendants as future criminals, at almost twice the rate as white defendants (Angwin et al., 2016).
2. **Amazon's AI Recruiting Tool**: In 2018, Amazon scrapped an AI recruiting tool that showed bias against women. The system, trained on resumes submitted to the company over a 10-year period, had learned to prefer male candidates because most resumes came from men, a reflection of male dominance in the tech industry (Dastin, 2018).
3. **Gender Bias in Language Models**: Research has shown that LLMs like GPT-3 can perpetuate gender stereotypes. For instance, these models are more likely to associate certain professions (like "doctor" or "engineer") with male pronouns and others (like "nurse" or "teacher") with female pronouns (Bender et al., 2021).

These examples highlight the potential for AI systems to perpetuate and even amplify societal biases if not carefully designed and monitored. Transparency is crucial in identifying these biases. By making AI systems more transparent, we can shine a light on potential biases and work toward eliminating them.

The Biden-Harris Administration recognized this critical need when they issued an Executive Order on October 30, 2023, directing the Department of Homeland Security to lead the responsible development of AI. This order emphasizes the importance of transparency in AI systems, particularly in identifying and mitigating bias (US Department of Homeland Security, 2023).

In practice, transparency in AI systems might involve:

- Providing clear explanations of the factors considered in decision-making processes
- Allowing for audits of training data and algorithms

- Implementing ongoing monitoring and testing for bias
- Establishing clear processes for addressing identified biases

By exposing the inner workings of AI systems, transparency allows us to identify if a system is unfairly weighing certain factors or producing biased outcomes. This awareness is the first step in correcting these issues and ensuring fairer outcomes for all.

10.3.3 Fostering Consumer Trust

Beyond mitigating bias, transparency in AI use is crucial for building and maintaining consumer trust. As AI becomes more pervasive in consumer-facing applications, users are increasingly concerned about how their data is being used and how decisions affecting them are being made.

Ricky Spears, founder and CMO of RickySpears.com, explains in a CMSWire article that customers see non-transparent AI systems as "black boxes" making choices without clear reasons. Spears notes, "This lack of transparency leads to a lack of trust in brands, which can have negative consequences. People are concerned about privacy and are curious about why their likes and dislikes are tracked" (Clark, 2023).

This lack of trust can have significant implications for businesses. A 2023 study by Cognizant found that only one-third of consumers trust generative AI, and 73% of respondents believe the economic gains of AI will primarily boost corporate profits, while just 23% think they'll personally benefit from these advances (Cognizant, 2023).

Transparency can help address these concerns. By providing clear explanations of how AI systems make decisions based on user data and behavior, brands can build trust and encourage responsible use of their AI-powered services. This transparency also empowers users, giving them a sense of control and understanding over the AI systems they interact with.

Several companies have taken steps to be more transparent about their use of AI:

1 **Spotify**: The music streaming platform provides explanations for its song recommendations, often stating something like "We recommended this because you listened to [Artist X]." This simple explanation helps users understand the logic behind the AI's decisions.
2 **Netflix**: The streaming giant not only provides reasons for its recommendations but also allows users to remove titles from their viewing history, which affects future recommendations. This gives users a degree of control over the AI system and its outputs.
3 **Google**: In its AI Principles, Google commits to making AI systems accountable to people, stating that they will "design AI systems that provide appropriate opportunities for feedback, relevant explanations, and appeal" (Google, 2018).

By being transparent about their use of AI, these companies are working to build trust with their users, showing that AI is not a mysterious force making opaque decisions, but a tool that can be understood and, to some extent, controlled by users.

10.3.4 Enabling Informed Decision-Making

Transparency in AI systems is not just about building trust; it's also about enabling informed decision-making. When users understand how an AI system works and what factors it considers, they can make more informed choices about whether and how to use that system.

For instance, in the context of AI-powered financial advice, transparency allows users to understand the basis of the advice they're receiving. They can then decide whether they agree with the system's reasoning and whether to follow its recommendations.

Similarly, in healthcare, where AI is increasingly being used for diagnostics and treatment recommendations, transparency is crucial. Doctors and patients need to understand the basis of an AI's diagnosis or recommendation to make informed decisions about treatment plans.

It is also becoming increasingly common to see generative AI-powered question and answer tools, such as Perplexity.AI and ChatGPT, transparently display the internet sources they use to generate answers, allowing users to see exactly where the information comes from.

Transparency also enables users to provide more meaningful feedback on AI systems. When users understand how a system works, they can offer more targeted and useful feedback, which in turn can help improve the system.

10.3.5 Facilitating Regulatory Compliance and Ethical Development

As AI becomes more prevalent, governments around the world are developing regulations to ensure its responsible use. Transparency is often a key requirement in these regulations.

For example, the European Union's proposed AI Act includes requirements for transparency, particularly for high-risk AI systems. The Act requires that high-risk AI systems be designed and developed in a way that ensures their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately (European Commission, 2021).

In the United States, the Biden-Harris Administration's Executive Order on AI emphasizes the importance of transparency in AI development and deployment. It calls for guidelines to ensure that AI systems used by federal agencies are transparent and accountable (U.S. Department of Homeland Security, 2023).

Transparency is not just about compliance; however, it's also crucial for the ethical development of AI. By making AI systems more transparent, developers can better identify and address ethical concerns throughout the development process. This can help ensure that AI systems are aligned with human values and societal needs.

10.4 Implementing Transparency in AI Systems

10.4.1 XAI

One of the key approaches to implementing transparency in AI systems is through XAI. XAI refers to methods and techniques in the application of AI such that the results of the solution can be understood by humans. It contrasts with the concept of the “black box” in machine learning where even their designers cannot explain why the AI arrived at a specific decision.

XAI is particularly crucial in the context of deep learning and other complex AI models, where the decision-making process can be opaque even to the system’s creators. Several techniques have been developed to make these systems more explainable:

1. **LIME (Local Interpretable Model-agnostic Explanations)**: This technique explains the predictions of any classifier in an interpretable and faithful manner by learning an interpretable model locally around the prediction (Ribeiro et al., 2016).
2. **SHAP (SHapley Additive exPlanations)**: SHAP uses game theory to assign each feature an importance value for a particular prediction (Lundberg & Lee, 2017).
3. **Attention Mechanisms**: In neural networks, especially those used for natural language processing, attention mechanisms can highlight which parts of the input the model is focusing on when making a decision (Vaswani et al., 2017).
4. **Counterfactual Explanations**: These explanations show how the model’s output would change if the input were slightly different, helping users understand what factors are most important in the model’s decision-making (Wachter et al., 2017).

10.4.2 Practical Approaches to Transparency

Beyond technical solutions like XAI, there are several practical approaches that organizations can take to increase the transparency of their AI systems:

1. **Clear Communication**: Organizations should clearly communicate when and how they are using AI. This includes informing users when they are interacting with an AI system, such as a chatbot.
2. **Explainable Outputs**: AI systems should provide explanations for their outputs in user-friendly terms. For example, a loan approval AI might explain which factors most influenced its decision.

3 **Data Transparency**: Organizations should be clear about what data they are collecting and how it's being used to train and operate AI systems.
4 **Model Cards**: Proposed by Google researchers, model cards are short documents accompanying trained machine learning models that provide benchmarked evaluation in a variety of conditions and are intended to be publicly available (Mitchell et al., 2019).
5 **AI Ethics Boards**: Many companies have established AI ethics boards to oversee the development and deployment of AI systems, ensuring they align with ethical principles including transparency.
6 **Open-Source Initiatives**: Some organizations are making their AI models open source, allowing for public scrutiny and improvement. For example, Meta has released its LLM, LLaMA, to the research community (Meta AI, 2023).

10.4.3 Case Studies in AI Transparency

Several companies have taken significant steps toward making their AI systems more transparent:

1 **IBM Watson for Oncology**: IBM has worked to make its Watson for Oncology system more transparent by providing explanations for its treatment recommendations. The system provides links to medical literature that support its suggestions, allowing doctors to understand the basis of the AI's recommendations (Chen et al., 2016).
2 **Audi"s AI:Trail quattro concept**: This concept car uses AI for autonomous driving and clearly displays what the AI "sees" and how it's interpreting its environment, helping passengers understand and trust the system's decisions (Audi, 2019).
3 **LinkedIn's AI-driven job matching**: LinkedIn provides explanations for why it recommends certain jobs to users, typically based on their skills, experience, and career interests. This transparency helps users understand and trust the recommendations (LinkedIn, 2021).
4 **COMPAS Risk Assessment Tool**: Following criticism of bias, the creators of COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) made efforts to increase transparency by releasing detailed documentation about how the tool works and the factors it considers (Equivant, 2018).

These case studies demonstrate that transparency is not just a theoretical concept, but a practical approach that companies are increasingly adopting to build trust and improve their AI systems.

10.4.4 Challenges in Achieving AI Transparency

While the importance of transparency in AI is clear, achieving it in practice can be challenging. Several obstacles stand in the way:

1 **Technical Complexity:** Many modern AI systems, particularly deep learning models, are inherently complex. The decision-making process in a neural

network with millions or billions of parameters is not easily reducible to simple, human-understandable rules. This "black box" nature of advanced AI systems makes transparency challenging.

2 **Trade-off with Performance:** There can be a trade-off between model performance and explainability. Often, the most accurate models are also the most complex and least interpretable. Simpler, more interpretable models might not achieve the same level of performance in complex tasks.
3 **Intellectual Property Concerns:** Companies may be reluctant to provide full transparency about their AI systems due to intellectual property concerns. Revealing too much about how a system works could potentially allow competitors to replicate proprietary technology.
4 **Security Risks:** Full transparency about an AI system's workings could potentially make it more vulnerable to attacks or manipulation. There's a delicate balance to strike between providing enough information for accountability and not exposing vulnerabilities.
5 **User Understanding:** Even when explanations are provided, they may not always be easily understood by users. There's a challenge in translating complex technical concepts into language that is accessible to a general audience without oversimplifying to the point of inaccuracy.
6 **Dynamically Changing Systems:** Many AI systems, especially those that engage in online learning, are constantly evolving based on new data. This dynamic nature can make it challenging to provide consistent, up-to-date explanations of how the system is making decisions at any given moment.

10.4.5 Future Directions in AI Transparency

Despite these challenges, there are several promising directions for improving AI transparency in the future:

1 **Advances in XAI Techniques:** Research into new XAI techniques continues to advance. Future developments may provide better ways to explain complex AI systems without sacrificing performance.
2 **Standardization Efforts:** There are ongoing efforts to develop standards for AI transparency. For example, the Institute of Electrical and Electronics Engineers has a standard for Transparency of Autonomous Systems (IEEE 7001-2021) that provides a framework for measuring and ensuring transparency in autonomous systems (IEEE, 2021).
3 **Regulatory Developments:** As AI becomes more prevalent, we can expect more detailed regulations around AI transparency. These regulations may provide clearer guidelines and requirements for companies developing and deploying AI systems.
4 **Education and AI Literacy:** Efforts to improve public understanding of AI may help bridge the gap between technical explanations and user comprehension. Increased AI literacy could enable more meaningful transparency.
5 **Human-AI Collaboration:** Future AI systems may be designed with transparency and human collaboration in mind from the start. This could lead to

systems that are inherently more interpretable and align better with human decision-making processes.

10.5 Conclusion

Transparency in AI systems, particularly in the era of generative AI, is not just a technical challenge but a societal imperative. As AI continues to play an increasingly significant role in our lives, understanding how these systems work, what they're capable of, and what limitations they have becomes crucial.

Transparency serves multiple purposes: it helps in identifying and mitigating bias, fosters trust between AI systems and their users, enables informed decision-making, and facilitates regulatory compliance and ethical development. While achieving transparency in complex AI systems presents significant challenges, ongoing research, regulatory efforts, and industry initiatives are paving the way for more transparent and accountable AI.

As we move forward, it's clear that transparency will be a key factor in determining the public's acceptance and trust of AI systems. By prioritizing transparency, we can work toward a future where AI is not just a powerful tool, but one that is understood, trusted, and aligned with human values and societal needs.

The path to truly transparent AI is still being forged, but the importance of this journey cannot be overstated. It is through transparency that we can ensure AI remains a technology that augments and empowers humanity, rather than one that confuses or controls. As we continue to push the boundaries of what's possible with AI, let us also push for greater understanding, accountability, and trust in these powerful systems.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks.*ProPublica*.https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Audi. (2019). (2019, September 10). Audi AI quattro – The off-roader of the future. *Audi MediaCenter*. https://www.audi-mediacenter.com/en.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.

Chen, Y., Elenee Argentinis, J., & Weber, G. (2016). IBM Watson: How cognitive computing can be applied to big data challenges in life sciences research. *Clinical Therapeutics*, *38*(4), 688–701. https://doi.org/10.1016/j.clinthera.2015.12.001

Clark, S. (2023, July 10). Why transparency is vital when brands use AI. *CMSWire*. https://www.cmswire.com/customer-experience/why-transparency-is-vital-when-brands-use-ai/.

Cognizant (2023, June 18). Turning potential to profit: Building consumer trust in AI. https://www.cognizant.com/us/en/insights/insights-blog/building-consumer-trust-in-ai-wf2729750

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

Equivant. (2018). *Practitioner's guide to COMPAS Core*. Northpointe Inc.

European Commission. (2021). Proposal for a regulation laying down harmonised rules on artificial intelligence. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

GitHub. (2021). *GitHub Copilot*. GitHub. https://github.com/features/copilot.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., …& Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*, 2672–2680.

Google. (2018). Artificial intelligence at Google: Our principles. Google AI. https://ai.google/principles/.

IEEE. (2021). IEEE 7001-2021- IEEE standard for transparency of autonomous systems. *IEEE Standards Association*. https://standards.ieee.org/standard/7001-2021.html.

LinkedIn. (2021). How LinkedIn's AI-powered job matching helps recruiters find the most relevant candidates. *LinkedIn Engineering*. https://engineering.linkedin.com/blog/2021/how-ai-powers-linkedin-recruiter-search.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774.

Meta AI. (2023). LLaMA: Open and efficient foundation language models. *Meta AI Research*. https://ai.facebook.com/blog/large-language-model-llama-meta-ai/.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., …& Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparenc*y, 220–229. https://doi.org/10.1145/3287560.3287596.

Open AI. (2022). ChatGPT: Optimizing language models for dialogue. *OpenAI Blog*. https://openai.com/blog/chatgpt/.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Minin*g, 1135–1144. https://doi.org/10.1145/2939672.2939778.

U.S. Department of Homeland Security. (2023, October 30). Fact sheet: Biden-Harris administration executive order directs DHS to lead the responsible development of artificial intelligence. https://www.dhs.gov/archive/news/2023/10/30/fact-sheet-biden-harris-administration-executive-order-directs-dhs-lead-responsible.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*, 5998–6008.

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, *31*, 841.

11 What Will We Learn From Failures and Trust Artificial Intelligence Applications in Organizations?

Hakkı Okan Yeloğlu

11.1 Introduction

In recent years, the importance of artificial intelligence applications within organizations has been examined in the literature in the context of improving the decision-making process, optimizing supply chains, and increasing efficiency and productivity. Adapting and integrating artificial intelligence (AI) applications within organizations may pose some difficulties, especially in predicting market trends so that organizations can compete in industries. In addition, AI can improve employee performance using highly intelligent tools and complex systems. On the other hand, AI enables the explanation of relationships between the organization and the customer while differentiating work environments, encouraging innovation, and introducing new business models. From a theoretical perspective, AI applications focus on the opportunities, threats, and risks they provide to organizations.

This chapter acknowledges the strengths, weaknesses, opportunities, and threats of AI applications in organizations and questions their developments at every level. On the other hand, it also emphasizes the failures that inadequate AI applications may cause. Poor performance and unreliable results resulting from failures can cause resistance to change and manifest as threats and obstacles at all levels.

11.2 Literature Review

11.2.1 The Integration of AI Applications in Organizations

Integrating AI and AI applications into organizations brings significant challenges. However, this integration also helps organizations compete within sectors by increasing efficiency, effectiveness, and innovation. The literature discusses how AI applications in various sectors can increase organizational learning by facilitating data analysis, identifying relevant patterns, and making predictions. It also emphasizes that AI tools can help organizations learn

DOI: 10.4324/9781003586944-13

from their experiences and develop a more effective organizational culture and climate.

There is ongoing debate regarding competition and the vast amounts of data organizations invest in. Analyzing and processing this data can enhance organizational learning, benefiting organizations from the analysis's results. These insights can be integrated into daily operations, addressing needs at every level of the organization. This process may lead to the effective utilization and adaptation of AI, ultimately fostering a culture of continuous improvement and development to ensure organizational survival.

AI applications within organizations' workflows will decrease value, especially in the decision-making scope, the details of employees' roles, and other factors that will increase their time. One of the most important reasons for this is that AI programs are tried to be implemented without making the necessary effort to spread institutional knowledge and talent, and the risk of failure may increase as people benefit from technology effectively.

Adapting AI applications within the organization may take time. Organizational managers need to inform technical and administrative staff about what the applications include. Failure to perceive the advantages of practices that take a long time to adapt to the organization's employees may lead to failure.

How organizations compete in their industry is closely related to the products and services they produce. To meet customer needs, products and services offered to the sector in quality, in the required quantities, and at affordable prices require significant management skills. It is necessary to discuss how this management will be done, especially in adapting AI within organizations.

11.2.2 Organizational Learning and Trust

Operating in various organizational contexts, organizational learning is a continuous and dynamic process designed to align with organizational objectives. Scholars argue that organizational learning is an outcome of organizational innovation. Pilar et al., (2005) define organizational learning as "the activities organizations engage in to transform their learning capabilities, including individuals and competitors." Organizational innovation is crucial for organizations to facilitate and develop internal innovation, and organizational learning is essential for survival. It has been widely studied across diverse fields.

Fiol and Lyles (1985) define "organizational learning as improving actions through better knowledge and understanding." DiBella et al., (1996) and Parris (2000) explain that organizational learning can be understood as a capacity within organizations to maintain or improve performance based on experience. The authors argue that organizational learning is both a process and a capability for processing knowledge to enhance organizational performance and inform decision-making.

Organizations acquire, distribute, interpret, and store knowledge to facilitate change. These actions are organized to improve workforce capabilities with the help of employees. Gieskes et al., (2002) explain that interrupted learning processes, psychological and cultural barriers, and obstacles related to organizational structure and leadership are significant barriers to learning.

Jerez-Gomez et al., (2005) identify acquisition, transfer, and integration as the three dimensions of organizational learning. Meanwhile, Yang (2011) expands this by identifying, assimilating, and applying knowledge as the subprocesses of organizational learning. In this sense, organizational learning has a unique value in coping with environmental changes, ensuring organizational continuity, and achieving sustainable competitive advantage.

As Garvin (1993) argues, organizational learning is the process of modifying behaviors or improving existing ones by obtaining and analyzing information more effectively. Although there are many explanations for organizational learning, academics have no consensus regarding its definition. In the long term, organizational learning involves systems that retain and create knowledge and transform it into organizational subprocesses.

Organizational learning allows one to determine which technologies will provide job opportunities, which technologies will be developed, and which technologies will be retained, protected, or commercialized within the organization (AlSaied & Alkhoraif, 2024). The issues that technology management deals with are widely discussed in the literature.

Organizational learning supports the organization's industry structure by combining management strategy with technological capabilities and resources to harmonize its internal and external environment (Chen & Lin, 2023). When organizational learning continuously occurs in an organization, managers must develop different strategies for perceiving technological change. At this point, organizations that keep up with technological change can make their presence sustainable in the sector by following different strategies. At the same time, organizational learning also impacts the life cycle of organizations. When organizations effectively learn what is happening in the environment, their life cycles can be extended.

Organizations with a high learning capacity can more easily adapt to the changing conditions of the external environment. Organizations that incorporate knowledge into their operations in various ways can be more effective in producing new knowledge and transforming this knowledge into economic value.

The learning abilities of organizations are also closely related to the absorptive capacity of organizations. To increase the absorptive capacity of organizations, it is necessary first to increase the learning abilities of employees. In addition, it is essential to improve employees' problem-solving skills, select the right people for the job, increase the willingness of employees to learn and adopt technologies, increase the experience of employees, and protect and store the knowledge they have acquired. On the other hand, creating organizational memory, discovering external information, developing research, investing in research and development, and sharing information and technology are among the factors that will positively affect organizational learning.

When evaluating organizations' capabilities, human, technology management, organizational learning, and strategic perspectives can be examples. Every organization's talent management must be evaluated in terms of creating competitive advantage and sustainability.

When the organizational learning process is examined, it is seen to have three important features. The first is the positive or negative behavioral change created by learning. The second is that the changes occur because of experience or training. The third is that the change must continuously be defined as learning.

Organizational learning is contingent upon an organization's capacity to learn and its ability to cultivate and foster this skill. Moreover, an organization's cultural elements significantly influence its learning potential, including norms, values, rules, roles, and traditions. Fundamentally, organizational learning is a dynamic process. When analyzing its constituent components, we can categorize them into inputs, subprocesses, and outputs. The inputs driving this process include human capital, the organization's learning aptitude, tangible and intangible resources, financial resources, and experiential knowledge. Learning to learn and its subsequent implementation are considered subprocesses. The final stage involves disseminating relevant information or experiences from organizational learning to appropriate organizational channels, which constitutes the output.

At this juncture, it is imperative to briefly explore the concepts of organizational learning and learning organizations. While organizational learning encompasses the collective learning of individuals, groups, and the organization, a learning organization is a dynamic system that actively engages in and benefits from this process. A pivotal concern within this context is the organization's effective acquisition, dissemination, and management of information. According to Peng et al., (2023), organizational learning and knowledge management are interconnected, and when learning organizations effectively leverage both, they can achieve superior performance.

Organizational learning can yield multifaceted outcomes, including individual and group reactions to the learning process, acceptance behaviors, and adapting acquired knowledge. These responses can vary in complexity and speed. It is important to note that organizational learning is only sometimes a positive endeavor. Ambiguous goals, internal conflicts, bureaucratic regulations, and formal procedures can introduce uncertainty into the learning process. This can lead to prolonged learning cycles and hinder organizational management.

Organizations actively engaging with their internal and external environments can expedite organizational learning. When examining the literature, a critical issue in explaining the relationships between organizational learning and trust is how organizations choose and use technology to provide strategic advantage. Additionally, individuals' willingness to share information, experiences, and lessons learned within and among themselves contributes to a trusting organizational culture. Once this trust is established, it makes it easier for new practices, ideas, thoughts, or systems to be adopted within the organization. When organizational learning reflects the collective experiences gained by an organization, team-based learning, individual empowerment, effective leadership, coordination, supervision, and control emerge as the most critical elements in increasing organizational trust (Gustafsson et al., 2021).

At this point, individuals and groups must trust the organization they work for during the learning process. When individuals experience adverse outcomes during this process, their performance, creativity, and innovation levels may decline (Ferreira et al., 2020). Focusing solely on disseminating knowledge within the organization can be a narrow approach to organizational learning. As a result, organizations may need to adopt practices that help them adapt to internal and external environments. While these practices can sometimes be simple, when they become complex, the process of "unlearning" may be necessary, or the organization may attempt to enforce these practices. This can lead to resistance and a decline in trust due to individuals' negative experiences. When managers push for strict enforcement of these practices, it can also lead to issues with organizational trust, as employees may feel that their concerns and input are being ignored.

The role of managers in building organizational trust is crucial. This trust develops through mutual communication and interaction between both managers and employees. Organizational trust is a key factor in helping employees achieve their goals. Additionally, the roles, job descriptions, responsibilities, and areas of authority that individuals hold within the organization play an important part in fostering this trust. The trust relationship between managers and employees in disseminating information, implementing practices, or sharing ideas within the organization will also directly or indirectly influence the company's ability to be innovative and creative.

An important factor affecting organizational trust is the consistent participation of employees in business processes within the organization. Creating an impact that will increase employees' motivation, internal satisfaction, and sense of belonging by actively involving them in organizational activities emerges as a critical situation from an organizational perspective. When employees at all levels of the organization feel valued and influential, they are more likely to trust the leaders in the organization, contribute directly or indirectly to decision-making processes, and share their knowledge and expertise. As a result, a culture of collaboration, innovation, and learning is encouraged (Adomako & Nguyen, 2024). While every interaction of learning organizations with their environment is essential for organizational learning, individual trust also allows learning to occur more quickly. Employees who trust the organization are likelier to take risks, try new ideas, and learn from their mistakes. Creating a safe and supportive environment where employees can freely express their thoughts and opinions will also affect the organizational culture and create a harmonious and dynamic organization.

Organizational trust and learning influence the successful implementation and adoption of AI applications. Existing research defines the adoption of AI within the organization by factors such as technological, organizational, and environmental elements. In addition, the literature states that the role of organizational learning and trust in solving the complexity levels of AI applications is an important research area (Gkinko & Elbanna, 2023).

11.3 AI Failures and Organizational Learning

The review literature focuses on technology management within the organization. This is shown by the relationship between technology management failure and AI programs, both in terms of the extent to which AI results improve within organizations and the integration of diversity within the organization.

On the other hand, organizations that do not have strong learning capabilities may experience difficulties in troubleshooting or redesigning their AI systems to perform as expected. Organizations with more focused learning can use failures as growth and opportunities. Organizations that fail often experience technological abandonment. There needs to be a solution to identify the root causes of the failure of the operation and evaluate the identified obstacles.

Another critical factor is the adaptation of decision-making processes of AI applications. AI applications contain complex patterns that provide more precise technological and work process coverage. In learning in organizations, where difficulties emerge, and knowledge sharing between departments within organizations is weak, AI needs to have the ability to interpret and implement focused goals effectively (Webster & Martocchio, 1995). As a result, missing and misunderstood fundamentally flawed parts cause negative consequences within the organization. In this way, the failure that causes the development and failure of AI applications can be revealed more clearly.

Conversely, organizations that demonstrate continuous learning are better equipped to see that AI applications can be embraced. This organization creates an environment where employees can improve their technical and writing competencies, allowing them to develop AI recordings faster and better when encountering real situations. This adaptation and learning capacity help prevent errors in dynamic business environments.

11.3.1 The Role of Technology Management and AI Applications in Organizations

Adapting and implementing AI within the organization requires a practical management approach. At this point, the role of technology management is vital. Technology management is the effective and efficient management of the technology produced by an organization with its material or intangible resources (Lee et al., 2023). Assets such as human capital, financial capital, physical infrastructure, brand, and patents owned by the organization are important in planning, coordinating, supervising, and controlling the technology they produce (Schuh et al., 2021). When sub-components of technology management are examined, the definition, selection, acquisition, protection, or abandonment of technology within the organization occur as subprocesses. The management of technology, which will change depending on the organizational structure, is vital in determining the organization's road map, the technology readiness levels of its products, and how long the product life cycle will be (Aljawder & Al-Karaghouli, 2024; Solaimani & Swaak, 2023). The adaptation of technology within organizations will vary depending on the organization's sector, its relationships with customers and stakeholders, and the economy. Therefore, taking the organizational learning process and technology management will be the right approach.

Technology management is an organizational strategic approach. Organizations develop many activities simultaneously to plan, develop, supply, and protect the technologies they acquire in the long term. Lakshmi et al., (2023) emphasize that the fact that technology management has a multidisciplinary structure emerges as necessary for technology management activities to occur together. Technology management, which is nourished by many management branches such as information management, innovation management, R&D management, business management, and economic management, receives support from these management activities to ensure the organization's sustainability and competition within the sector (Kim & Seo, 2023).

Finally, the effectiveness of the projects carried out within the organization is an important factor in the success of AI applications, and technology management plays a central role here. As it is known, many AI applications start as small and pilot projects. However, problems such as poor technology management, lack of infrastructure, and financial inadequacy may need to be improved to integrate AI applications into existing systems. Apell and Eriksson (2023) argue that effective management of technological resources, infrastructure planning, and corporate strategies are essential for AI applications to develop and contribute to the organization long-term.

Planning, organization, supervision, and coordination of managerial activities emerge as successive processes for organizations to manage their limited resources and achieve their goals effectively and efficiently.

One of the most important stages of management activities in the planning phase is identifying problems or opportunities, determining business objectives, identifying and evaluating options for solving problems, selecting the most appropriate alternative, and preparing auxiliary plans.

In the organizing process, to carry out the planned activities, it is necessary to establish the organizational structure correctly, determine the relationships within the organization, make relevant job descriptions, and determine the right workforce.

In addition, the organization should ensure that employees fulfill the duties assigned to them appropriately (Allen, 2013). At this point, clear instructions to individuals, continuous educational activities, and a balance of discipline, reward, and punishment are necessary.

Comparing the goals and actual results within the organization and determining at which points the goals are not achieved is an important auditing issue. At this stage, the managers' determination of the standards within the organization, the current situation's determination, and studies to eliminate the differences detected through comparison contribute to the organizational learning process.

At the same time, technology management is closely related to managing the talents of individuals within the organization. While determining the strategic and operational capabilities of organizations, the willingness, motivation, reactions, and resistance of individuals and groups of individuals toward learning are a subject that is studied in detail in determining organizational resources. The concept of leadership, which is considered in an organizational sense with technology management, is of critical importance for the success of AI applications.

Organizational technologies are used in producing and marketing products and services. Product and service technologies depend on the products and services provided by the organization. These technologies focus on usability and effectiveness and support gaining competitive advantage, improving organizational and brand image, and business development.

Depending on organizational learning, the role of technology within the organization, definition of technology, matching of technologies with business and sector needs, and continuous development of projects within the organization can be shown as other technology management functions.

Ensuring the continuous development of technology for the products and services produced within the organization will add value. At this point, the role of technology management, how much investment should be made in technology, or how and with what methods existing technologies should be followed.

Organizations have multiple options for acquiring technology. One is developing products and services through in-house research and development activities, transferring technology from other organizations or sectors, or using existing technology.

Technologies add value to organizations that relevant employees in the organization should conduct resource and cost research, allowing other employees to learn the process for sustainability. With the trust that comes from organizational learning, the products and services produced can be continuously improved.

When AI applications intended to be integrated into the organization are considered a technology, they must be defined as a resource. When this resource adds economic value to the organization, the higher the value, the higher the competitive advantage it will provide.

11.3.2 Technology Management and Failures of AI Applications in Organizations

The relationship between technology management and the failure of AI applications in organizations is crucial, as effective technology management determines how well AI systems are integrated, maintained, and optimized. One of the most important reasons for failures in AI applications in organizations is the compatibility between the capabilities of AI and the technologies offered. An organization needs the information technology systems, relevant management applications, or computing resources necessary to support AI storage to be used and managed effectively. For this reason, technology management plays a vital role in maintaining the basic systems required to develop AI applications in organizations.

Another important issue is the management of the life cycle processes of AI applications. To adapt to changing environments, AI applications require constant monitoring, updating, and improvement. Technology management that is considered inadequate causes failures. The most important reason for this is that organizations must establish robust systems to protect and develop models of AI applications. In addition, when the model is updated, the quality of the data, the evaluation of this data, and the review of the algorithms are systematically inspected within the organization, these AI applications can quickly become outdated and fail to produce the expected results.

When the literature on AI is examined, it is seen that poor data management within organizations is an essential reason for the failures of AI applications. This is also directly related to how well the technology in organizations is managed. As Reddy and Dyaram (2014) state, AI systems may be built on biased, incomplete, or outdated data if data management is not prioritized within the organization. Such a situation may lead to erroneous predictions and inaccurate predictions. Effective technology management ensures that data is up-to-date and secure, thus providing accurate and timely information that AI systems need to perform well.

Monitoring sectoral technological products and services, predicting which technologies to invest in, determining which technology suits customers' needs, and predicting which technology will give the organization a competitive advantage emerge as benefits of technology management for organizational learning.

It is argued that using AI applications in organizations has complex systems. According to the literature, these applications are generally defined as expert systems. These systems can provide consultancy support on specific specialist topics. At the same time, they are used within the organization as programs that solve problems in a specific field using expert knowledge.

At the same time, expert systems also aim to suggest solutions using AI techniques. The information transferred to the computer regarding the design of expert systems by people with specific qualifications provides different services.

The failure of AI applications in organizations can be grouped under many reasons. Failure to fulfill the production purpose and cleanliness of the technical part within the organization, or communication problems between technical personnel and administrative personnel, may prevent AI programs from failing. However, different perspectives between those used within the organization and those working in the business world may cause the AI services in organizations to deteriorate.

Infrastructure problems may prevent the integration of AI applications into the organization. If the necessary infrastructure is not provided for the applications to become operational, the organization's employees may not use the applications' capacity sufficiently. Underused capacity will increase resistance to AI applications and make it difficult to use them. Failure to effectively use AI applications whose content and scope are not understood within the organization will be reflected as failure in the future.

Organizations in the information and communication sector usually have a data science team. AI applications can be developed within this team group and emerge over long periods. However, as time passes, the need for more impact on AI applications, whether monitored or supported, will cause a communication gap within the well-established organization. This situation is likely to fail. One of the reasons for this lack of communication is that organizational managers need more technical detail and knowledge on subjects such as data science and data analytics. Even if the AI applications created because of mutual disagreements are highly efficient and effective, the use of these applications by organizational managers within the organization or marketing them outside the organization will still be one of the reasons for organizational failure.

11.4 Conclusion

One of the key aspects of AI applications that are considered reliable is the need for multidisciplinary collaboration and awareness among stakeholders. Successful AI applications address technical, ethical, and managerial issues holistically. Non-technical factors, including standardization and management processes, also increase trust in AI. These challenges can create vulnerabilities for different stakeholders, underscoring the need for a multi-stakeholder approach to address the complexities of trust in AI applications.

Determining the inputs, processes, and outputs that affect the organizational learning process is also essential in controlling the applications that will be adapted to or used within the organization. The harmony of two critical issues, the kind of technology the products and services have and the business technology, will directly or indirectly affect the organizational learning process. Learning the technologies and applications used within the organization effectively and efficiently has essential effects on entering products or services into the market. The first of these is continuity in quality, and the second is the expectation of low cost.

Effective organizational integration of knowledge, practices, innovations, or approaches requires strong communication and interaction among employees and between employees and managers. These interpersonal dynamics are crucial for fostering trust within teams. Building trust among individuals can increase satisfaction, performance, motivation, and efficiency, ultimately improving relationships with management.

Effective and efficient organizational learning is seen as necessary for determining technology needs by analyzing sectoral needs, conducting competitor analysis, and determining what business technologies will be used in determining the development, production, and support stages. At this point, auditing the practices to be adapted into the organization with a good feasibility study will increase efficiency.

Technology managers in the organization should cooperate with all levels to ensure that AI systems are compatible with organizational goals. Failure to develop and implement AI applications in alignment with strategic goals may result in investments that will not create value for the organization. This will lead to organizational failure. It is only possible to achieve the organization's vision for adopting and implementing AI applications within the organization through technology management. It must be investigated whether the strategies that will meet the needs of the organization meet the needs of the sector.

References

Adomako, S., & Nguyen, N. P. (2024). Digitalization, inter-organizational collaboration, and technology transfer. *The Journal of Technology Transfer*, *49*(4), 1176–1202.

Aljawder, A., & Al-Karaghouli, W. (2024). The adoption of technology management principles and artificial intelligence for a sustainable lean construction industry in the case of Bahrain. *Journal of Decision Systems*, *33*(2), 263–292.

Allen, D. K. (2013). *Diagnosing social support and performance management: A case study of contextual ambidexterity in a manufacturing company*. Georgia State University.

AlSaied, M. K., & Alkhoraif, A. A. (2024). The role of organizational learning and innovative organizational culture for ambidextrous innovation. *The Learning Organization*, *31*(2), 205–226.

Apell, P., & Eriksson, H. (2023). Artificial intelligence (AI) healthcare technology innovations: the current state and challenges from a life science industry perspective. *Technology Analysis & Strategic Management*, *35*(2), 179–193.

Chen, C. N., & Lin, J. Y. (2023). Organizational learning and the evolution of firms' competitive advantage. *Journal of Engineering and Technology Management*, *70*, 101780.

DiBella, A. J., Nevis, E. C., & Gould, J. M. (1996). Understanding organizational learning capability. *Journal of Management Studies*, *33*(3), 361–379.

Ferreira, J. J., Fernandes, C. I., & Ferreira, F. A. (2020). Wearing failure as a path to innovation. *Journal of Business Research*, *120*, 195–202.

Fiol, C. M., & Lyles, M. A. (1985). Organizational learning. *Academy of Management Review*, *10*(4), 803–813.

Garvin, D.A. (1993). Building a learning organization. *Harvard Business Review*, *73*(4), 78–91.

Gieskes, J. F., Hyland, P. W., & Magnusson, M. G. (2002). Organizational learning barriers in distributed product development: Observations from a multinational corporation. *Journal of Workplace Learning*, *14*(8), 310–319.

Gkinko, L., & Elbanna, A. (2023). Designing trust: The formation of employees' trust in conversational AI in the digital workplace. *Journal of Business Research*, *158*, 113707.

Gustafsson, S., Gillespie, N., Searle, R., Hope Hailey, V., & Dietz, G. (2021). Preserving organizational trust during disruption. *Organization Studies*, *42*(9), 1409–1433.

Jerez-Gomez, P., Cespedes-Lorente, J., & Valle-Cabrera, R. (2005). Organizational learning capability: A proposal of measurement, *Journal of Business Research, 58*, 715–725.

Kim, J. S., & Seo, D. (2023). Foresight and strategic decision-making framework from artificial intelligence technology development to utilization activities in small and medium enterprises. *Foresight*, *25*(6), 769–787.

Lakshmi, A. J., Kumar, A., Kumar, M. S., Patel, S. I., Naik, S. L., & Ramesh, J. V. N. (2023). Artificial intelligence in steering the digital transformation of collaborative technical education. *The Journal of High Technology Management Research*, *34*(2), 100467.

Lee, M. C., Scheepers, H., Lui, A. K., & Ngai, E. W. (2023). The implementation of artificial intelligence in organizations: A systematic literature review. *Information & Management*, *60*(5), 103816.

Parris, U. (2000). An access to care center as a learning organization. *Healthcare Management Forum*, *13*(1), 50–52. https://doi.org/10.1016/S0840-4704(10)60733-8.

Peng, J., Quan, J., & Qin, Q. (2023). R&D investment, intellectual capital, organizational learning, and firm performance: A study of Chinese software companies. *Total Quality Management & Business Excellence*, *34*(9–10), 1196–1216.

Pilar, Jerez-Gomez, Jose, Cespedes-Lorente, & Ramon, Valle-Cabrera (2005). Organizational learning capability: A proposal of measurement. *Journal of Business Research*, *58*(6), 715–725.

Reddy, V. R., & Dyaram, L. (2014). Dimensions of absorptive capacity across multiple generations in organizational context. In *European Conference on Management, Leadership & Governance* (pp. 379–387). Academic Conferences International Limited.

Schuh, G., Boßmeyer, H. J., & Brälking, A. (2021). Data-driven technology management supported by artificial intelligence solutions. *Journal of Production Systems and Logistics*, *1*, 1–28. https://doi.org/10.15488/10528.

Solaimani, S., & Swaak, L. (2023). Critical success factors in a multi-stage adoption of artificial intelligence: A Necessary Condition Analysis. *Journal of Engineering and Technology Management*, *69*, 101760.

Webster, J., & Martocchio, J. J. (1995). The differential effects of software training previews on training outcomes. *Journal of Management*, *21*(4), 757–787.

Yang, D. (2011). The effect of knowledge management on product innovation-Evidence from the Chinese software outsourcing vendors. *Business*, *3*, 16–22.

12 Generative AI in Finance – New Challenges across Trust-Building Needs

Magdalena Mądra-Sawicka

12.1 Introduction

Generative AI (GenAI) in the aspect of changes in the financial sector could be analyzed from two sides: the customer and investor approach, and the second view represents financial and non-financial institutions that are present actively on the financial market. From the first side, GenAI transforms the finance sector by enhancing decision-making, automating processes, and improving customer experiences. GenAI transforms the finance sector into mechanisms based on various impactful applications that enhance processes and outcomes. It should be underlined that risk management and sentiment analysis tools like ChatGPT evaluate corporate sentiments from financial reports to forecast companies' risk-handling capabilities and stock performance (Chen et al., 2023). The second approach, from a company point of view, covers problems like fraud detection, risk management, or personalized financial services that could be solved by AI-driven models that offer innovative solutions to optimize the efficiency and accuracy of financial operations. However, its adoption still raises ethical and regulatory challenges concerning transparency, fairness, and data security (Ayub & Banday, 2023; Firmansyah et al., 2024; Jedličková, 2024), thus its impact on the trust of AI tools by users who should be aware that AI can be a final decision-making authority.

The financial services industry has recognized the potential of technology-focused initiatives, such as payment platforms and trade and investment technologies, to build trust and increase transparency. Payment platforms, for instance, provide real-time transaction tracking, reducing the risk of errors and fraud while ensuring greater clarity for consumers and businesses. Trade and investment technologies play a crucial role in democratizing access to markets by offering automated advisory services (robo-advisors), data-driven portfolio management, and predictive analytics. These tools help reduce information asymmetry and empower investors with insights that were previously accessible only to large institutions. Additionally, tailored GenAI models streamline tasks such as dialogue

DOI: 10.4324/9781003586944-14

summarization in customer service, reducing processing time and improving service efficiency and client satisfaction (Yun et al., 2023). By providing clearer, unbiased recommendations, they foster confidence in the fairness and integrity of financial decisions.

GenAI, including large language models (LLMs), facilitates the generation of new data from existing datasets, supporting better decision-making, risk assessments, and personalized banking solutions (Reshmi et al., 2024). Models used by GenAI stimulate fraud scenarios and training systems to recognize and counteract evolving fraud tactics. Overly sensitive models may incorrectly flag legitimate transactions, causing customer frustration and diminishing confidence in financial services (Barde & Kulkarni, 2023a; Shafik, 2024). However, GenAI can significantly enhance fraud detection in finance, reinforcing trust by providing more secure and reliable services.

This chapter studies the need for a comprehensive examination of current trends and challenges and identifies the broad application of GenAI in finance and trust-building needs. Given the growing role of AI in predictive analysis, automated decision-making, and the delivery of personalized solutions, it is crucial to develop trust by developing ethical and regulatory standards that will support data security in the finance sector.

This chapter contributes to a better understanding of the transformative processes occurring in the financial sector due to GenAI implementation. It offers analytical and theoretical frameworks to assist researchers and practitioners in evaluating its impact on the sector by underlying the trust approach. The literature review outlined the rapid adoption of GenAI in the finance sector and its growing market value. It provides insights into its usage for productivity enhancement, customer service, and software development. In the next section, the methodology was presented based on secondary data derived from sector reports and databases like Statista and EMIS. This section details the scope, sources, and characteristics of the data analyzed, emphasizing regional and temporal variations in the studies. The analysis integrates insights from various surveys and reports to discuss the state of AI adoption in the financial sector. Key themes include productivity improvements, investment trends, workforce transformation, ethical considerations, and the evolving role of GenAI in areas like fraud detection, credit risk assessment, and personalized customer services. The section, New Challenges for GenAI in Finance, addresses the challenges of integrating GenAI into financial systems, including regulatory compliance, ethical concerns, and technical barriers. It emphasizes the need for robust frameworks to ensure transparency, fairness, and compatibility with existing operations. The conclusion synthesizes the findings, highlighting the disparity in AI adoption across industries due to varying levels of digital infrastructure and sector-specific needs. This chapter includes the study's limitations and further research directions.

12.2 Literature Review

Personal psychological traits, social dynamics, and confidence in financial institutions and AI-driven technologies influence trust. Trust can outweigh the importance of gathering detailed information. Even when individuals have access to financial data, their decisions are driven more by their level of trust than by the specifics of the information available. Trust in financial institutions plays a vital role, particularly during financial crises. When trust erodes, it can spark a broader loss of confidence in financial markets, impacting investment behavior and undermining the advisory role.

GenAI brings substantial financial sector advantages, particularly efficiency, predictive analysis, and personalized services (Barney & Reeves, 2024). Financial institutions use GenAI to enhance predictive analysis, improving credit scoring, risk assessment, and trading strategies for more precise and informed financial outcomes (Botunac et al., 2024; Lăzăroiu et al., 2024). Furthermore, AI-powered tools like chatbots and other applications enable tailored customer service, boosting client engagement and satisfaction (Botunac et al., 2024; Shalini & Bagrecha, 2023a). Chatbot automation can make clients feel undervalued or ignored, especially in complex or emotional cases requiring empathy and a personalized approach. Automating repetitive tasks allows human workers to focus on higher-level decisions that involve ethics and empathy, as well as more advanced communication skills.

Implementing GenAI in the finance sector presents several challenges and key considerations that must be carefully managed. It is related to regulatory compliance that remains a significant issue, as financial institutions must adhere to established regulations and require strategic approaches to balance compliance with AI innovation (Botunac et al., 2024; Shalini & Bagrecha, 2023a). Ethical and privacy concerns, including data quality, interpretability, and algorithmic biases, highlight the need for clear guidelines and robust accountability measures to promote transparency and fairness to match the trust of technology users. Additionally, integrating GenAI systems into existing financial infrastructure brings technical challenges, such as upgrading data systems and ensuring smooth compatibility with current operations. Overcoming these hurdles is crucial for realizing the full benefits of AI in financial services (Balavenu et al., 2022). Trust is a key psychological element influencing financial decision-making, significantly shaping how individuals perceive risk and impacting overall market behavior. That is why creating a trust-building environment for GenAI is essential to ensure its adoption, foster confidence in its outputs, and mitigate fears of bias, errors, and lack of transparency in financial processes.

12.3 Methodology and Data Source

This chapter utilized secondary data derived from sector reports based on market and survey data directly or indirectly (like reports about the sector with additional information on AI) GenAI in finance. The analysis of these data was conducted for the period 2022-2024. The scope of the data varies depending on the study conducted by a given institution. Table 12.1 presents each study's name, the research details, and supplementary information. The presented data differed between region and period of data collection and the main aim for which they were conducted. The selected reports were reviewed based on the Statista and EMIS databases.

Data were analyzed from different reports based on the institutions' research data to describe the phenomenon studied. To comment actual situation, also use the latest publications from the Financial Times and Wall Street Journal to outline current trends and changes.

12.3.1 Report, Surveys, and Statistical Data Review

Venture capital has been instrumental in driving AI technology development and widespread adoption. By offering essential financial resources and support, venture capitalists have empowered startups to innovate and introduce cutting-edge AI solutions to the market. In 2023, Capital One led the way in AI readiness among banks in the Americas and Europe, achieving a top score of 90.9 on the AI readiness index. JPMorgan Chase ranked second with 89.5, followed by the Royal Bank of Canada with a score of 73.7 (CB Insights, 2023). Thus, the banking sector dominated AI investments in Europe, allocating approximately $ 5.37 billion, representing 15.7% of the region's total AI spending (Statista, 2024b). AI adoption in finance was widespread, with over two-thirds of institutions utilizing AI for data analytics in 2023, alongside other applications such as data processing and natural language processing (Statista, 2024a). In commercial payments, European banks demonstrated a stronger inclination toward blockchain investments (47%) compared to just 14% in Latin America (Accenture, 2023). Globally, the financial services sector invested around $ 35 billion in AI during 2023, highlighting its commitment to embracing AI-driven innovation and operational efficiency (Statista, 2024a).

According to Statista and Juniper, "Research conducted in 2023 presents the banking sector challenges and new technologies focused on three main areas: generative AI, blockchain, and cloud computing. It showed a growing focus on digital transformation and innovation" (Statista and Juniper Research, 2024).

GenAI is revolutionizing the finance sector at an accelerated pace, with its market value projected to surge from $ 1.09 billion in 2023 to over $ 12 billion by 2033, reflecting a 28.1% compound annual growth rate (MarketResearch.biz, 2024). In Europe, 77% of financial sector professionals expect GenAI to impact productivity substantially, and 75% intend to boost their investments in AI technologies (EY, 2023a). In the financial sector, GenAI is primarily utilized for employee training and collaboration, enhancing customer service, and advancing software development (EY, 2024). The finance sector analytics demonstrates

Table 12.1 Overview of reports used for data analysis

Report	*Survey details*	*Further information*
Accenture. Main technology invested in by banks worldwide to help innovate their commercial payments in 2023	Survey period: July and August 2023; region: worldwide; type of survey: Face-to-face interview and online survey; number of respondents: 223.	The source phrased the question: "To what degree has your organization adopted these technologies in your commercial payments division?
Accenture. Investments, and implementation of generative artificial intelligence in commercial payments offered by banks as of 2023 by use case	Survey period: July and August 2023; region: worldwide; type of survey: Face-to-face interview and online survey; number of respondents: 223.	The source phrased the question: "What are your plans to use generative AI for the following activities within commercial payments?"
CB Insights. Leading banks in artificial intelligence (AI) readiness in the Americas and Europe in 2023	Survey period: 2023; region: North America, Europe, LAC	The index covers the 50 largest retail banks in the Americas and Europe by market cap. The banks were analyzed across three metrics: talent, execution, and innovation. The talent score examined a bank's ability to attract and retain AI specialists.
Crisil. Asset Management in Europe AI	Survey period: 2022 to 2030; region: Europe; special characteristic: Institutional asset managers	Managers engaged in AI have either put AI solutions into practice or are currently working on AI applications. The remaining asset managers have not yet engaged in AI.
Deloitte. Financial services processes using artificial intelligence (AI) in day-to-day use worldwide in 2022, by business segment	Survey period: April to May 2022; region: worldwide; number of respondents: 2620; special characteristic: business leaders	All participating companies adopted AI technologies and were AI users. Respondents were required to meet one of the following criteria: responsible for AI technology spending or approval of AI investments, developing AI technology strategies, managing or overseeing AI technology implementation, serving as an AI technology subject matter specialist, or making or influencing decisions around AI technology.
EY. Attitude of financial services industry leaders toward generative artificial intelligence (GenAI) in Europe in 2023	Survey period: October 2023; region: America, Europe; number of respondents: 60	Respondents were executives from 60 European financial institutions, including listed firms representing an aggregate market cap of 507.7 billion British pounds

Heidrick & Struggles. 2023 Europe and US Data, Analytics, and Artificial Intelligence Executive Organization and Compensation Survey	Survey period: 2023; region: North America, Europe, United States; number of respondents: 158; special characteristic: Executives self-reporting anonymously	The source phrased the question as: "Who at your company owns the AI strategy today?
MarketResearch.biz. Market size of generative artificial intelligence (AI) in the financial services sector from 2022 to 2023, with a forecast until 2033 (in billion US dollars)	Survey period: 2022 to 2023; region: worldwide	-
McKinsey & Company. (July 6, 2023). Potential impact of generative artificial intelligence (AI)on sector revenues worldwide in 2023, by sector (in billion US dollars)	Survey period: 223; region: worldwide	Survey by IHS Markit; Oxford Economics; McKinsey & Company; S&P Global
Stanford University. Artificial intelligence (AI) adoption worldwide 2022, by sector and function, conducted by McKinsey & Company	Survey period: 2022	It categorizes adoption across various organizational functions, including Human Resources, Manufacturing, Marketing & Sales, Product/Service Development, Risk, Service Operations, Strategy & Corporate Finance, and Supply Chain Management.
Statista, & Juniper Research. (March 20, 2024). The estimated value of the banking sector's generative artificial intelligence (AI) spending worldwide in 2023, with forecasts from 2024 to 2030 (in billion US dollars)	Survey period: 2022 to 2030; region: worldwide	The sources provided the banking sector's annual generative AI spending for 2024 and 2030. Statista calculated the rest of the figures based on the compound annual growth rate.
WEKA. (August 15, 2023). Leading infrastructure challenges for AI developments worldwide in 2023	Survey period: 2023; region: worldwide; number of respondents: 1516	Survey by S&P Global Market Intelligence

Source: Own elaboration based on Statista database strong optimism about GenAI, with 87% of US financial services executives confident in its potential to enhance customer experiences (EY, 2023b).

According to the EY European Financial Services AI Survey conducted in October 2023, nearly 80% of European financial sector professionals expected GenAI to boost productivity significantly. Furthermore, 68% predicted that would require upskilling, though 35% acknowledged lacking concrete plans to implement such training. Regarding investment, 75% of respondents indicated they plan to increase spending on AI technologies within the following year. Furthermore, 60% of respondents have allocated capital to GenAI technologies over the past year, demonstrating proactive investment and adoption. Similarly, 60% predict that AI adoption will significantly affect entry-level roles, indicating a shift in workforce dynamics as automation and AI integration reshape job opportunities. However, only 7% of respondents acknowledged ethical issues, highlighting a lack of focus on the ethical implications of AI deployment directly related to trust. These findings suggest that while the sector is optimistic about the benefits of GenAI, there is a pressing need for strategic workforce development to meet upskilling demands. The anticipated impact on entry-level roles highlights the importance of balancing technological advancements with sustainable job transitions and workforce readiness. Addressing these challenges will require targeted efforts to implement structured training programs and to prioritize ethical standards, ensuring a balanced and responsible approach to GenAI adoption.

According to a Stanford University survey conducted in 2023 by McKinsey & Company, different industries leverage AI across various operational areas, including human resources, manufacturing, marketing and sales, product/service development, risk, service operations, strategy and corporate finance, and supply chain management. The survey presents that AI adoption is most prominent in risk management (19%), service operations (19%), and strategy and corporate finance (21%), indicating a strong focus on operational efficiency and risk mitigation. Other areas, such as human resources (11%), product/service development (10%), and supply chain management (9%), show moderate levels of adoption. AI application in financial services is concentrated in product/service development (31%), service operations (24%), and strategy and corporate finance (23%), highlighting a focus on innovation and operational efficiency (Stanford University, 2023).

Insurance companies increasingly adopt AI technologies to improve financial market decision-making and streamline operations. AI applications span key areas such as underwriting, claims processing, fraud detection, and customer service, driving greater efficiency and accuracy. Machine learning and data analytics are central to automating repetitive tasks, processing large volumes of data, and identifying fraudulent activities, helping insurers optimize performance and reduce risks (Akoglu et al., 2024). Combining Robotic Process Automation with AI and advanced analytics revolutionizes traditional processes like risk assessment and compliance management. Moreover, sophisticated algorithms enable insurers to analyze transactional data faster and combine information innovatively. This leads to improved risk evaluation and more precise pricing aligning with insured businesses' value and risk profiles.

Based on the Accenture survey, "Reinventing Commercial Payments Survey 2023," conducted among banks, the most important was security as the top priority globally, with a 54% investment rate reflecting banks' commitment to safeguarding systems and transactions. Cloud services follow closely at 43%, highlighting the ongoing adoption of scalable infrastructure to enhance operational efficiency. Blockchain ranks third globally at 38%, emphasizing its secure and transparent transaction processing role. Data technologies and analytics stand at 33%, showing the growing importance of data-driven decision-making. AI and automation receive 30% investment, reflecting efforts to improve process efficiency and innovation. GenAI, as an emerging technology, has a global average investment of 13%. The highest adoption of AI was seen in Europe and Asia-Pacific (14%), North America trails slightly at 12%, and Latin America records the lowest at 11% (Accenture, 2023).

Crisil's report data concerns the institutional asset managers implementing AI in Europe in 2024. A survey indicates that 45% of asset managers are engaged in implementing AI, while 55% have not yet adopted AI (Crisil, 2024). This suggests that while AI adoption is growing, there remains a considerable opportunity for further implementation in the asset management sector. Encouraging factors such as improved efficiency, data-driven decision-making, and competitive advantages could drive the remaining organizations toward AI adoption in the near future.

Based on the Heidrick & Struggles survey that analyzed leaders in charge of AI strategies in the United States and Europe in 2023, by sector, it can be stated that Financial Services or Fintech AI executives lead AI strategies in 37% of cases, followed by Chief Information or Technology Officers (CIO/CTO) at 26%. Other roles, such as the executive committee (9%) and technology function executives (9%), play more minor roles. CEOs (2%) and COOs/CFOs (2% each) have minimal involvement (Heidrick and Struggles, 2023). The survey highlights that AI strategy leadership is predominantly concentrated in specialized roles. It indicates that organizations prioritize domain-specific expertise and technical leadership over general executive oversight in shaping AI strategy.

According to the study of McKinsey & Company related to the potential impact of GenAI on sector revenues worldwide in 2023, the industries expected to experience the most significant effect include High Tech, with an estimated range of \$ 240–\$ 460 billion, followed by Retail at \$ 240–\$ 390 billion, and Banking at \$ 200–\$ 340 billion. These sectors stand out due to their reliance on data, automation, and advanced digital technologies, which GenAI can enhance substantially (McKinsey & Company, 2023).

A report published by WEKA in 2023 presents leading infrastructure challenges for AI developments worldwide. The most significant challenge is data management, affecting 32% of organizations, as handling and processing large volumes of complex datasets remains critical for AI operations. Security follows closely at 26%, reflecting growing concerns about data privacy, system protection, and cybersecurity threats. Ensuring sufficient compute performance is another notable issue, affecting 20% of respondents, as the demand for computational power to train and operate AI models continues to grow. Additionally, networking poses challenges for 13% of organizations, indicating issues with connectivity, speed, and efficiency needed for AI deployment. Finally, storage emerges as a concern for 8%, particularly regarding scalability and performance as AI systems expand. Solving these challenges will require investments in advanced data systems, robust cybersecurity measures, high-performance computing infrastructure, and efficient networking and storage solutions to fully support AI development on a global scale (WEKA, 2023).

Based on Deloitte data, financial services processes using AI in day-to-day use worldwide in 2022 noticed the highest adoption in voice assistants, chatbots, and conversational AI, with 42% of companies utilizing these technologies. This indicates the decisive importance of AI-powered communication tools among business leaders in providing immediate and personalized support (Deloitte, 2022). Personalization follows closely at 40%, reflecting the importance of customer experiences to individual preferences. Contact center optimization (39%), customer feedback analysis (38%), and customer service operations (38%). Companies investing in AI will likely gain a competitive advantage by improving service quality and operational performance.

From the investor's point of view, GenAI could help optimize risk. GenAI tools, such as ChatGPT, can evaluate corporate financial statements to assess companies' risk management effectiveness and forecast stock return performance by analyzing their sentiments toward environmental policies (Financial Times, 2024; Wall Street Journal, 2024). AI model can process earnings call transcripts to evaluate company-specific exposures to various risks, including those related to environmental policies. The findings indicate that AI-generated risk assessments significantly enhance the prediction of firm-level volatility and influence corporate decisions, such as investment and innovation strategies. Integrating GenAI into financial analysis can provide investors' and stakeholders' valuable insights into a company's risk management effectiveness and potential stock performance, particularly concerning environmental policy sentiments. It enhances trust by offering deeper transparency into a company's risk management and environmental policies, fostering informed decision-making.

12.3.2 New Challenges and New Vision in Adopting GenAI in Finance – Discussion

Implementing GenAI in the financial sector introduces several key challenges and considerations that must be carefully managed to ensure its effective and responsible use. A significant challenge is regulatory compliance, as financial institutions must adhere to existing standards such as the General Data Protection Regulation while preparing for new regulations like the EU AI Act innovation (Botunac et al., 2024; Shalini & Bagrecha, 2023a).

GenAI fosters substantial innovation, enhances efficiency in banking services, and supports banks to achieve greater operational consistency and reliability (Botunac et al., 2024). Looking to the future, the financial sector is set for sustained growth as AI technologies continue to drive innovation, resulting in more sophisticated and diverse financial services. To unlock the full potential of GenAI, financial institutions need to prioritize long-term strategic planning and invest in AI education, enabling them to overcome challenges and capitalize on transformative opportunities (Al Naqbi et al., 2024; Go et al., 2020; Panda et al., 2024). GenAI plays an increasingly vital role in the financial sector by improving risk assessment, with its primary application being fraud detection and prevention, where GenAI identifies irregularities and suspicious activities within transaction data (Barde & Kulkarni, 2023b; Shalini & Bagrecha, 2023b; Zhang et al., 2024). In credit risk assessment, GenAI models improve the accuracy of default probability predictions for credit applicants. These systems use deep learning and neural networks to analyze extensive datasets for more precise risk evaluations. Additionally, market and systemic risk management utilizes generative models to predict and measure systemic financial risks (Chen, 2023; Kamruzzaman et al., 2024; Ramesh & Jeyakarthic, 2024; Rhzioual Berrada et al., 2022). These models process diverse financial inputs, enabling institutions to manage market volatility and defaults effectively while presenting insights through interactive dashboards. Moreover, predictive analytics powered by GenAI generates insights by creating new data from existing sources, facilitating forecasting of financial risks and trends. These models analyze vast amounts of data to uncover patterns and trends, empowering financial institutions to anticipate risks and proactively address them (Goel et al., 2023; Reshmi et al., 2024).

Investors increasingly utilize AI to improve their decision-making processes in financial markets across multiple applications. Data analysis and pattern recognition enable AI to process vast datasets and identify trends or patterns that traditional methods often overlook. Models like Long Short-Term Memory and Convolutional Neural Networks are particularly effective for forecasting stock market behavior, delivering high accuracy in predictions. In the area of predictive analytics, AI analyzes both historical data and real-time signals to predict market trends and asset prices, supporting strategic planning and risk management (Albaooth, 2023; Boggavarapu et al., 2024; Pillai & Bi, 2024; Venkatarathnam et al., 2024; Zhou, 2024). Reinforcement learning techniques, such as Deep Q-learning networks, are used to design automated trading strategies that maximize profit potential (Goel et al., 2023; Kalva & Satuluri, 2023). AI systems capable of analyzing and quantifying news can respond to market-impacting events faster than humans, affecting stock market dynamics such as price, trading volume, and volatility (Kayathri & Prabakaran, 2024). Moreover, Generative Adversarial Networks and Recurrent Neural Networks are employed to uncover patterns and connections within time series data, playing a vital role in forecasting stock market trends and developing trade execution strategies (Salama, 2024). For financial forecasting, AI models enhance the accuracy of stock price predictions and market forecasts, improving investment decision-making reliability (Venkatarathnam et al., 2024). It can reduce the information gap between institutional investors and smaller market participants, fostering greater market transparency and fairness. As AI-driven models reveal hidden patterns and improve forecasting accuracy, they mitigate the risks associated with incomplete or unequal information, ultimately building trust among investors by promoting more equitable access to critical financial data and improving the reliability of market predictions. AI also transforms risk management by evaluating systemic and systematic risks through advanced data analysis, helping investors mitigate financial risks (Kamruzzaman et al., 2024; Khattak et al., 2023; Lan, 2024; Zhu et al., 2024). AI can significantly optimize portfolios and select profitable assets. In behavioral finance, AI incorporates insights from investor behavior to predict market movements, reducing decision-making biases and minimizing emotional influences that can destabilize markets (Lin et al., 2024; Sarin & Sharma, 2023). By offering rational, data-driven insights, AI supports more efficient and balanced investment decisions. The use of robo-advisors and automated trading has further revolutionized the financial landscape. AI-powered platforms provide personalized investment advice, portfolio optimization, and automated portfolio management, allowing investors to manage assets with minimal intervention (Devapitchai et al., 2024; Kalva & Satuluri, 2023; Wang et al., 2024; Zhou, 2024). Additionally, algorithmic trading automates processes, improving trading efficiency and execution speed. To enhance trust, it is essential to introduce AI auditing mechanisms, develop explainability methods, and establish international regulatory standards. The key role also plays in education in that field.

12.4 Conclusion

Adopting AI technologies differs significantly across industries, with some sectors embracing AI more quickly than others. Factors like the maturity of digital infrastructure and the unique demands of each sector drive this disparity. Thus, the increasing focus on GenAI, blockchain, and cloud computing highlights the financial sector's commitment to digital transformation. However, challenges related to workforce upskilling, regulatory compliance, and ethical considerations remain under-addressed.

As AI reshapes workforce dynamics and automates complex processes, financial institutions must balance innovation with sustainable job transitions, ethical AI deployment, and structured training programs to ensure long-term sector growth and resilience. To leverage the capabilities of GenAI, financial organizations must adopt strategic approaches that align technological innovation with legal and regulatory requirements and trust importance (Kumar et al., 2023; Panda et al., 2024). Equally important for trust creation are the ethical and privacy concerns associated with GenAI, including data quality, algorithmic transparency, and the potential for bias. Addressing these challenges requires clear guidelines and accountability measures to foster trust and ensure fairness in AI systems. On a technical level, integrating GenAI involves overcoming infrastructure limitations and ensuring compatibility with existing financial systems. Tackling these issues is essential for maximizing the benefits of GenAI while upholding ethical, regulatory, and operational standards.

Despite its advantages, integrating AI into financial markets poses challenges. Data quality and algorithmic bias are ongoing concerns, as accurate outputs rely on clean, unbiased data. Ethical and regulatory factors also play a critical role, requiring transparency and strategies to mitigate systemic risks. Addressing these challenges is essential for AI's sustainable adoption and ethical use in financial decision-making.

AI can deliver standardized insights and enhance communication efficiency, particularly in asset management. Furthermore, GenAI fosters innovation by uncovering new product opportunities and business models, driving growth, and enabling the financial sector to evolve with creative, technology-driven solutions. These capabilities position GenAI as a key driver of progress and transformation within the finance sector. Financial institutions must implement clear governance frameworks to foster long-term confidence, prioritize fairness and accountability, and communicate transparently with stakeholders, ensuring that AI-driven innovation aligns with global ethical standards and public expectations. AI continues to evolve, fostering a culture of responsibility and shared accountability in the finance sector. GenAI will be essential in bridging the gap between innovation and societal acceptance, ensuring that AI enhances stability, equity, and long-term growth in the global financial ecosystem. Building trust involves demonstrating that AI can deliver profits and efficiencies and is critical in safeguarding financial stability. Institutions prioritizing digital literacy and AI upskilling will foster mutual trust and reduce fears associated with technological disruption. To sum up, trust in AI is not a static goal but an ongoing process that must evolve alongside technological advancements.

The study presented here has several limitations that should be considered. It relies exclusively on secondary data sourced from sector reports and surveys, which, while valuable for broad insights, limits the ability to validate or critically assess the primary data's accuracy, methodologies, and potential biases. Additionally, the analysis is constrained by its geographical and temporal focus, emphasizing specific regions, such as Europe and North America, and a limited time frame of 2022–2024. As a result, broader global trends and long-term developments in AI adoption may not be fully captured. The study's emphasis on the financial sector further narrows its applicability, as it does not comprehensively address how GenAI impacts other industries. Future research on GenAI in the financial sector can focus on the ethical and regulatory implications of using GenAI, particularly in areas such as algorithm transparency, bias prevention, and data privacy protection. Developing and evaluating the effectiveness of new legal frameworks and ethical guidelines would significantly contribute to the sustainable development of AI. Another direction could be related to sustainable development and responsible investing, including assessing AI's potential in analyzing environmental, social, and governance risks and creating new financial products to support sustainable initiatives.

References

Accenture. (September 12, 2023). Main technology invested in by banks worldwide to help innovate their commercial payments in 2023 [Graph]. In *Statista*. Retrieved December 18, 2024, from https://www.statista.com/statistics/1254749/ai-challenges-financial-services/.

Akoglu, L., Chawla, N., Domingo-Ferrer, J., Kurshan, E., Kumar, S., Naware, V., Rodriguez-Serrano, J. A., Chaturvedi, I., Nagrecha, S., Das, M., Das, M., & Faruquie, T. (2024). Machine learning in finance. *Proceedings of the ACM SIGKDD* International *Conference on Knowledge Discovery and Data Mining, 24*, 6703. https://doi.org/10.1145/3637528.3671488.

Albaooth, B. (2023). The role of artificial intelligence prediction in stock market investor's decisions. *Proceedings of the 2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering, CSDE 2023*. https://doi.org/10.1109/CSDE59766.2023.10487719.

Al Naqbi, H., Bahroun, Z., & Ahmed, V. (2024). Enhancing work productivity through generative artificial intelligence: A comprehensive literature review. *Sustainability (Switzerland)*, *16*(3), 1166. https://doi.org/10.3390/su16031166.

Ayub, Z., & Banday, M. T. (2023). Ethics in Artificial Intelligence: An analysis of ethical issues and possible solutions. *Proceedings - 3rd International Conference on Smart Technologies, Communication and Robotics 2023, STCR 2023, 1*, 1–6. https://doi.org/10.1109/STCR59085.2023.10396966.

Balavenu, R., Khan, A. K., Faisal, S. M., Sriprasadh, K., & Sisodia, D. R. (2022). An empirical investigation in analysing the proactive approach of artificial intelligence in regulating the financial sector. In V. E. Balas, G. R. Sinha, B. Agarwal, T. K. Sharma, P. Dadheech, & M. Mahrishi (Eds.), *Communications in computer and information science* (Vol. 1591, pp. 90–98) *CCIS*. https://doi.org/10.1007/978-3-031-07012-9_8.

Barde, K., & Kulkarni, P. A. (2023a). Applications of generative AI in fintech. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3639856.3639893.

Barde, K., & Kulkarni, P. A. (2023b). Applications of generative AI in fintech. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3639856.3639893.

Barney, J. B., & Reeves, M. (2024). AI won't give you a new sustainable advantage. *Harvard Business Review*, *2024-September, 103*(9–10), 71–79.

Boggavarapu, S., Ramkumar, G., Gedamkar, P. R., Kaneria, A., Pundir, S., & Selvameena, R. (2024). Research on unmanned artificial intelligence based financial volatility prediction in international stock market. *5th International Conference on Recent Trends in Computer Science and Technology, ICRTCST 2024 - Proceedings*, 16–20. https://doi.org/10.1109/ICRTCST61793.2024.10578397.

Botunac, I., Parlov, N., & Bosna, J. (2024). Opportunities for gen AI in the banking sector regarding the AI Act, GDPR, data act, and DORA. *2024 13th Mediterranean Conference on Embedded Computing, MECO 2024*. https://doi.org/10.1109/MECO62516.2024.10577936.

CB Insights. (June 28, 2023). Leading banks in artificial intelligence (AI) readiness in the Americas and Europe in 2023 [Graph]. In *Statista*. Retrieved December 23, 2024, from https://www.statista.com/statistics/1446366/leading-banks-in-ai-adoption-worldwide/.

Chen, B., Wu, Z., & Zhao, R. (2023). From fiction to fact: the growing role of generative AI in business and finance. *Journal of Chinese Economic and Business Studies*, *21*(4), 471–496. https://doi.org/10.1080/14765284.2023.2245279.

Chen, W. (2023). The GSO-deep learning-based financial risk management system for rural economic development organization. *International Journal of Advanced Computer Science and Applications*, *14*(10), 670–678. https://doi.org/10.14569/IJACSA.2023.0141071.

Crisil. (April 9, 2024). Portion of institutional asset managers implementing Artificial Intelligence (AI) in Europe in 2024 [Graph]. In *Statista*. Retrieved December 18, 2024, from https://www.statista.com/statistics/1466460/ai-in-european-institutional-asset-management/.

Deloitte. (October 27, 2022). Financial services processes using artificial intelligence (AI) in day-to-day use worldwide in 2022, by business segment [Graph]. In *Statista*. Retrieved December 19, 2024, from https://www.statista.com/statistics/1374553/ai-in-financial-services-day-to-day-use/.

Devapitchai, J. J., Krishnapriya, S. V., Karuppiah, S. P., Robert, P. W., & Saranya, S. (2024). Using AI-driven decision-making tools in corporate investment planning. In J. G. Sankar & A. David (Eds.), *Generative AI for transformational management* (pp. 137–160). IGI Global. https://doi.org/10.4018/979-8-3693-5578-7.ch006.

EY. (October 25, 2023). Attitude of financial services industry leaders toward generative artificial intelligence (GenAI) in Europe in 2023 [Graph]. In *Statista*. Retrieved December 23, 2024, from https://www.statista.com/statistics/1446207/europe-attitude-toward-genai-in-financial-services/

EY. (2023b). *Attitude of financial services sector leaders toward generative artificial intelligence (GenAI) in the United States in 2023*. https://www.statista.com/statistics/1446190/us-attitude-toward-genai-in-financial-services/

EY. (February 8, 2024). Which are, or will be, the most significant AI and GenAI application types for your organization? [Graph]. In *Statista*. Retrieved December 23, 2024, from https://www.statista.com/statistics/1482550/firms-demand-for-ai-by-application-type-global/.

Financial Times, AI can pick stock better than you can. Retrieved December 19, 2024, from https://www.ft.com/content/a680e7a2-2799-4fc9-ace9-b01770b260df?utm_source=chatgpt.com.

Firmansyah, G., Bansal, S., Walawalkar, A. M., Kumar, S., & Chattopadhyay, S. (2024). The future of ethical AI. In B. Gupta (Ed.), *Challenges in large language model development and AI ethics* (pp. 145–177). IGI Global. https://doi.org/10.4018/979-8-3693-3860-5.ch005.

Go, E. J., Moon, J., & Kim, J. (2020). Analysis of the current and future of artificial intelligence in the financial sector with big data techniques. *Global Business and Finance Review*, *25*(1), 102–117. https://doi.org/10.17549/gbfr.2020.25.1.102.

Goel, M., Tomar, P. K., Vinjamuri, L. P., Swamy Reddy, G., Al-Taee, M., & Alazzam, M. B. (2023). Using AI for predictive analytics in financial management. *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2023*, 963–967. https://doi.org/10.1109/ICACITE57410.2023.10182711.

Heidrick & Struggles. (October 17, 2023). American and European companies' responsible person for the artificial intelligence (AI) strategy in 2023, by sector [Graph]. In *Statista*. Retrieved December 18, 2024, from https://www.statista.com/statistics/1455283/ai-strategy-leaders-by-industry/

Jedličková, A. (2024). Ethical approaches in designing autonomous and intelligent systems: A comprehensive survey towards responsible development. *AI and Society*, 1–14. https://doi.org/10.1007/s00146-024-02040-9.

Kalva, S., & Satuluri, N. (2023). Stock Market Investment Strategy Using Deep-Q-Learning Network. In J. Hartmanis & G. Goos (Eds.), *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and bioinformatics)* (Vol. 14078, pp. 484–495) LNAI. https://doi.org/10.1007/978-3-031-36402-0_45.

Kamruzzaman, M. M., Alruwaili, O., & Aldaghmani, D. (2024). Measuring systemic and systematic risk in the financial markets using artificial intelligence. *Expert Systems*, *41*(5), e12971, https://doi.org/10.1111/exsy.12971.

Kayathri, V., & Prabakaran, P. (2024). Deep learning-based algorithmic trading based on news and events strategies. In J. Kacprzyk (Ed.), *Lecture Notes in Networks and Systems: LNNS* (Vol. 970, pp. 311–320). Springer. https://doi.org/10.1007/978-981-97-2031-6_27.

Khattak, B. H. A., Shafi, I., Khan, A. S., Flores, E. S., Lara, R. G., Samad, M. A., & Ashraf, I. (2023). A systematic survey of AI models in financial market forecasting for profitability analysis. *IEEE Access*, *11*, 125359–125380. https://doi.org/10.1109/ACCESS.2023.3330156.

Kumar, K., Kuldeep, & Bhushan, B. (2023). Augmenting cybersecurity and fraud detection using artificial intelligence advancements. *Proceedings - 4th IEEE 2023 International Conference on Computing, Communication, and Intelligent Systems, ICCCIS 2023*, 1207–1212. https://doi.org/10.1109/ICCCIS60361.2023.10425069.

Lan, M. (2024). Financial market risk prediction systems based on data analysis. *Proceedings - 2024 3rd International Conference on Artificial Intelligence and Autonomous Robot Systems, AIARS 2024*, 421–426. https://doi.org/10.1109/AIARS63200.2024.00084.

Lăzăroiu, G., Gedeon, T., Rogalska, E., Valaskova, K., Nagy, M., Musa, H., Zvarikova, K., Poliak, M., Horak, J., Crețoiu, R. I., Avram, L. G., & Braga, V. (2024). Digital twin-based cyber-physical manufacturing systems, extended reality metaverse enterprise and production management algorithms, and internet of things financial and labor market technologies in generative artificial intelligence economics. *Oeconomia Copernicana*, *15*(3), 837–870. https://doi.org/10.24136/oc.3183.

Lin, C., Chang, M., & Sun, Y. (2024). Assessing the efficacy of artificial intelligence in mitigating stock market volatility induced by emotional decision-making. *2024 4th International Conference on Computer Communication and Artificial Intelligence, CCAI 2024*, 320–331. https://doi.org/10.1109/CCAI61966.2024.10603355.

MarketResearch.biz. (April 30, 2024). Market size of generative artificial intelligence (AI) in the financial services sector from 2022 to 2023, with a forecast until 2033 (in billion U.S. dollars) [Graph]. In *Statista*. Retrieved December 18, 2024, from https://www.statista.com/statistics/1449285/global-generative-ai-in-financial-services-market-size/.

McKinsey & Company. (July 6, 2023). Potential impact of generative artificial intelligence (AI) on sector revenues worldwide in 2023, by sector (in billion U.S. dollars) [Graph]. In *Statista*. Retrieved December 18, 2024, from https://www.statista.com/statistics/1424706/impact-generative-ai-in-law-worldwide/.

Panda, R. S., Mishra, A., & Mohanty, A. (2024). Innovating reality: The pioneering applications of generative AI. In R. Kumar, S. Sahu, & S. Bhattacharya (Eds.), *The Pioneering Applications of Generative AI* (pp. 85–105). IGI Global. https://doi.org/10.4018/979-8-3693-3278-8.ch004.

Pillai, R. H., & Bi, A. (2024). Revolutionizing stock market predictions: Exploring the role of artificial intelligence. In R. Sharma & K. Mehta (Eds.), *Deep learning tools for predicting stock market movements*. Wiley. https://doi.org/10.1002/9781394214334.ch10.

Ramesh, R., & Jeyakarthic, M. (2024). Enhancing credit risk prediction with hybrid deep learning and sand cat swarm feature selection. *Multimedia Tools and Applications*, *83*(21), 60243–60263. https://doi.org/10.1007/s11042-023-17974-3.

Reshmi, L. B., Raj, R. V., Balasubramaniam, S., & Kumar, K. S. (2024a). Generative AI and LLM: Case study in finance. In S. Balasubramaniam, S. Kadry, A. Prasanth, & R. K. Dhanaraj (Eds.), *Generative AI and LLMs: Natural language processing and generative adversarial networks* (pp. 231–252). De Gruyter. https://doi.org/10.1515/9783111425078-012.

Rhzioual Berrada, I., Barramou, F. Z., & Bachir Alami, O. (2022). A review of Artificial Intelligence approach for credit risk assessment. *2022 2nd International Conference on Artificial Intelligence and Signal Processing, AISP 2022*, 1–5. https://doi.org/10.1109/AISP53593.2022.9760655.

Salama, R. (2024). Integrating spotted hyena optimization technique with generative artificial intelligence for time series forecasting. *Expert Systems*, *42*(2), e13681. https://doi.org/10.1111/exsy.13681.

Sarin, A. B., & Sharma, S. (2023). Artificial intelligence in behavioral finance for investment decision-making. In B. Bhardwaj, D. Sharma, & M. C. Dhiman (Eds.), *AI and emotional intelligence for modern business management* (pp. 194–212). IGI Global. https://doi.org/10.4018/979-8-3693-0418-1.ch013.

Shafik, W. (2024). The role of artificial intelligence in the emerging digital economy era. In *Artificial intelligence enabled management: An emerging economy perspective*. https://doi.org/10.1515/9783111172408-003.

Shalini, R., & Bagrecha, C. (2023a). A study on generative AI and its impact on banking and financial services sector: Data privacy & sustainable perspective. *Proceedings of 2023 IEEE Technology and Engineering Management Conference - Asia Pacific, TEMSCON-ASPAC 2023*, 1–5. https://doi.org/10.1109/TEMSCON-ASPAC59527.2023.10531592

Stanford University. (April 15, 2023). Artificial intelligence (AI) adoption worldwide 2022, by sector and function [Graph]. In *Statista*. Retrieved December 18, 2024, from https://www.statista.com/statistics/1378695/ai-adoption-rate-selected-countries/.

Statista. (2024a). *Artificial intelligence (AI) in finance*. In *Statista*. Retrieved December 18, 2024, from https://www.statista.com/forecasts/1428631/ai-market-share-by-industry.

Statista. (2024b). *Artificial intelligence - Europe*. In *Statista*. Retrieved December 18, 2024, from https://www.statista.com/study/136791/artificial-intelligence-in-europe/.

Statista, & Juniper Research. (March 20, 2024). Estimated value of the banking sector's generative artificial intelligence (AI) spending worldwide in 2023, with forecasts from 2024 to 2030 (in billion U.S. dollars) [Graph]. In *Statista*. Retrieved December 18, 2024, from https://www.statista.com/statistics/1557315/americas-banking-sector-ai-spending-forecast/.

Venkatarathnam, N., Goranta, L. R., Kiran, P. C., Raju, B. P. G., Dilli, S., Basha, S. M., & Kethan, M. (2024). An empirical study on implementation of AI & ML in stock market prediction. *Indian Journal of Information Sources and Services*, *14*(4), 165–174. https://doi.org/10.51983/ijiss-2024.14.4.26.

Wall Street Journal, When IR met AI: How the technology is shaping earnings-day prep. Retrieved December 19, 2024, from https://www.ft.com/content/a680e7a2-2799-4fc9-ace9-b01770b260df?utm_source=chatgpt.com.

Wang, O., Ye, H., & Li, R. (2024). Application of AI technology in internet finance and analysis of security risks. In J. C. Hung, N. Yen, & J. W. Chang (Eds.), *Frontier Computing on Industrial Applications Volume 2. FC 2023. Lecture notes in electrical engineering* (Vol. 1132, pp. 354–359). Springer. https://doi.org/10.1007/978-981-99-9538-7_55.

WEKA. (August 15, 2023). Leading infrastructure challenges for AI developments worldwide in 2023 [Graph]. In *Statista*. Retrieved December 18, 2024, from https://www.statista.com/statistics/1412761/ai-infrastructure-challenges-worldwide/.

Yun, J., Sohn, J. E., & Kyeong, S. (2023). Fine-tuning pretrained language models to enhance dialogue summarization in customer service centers. *ICAIF 2023 - 4th ACM International Conference on AI in Finance*, 365–373. https://doi.org/10.1145/3604237.3626838.

Zhang, Y., Jiang, Z., Peng, C., Zhu, X., & Wang, G. (2024). Management analysis method of multivariate time series anomaly detection in financial risk assessment. *Journal of Organizational and End User Computing*, *36*(1), 1–19. https://doi.org/10.4018/JOEUC.342094.

Zhou, J. (2024). The use of artificial intelligence in investment selection: Robo-advisory platform. *Exploring the Financial Landscape in the Digital Age - Proceedings of the International Conference on Financial Management and the Digital Economy, ICFMDE 2023*, 661–669. https://doi.org/10.1201/9781003508816-95.

Zhu, X., Hu, C., & Luo, L. (2024). Research on intelligent portfolio optimization model based on big data and artificial intelligence. *Proceedings - 2024 International Conference on Interactive Intelligent Systems and Techniques, IIST 2024*, 160–166. https://doi.org/10.1109/IIST62526.2024.00101.

Part III

Case Studies and Comparative Analyses

13 Perception of ChatGPT by University Students in Poland

Piotr Pietrzak

13.1 Introduction

The rapid development of artificial intelligence (AI) and natural language processing (NLP) technologies has significantly transformed how information is accessed, processed, and utilized in educational settings. Many authors have written about the challenges and opportunities of using AI and NLP in higher education institutions (HEIs), e.g. Demartini et al., 2024; Katsamakas et al., 2024; Kshetri and Voas, 2024 Kuleto et al., 2021; Michel-Villarreal et al., 2023; Wang et al., 2023. Among the most notable advancements in this field is ChatGPT (*Chat Generative Pre-Trained Transformer*), an AI-powered conversational agent developed by OpenAI (Ray, 2023). As an AI tool capable of generating human-like responses and engaging in complex dialogues, ChatGPT has garnered widespread attention for its potential applications in various domains, including education (Farrokhnia et al., 2024). In this context, understanding how university students perceive and utilize such technologies is crucial, particularly as these tools become more prevalent in academic environments.

In Poland, where digitalization in higher education is advancing rapidly (Rosak-Szyrocka et al., 2023), students' attitudes toward AI tools like ChatGPT can provide valuable insights into the readiness of educational institutions to integrate these technologies into their learning processes (at all levels of study). Exploring students' perceptions can help identify both the opportunities and challenges associated with using AI in academia, such as enhancing learning experiences, improving study efficiency, and fostering critical thinking skills, as well as concerns related to privacy, data security, and ethical implications.

This study aims to examine the perception of ChatGPT by university students in Poland. By doing so, the research seeks to contribute to a deeper understanding

DOI: 10.4324/9781003586944-16

of the role that AI technologies can play in shaping the future of higher education in Poland. In addition, one research hypothesis was stated:

- H_1: The assessment of ChatGPT attributes is not influenced by the level of studies.

ChatGPT is designed to provide information, answer questions, and assist with tasks in a way that is broadly applicable to a wide range of users, regardless of their academic level. Whether a student is an undergraduate, or graduate, the core functionalities of ChatGPT – such as providing explanations, generating content, or assisting with research – remain the same. This universal approach can make the perceived attributes of ChatGPT, like versatility, adaptability, availability and scalability, interactivity, user-friendly interface, and conversational ability.

It should be kept in mind that the findings of the study may offer practical recommendations for educators and policymakers on effectively integrating AI tools like ChatGPT into the academic setting to maximize their benefits while addressing potential challenges.

This chapter was organized into six sections. Following the introduction, there was a literature review on the application of AI tools in higher education, with a specific emphasis on ChatGPT. The third section outlined the research methodology. The fourth section focused on the data analysis. In the fifth section, the study's implications were discussed, along with the limitations of the research and potential directions for future studies. This chapter concluded with a final summary in the last section.

13.2 Review of Literature

AI technologies are reshaping the landscape of higher education by offering innovative solutions to enhance learning (Jia & Tu, 2024) and improve student engagement (Nguyen et al., 2024). AI-driven tools, such as intelligent tutoring systems (Rybina & Grigoriev, 2023), adaptive learning platforms (Tretow-Fish & Khalid, 2023), and automated grading systems (Matthews et al., 2012), are helping educators tailor their teaching methods to meet the diverse needs of students. urthermore, AI technologies are enhancing administrative processes like managing enrollment, allocating resources, and facilitating communication with students, enabling HEIs to function more efficiently (Katsamakas et al., 2024).

Currently, one of the most widely used AI tools in higher education is ChatGPT. Utilizing NLP, ChatGPT can hold meaningful conversations, respond to questions, offer explanations, and even produce creative content. This versatility allows it to support multiple areas within higher education, including teaching (Avila et al., 2024), or research (Nachalon et al., 2024). However, it is worth noting that ChatGPT is not only used by educators/researchers, but also by students. Therefore, the use of ChataGPT in the learning process will be presented in the following section.

ChatGPT has the potential to play a pivotal role in enhancing learning in HEIs. Its capabilities can be utilized in several key areas:

- Personalized learning assistance. ChatGPT can act as a personalized tutor, delivering customized explanations and support across various subjects. Unlike conventional learning materials, it provides immediate, interactive responses to student questions, helping to clear up confusion, strengthen comprehension, and encourage independent learning (Chan & Hu, 2023; Sallam et al., 2023).
- Improved accessibility and inclusivity. ChatGPT can make education more accessible by offering round-the-clock support to students, regardless of their location or schedule. This is especially advantageous for non-traditional learners, such as working professionals or students with disabilities, who may need flexible learning options (Ayala, 2023; Esplugas, 2023).
- Support for language learning and communication skills. ChatGPT's proficiency in understanding and generating text in multiple languages makes it an effective tool for language learning. It can engage students in real-time conversations, provide feedback on language use, and suggest improvements, thereby enhancing language acquisition and communication skills. Moreover, it can assist international students in overcoming language barriers and integrating more smoothly into academic life (Lee et al., 2023; Qu & Wu, 2024).

The key characteristics (features/ attributes) of ChatGPT include: (1) versatility (ChatGPT can be used for a wide range of tasks, such as answering questions, generating content, or summarizing text); (2) adaptability (ChatGPT can tailor its responses according to the context of the conversation, making it appropriate for a broad spectrum of topics, and complexities, from straightforward questions to more detailed discussions); (3) availability and scalability (ChatGPT is available 24/7, providing continuous support without downtime); (4) interactivity (ChatGPT facilitates lively interactions, allowing users to ask additional questions, resolve uncertainties, and delve deeper into topics, thereby enriching the learning experience); (5) user-friendly interface (ChatGPT is designed to be easy to use, requiring no specialized knowledge to interact with, which makes it accessible to a broad range of users/students); and (6) conversational ability (ChatGPT is built to participate in natural, human-like conversations, enabling it to grasp context, provide suitable responses to various questions, and sustain the flow of dialogue over several exchanges). The features of ChatGPT have been extensively analyzed by: Acosta-Enriquez et al., 2024; Chan & Lee, 2023; Chellappa & Luximon, 2024; Espartinez, 2024; Singh et al., 2023; and

Tiwari et al., 2024. The highlighted properties were also examined in a study conducted by the author.

In addition to the benefits of ChatGPT, it is important to consider its weaknesses and limitations. Among them, the following can be emphasized: inaccuracy of information (Ulla et al., 2023); forged citations and reference (Branum & Schiavenato, 2023); bias in results (Kooli, 2023); and lack of motivational improvement in challenging tasks (Yilmaz & Karaoglan Yilmaz, 2023).

13.3 Methodology

13.3.1 The Instrument /Survey

The research utilized the CAWI (*Computer-Assisted Web Interviewing*) method, which is a type of survey that the respondent completes online, "using a computer or mobile device" (Balińska et al., 2024, p. 6213). This method is often employed in market and opinion research because it enables the rapid and convenient collection of data from a substantial number of respondents (Sowa et al., 2015).

The initial design of the questionnaire was informed by prior studies (e.g. Chan & Hu, 2023), and pre-existing surveys on students' attitudes toward educational technologies in higher education. To enhance the clarity and relevance of the questions, a pilot study was carried out before the main data collection (June 2024). Feedback from this preliminary phase was used to refine and adjust the questionnaire. The final version included 12 items using a 5-point Likert scale, ranging from "Strongly agree" to "Strongly disagree." The features (attributes) of the ChatGPT were verified, i.e., versatility, adaptability, availability and scalability, interactivity, user-friendly interface, and conversational ability. In addition, the limitations of ChatGPT were assessed.

The survey was addressed to students at all levels of study: undergraduate and graduate (both full-time and part-time). A convenience sampling method was used to choose respondents based on their accessibility and willingness to participate in the study. The survey link was distributed via online platforms, including social media. The snowball sampling technique was utilized to gather participants (Vincent & Thompson, 2022). Respondents were encouraged to share the questionnaire link on their own social media profiles. Participation was entirely voluntary, and all responses were kept anonymous. A proper study was conducted from June to July 2024.

13.3.2 Participants

After a two-month survey (June-July 2024), a total of 326 undergraduate and graduate students, from various disciplines of five universities in Poland, completed the survey. Among respondents, 178 (54.6%) were men and 148 (45.4%) were women. There were (69.9%, n = 228) undergraduate students and (30.1%, n = 98) graduate students. The majority of respondents lived in large cities with more than 500,000 inhabitants (58.9%); 16.2% of respondents lived in rural areas; residents of cities of 50,000 to 100,000 people made up 10.7% of the sample; residents of cities with up to 20,000 inhabitants made up 8.6%; residents of cities of 20,000–50,000 people made up 5.6% of the sample. None of the respondents indicated that they live in a city with a population of over 100,000 but below 500,000 residents. Additionally, 77.0% participants (251 students) have reported using ChatGPT in the general context (not specifically for learning) at least once. Specifically, 20.9% (68 students) reported rarely using it, 28.2% (92 students) using it sometimes, 20.5% (67 students) often using it, and 7.4% (24 students) reported always using it.

13.4 Results

Students who indicated that they had used ChatGPT at least once ($n = 251$) were asked to rate 12 statements on a scale from one to five. The first six statements were related to the features of this tool (versatility, adaptability, availability and scalability, interactivity, user-friendly interface, and conversational ability), while the next six concerned its potential limitations.

As illustrated in Table 13.1., participants had a generally good understanding of features of ChatGPT, with mean scores ranging from 4.17 to 4.78 (for undergraduate students), and from 4.20 to 4.70 (for graduate students). Specifically, both undergraduate and graduate students had the highest mean score for the statement "I think AI technologies such as ChatGPT is a great tool as it is available 24/7" and the lowest mean score for the statement "I think AI technologies such as ChatGPT can grasp the context of my queries, provide appropriate responses to various questions, and maintain the flow of dialogue."

To verify H_1 (*The assessment of ChataGPT attributes is not influenced by the level of studies*), the Mann-Whitney Z test was used. This is a non-parametric statistical test used to compare differences between two independent groups. It assesses whether the distributions of the two groups are different from each other (Fadeikina et al., 2019). In other words, the result is a *p-value*, which helps determine whether the differences observed between the groups are statistically significant.

There was no statistically significant difference between the ratings given by undergraduate and graduate students regarding the different features of ChatGPT. The values of the Z test statistic and the *p-value* were as follows: versatility ($Z = 0.702$, $p = 0.483$); adaptability ($Z = 0.702$, $p = 0.483$); availability and scalability ($Z = 0.205$, $p = 0.388$); interactivity ($Z = -1.140$, $p = 0.254$); user-friendly interface ($Z = 0.029$, $p = 0.977$); conversational ability ($Z = 0.292$, $p = 0.770$).

Table 13.1 Differences in the evaluation of ChatGPT attributes based on respondents' level of study

Attributes of ChatGPT	*Level of study*	*Mean*	*Median*	*Standard deviation*	*Z*, **p**
I think AI technologies such as ChatGPT can help me with wide range of tasks, such as answering questions, generating content, or summarizing text (Versatility)	Undergraduate Graduate	4.33 4.55	4.50 5.00	0.75 0.80	Z = -1.009 p = 0.313
I think AI technologies such as ChatGPT can provide me with personalized and immediate feedback and suggestions for my assignments (Adaptability)	Undergraduate Graduate	4.67 4.40	5.00 5.00	0.47 0.80	Z = 0.702 p = 0.483
I think AI technologies such as ChatGPT is a great tool as it is available 24/7 (Availability and Scalability)	Undergraduate Graduate	4.78 4.70	5.00 5.00	0.42 0.56	Z = 0.205 p = 0.388
I think AI technologies such as ChatGPT allow me to ask additional questions, resolve my uncertainties, and explore topics in-depth, thereby enriching my educational experience (Interactivity)	Undergraduate Graduate	4.33 4.45	4.00 5.00	0.47 0.86	Z = -1.140 p = 0.254
I think AI technologies such as ChatGPT have a user-friendly interface, which makes them easy to use (User-friendly interface)	Undergraduate Graduate	4.56 4.55	5.00 5.00	0.50 0.50	Z = 0.029 p = 0.977
I think AI technologies such as ChatGPT can grasp the context of my queries, provide appropriate responses to various questions, and maintain the flow of dialogue (Conversational ability)	Undergraduate Graduate	4.17 4.20	5.00 4.00	1.07 0.75	Z = 0.292 p = 0.770

Z – Mann-Whitney test result, p – statistical significance of the test result, significant result at $p < 0.05$.

Source: Own elaboration based on the conducted research.

Based on the obtained results, H_1 was confirmed.

In the next stage, respondents were asked to evaluate six statements regarding the limitations of ChatGPT. All of these were taken from the study by Chan and Hu (2023): "I understand generative AI technologies like ChatGPT have limitations in their ability to handle complex tasks"; "I understand generative AI technologies like ChatGPT can generate output that is factually inaccurate"; "I understand generative AI technologies like ChatGPT can generate output that is out of context or inappropriate"; "I understand generative AI technologies like ChatGPT can exhibit biases and unfairness in their output"; "I understand generative AI technologies like ChatGPT may rely too heavily on statistics, which can limit their usefulness in certain contexts"; "I understand generative AI technologies like ChatGPT have limited emotional intelligence and empathy, which can lead to output that is insensitive or inappropriate."

Regarding the limitations of ChatGPT, both undergraduate and graduate students had a fairly good understanding. The average scores for the individual statements ranged from 3.95 to 4.67 (average scores across the entire group of respondents). Specifically, students had the highest mean score for the statement "I understand generative AI technologies like ChatGPT can generate output that is out of context or inappropriate" (*mean* = 4.67), and the lowest mean score for the statement "I understand generative AI technologies like ChatGPT have limited emotional intelligence and empathy, which can lead to output that is insensitive or inappropriate" (*mean* = 3.95).

13.5 Discussion

13.5.1 Contributions

The study filled knowledge gaps in multiple areas. First, the application of AI in higher education, especially ChatGPT, is a relatively new phenomenon and, consequently, academic research on the subject (especially empirical studies) is limited. Specifically, there are a limited number of studies on this topic involving Polish students. Among them are: Strzelecki, 2023; Strzelecki, 2024; Strzelecki et al., 2024; Ziemba et al., 2024. Second, the survey sheds light on students' perceptions of particular attributes of the ChatGPT, such as versatility, adaptability, availability and scalability, interactivity, user-friendly interface, and conversational ability. By identifying these features, the study offers valuable information for educators and policymakers aiming to integrate AI technologies into educational settings more effectively. It also highlights potential areas for improvement in AI tools to better meet students' needs and enhance their learning experiences.

13.5.2 Limitations

The limitations of the conducted study pertain to the adopted assumptions and the research methods used. In the theoretical section of this chapter, certain studies that might be considered crucial by some scholars in relation to the topic under discussion could have been overlooked. The author, in selecting the literature, prioritized sources based on their accessibility and the prestige of the journal or academic publisher.

Nevertheless, most of the limitations relate to the methodology and results sections. First, the use of convenience sampling made it impossible to generalize the results. Since participants were selected based on their accessibility and willingness to partake in the study, the sample may not have been representative of the broader population of university students in Poland. Second, the snowball sampling technique helped in reaching a larger number of respondents by encouraging participants to share the survey link, it may also have introduced bias. This method often relies on the social networks of initial respondents, which could lead to a homogenous group of participants with similar characteristics or views, reducing the diversity of the sample. Third, the study was conducted exclusively among university students in Poland, which limits its applicability to other geographic or cultural contexts. The perceptions and attitudes of students in different countries or regions may vary, and the findings cannot be generalized globally without further cross-cultural research. Finally, the study was conducted over a relatively short period (June–July 2024), which may have impacted the number of participants and the depth of responses.

13.5.3 Future Research Directions

Future studies should aim to include a larger and more diverse sample, possibly using random sampling techniques to capture a more representative picture of university students' perceptions of ChatGPT across different countries and types of institutions. Another promising direction for future research would be to broaden the focus to include various AI tools and their integration into higher education, extending beyond just ChatGPT. This could help in understanding the overall impact of AI technologies on learning, teaching, and administration in HEIs. Finally, in-depth qualitative research could be conducted to explore students' ethical concerns about the use of AI tools, such as ChatGPT, in education. This could include focus groups or interviews to gain deeper insights into their attitudes.

13.6 Conclusions

This study investigated Polish students' (both undergraduate and graduate) perceptions of ChatGPT. Based on the results obtained, the following conclusions could be drawn:

- 77.0% participants have reported using ChatGPT in the general context.
- On average, the highest rated features of ChatGPT were availability and scalability, and the lowest was conversational ability.

- Both undergraduate and graduate students rated all the features of ChatGPT similarly.
- Both undergraduate and graduate students had a good understanding of the limitations of ChatGPT; among them, the highest rated was the generation of out-of-context responses, incorrect.

In summary, the author effectively achieved the primary objective of the study. In turn, the research hypothesis (*the assessment of ChataGPT attributes is not influenced by the level of studies*) was confirmed.

The author considers this study to be a foundational piece that can stimulate further academic dialogue and exploration.

References

Acosta-Enriquez, B. G., Arbulú Ballesteros, M. A., Huamaní Jordan, O., López Roca, C., & Saavedra Tirado, K. (2024). Analysis of college students' attitudes toward the use of chatGPT in their academic activities: Effect of intent to use, verification of information and responsible use. *BMC Psychology*, *12*(1), 255.

Avila, K. E., Steinert, S., Ruzika, S., Kuhn, J., & Küchemann, S. (2024). Using chatGPT for teaching physics. *Physics Teacher*, *62*(6), 536–537.

Ayala, S. (2023). ChatGPT as a universal design for learning tool supporting college students with disabilities. *Educational Renaissance*, *12*, 22–41.

Balińska, A., Jaska, E., & Werenowska, A. (2024). The importance of the vinted application in popularizing sustainable behavior among representatives of generation Z. *Sustainability*, *16*, 6213.

Branum, C., & Schiavenato, M. (2023). Can chatGPT accurately answer a PICOT question? assessing AI response to a clinical question. *Nurse Educator*, *48*, 231–233.

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, *20*, 1–18.

Chan, C. K. Y., & Lee, K. K. W. (2023). The AI generation gap: Are gen Z students more interested in adopting generative AI such as chatGPT in teaching and learning than their Gen X and millennial generation teachers?. *Smart Learning Environments*, *10*(1), 60.

Chellappa, V., Luximon, Y. (2024). Understanding the perception of design students towards chatGPT. *Computers and Education: Artificial Intelligence*, *7*, 100281.

Demartini, C. G., Sciascia, L., Bosso, A., & Manuri, F. (2024). Artificial intelligence bringing improvements to adaptive learning in education: A case study. *Sustainability*, *16*, 1347.

Espartinez, A. S. (2024). Exploring student and teacher perceptions of chatGPT use in higher education: A Q-methodology study. *Computers and Education: Artificial Intelligence*, *7*, 100264.

Esplugas, M. (2023). The use of artificial intelligence (AI) to enhance academic communication, education, and research: A balanced approach. *Journal of Hand Surgery*, *48*(8), 819–822.

Fadeikina, O. V., Volkova, R. A., & Karpova, E. V. (2019). Statistical analysis of results from the attestation of biological standard samples: Use of the mann-whitney test. *Pharmaceutical Chemistry Journal*, *53*(7), 655–659.

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of chatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, *61*(3), 460–474.

Jia, X. H., & Tu, J. C. (2024). Towards a new conceptual model of AI-enhanced learning for college students: The roles of artificial intelligence capabilities, general self-efficacy, learning motivation, and critical thinking awareness. *Systems*, *12*, 74.

Katsamakas, E., Pavlov, O. V., & Saklad, R. (2024). Artificial intelligence and the transformation of higher education institutions: A systems approach. *Sustainability*, *16*, 6118.

Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability*, *15*(7), 5614.

Kshetri, N., & Voas, J. (2024). Adapting to generative artificial intelligence: Approaches in higher education institutions. *Computer*, *57*(9), 128–133.

Kuleto, V., Ilić, M., Dumangiu, M., Ranković, M., Martins, O. M. D., Păun, D., & Mihoreanu, L. (2021). Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions. *Sustainability*, *13*, 10424.

Lee, M., Contreras Alejandro, J., & IJsselsteijn, W. (2023). Cultivating gratitude with a chatbot. *International Journal of Human-Computer Interaction*, *17*, 1–16.

Matthews, K., Janicki, T., He, L., & Patterson, L. (2012). Implementation of an automated grading system with an adaptive learning component to affect student feedback and response time. *Journal of Information Systems Education*, *23*(1), 71–84.

Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by chatGPT. *Education Sciences*, *13*, 856.

Nachalon, Y., Broer, M., & Nativ-Zeltzer, N. (2024). Using chatGPT to generate research ideas in dysphagia: A pilot study. *Dysphagia*, *39*(3), 407–411.

Nguyen, A., Kremantzis, M., Essien, A., Petrounias, I., & Hosseini, S. (2024). Enhancing student engagement through artificial intelligence (AI): Understanding the basics, opportunities, and challenges. *Journal of University Teaching and Learning Practice*, *21*(6), 1–13.

Qu, K., & Wu, X. (2024). ChatGPT as a CALL tool in language education: A study of hedonic motivation adoption models in english learning environments. *Education and Information Technologies*, *29,* 1–33.

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154.

Rosak-Szyrocka, J., Apostu, S. A., & Akkaya, B. (2023). Higher education and digitalization in perspective of use of internet, integration of digital technology, digital public services: Panel study of EU nations. Scientific Papers of Silesian University of Technology, *181*, 469–491.

Rybina, G. V., & Grigoriev, A. A. (2023). Modern architectures of intelligent tutoring systems based on integrated expert systems: Features of the approach to the automated formation of the ontological space of knowledge and skills of students. *Pattern Recognition and Image Analysis*, *33*(3), 491–497.

Sallam, M., Salim, N., Barakat, M., & Al-Tammemi, A. (2023). ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive study highlighting the advantages and limitations. *Narra J*, *3*(1), 103.

Singh, H., Tayarani-Najaran, M. H., & Yaqoob, M. (2023). Exploring computer science students' perception of chatGPT in higher education: A descriptive and correlation study. *Education Sciences*, *13*(9), 924.

Sowa, P., Pędziński, B., Krzyżak, M., Maślach, D., Wójcik, S., & Szpak, A. (2015). The computer-assisted web interview method as used in the national study of ICT use in primary healthcare in poland - reflections on a case study. *Studies in Logic, Grammar and Rhetoric*, *43*(56), 137–146.

Strzelecki, A. (2023). To use or not to use chatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments*, *32*(9), 1–14.

Strzelecki, A. (2024). Students' acceptance of chatGPT in higher education: An extended unified theory of acceptance and use of technology. *Innovative Higher Education*, *49*(2), 223–245.

Strzelecki, A., Cicha, K., Rizun, M., & Rutecka, P. (2024). Acceptance and use of chatGPT in the academic community. *Education and Information Technologies*, *29*, 1–26.

Tiwari, C. K., Bhat, M. A., Khan, S. T., Subramaniam, R., & Khan, M. A. I. (2024). What drives students toward chatGPT? An investigation of the factors influencing adoption and usage of chatGPT. *Interactive Technology and Smart Education*, *21*(3), 333–355.

Tretow-Fish, T. A. B., & Khalid, M. S. (2023). Methods for evaluating learning analytics and learning analytics dashboards in adaptive learning platforms: A systematic review. *Electronic Journal of E-Learning*, *21*(5), 430–449.

Ulla, M. B., Perales, W. F., & Busbus, S. O. (2023). "To generate or stop generating response": exploring EFL teachers' perspectives on chatGPT in english language teaching in Thailand. *Learning*, *9*, 168–182.

Vincent, K., & Thompson, S. (2022). Estimating the size and distribution of networked populations with snowball sampling. *Journal of Survey Statistics and Methodology*, *10*(2), 397–418.

Wang, T., Lund, B. D., Marengo, A., Pagano, A., Mannuru, N. R., Teel, Z. A., & Pange, J. (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*, 6716.

Yilmaz, R., & Karaoglan Yilmaz, F. G. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, *4*, 100147.

Ziemba, E., Maruszewska, E. W., Grabara, D., & Renik, K. (2024). Acceptance and use of ChatGPT among accounting and finance higher education students. In *Proceedings of 26th European Conference on Artificial Intelligence ECAI 2023 IMIS 2023 Workshop*, 185–202.

14 The Potential of Using Solutions Based on Generative Artificial Intelligence in Agriculture in European Union Countries

Henryk Runowski and Paulina Kramarz

14.1 Introduction

Socio-economic development is determined mainly by the effects of technological progress. Groundbreaking changes in this area are referred to as successive industrial revolutions. Historically, the first industrial revolution was driven by mechanical production powered by steam and water. This was followed by the second industrial revolution, characterized by mass production enabled by electricity and fossil fuels. The third industrial revolution then introduced computer-controlled and automated production. The current era of technological progress, characterized by the integration of digital technologies, artificial intelligence (AI), robotics, and automation in various industries and sectors, is called the Fourth Industrial Revolution (Lund, 2021). It is characterized by the fusion of physical and virtual reality in the so-called homogeneous cyber-physical system.

The term Fourth Industrial Revolution was first used at the Hanover Fair in 2011 as part of the presentation of the German government's high-tech strategy project, promoting the computerization of manufacturing processes (Da XU et al., 2018). It is characterized by using advanced digital technologies, including AI, combined with automation to transform production processes into more autonomous and efficient ones. This applies to industry and other sectors of the economy, including agriculture, which is responsible for feeding the growing world population (Runowski, 2023). It is predicted that by 2050, the world's population will increase from the current eight to almost ten billion, which means a significant increase in the demand for food. In meeting this task, it will be helpful to use the achievements of the Fourth Industrial Revolution, including AI and generative AI. This justifies an attempt to determine the potential of using generative AI in agriculture. Due to the complexity of the problem, the area of analysis will be limited to the countries of the European Union (EU). EU agriculture is characterized by different levels of agricultural development, depending on the countries. In some EU countries, agriculture still uses traditional methods of farming to a large extent, while, in other countries, we are dealing

DOI: 10.4324/9781003586944-17

with very modern agriculture using the latest available technologies, including digital technologies and generative AI.

This chapter aims to indicate the potential for implementing technologies based on generative AI in agriculture in individual EU countries. They differ in terms of the level of socio-economic development, factors determining the formation of a culture of trust and the state of agriculture. According to Eurostat data in the European Union, in 2018, the share of people employed in agriculture, forestry, and fisheries who use computers, laptops, smartphones, tablets, or other portable devices at work – which is necessary to operate some technologies using generative AI – amounted to 29% and was enormously diversified in individual EU countries (from 11% in Greece to 88% in the Netherlands). These factors, combined with technological advancement, the level of digital skills, and the shape of the agricultural development policy pursued in individual countries, may differentiate the EU countries regarding the potential for implementing solutions based on AI in agricultural activities. This chapter consists of the following elements: a review of the literature on the concept, types, and possibilities of using AI in the economy and agriculture, a discussion of research methodology, a presentation of the results of statistical analysis and results from the empirical study, discussion of results, and conclusions.

14.2 AI – Concept, Types, Legal Regulations, Use in Economy and Agriculture

The initial interest in AI appeared several decades ago. The first research works on AI are associated with two scientists (McCulloch and Walter Pitts), who in the early 1940s proposed using artificial neural networks, which were modeled on the human way of thinking (McCulloch & Pitts, 1943). The term AI was first used in 1956 during the "Summer Research Project on Artificial Intelligence at Dartmouth" conference by John McCarthy (Russell et al., 2010). In the 1950s, McCarthy wrote: "… force machines to use language, create abstractions and concepts, solve types of problems currently reserved for humans, and improve themselves" (McCarthy et al., 1955). Later, McCarthy defined AI as "the science and technology of producing intelligent machines" (McCarthy, 2007). AI is "the simulation of human intelligence in machines, enabling them to perform tasks and make decisions that typically require human intelligence" (Kok et al., 2009). Other authors define AI as "the ability of a system to correctly interpret external data, learn from it, and use these insights to adapt to achieve specific goals and tasks flexibly" (Kaplan & Haenlein, 2019). According to the definition by the European Parliament, "artificial intelligence refers to systems that exhibit intelligent behaviour, analyzing their environment and taking action – with a certain degree of autonomy – to achieve specific goals" (Boucher, 2020). Due to the high dynamics of research on AI, many different proposals for defining this concept appear, and new ones are constantly being developed. A common feature of most definitions is that AI perceives the surrounding environment, processes the information, and makes autonomous decisions to achieve a goal (Samoili et al., 2021).

Intensive work on these problems lasted until the end of the 1970s. Then, until the 1990s, a pessimistic approach to the potential of AI became visible (Francesconi, 2022). However, at the end of the 1990s, a noticeable development of computer technology and network structures was noted. These significant advances were due, among other things, to the increase in the computing power of computers. An additional impetus was provided by progress in cloud computing and big data (Chan et al., 2022). The development of AI is taking place at a rapid pace. AI is one of the strategic technologies of the 21st century, both in Europe and worldwide, which brings positive changes to the world economy, increasing innovation, productivity, competitiveness, and prosperity. It stimulates innovation and underpins new business models, pivotal in social transformation and the digitalization of economies. AI is widely recognized as a critical factor of technological change in the Fourth Industrial Revolution. Hence, it attempts to systematically organize the conditions for the functioning of AI and its use in the socio-economic processes of individual countries and their groups.

14.2.1 Legal Conditions for the Development of AI in the EU

Recognizing the need for legal regulations for the development of AI, the European Commission published the Digital Europe program for 2021–2027 on June 6, 2018 (Digital Europe, 2018), and on February 12, 2019, the European Parliament adopted a comprehensive European industrial policy in the field of AI and robotics (European Parliament, 2019). Subsequently, in April 2021, the European Commission proposed the first EU legislative framework for AI. The project analyzes and classifies AI systems that can be used in various applications according to the risks they pose to users. On March 13, 2024, the European Parliament adopted a law regulating AI. This is the world's first concrete attempt to regulate AI, establishing harmonized rules for developing, marketing, and using AI in its territory. The Act aims to ensure the safety of AI systems in the EU and respect for fundamental rights and values. It also aims to promote investment and innovation in AI, improve governance and law enforcement, and enhance the EU's internal market for AI (Digital Europe, 2018). It is a priority for the European Parliament to ensure that AI systems used in the EU are safe, transparent, traceable, non-discriminatory, and environmentally friendly.

14.2.2 Generative AI

AI has made significant progress recently, especially in generative AI. Generative AI is a subset of AI in which machines create new content in the form of text, code, voice, images, videos, processes, and predictions (Aydın & Karaarslan, 2023; Banch & Strobel, 2023). The fundamental difference between traditional AI and generative intelligence is that the latter can create novel content that appears to have been generated by humans. Coherent substantive arguments, texts, and realistic images that arouse public and business interest are examples of Gen AI models. These models provide data in a way previously only possible through the involvement of the human mind and its creativity. The progress achieved in generative AI means these systems can solve previously considered unsolvable problems (Bubeck et al., 2023). While recognizing the potential for new solutions and forms of access to information, it is also important to note the risks associated with increased disinformation and the erosion of trust in digital content. Another central concern is that individuals may not know whether they are interacting with a human or a machine, which may influence their behavior (March, 2021; Natale, 2021).

The ultimate goal should be to harness the potential of generative AI in a way that promotes human development while maintaining a balance between technological progress and societal well-being. AI can also help automate the fact-checking process, aiding the dissemination of accurate information (Hoes et al., 2023; Zovolokina et al., 2024).

Generative AI has the potential to expand access to information, but it also poses significant challenges, such as misusing data, falsifying data, changing human-machine interactions, and spreading disinformation. However, generative AI has many benefits across a variety of industries. The use of generative AI in agriculture may be particularly beneficial. The agricultural sector has always been at the forefront of technological advances, from mechanization to precision farming. Now, generative AI has the potential to revolutionize the way agricultural businesses operate, communicate, train, and innovate. It suffices to highlight only a selection of the available functions of rapidly advancing technology that can be applied in agriculture. (Sathishkumar, 2024):

1. Generative AI systems can imitate texts written by humans. Some translation programs use generative AI models, making it easier to break down language barriers in communication between people who speak different languages. This improves communication capabilities and makes tracking international trends in production, technology, marketing, or the agri-food market easier. Generative AI tools can respond to prompts, creating high-quality content on a given topic.
2. Speech and music generation. Using written text and audio samples of a person's voice and voice tools, AI can create narration or songs that imitate the sounds of real people. This can be used in contact with animals that respond to the voice of a person they know well, who, for some reason, needs to be replaced periodically. This reduces animal stress due to changing the staff who handle them.

3 Generating images of AI computer tools can synthesize high-quality images of agricultural production systems, animal behavior, plant condition, or their response to agrotechnical treatments.

4 Generating movies and new experimental services with generative AI techniques to create moving graphics.

5 Generating a large amount of synthetic data, their processing, suggesting decisions regarding the production process, and also in situations where accurate data is difficult or practically impossible. Modern agriculture, for its management, requires a large amount of data that can be used after subjecting them to specific synthesis processes, mutual connection, and creation of a model that reflects actual agricultural processes in both plant and animal production. This is especially important when an incomplete data set can be expanded with an additional set of synthetic data for the required purposes.

6 Improving agricultural education with AI-generated content. AI can generate lesson plans and course materials. This allows for improving agricultural education by creating customized lesson plans and course materials. Generative AI, such as Typetone, can analyze the needs and interests of the audience and create content tailored to their specific learning styles. Creating virtual field trips or providing interactive learning experiences.

7 Generative AI can also create virtual field trips and interactive learning experiences that immerse students in real-world agricultural scenarios. AI generates engaging, dynamic content and helps students understand complex agricultural concepts and techniques.

8 Using AI-generated content in precision farming, making decisions based on such information.

9 AI-generated content can help farmers decide based on current crop and animal health information, weather conditions, and market conditions. By processing large amounts of data, generative AI can generate valuable information that allows farmers to optimize operations and maximize yields while considering environmental protection, climate, and animal welfare principles.

10 Generative AI saves time and makes it easier for farmers to perform administrative and reporting work by automatically generating reports on the use of resources, crop structure, the number of individual species and groups of animals, the performance of crops and livestock, calculating financial results, creating applications, business plans and reports. This not only improves production processes on farms and optimizes the use of production resources but also allows the opportunity to share them with interested stakeholders.

11 AI-generated content can be used to create personalized marketing campaigns. By analyzing data about customers or consumers, generative AI

can understand their preferences and generate content that appeals to them, increasing the effectiveness of farmers' marketing activities.

12 Generative AI can facilitate collaboration among agricultural researchers, conducting research combining multiple stakeholders' knowledge and insights. This can help speed up the research process and lead to agricultural technology and practice breakthroughs.

13 Also, companies in the broad sense of agribusiness can benefit significantly from generative AI in their operations and improve the conditions for good communication between them and farmers.

14.3 Methodology

14.3.1 The Instrument/Survey

The phenomenon was assessed based on the analysis of data collected by Eurostat, the information contained in EU directives and EU documents presenting the policy towards the use of AI in agriculture. The results of empirical research on implementing the empirical research objective were also used. The survey questionnaire method was also used to obtain research material. Statistical methods were used in the analysis of the obtained results. First, the level of socio-economic development of the 27 EU countries was determined. The zero unitarization method was used, making calculations according to the following formula (Kukuła, 2000):

$$Q_{i=\frac{1}{k}} \sum_{k}^{j=1} u_{ij}, \text{ where}$$

k – number of diagnostic features,

u_{ij} – the value of the jth standardized feature for the ith subregion according to the formula,

$$u_{ij} = \begin{cases} \dfrac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}}, X_j \in S \\ \dfrac{\max x_{ij} - x_{ij}}{\max x_{ij} - \min x_{ij}}, X_j \in D \end{cases}$$

In order to define the combination of socio-economic development factors that support the adaptation of digital technologies – including tools based on generative AI, a synthetic socio-economic development indicator was defined for the 27 European Union countries. The choice of variables was determined by their availability, which was incomplete for some data. After statistical verification, the following seven variables were used:

- GDP PPS per capita;
- Percentage of persons at risk of poverty or social exclusion;
- Healthy life years in absolute value at birth;
- Gross domestic expenditure on R&D – percentage of gross domestic product;
- Net greenhouse gas emissions – tons per capita;
- Percentage of farms with an economic size below 4,000 euros;
- Mobile internet access – percentage of individuals used a laptop, notebook, netbook, or tablet computer to access the internet away from home or work.

Based on the calculated value of the synthetic indicator, a ranking of the 27 EU countries was built according to the level of socio-economic development, considering four groups of countries designated using the quartile method. In addition to socio-economic development, three additional indicators were adopted for comparisons between countries, which may determine the potential for the use of digital technologies, including generative AI, in individual EU countries:

- Use of Information and Communication Technology (ICT) at work – individuals in rural areas (%);
- Broadband internet coverage by speed More than one gigabit per second (% of households);
- Individuals working in agriculture, forestry, or fishing who used computers, laptops, smartphones, tablets or other portable devices at work (%).

In addition to research on the potential of using generative AI, surveys were conducted among farmers in southern Poland (Małopolskie and Podkarpackie provinces) regarding farmers' trust in AI. These studies were conducted as part of the MINIATURA7 scientific activity, financed by the National Science Center in Poland. It was recognized that trust in AI is a primary condition for its subsequent use in production processes and the functioning of farms.

Participants: Data from 27 EU member states were used in the research on the EU. As part of the empirical research, research was conducted among a representative group of 389 farmers running farms in southern Poland (Małopolskie and Podkarpackie provinces). Among the farmers surveyed, 36% were women and 64% were men. People over 50 dominated -50% of the respondents. It reflects the demographic characteristics of farmers from the surveyed region. The leading direction of production was plant production (in the case of 80% of farms). The average area of the surveyed farms was 8.4 ha.

Procedure: The survey was conducted in July 2024 in randomly selected 389 farms. Farm managers provided the answers to the questions. The survey was entirely anonymous. It was conducted using the PAPI method, in direct contact with respondents. Paper questionnaires were used.

14.4 Results

In the ranking that considered the socio-economic development level for the 27 EU countries, Sweden came first with an index value of 0.771, and Lithuania last with an index value of 0.172 (Table 14.1). When dividing countries into four groups of socio-economic development levels, the boundaries were set using quartile values: $Q1$ = 0.385, $Q2$ = 0.434, $Q3$ = 0.585, $Q4$ = 0.771. The group of countries with the highest level of socio-economic development included Sweden, Belgium, Denmark, Germany, Austria, France, and the Netherlands. In this group of countries, the percentage of households with access to fast internet was, on average, 78%, and in countries such as Denmark and the Netherlands, it exceeded 90%. This was accompanied by a 47% share of rural residents using ICT technologies in their professional activity. Even though in the remaining socio-economic development groups, there were countries providing citizens with high access to fast internet (including Malta and Luxembourg), the percentage of people using ICT in rural areas that accompanied this possibility was lower. In group two, which included countries such as Ireland, Finland, Luxembourg, Slovenia, Malta, and Czechia, it was 39%. At the same time – despite the lack of data available for all the countries mentioned, one can notice in this group that a smaller share of people working in agriculture, forestry, or fishing who used computers, laptops, smartphones, tablets, or other portable devices at work. Groups two and three, distinguished by the level of socio-economic development, had a percentage of around 30%. The percentage of people from rural areas using ICT at work was significantly lower in the third group, which included Hungary, Spain, Italy, Poland, Portugal, Slovakia, and Greece. It amounted to an average of 22%.

At the same time, the percentage of households with access to fast internet was at a similar level as in group 2–63%. In these countries, only 19% of individuals working in agriculture, forestry, or fishing use computers, laptops, smartphones, tablets, or other portable devices at work. The group with the lowest level of socio-economic development included Croatia, Cyprus, Lithuania, Estonia, Bulgaria, Romania, and Latvia. Despite the average lowest availability of fast internet compared to the other groups, the percentage of people from rural areas using ICT at work was at a similar level as in group three. The percentage of individuals working in agriculture, forestry, or fishing who used computers, laptops, smartphones, tablets, or other portable devices at work was higher and amounted to 27%. This may be related to the perception of the role of ICT in

Table 14.1 The level of socio-economic development in EU countries and selected indicators presenting the potential for implementing digital technologies

EU countries	*Synthetic index of socio-economic development*	*Use of ICT at work – individuals in rural areas [%]*	*Broadband internet coverage by speed more than 1 gigabit per second [% of households]*	*Individuals working in agriculture, forestry or fishing who used computers, laptops, smartphones, tablets, or other portable devices at work [%]*
		Group 1		
Sweden	0.771	38.91	81.60	no data
Belgium	0.707	41.06	78.00	60.19
Denmark	0.662	47.73	91.60	no data
Germany	0.640	50.05	68.60	61.50
Austria	0.636	47.77	54.80	46.73
France	0.623	41.15	73.20	65.31
Netherlands	0.607	61.52	97.80	88.04
Average	0.664	46.88	77.94	64.35
		Group 2		
Ireland	0.563	32.42	72.30	23.46
Finland	0.562	46.43	60.00	no data
Luxembourg	0.540	44.16	93.30	no data
Slovenia	0.524	39.21	6.60	no data
Malta	0.508	40.74	100.00	no data
Czechia	0.506	30.51	42.50	31.03
Average	0.534	38.91	62.45	27.25
		Group 3		
Hungary	0.434	19.17	81.90	12.29
Spain	0.427	25.17	86.70	20.50
Italy	0.424	29.72	53.50	32.84
Poland	0.423	19.03	62.20	12.46
Portugal	0.408	25.98	88.40	18.40
Slovakia	0.406	22.32	40.30	27.50
Greece	0.391	14.74	27.90	10.92
Average	0.416	22.30	62.99	19.27
		Group 4		
Croatia	0.379	22.73	57.60	19.37
Cyprus	0.372	26.61	60.00	14.23
Lithuania	0.341	25.60	77.80	28.26
Estonia	0.329	45.02	56.50	65.01
Bulgaria	0.262	9.52	21.40	13.29
Romania	0.219	7.56	91.80	16.35
Latvia	0.172	29.09	0.00	32.47
Average	0.296	23.73	52.16	27.00

Source: own study based on Eurostat data (2018–2022).

facilitating communication, knowledge, and data collection. This may be the first step towards implementing technologies based on AI. It can be assumed that factors associated with shaping socio-economic development create infrastructural conditions for popularizing digital technologies. However, using available tools is determined by social factors combined with knowledge, level of trust, or social awareness.

Poland, Romania, Hungary, Greece, and Bulgaria are among the five EU countries with the lowest percentage of people functioning in rural areas using ICT technologies in their professional activity. The percentage of individuals working in agriculture, forestry, or fishing who used computers, laptops, smartphones, tablets, or other portable devices at work was also among the lowest. It amounted to 12% in Poland, with a minimum of 10% recorded in Greece. The fragmentation of the agrarian structure can explain this. This is confirmed by research conducted by Stępień et al., (2023) in Poland, Romania, and Lithuania, which showed that in these countries, the use of AI technology is rare, and the main reason for not making attempts in this area is the low-scale of production and the too small size of many farms.

A survey conducted in the Małopolskie and Podkarpackie provinces – the most fragmented areas of Poland, where the possibilities of increasing the production area due to the high fragmentation of land and the diversification of its quality are also limited, showed that the scale of application of solutions based on AI is small. Among the 389 farms surveyed, with an area from 1 ha to 390 ha, only 5% used sensors to collect real-time data, which could then be processed by generative AI. However, a tendency to use software for data processing and forecasting was revealed – applications supporting decision-making and farm management systems in the cloud (27%). However, the problem of providing these systems with the most up-to-date, complete data remains. The trust of the surveyed farmers in AI can be considered high – it was declared by 49% of the respondents. This group was dominated by farmers up to 35 years of age (77%). There was no clear correlation between the declared level of trust and the direction and level of education – among those trusting AI were 49% of farmers with agricultural education and 51% with non-agricultural education. Respondents showing trust in AI as the main barrier to implementing solutions, such as systems enabling automation of crop production based on AI actions, indicated financial constraints – 72%. In the group of respondents who did not trust AI (34%) or were undecided (17%), the main barrier was the lack of need for implementation – 72% of indications. In this group, a financial barrier was revealed by 41% of respondents. Farmers trusting AI indicated the lack of need to implement crop production automation systems based on AI action at the level of 20%. The source of information used by farmers may also be important in shaping trust in AI. Among the five most important sources of information disclosed by the surveyed farmers, those trusting AI indicated materials presented in social media by other farmers (89%), internet forums (83%), social media groups (67%), websites of institutions supporting the development of agriculture (56%), agricultural manuals and guides (52%). Among respondents who did not trust AI and were undecided, the dominant sources of information were direct conversation with other farmers (57%), television programs on agriculture (45%), materials presented on social media by other farmers (31%), agricultural manuals, guides (29%), and internet forums (28%).

14.5 Discussion

AI, including generative AI, is increasingly entering various areas of socio-economic life (Bhat & Huang, 2017; Garske et al., 2021; Runowski & Kramarz, 2025). This brings with it both opportunities and threats. Without the necessary infrastructural, social, and legal support, AI may encounter development barriers and deepen the economic divide between countries (Yu, 2020). AI systems should be adequately supervised to prevent possible harmful effects of their use (Härtel, 2020). Multimedia content created or processed using AI, such as images, audio, or video files, should be marked as AI-generated content so that recipients know and interpret such materials appropriately. Using generative AI raises questions about copyright, liability, and data protection. These issues have not yet been finally clarified. Among others, the following questions arise. Can generative AI easily use copyrighted content, such as texts from the Internet? Who is ultimately the author of new generative content? (Scheufen, 2023). Another question: Who is responsible for the spread of disinformation if it is generative? Perhaps, for these reasons, potential users of systems based on AI are concerned about the need to provide data (Eastwood et al., 2023; Linsner et al., 2021).

Lack of trust may translate not only into the inclination to use technologies but also into their effectiveness. Research conducted in Ireland has shown that the problem with learning the principles of operation of various technological solutions was significantly greater among people who lack confidence in technology (Irish Farm Center 2019). The digitalization of agriculture is part of the country's policy, and the level of support is essential in popularizing agricultural digital technologies and shaping and trusting them (Smidt Yokonya, 2021). Raising farmers' knowledge level is also essential (Dibbern et al., 2024; Dhillon et al., 2023; Gebresenbet et al., 2023). FAO distinguishes the inclusion of ICT in the educational system among the factors creating the conditions for using digital solutions (Trendov et al., 2019). ICT undoubtedly also facilitates the expansion of access to information. According to Caffaro et al. (2020), they can affect the perceived usefulness of digital technologies. He distinguished informal and formal sources of information, assuming that perceived usefulness is increased by formal sources and decreased by informal sources. It is worth noting, however, that respondents from the Małopolskie and Podkarpackie voivodeships who expressed trust in AI used information presented by other farmers in social media more often than others – i.e. from an informal source, and at the same time, a high percentage of them declared using formal sources of knowledge. Using social media to search for information, they may have at their disposal a broader picture of farmers' experiences and greater knowledge about how technology works in practice, and modern technology can provide material in the form of video recordings or enable direct contact with farmers from even distant regions (Kaur, 2022; Kramarz & Runowski, 2025). The lower tendency to trust AI in some countries or regions may also be due to the low popularity of modern technological solutions and the lack of contact with them (Runowski & Kramarz, 2025). In addition, the willingness of individuals to share experiences is also conditioned by trust (Grudzewski et al., 2007; Paliszkiewicz, 2013). This may be difficult in areas with a low level of social trust, which concerns, among others, the countries of Central and Eastern Europe affected by difficult historical conditions.

14.6 Conclusion

The potential for generative AI in developing sustainable agriculture in the European Union is significant, although it varies between countries. To develop generative AI in agriculture, it is necessary to provide IT infrastructure, including broadband Internet, trust in this technology, and a legal framework to convince farmers and consumers of the benefits that may result from it. Technical and organizational data protection measures for AI must be improved to ensure the development of model conditions for access rights to agricultural data and to limit the risk of damage caused by AI in agriculture. EU product safety law and liability law require reform.

Implications of the Study: A higher level of socio-economic development provides opportunities for broader use of AI, thanks to extensive infrastructure, a higher level of expenditure on research and development, better availability of digital tools, and increased access to infrastructure. Therefore, the differentiation of the level of socio-economic development of EU countries will deepen the inequalities between them in terms of the possibility of using digital technologies. Future possibilities of using AI in agricultural practice are, however, determined mainly by social acceptance and the level of knowledge of farmers about the functioning of technologies based on their operation. Where the possibility of using technologically advanced digital tools is challenging – e.g. in countries with a low level of socio-economic development, the first step towards digital inclusion may be an interest in ICT technologies. They facilitate improving digital skills, equipping farms with computer equipment, or greater access to knowledge and information. Social media, which is rich in information about the practice of agricultural technologies, is also an element of ICT.

Limitations of the Study and Future Directions: The conducted analysis is limited by the low availability and currency of databases. This results from the initial development phase of AI. Statistical data are not able to reflect intermittent changes related to social phenomena. The presented results of the questionnaire study refer to one of the regions of Poland (with high agrarian fragmentation), which cannot explain all aspects of shaping trust in AI on a European scale. However, because small farms constitute a significant percentage of all EU farms, the presented results may help plan further research on this subject – especially in countries with a similar level of socio-economic development to Poland. Further studies should also include a broader analysis of the impact of sources of information on trust in AI, the role of disinformation in this area, and the barriers to its use perceived by farmers. Recognizing the possibilities of supporting agriculture with elements of digital technologies – including those based on the operation of AI in those regions that are less developed may be of particular importance.

References

Aydın, Ö., & Karaarslan, E. (2023). Is chatGPT leading generative AI? What is beyond expectations?. *Academic Platform Journal of Engineering and Smart Systems*, *11*(3), 118–134.

Banch, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, *33963*, 1–17.

Bhat, S. A, & Huang, N. (2017). Big data and AI revolution in precision agriculture: Survey and challenges. IEEE Access, *9*(99), 110209–110222. https://doi.org/10.1109/ACCESS.2021.3102227.

Boucher, P. (2020). *Artificial intelligence: How does it work, why does it matter, and what we can do about it?*. European Parliamentary Research Service.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, S., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. Retrieved from: https://www.researchgate.net/publication/369449949_Sparks_of_Artificial_General_Intelligence_Early_experiments_with_GPT-4#fullTextFileContent

Caffaro, F., Cremasco, M. M. Roccato, M., & Cavallo, E. (2020). Drivers of farmers' intention to adopt technological innovations in Italy: The role of information sources, perceived usefulness, and perceived ease of use. *Journal of Rural Studies*, *76*, 264–27.

Chan, L., Hogaboam, L., & Cao, R. (2022). *Applied artificial intelligence in business: Concepts and cases.* Springer International Publishing.

Da Xu, Li., Xu, E. L., & Li, L. (2018). Industry 4.0: State of the art and future trends. *International Journal of Production Research*, *56*(8), 2941–2962.

Dhillon, R., Moncur, Q., Lowell, C., Kumaran, S., Folck, A., & Cao, D. (2023). Precision Agriculture (PA) techniques for smallholder farmers in the US: Status and potential opportunities. Proceedings of the National Conference on Next-Generation Sustainable Technologies for Small-Scale Producers. *Springer Nature, 34*, 166–175.

Dibbern, T., Santos Romani, L. A., Silveira, S., & Massruh, M. F. (2024). Main drivers and barriers to the adoption of digital agriculture technologies. *Smart Agricultural Technology 8*, 1–10.

Digital Europe programme for the period 2021–2027. (2018). Retrieved from https://www.eumonitor.eu/9353000/1/j4nvhdfcs8bljza_j9vvik7m1c3gyxp/vkp1fqrgymox.

Eastwood, C., Turner, J. A., Selbie, D., Henwood, R., Espig, M., & Wever, M. (2023). A review of multi-scale barriers to transitioning from digital agriculture to a digital bioeconomy. Retrieved from https://www.cabidigitallibrary.org/doi/full/10.1079/cabireviews.2023.0002.

European Parliament. (2019). European parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics (2018/2088(INI)). Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52019IP0081.

Eurostat. (2018–2022). Database - Eurostat. Agriculture, forestry and fisheries. https://ec.europa.eu/eurostat/web/main/data/database

Francesconi, E. (2022). The winter, the summer and the summer dream of artificial intelligence in law. *Artificial Intelligence and Law*, *30*(2), 147–161.

Garske, B., Bau, A., & Ekardt, F. (2021). Digitalization and AI in European agriculture: A strategy for achieving climate and biodiversity targets?. *Sustainability*, *13*(9), 1–21.

Gebresenbet, G., Techane, B. D. Patterson, P. Henrik, B. Fischer, N. Mandaluniz, G. Chirici, A. Zacepins, V. Komasilovs, T., Pitulach, T., & Nasirahmadi, A. (2023). A concept for application of integrated digital technologies to enhance future smart agricultural systems. *Smart Agricultural Technology*, *5*, 1–12.

Grudzewski, W. M., Hejduk, I. K., Sankowska, A., & Wańtuchowicz, M. (2007). *Zarządzanie zaufaniem w organizacjach wirtualnych*, [*Trust Management in Virtual Work Environments*]. Difin.

Härtel, I. (2020). Künstliche intelligenz in der nachhaltigen landwirtschaft – datenrechte und haftungsregime. *AUFSÄTZE*, *42*, 439–453.

Hoes, E., Altay, S., & Bermeo, J. (2023). Leveraging chatGPT for efficient fact-checking. Retrieved from https://osf.io/preprints/psyarxiv/qnjkf_v1.

Farm Centre 2019 Digital Agriculture Technology. Adoption & Attitudes Study. Retrieved from https://www.ifa.ie/wp-content/uploads/2020/11/Digital-Ag-Tech-Adoption-Attitudes.pdf.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri in my hand, who is the fairest in the land? *On the interpretations, illustrations and implications of artificial intelligence. Business Horizons*, *62*(1), 15–25.

Kaur, A. (2022). Agricultural communication: A theoretical perspective. *International Journal of Advanced Mass Communication and Journalism*, *3*(1), 73–77.

Kok, J. N., Boers, E. J., Kosters, W. A., & Van der Putten, P. (2009). Artificial intelligence: Definition, trends, techniques, and cases. *Artificial Intelligence*, *1*, 270–299.

Kramarz, P., & Runowski, H. (2025). Trust and communication in agriculture. In Paliszkiewicz J., Chen K., & Mendel M. (Eds.), *Trust in social and business relations: Theory and practice*, Routledge.

Kukuła, K. (2000). *Metoda unitaryzacji zerowanej*, [*Zero unitarization method*]. PWN.

Linsner, S., Kuntke, F., Steinbrink, E., Franken, J., & Reuter, Ch. (2021). The Role of Privacy in digitalization – analyzing perspectives of German farmers. *Proceedings on Privacy Enhancing Technologies*, *3*, 334–350.

Lund, D. B. (2021). The fourth industrial revolution. *Information Technology and Libraries, 40*(I), 1–4.

March, C. (2021). Strategic interactions between humans and artificial intelligence: lessons from experiments with computer players. *Journal of Economic Psychology*, *87*, 102426.

McCarthy, C. (2007). From here to human-level AI. *Artificial Intelligence*, *171*(18), 1174–1182.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A proposal for the dartmouth summer research project on artificial intelligence. Retrieved from https://dl.acm.org/doi/10.1609/aimag.v27i4.1904

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133.

Natale, S. (2021). *Deceitful media: Artificial intelligence and social life after the turing test*. Oxford University Press.

Paliszkiewicz J. O. (2013). *Zaufanie w zarządzaniu*, [*Trust in management*]. Wydawnictwo Naukowe PWN.

Runowski, H., & Kramarz, P. (2025). Trust in artificial intelligence in agriculture. In Paliszkiewicz J., & Gołuchowski J. (Eds.), *Trust and artificial intelligence: Development and application of AI technology*, Routledge.

Runowski, H. (2023). Możliwości wzrostu produkcji żywności, [*Possibilities of increasing food production*]. In Szymańska E. J. (Ed.), *Współczesne tendencje w rozwoju rolnictwa i obszarów wiejskich*. IERiGŻ-PIB, SGGW.

Russell, S. J., Norvig, P., & Davis, E. (2010). Artificial intelligence: A modern approach. *Artificial Intelligence*, *175*, 935–937.

Samoili, S., Lopez Cobo, M., Delipetrev, B., Martinez-Plumed, F., Gomez Gutierrez, E., & De Prato, G. (2021). *AI Watch. Defining Artificial Intelligence 2.0*. Publications Office of the European Union.

Sathishkumar, P. (2024). Generative AI in agriculture: Revolutionizing crop management. Retrieved from https://www.bitsathy.ac.in/blog/generative-ai-in-agriculture-revolutionizing-crop-management/.

Scheufen, M. (2023). Generative KI und autorenrechte. *IW-Kurzbericht*, *79*, 1–3.

Smidt, H. J., & Yokonya, O. (2021). Factors affecting digital technology adoption by small-scale farmers in agriculture value chains (AVCs) in South Africa. *Information Technology for Development*, *28*(3), 558–584.

Stępień, S., Smędzik-Ambroży, K., Polcyn, J., Kwiliński, A., & Maican, I. (2023). Are small farms sustainable and technologically smart? Evidence from Poland, Romania, and Lithuania. *Central European Economic Journal*, *10*(57), 116–132.

Trendov, N. M., Varas, S., & Zeng, M. (2019). *Digital technologies in agriculture and rural areas. Briefing paper*. Food and Agriculture Organization of the United Nations, Rome.

Yu, P. K. (2020). The algorithmic divide and equality in the age of artificial intelligence. *Florida Law Review*, *72*, 331–389.

Zavolokina, L., Sprenkamp, K., Katashinskaya, Z., Jones, D. G., & Schwabe, G. (2024). Think fast, think slow, think critical: Designing an automated propaganda detection tool. *Proceedings of the CHI Conference on Human Factors in Computer Systems* (CHI ‘24). Association for Computing Machinery.

15 Trust in Artificial Intelligence – Generational Differences

Katarzyna Łukasiewicz, Elżbieta Kacperska and Jakub Kraciuk

15.1 Introduction

In recent years artificial intelligence (AI) has revolutionized many areas of our lives. It is contributing to rapid technological progress, while also raising questions related to trust in this technology. Trust in AI is not uniform and varies depending on several factors, such as age, previous experiences with technology, or belonging to a specific generational group. Generations growing up during different technological eras will consequently be diverse in how they perceive both opportunities and threats connected with the development and use of AI. Older generations, such as Baby Boomers or representatives of Generation X, were growing up in times when the digital transformation was only starting and AI technologies were at the stage of theoretical concepts or very few simple applications. Many of these individuals still perceive AI as a novelty, something unknown and thus potentially dangerous. Frequently users, representatives of different generations, express their concerns related to privacy, data security, or even the potential of people being replaced by machines in the labor market. Trust in technology in those generational groups is often based on knowledge, education, as well as the level of transparency and usefulness of technology. On the other hand, younger generations, such as Millennials and Generation Z, were brought up in the age of the Internet, smartphones, as well as commonly used automation systems. For them, AI is a natural element of the surrounding technological world, which plays a key role in their everyday lives – starting from personalized recommendations in various applications, to smart voice assistants and learning support tools, etc. As a rule the younger generations exhibit greater trust in AI (in the context of entertainment and applicability), although they may also be more aware of threats related to the protection of privacy in the digital sphere. This chapter aims at presenting an in-depth analysis of diverse attitudes to AI depending on the generation, while attempting to answer the following questions: Why do representatives of certain age groups trust technology more than it is observed in other generational groups? What factors play a role when determining the level of trust? It is crucial to understand these differences in terms of diverse perspectives (sociological, psychological), but also

DOI: 10.4324/9781003586944-18

for firms and institutions involved in the implementation of the AI technology. It is necessary to adapt the method of communication and designing of AI systems to take into account preferences and concerns of each of the diverse generational groups. Older generations expect AI first of all to be predictable, safe, and comprehensible, whereas younger generations focus more on functionality, convenient use, and tremendous innovativeness. These differences offer huge opportunities for creators and users of AI systems, who need to find methods to build trust, adapting their products and services to the diverse needs of end-users. In order to properly understand and manage social potential for the implementation of AI, it is necessary to take into consideration these generational differences in the approach to trust. The conducted analysis of the disparity in the generational trust in AI will indicate what methods of implementation of this technology may enhance its acceptance among the general public and how to design systems to be more universal, transparent, and adapted to the needs of various generational groups.

15.2 Trust in Technology – A Review of Literature

Trust is a complex, multifaceted, and multidimensional category (Castelfranchi & Falcone, 2010; Faulkner & Simpson, 2017; Hosmer, 1995). It is described by representatives of various fields of science, psychology, economics, sociology, management science, and the humanities (Boddington, 2017; Kramer & Tyler, 1996; Paliszkiewicz 2013;Sztompka 2007). Based on findings described within these disciplines, trust may clearly be defined as a belief that actions of the other party will be consistent with the expectations of the trusting party (trustor).

Approaches to the understanding of trust may be divided into at least four categories, treating trust as (1) a personality trait, (2) an individual expectation or belief, (3) the foundation for interpersonal relationships, and (4) the foundation for economic and social cooperation (Paliszkiewicz, 2019).

Within the first category, we may distinguish definitions presented by Wrightsman (1966) and Luhmann (1979), stating that trust is a personality trait, which is reflected in general expectations concerning intentions of other individuals. This concept was also interpreted by Rotter (1967), who acknowledged that trust is a personality trait, which reflects the general expectation concerning reliability of other people. In turn, Gibba (1978) indicated that trust is instinctive and as a feeling it resembles love.

The second category comprises definitions presented, e.g. by such authors as Sako (1992), Lewicki and Bunker (1996), Mayer et al., (1995), and Das and Bing-Sbeng (1998). For example, Sako (1992) described trust as a state of mind, expectations in relation to a partner that he or she will behave in a predictable and mutually acceptable manner. In turn, Lewicki and Bunker (1996) stated that trust is a positive expectation concerning motivations of other people. Mayer et al., (1995) defined trust as the willingness of the trusting party to be dependent on actions of another person, based on the expectation that the trustee will behave in a predictable manner from the point of view of the trustor regardless of the potential to monitor or control the trustee. In turn, Das and Bing-Sbeng (1998) presented trust as a positive expectation in relation to motivations for actions in situations burdened with a risk.

At present, in the era of tremendous technological progress, analyses are conducted focused on trust in technology, which is manifested in people's readiness to be affected by technology, because they need to use it, as well as predict and diagnose the resulting threats (Ejdys 2017). According to McKnight et al. (2011), the concept of trust refers to the belief that the other party in the relationship (here: technology) will act in a predictable and reliable manner, ensuring positive results.

Trust in the context of technology, such as AI, may be defined as the conviction of users that systems based on AI will act in a predictable, reliable, and ethical manner, which is consistent with the expectations of users (McKnight et al., 2002). Trust is dynamic and may be modified by experiences of users, transparency of systems, as well as communications between users and technology (Mayer et al., 1995).

Among the multitude of factors influencing trust in AI we may distinguish first of all transparency and explainability. Users are more willing to trust systems, which are capable of explaining their decisions and actions in a comprehensible manner (Doshi-Velez & Kim, 2017). The capability to explain ideas within the framework of AI promotes the building of trust, because it makes it easier for users to understand the mechanisms regulating the operation of AI systems (Gunning et al., 2017). Among factors of the influence, we may also distinguish credibility and reliability. Credibility in relation to AI refers to its ability to provide accurate and cohesive results. Reliability of AI systems is crucial for the building of trust, particularly in critical applications such as medicine or finance (Lee & See, 2004). It is indicated that greater credibility and reliability of AI systems are reflected also in an enhanced trust of users (Zhang & Dafoe, 2019). Among the abovementioned elements we may also point to ethics and safety. Trust vested by users is also influenced by the ethical aspects of AI, such as avoiding bias or prejudice, protection of privacy, as well as ensuring data security. Users expect AI systems to operate in accordance with generally accepted social values and ethical standards (Jobin et al., 2019). Incidents connected with breach of privacy or algorithmic discrimination may considerably reduce the level of trust in AI (O'Neil, 2016). We also need to consider the aspect related to the human-machine interaction. It will also affect trust. An intuitive user interface or personalization of user experiences may increase the level of trust (Lankton et al., 2015). Good experiences related to interactions with AI promote a positive perception of this technology.

AI is perceived differently by various authors. According to John McCarthy, it is "a discipline of art and engineering connected with designing intelligent machines, particularly intelligent computer programs. Intelligence is the computational aspect of the capability to realize objectives in the real world" (McIlwraith et al., 2017). According to Duch, AI is defined as a discipline of art concerned with effectively solving non-algorithmic problems based on models of knowledge (Duch, 1997, p. 54). According to Tadeusiewicz (2020, p. 27), we deal with AI when a machine (computer or electronically controlled device: robot, autonomous vehicle, self-organizing network) exhibits behaviors, which when observed in a human would be considered a consequence of his or her intelligence.

Key components of AI include (Piecuch, 2023):

- machine learning (ML), which is a subset of AI requiring no explicit software. It learns automatically based on previous experiences. Accuracy of ML is increasing with time and the amount of data;
- deep learning (DL) is characterized by deep processing of data using artificial neural networks (NMs). DL is a subset of ML.
- NN is a system designed to process information, which structure and principle of action to a certain extent imitate the functioning of fragments of an actual (biological) nervous system;
- natural language processing is a tool facilitating communication with humans. It consists of the recognition, understanding and interpreting a natural language. Communication may be executed in the text form and in speech generation;
- computer vision. Visual processing uses the DL process. It facilitates distinguishing diverse graphic patterns such as tables, images, graphs, and video;
- cognitive computing, which consists of learning, understanding tasks and interpretation of data. This group includes identification of images and processing of a natural language.

The digital transformation process poses a huge developmental challenge for the present-day society. All services, both public and commercial, have to be data-intensive to fit in the AI era. Gathering, accumulation, analysis, processing, and continuous development of AI algorithms are becoming a fundamental competence of economies and states (Polityka, 2020). In the nearest future, we may expect a dramatic increase in applications for AI solutions in all areas of our lives and the national economy.

15.3 Methodology

The main aim of this chapter was to indicate differences in the level of trust in AI in various generational groups – from Baby Boomers to Generation Alfa – and identify factors influencing the perception and acceptance of AI in the context of everyday life.

The following research hypotheses were proposed in this study:

Table 15.1 Structure of the population of respondents

Generation	*Division in terms of age*	*Number of responses*	
		N	*w %*
Baby Boomers	1946–1964 – individuals aged 60–78 years	33	5.8
Generation X	1965–1979 – individuals aged 45–59 years	120	21.2
Generation Y Millennials	1980–1994 – individuals aged 30–44 years	85	15.0
Generation Z	1995–2009 – individuals aged 15–29 years	302	53.5
Generation Alfa	2010–2024 – individuals aged 0–14 years	25	4.4
Total		565	100

Source: the authors' study based on research.

H1. Younger generations such as Generation Y, Z, and Alfa exhibit a higher level of trust in AI than older generations, i.e. Baby Boomers and Generation X.

H2. Older generations exhibit greater concerns related to the effect of AI on the job market and on privacy, which results in a lower level of trust in this technology.

H3. Younger generations are more willing to cooperate with AI used in daily life compared to older generations (BB and X).

H4. A higher standard of technological education and frequent experiences with tools based on AI in younger generations are correlated with a higher level of trust in this technology.

H5: Older generations, to a greater extent than younger ones, underline the need to implement legal regulations related to the development of AI.

A survey questionnaire was the tool used when collecting data. The CAWI (Computer-Assisted Web Interviewing) method was applied. The questionnaire consisted of 20 questions. The survey was conducted in September 2024 in Poland among 565 respondents (Table 15.1).

Questions contained in the questionnaire made it possible to investigate generational differences in terms of knowledge, concerns, and trust in AI, as well as verify areas, in which various age groups may have different perspectives on the subject. The first research area consisted in the analysis of knowledge, respondents' experiences and opinions on AI. The first question aimed at determining the level of basic knowledge on AI in order to specify how accurately respondents understand the concept itself and the functions of this technology. Another aspect was connected with their experiences in using AI-based tools, such as voice assistants (Alexa, Siri) or systems recommending films and chats. Frequent use of these tools may indicate greater trust in AI and increased willingness to integrate this technology into one's daily life. Respondents were also asked about areas, in which they see the greatest potential for AI applications, including the future use of AI in education or transport.

The second research area included questions related to concerns about the application of AI in everyday life. Questions asked were aimed at identifying key threats, perceived by respondents in connection with the development of AI. General areas of concern were investigated, such as potential negative consequences of development of this technology, which might raise fears in various aspects of everyday existence. In particular, the questions referred to threats to privacy, which is connected with concerns over security of personal data. Additionally, analyses focused on potential abuse, such as disinformation or attempts to influence elections, which points to uncertainty related to the impact of AI on the society. Questions also referred to the need to introduce legal regulations, which could control the development of AI and prevent these threats. The last aspect was connected with concerns related to the potential replacement of humans by AI in certain professions, which raises fears over future employment and the labor market.

The last research area comprised questions concerning trust in AI. They aimed at an assessment of the attitudes of respondents representing different generations toward this technology, as well as their willingness to accept it in various aspects of their lives. The general attitude to AI was investigated to provide indications whether respondents perceive it as a positive or negative phenomenon. An important element was also connected with the assessment of whether AI may have a positive impact on the future, thus indicating their convictions concerning benefits of the development of this technology. An important aspect of this part of the questionnaire was investigated in questions related to the opinions on the capacity of AI to make better decisions than humans in certain areas. Additionally, respondents were asked whether they see differences between generations in their trust in AI, particularly in the context of older and younger generations.

15.4 Results and Discussion

In the survey concerning awareness and trust in AI as many as 98.8% of respondents declared that they know what AI is. It is a definite indication that AI is a commonly identified technology and we can observe high awareness of this phenomenon. Moreover, 61.7% respondents stated that they have adequate knowledge of AI, which suggests that a considerable proportion of the surveyed population not only knows basic terminology referring to this technology, but also understands its functioning and potential applications in various areas of life. These results may show a growing interest and education concerning AI in the general public. However, responses varied between the individual generations. Respondents born in the years 1946–1964, i.e. Baby Boomers, were the only group who declared that they have no idea what AI is. Such a response was given by slightly below 25% respondents, while over 50% heard about AI, but knew little about the subject. Over 63% respondents representing Generation X also declared limited knowledge on AI. In turn, responses provided by representatives of Generations Y, Z, and Alfa predominantly declared good knowledge of AI. Moreover, in the successive generations, the declaration of good familiarity with AI technology was increasingly common, as for Generation Y it was 56%, Generation Z – 75.5%, and Generation Alfa – 80%, respectively. The younger generations were growing up in a more technologically advanced world, as a result, they feel more comfortable using new technologies, such as AI. This analysis shows that 72% of respondents are of an opinion that the older generations exhibit lesser trust in AI compared to the younger age groups. When analyzing individual generations, it was found that apart from Generation Alfa (60%) the indications exceeded 70%, while among Baby Boomers it was as high as 80%. Results of this study indicate that the younger generations show greater optimism concerning the influence of AI on the future. Among respondents from Generation Alfa, as many as 76% were of an opinion that AI may have a positive impact on the future, while in Gen Z it was 83%. Slightly lower, still high values were recorded among representatives of Generations Y (68.8%) and X (62.5%). A marked decrease in positive opinions may be observed in the older age groups, as among individuals born in the years 1946–1964 (Baby Boomers), only 55% of respondents expressed similar opinions. These results suggest that acceptance and enthusiasm for the development of AI decrease in progressively older generations. Similar findings were presented in their study by Łapińska et al., (2020).

These results confirm hypothesis H1 assuming that younger generations (Millennials, Generation Z, Generation Alfa) exhibit a higher level of trust in AI than it is the case in older generations (Baby Boomers, Generation X).

It results from the analysis of data that older generations show greater concerns about the influence of AI on everyday life. Over 60% respondents from the Boomers population declared having concerns related to AI, while additional 27% declared a similar sense of unease while, at the same time, they admitted that they are not fully confident about their assessment. Among respondents from Generation X 44% declared they are concerned, with another 44% stating they have concerns, but they were not fully certain about their opinion. In the group of respondents aged below 14 years old, only 28% expressed concerns, which confirmed hypothesis 2 concerning lesser concerns among the younger generations.

In turn, older generations believe that AI may become a threat for privacy. This opinion is declared by 81% of respondents being Baby Boomers and 91% representing Gen X, but similar responses were given by respondents from Generation Y at 87% and Generation Z at 83%, respectively. The lowest level of fear in that respect was indicated by respondents from Generation Alfa, as it was declared by 76% of them. The greatest concerns related to the development of AI were connected with excessive control the technology has over our lives – here these responses were predominant in all the generations. Concerns about privacy and data security ranked second. Loss of employment raised the greatest fears among representatives of Generations Z at 60% and X at 50.6% responses, whereas they were not as prevalent in the youngest and the oldest groups (both groups are pre- and post-working age populations).

The second hypothesis was confirmed. Older generations frequently perceive AI as a threat to privacy and data security, which may in turn be manifested in their skeptical attitude to new technologies.

In contrast, younger generations show greater trust in decisions made by AI in everyday situations, such as online shopping or suggested itineraries. High trust was declared by 61% of respondents representing Generation Y, 69% from Generation Z, and 64% from Generation Alfa. Among the older generations the percentage was lower, amounting to 40% in Generation X and 36% among Baby Boomers. Such decisions are not trusted by as many as 48% of respondents being Baby Boomers, 28% of respondents from Generation X, 17% from Generation Y, 9% from Generation Z, and 12% from Generation Alfa, respectively.

Only 12% of respondents from Generation Alfa and 6% from Generation Z would definitely choose AI as an advisor when making personal decisions, while it was 9% among Baby Boomers. In contrast, AI is chosen as an additional advisor by as many as 60% of respondents from Generation Alfa, 46% from Generation Z, 50.6% among Millennials, at a decreasing share among respondents from Generation X with 44% declarations and 39.4% among Baby Boomers.

A similar trend may be observed for responses concerning openness to cooperation with AI in professional life. Younger generations exhibit greater readiness to cooperate with AI, both in the entire and limited extent. In Generation Y willingness for such a cooperation is declared by 65% of respondents, in Generation Z it is 76%, while in Generation Alfa, it is 56%, while 32% of respondents in that group stated that they are not certain of their position. In older generations, this percentage was lower, at 56.6% in Generation X and only 33.4% among Baby Boomers. Younger generations are accustomed to using AI in everyday activities, such as online shopping, suggested itineraries, or consumer preferences, which increase their acceptance of this technology also in the professional context, thus confirming the third hypothesis H3, similarly as it was with hypothesis H4. Younger generations more often use recommendations of AI and instructions of voice assistants, which are manifested in building greater trust in these technologies. As many as 76% of respondents from Generation Z use voice assistants, similarly to 64% from Generation Y. In Generation X this percentage amounts to 56%, whereas among Baby Boomers it was only 33%. For Generation Alfa, although its representatives also use these technologies, it was 56%, which may be connected to limited access to electronic devices in that age group.

Representatives of Generation Alfa in 48% responses declared that they have greater knowledge on AI compared to other people in their age group, which constitutes the highest result among all the generations. In the other generational groups the dominant response indicating a similar level of knowledge as that of their peers was given by 72% of respondents in Generation Z, 57% in Generation Y, 55% in Generation X, and 30.3% among Baby Boomers. What is more, in the latter group 39.4% of respondents stated that they have less knowledge on AI compared to most of their peers.

The analysis of data showed that over 70% of Baby Boomers, Generations X, and Y definitely agree that it is necessary to introduce adequate legal regulations in order to control the development of AI. Among Generation Z, this is declared by 55% of respondents, while, in Generation Alfa, it is as little as 28%. In individual generations, respondents also indicated the need to implement respective regulations, but only in selected areas. This analysis confirmed hypothesis H5. Older generations perceive AI as a potential threat and thus they support proposals to introduce greater controls and regulation of this technology in order to limit the risk of abuse, as opposed to the general opinion expressed by younger generations.

There are many threats concerning the use of AI, as indicated by the respondents. A similar opinion was also expressed by representatives of the largest tech companies. Bill Gates indicates threats posed by AI (Rawlinson, 2015). Elon Musk is also of an opinion that people should be concerned about AI (Gomez, 2021). In her study, Królikowska (2022) pointed to the aspect of threats resulting from the use of AI, presenting results of research conducted in a group of young people. Opinions on trust in AI were also formulated by Piecuch (2023).

15.5 Conclusion

At present, the use of AI is practically a necessity, which brings many benefits in various areas, but at the same time, its applications indicate many challenges related to information safety. The following issues need to be focused on (Frączek & Spaliński, 2023):

1. Cybersecurity;
2. Protection of data and privacy;
3. Disinformation;
4. Excessive dependence on AI;
5. Ethical application of AI;
6. Responsibility and accountability.

This study concerning awareness and trust in AI (SI) indicates a high level of recognizability of this technology among the general public. As many as 98.8% of respondents declared that they know what AI is, while 61.7% assessed their knowledge on the subject as good. It needs to be stated here that the level of knowledge and trust in AI varies among generations. Younger generations, such as the Millennials, Generation Z, and Generation Alfa, show greater trust in AI and more often declare good knowledge concerning this technology. In contrast, older generations, including Baby Boomers and Generation X, are more skeptical and express greater concerns about the influence of AI on daily life, particularly in the context of privacy and data security.

Younger generations more frequently use AI technology in daily activities, such as online shopping or using voice assistants, which was reflected in the greater level of trust in this technology. Respondents from Generations Z (76%) and Y (64%) are more open to cooperate with AI, both in the professional and personal life, whereas older generations are more reserved in this respect.

Conducted investigations indicate its limitations resulting from the fact that for example numbers of representatives in each generational group surveyed in this study were not identical, or differences between generational groups were not compared with representatives of other nationalities. A multinational population would make it possible to indicate many differences, which would be an interesting area for in-depth studies.

Over the years, trust in AI has been changing dynamically. This variability is also observed for factors, which influence this trust. As a result, it is a problem, which needs to be focused on in research and investigated on a regular basis.

References

Boddington, P. (2017). *Towards a code of ethics for artificial intelligence*. Springer.

Castelfranchi, C., & Falcone, R. (2010). *Trust theory: A socio-cognitive and computational model*. John Wiley and Sons, Ltd.

Das, T. K., & Bing-Sbeng, T. (1998). Between trust and control: Developing confidence in partner cooperation in alliances. *Academy of Management Review*, *23*(3), 491–512.

Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*. *arXiv preprint arXiv:1702.08608*. https://doi.org/10.48550/arXiv.1702.08608

Duch, W. (1997). *The fascinating world of computers*. Nakom.

Ejdys, J. (2017). Determinants of trust in technology. *Organizational Review*, *12*, 20–27.

Faulkner, P., & Simpson, T. (2017). *The philosophy of trust*. *Oxford University Press.*

Frączek, M., & Spaliński, K. (2023). Information security in the age of artificial intelligence: Risk analysis and challenges related to chatGPT. *Pro Publico Bono Scientific Papers*, *1*(1), s. 287–302.

Gibb, J. R. (1978). Trust, A new view of personal and organizational development, International College, Guild of Tutors Pre, Los Angeles.

Gomez, B. (2021). Elon Musk warned of a 'Terminator'-like AI apocalypse — now he's building a Tes la robot. CNBC. Pobrane 20 stycznia 2022 Retrieved at https://www.cnbc.com/2021/08/24/elon-musk-warned-of-ai-apocalypsenow-hes-building-a-tesla-robot.html.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G-Z. (2017). XAI explainable artificial intelligence. *Science Robotics*, *4*(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120.

Hosmer, L. T. (1995). Trust: The connecting link between organizational theory and philoso phical ethics. *Academy of Management Review*, *20*(*2*), 379–403.

Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399.

Kramer, R. M., & Tyler, T. R. (Eds.). (1996). *Trust in organizations: Frontiers of theory and research*. Sage Publications.

Królikowska, J. (2022). Perception of artificial intelligence by Poles aged 16–25. An attempt at assessment, [in:] *IT in business.* Dudycz H. (red.), Publishing House of the Wrocław University of Economics, 62–78.

Lankton, N. K., McKnight, D. H., Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, *16*(10), 880–918.

Łapińska, J., Sudolska, A., Gorka, J., Escher, I., Kadzielawski, G., Brzustewicz, P. (2020). Employee trust in artificial intelligence in industrial enterprises operating in poland. Research report. *Institute of Economic Research Reports*. Institute of Economic Research, Olsztyn, *1*, 1–20.

Lee, J. D., See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80.

Lewicki, R. J., Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In w: R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations, frontiers of theory and research* (pp. 114–39). Sage Publications. s.

Luhmann, N. (1979). *Trust and power*. John Wiley.

Mayer, R. C., Davis, J. H., Schoorman, D. F. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(*3*), 709–734.

McIlwraith, D., Marmaris, H., Babenko, D. (2017). *Smart grid. Future algorithms*. Helion.

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2), 12–32.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research, 13*(3), 334–359.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.

Paliszkiewicz, J. (2013). *Trust in management*. Scientific Publishing House PWN.

Paliszkiewicz, J. (2019). *Leadership, trust and knowledge management in innovative enterprises*. Publishing House CeDeWu.

Piecuch, A. (2023). Artificial intelligence in a social perspective. - *Adult Continuing Education*, *4*(123) 13–25.

Policy for the development of artificial intelligence in Poland from (2020). Annex to resolution no. 196 of the Council of Ministers of 28 December 2020 (item 23).

Rawlinson, K. (2015). Microsoft's Bill Gates insists AI is a threat. *BBC News*. z Retrieved from: https://www.bbc.com/news/31047780 (Accessed: 20 January 2024).

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*(4), 651–665.

Sako, M. (1992). *Prices, quality and trust, inter-firm relations in britain & japan*. Cambridge University Press.

Sztompka, P. (2007). *Zaufanie. Fundament społeczeństwa*. Trust. The Foundation of Society.Wydawnictwo Znak.

Tadeusiewicz, R. (2020). *Archipelago of artificial intelligence. part 1. drives and controls.* 12, Wyd. Druk-Art. S.C.

Wrightsman, L. S. (1966). Personality and attitudinal correlates of trusting and trustworthy behaviors in a two-person game. *Journal of Personality and Social Psychology*, *4*(3), 328–332.

Zhang, B., Dafoe, A. (2019). Artificial intelligence: American attitudes and trends. Retri eved from: *SSRN 3312874*.

16 Trust and Integration

AI Adoption and Socio-Technical Systems

Anezka Viskova-Robertson

16.1 Introduction

This chapter examines the relationship between trust and Artificial Intelligence (AI) integration within socio-technical systems (STSs), emphasizing that trust should be an inherent system characteristic, not an afterthought. The integration of AI into organizational processes presents both opportunities for increased efficiency and challenges related to trust, adoption, and the balance between technological advancement and human-centered design within STSs.

As AI becomes increasingly integrated into organizational work processes since the early 2010s, trust and AI adoption have emerged as critical considerations across various sectors. While some industries have embraced AI already, others are only now responding to the post-November 2022 generative AI boom. Current business challenges, such as mass layoffs and pandemic-induced shifts in work environments, have exacerbated leadership struggles, complicating AI integration efforts. Moreover, the growing consolidation of resources within large tech companies raises ethical questions about decision-making in AI development, with rules and regulations still in development. These dynamics have given rise to concerns about the pace and oversight of development, which in turn complicate AI adoption and affect trust in AI systems.

This chapter explores the interplay between STSs, AI integration, and trust. It offers insights into the complex nature of AI adoption and the critical role of trust in facilitating effective and responsible AI integration. The remainder of this chapter is structured as follows: First, we examine the concept of STSs, providing a foundation for understanding the organizational environment in which AI integration occurs. Next, we explore the factors influencing AI integration, including individual perspectives on AI adoption, organizational frameworks for integration, and various modalities of AI implementation. The following section then focuses on trust as a linchpin in the AI integration process. This section elaborates on the multi-layered nature of trust in AI, its role within the STS, and how it shapes attitudes and behaviors toward AI adoption. Finally, the conclusion synthesizes these insights, emphasizing the importance of a balanced

DOI: 10.4324/9781003586944-19

approach that considers both the comprehensive socio-technical framework and targeted interventions for successful AI integration.

16.2 Understanding STSs

The socio-technical systems (STS) framework began in the 1950s at the Tavistock Institute, focusing on balancing social and technical elements in organizations, especially in manufacturing (Trist & Bamforth, 1951). By the 1960s, Emery and Trist formally coined the term and integrated STS with the rise of computerization (Mumford, 1983). In the 1970s, the framework extended into office systems and non-industrial sectors (Cherns, 1987). However, the 1980s and 1990s saw a decline in STS focus due to the rise of lean production and business process reengineering (Clegg, 2000). By the 2000s, with the advent of globalization and network-based organizations, STS re-emerged as companies adopted decentralized structures (Clegg, 2000). In the 2020s, the integration of AI has once again pushed the boundaries of STS theory, renewing its relevance for addressing complex organizational and technological dynamics (Tarafdar et al., 2019).

In this chapter, AI is defined as a "machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments" (OECD, 2019, p. 15). AI encompasses a wide range of specialized branches, each with distinct applications. These include machine learning (ML), which allows computers to learn from data and make predictions or decisions, and natural language processing, enabling machines to interact with human language. Computer vision grants computers the ability to interpret visual information, while robotics combines AI and mechanical engineering to create machines capable of performing tasks in the physical world. Expert systems mimic human decision-making in specific domains, and planning and decision-making strategies enable AI systems to achieve set goals. Lastly, neural networks and deep learning, often considered a subset of ML, use artificial networks inspired by the human brain to handle complex tasks. These branches reflect AI's broad capabilities across various industries and align with traditional AI categories such as symbolic AI and evolutionary algorithms (Galbusera et al., 2019). Understanding AI through these branches provides a clearer view of its potential and challenges. AI can be deployed in different modalities, such as complementary AI, which supports and enhances human capabilities, or substitutional AI, which automates tasks traditionally handled by humans (Viskova-Robertson, 2023).

The STS framework offers a comprehensive approach to understanding and implementing AI integration in organizations. This framework conceptualizes organizations as complex systems composed of interconnected technical and social subsystems. The technical subsystem typically includes dimensions such as technology (including AI systems), processes, and infrastructure. The social subsystem encompasses people, culture, goals, and organizational structure (Viskova-Robertson, 2023). Some versions of the framework are visualized as a hexagonal model with each point housing one of the subsystem's domains, Clegg (2000).

16.3 STS and AI Integration

The integration of AI within the STS framework offers a holistic approach that encompasses the entire organizational environment. This integration affects and is affected by all aspects of the STS, including technology, people, processes, and organizational structures.

Within this framework, employee attitudes and trust toward AI emerge as critical components, playing a pivotal role in the success of AI integration efforts. These attitudes are trifold, comprising cognitive, affective, and behavioral aspects (Rosenberg & Hovland, 1960). They serve as key indicators of an organization's readiness for AI adoption and its potential impact on the system. Trust in AI systems, in particular, acts as a cornerstone that influences people's attitudes toward AI.

The relationship between AI technologies and the various subsystems of the STS is dynamic. As AI influences work processes, decision-making, and organizational structures, existing organizational cultures and employee attitudes simultaneously shape how AI is adopted and utilized. Achieving harmony within an AI-integrated STS depends on how well organizations align AI initiatives with their existing culture and values, viewing AI as a tool to enhance the organization's core mission rather than disrupt it, while also addressing ethical concerns and potential biases inherent in AI systems to ensure responsible and trustworthy integration. Organizational culture emphasizing learning, curiosity, and adaptability creates a system more receptive to AI integration alongside leadership that plays a crucial role in shaping these attitudes, as their openness to technology influences broader acceptance. Supporting this with training programs and safe environments for experimentation enhances integration.

16.4 Additional Factors Influencing AI Integration

While numerous factors affect AI integration across various domains, our discussion will center on three critical areas:

1. The individual perspective on AI adoption and integration, including the stages people may go through from initial resistance to potential enthusiasm, and relevant psychological theories that help explain these transitions.
2. The organizational perspective on AI adoption and integration, outlining common stages of adoption and strategic choices organizations often face.
3. The modalities of AI integration, examining various ways AI can be incorporated into existing systems and workflows, and how these choices impact the overall STS.

16.4.1 Individual Perspective on AI Adoption and Integration

The individual perspective on AI adoption and integration can be more comprehensively understood when viewed as a continuum of resistance rather than discrete stages. This conceptualization aligns with resistance to change theories in organizational psychology and technology acceptance models. The resistance continuum spans from high to low resistance.

16.4.1.1 High Resistance → Moderate Resistance → Low Resistance

High Resistance is exhibited by those who largely renounce AI technology altogether. This is often stemming from cognitive biases, low technological self-efficacy, or fear of change. Moderate Resistance is shown by those who cautiously engage with AI but harbor reservations and skepticism. Their adoption is primarily driven by extrinsic motivators. Low Resistance characterizes those who readily embrace and adopt AI innovations with enthusiasm and curiosity. They are motivated by intrinsic factors and growth. This continuum allows for a fluid understanding of individual positions, acknowledging that people may exhibit characteristics of multiple categories simultaneously or move back and forth along the scale as their experiences and perceptions evolve. Several theories provide insights into the psychological processes underlying these stages. Flow Theory (Csikszentmihalyi, 1990) explains the engagement levels at each stage. Those renouncing the adoption experience anxiety due to perceived high challenge and low skill, while those who approach it with intrinsic curiosity and motivation achieve flow through balanced challenge and skill. Self-Determination Theory (Ryan & Deci, 2000) outlines the role of intrinsic and extrinsic motivation in AI adoption. On the far end of high resistance, there is lack intrinsic motivation, moderate resistance allows adoption yet extrinsically motivated, and adoption with low resistance is intrinsically driven (Demerouti et al., 2001). The Job Demands-Resources Model (Bakker & Demerouti, 2007) provides a framework for understanding how the balance of demands and resources influences the transition between stages, particularly relevant for moderate adoption. The Technology Acceptance Model (Davis, 1989; Venkatesh & Davis, 2000) offers insights into the factors affecting the perceived usefulness and ease of use of AI technologies across all stages. Lastly, the Capability Approach (Nussbaum, 2011) provides a lens for examining how enhancing both individual capabilities and aligning social structures can facilitate progression through the stages, emphasizing that successful AI adoption is not just about individual readiness but also about creating supportive environments that allow people to achieve their full potential in the face of technological change. The transition from high to low resistance in AI adoption is shaped by a complex interplay of individual factors. Cognitive elements, such as technological literacy and perceived AI complexity, form the foundation of one's stance. Affective components, including attitudes toward change and self-efficacy beliefs, influence the emotional readiness for AI adoption. Behavioral aspects, like past technology experiences and adaptability, can either catalyze or impede movement through the resistance stages.

16.4.2 Organizational Perspective on AI Adoption and Integration

We can conceptualize organizational AI adoption and integration as an automation scale. This approach aligns with the progression from "No AI Integration" to "Full Automation."

16.4.2.1 No AI Integration → AI Co-operation → Domain Automation → Full Automation

The AI automation scale consists of four stages: No AI Integration (manual processes only), AI Co-operation (AI assisting human tasks), Domain Automation (specific functions fully automated), and Full Automation (AI embedded across all functions with minimal human involvement).

This automation scale is supported by several theoretical frameworks in organizational and technology adoption literature. STS Theory (Trist & Bamforth, 1951) emphasizes the interplay between people, technology, and organizational structure as AI automation increases. The Technology Acceptance Model (Davis, 1989) offers insights into factors affecting perceived usefulness and ease of use as organizations progress through the automation stages. Diffusion of Innovations Theory (Rogers, 2003) provides a framework for understanding how AI technologies are adopted and diffused throughout the organization. Full automation requires organizations to assess ethical considerations more rigorously, necessitating the development and implementation of robust ethical frameworks and governance structures (Floridi et al., 2018). As automation progresses, the workforce undergoes significant transformation, with each stage demanding different skills and roles. This shift highlights the importance of continuous training and development to ensure employees can effectively collaborate with and manage AI systems (Brynjolfsson & McAfee, 2017). Additionally, traditional organizational structures and processes may need to be redesigned to fully capitalize on AI's capabilities as automation permeates the organization (Davenport & Kirby, 2016). By conceptualizing organizational AI adoption as an automation scale, we can develop targeted strategies for each stage of AI integration. Transitioning from No AI Integration to AI Cooperation requires a foundational understanding of business needs and opportunities, establishing operational plans, implementing governance and ethical standards, and investing in the necessary infrastructure and workforce development. Moving from AI Cooperation to Domain Automation involves focusing on refining data management, preparing processes and tasks, improving infrastructure, advancing AI development, and designing user interaction strategies. The final stage, transitioning from

Domain Automation to Full Automation, demands a comprehensive approach that addresses social, ethical, cultural, and organizational considerations. This phase involves embedding AI across all functions, coupled with frameworks for oversight, monitoring, and continuous adaptation to ensure a seamless integration with minimal human intervention.

16.4.3 Modalities of AI Integration

Organizational AI integration can manifest through various modalities that determine how AI is incorporated into workflows. These include voluntary adoption, where employees are given the freedom to explore and adopt AI tools autonomously, typically seen in innovative environments that encourage experimentation and present AI as a tool to enhance creativity and problem-solving capabilities (Viskova-Robertson, 2023). Non-voluntary or mandated adoption involves organizations enforcing AI use through top-down directives, which can lead to passive resistance or disengagement if employees are not adequately involved in the decision-making process, especially when AI systems are perceived as replacements rather than augmentations. Complementary AI focuses on integrating systems that augment human capabilities rather than replace them, allowing employees to concentrate on complex, creative, or socially driven tasks while AI handles repetitive or data-intensive processes. This approach enhances organizational productivity and fosters a collaborative environment where AI is viewed positively (Makarius et al., 2020). Substitutional AI, on the other hand, involves implementing AI to fully take over tasks traditionally performed by humans, leading to changes in job roles and workforce dynamics. While this can boost efficiency, it requires careful handling of employee transitions, including re-skilling or job redesign, to mitigate fears of redundancy. These modalities shape how AI is embedded into organizational structures and processes, influencing employee perceptions, productivity, and the overall success of AI integration initiatives.

16.5 Trust as the Linchpin of AI Integration

Trust is a fundamental element in all interpersonal, organizational, and technological relationships. Mayer et al. (1995) define trust as

> the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party.

This definition highlights three key components of trust: competence, benevolence, and integrity. In the context of AI, trust extends beyond the basic competence of the systems themselves. It also encompasses benevolence, understood as the belief that the AI system will aim to benefit users and society, and integrity, which relates to the ethical principles underlying AI's development, deployment, and operation. Rousseau et al. (1998) further emphasize trust as a psychological state of accepting vulnerability based on positive expectations of another's behavior or intentions. This vulnerability is central to the concept of trust, as it implies reliance on another party without complete certainty of outcomes. When applied to AI systems, users expect AI to behave in ways that align with their goals and values, while also being competent and predictable in its operations not unlike the expectation in human relationships. As AI becomes increasingly integrated into various aspects of society, establishing and maintaining this trust becomes crucial for its successful adoption and ethical implementation. In the context of STSs, it operates at both the social level, where people must trust the organization implementing AI, and the technical level, where the systems themselves must be reliable and trustworthy.

16.5.1 Trust in AI: A Critical Component for Integration

Trust is fundamental to AI adoption because it determines how individuals and organizations interact with AI systems, particularly in environments characterized by uncertainty. Siau and Wang (2018) emphasize that trust in AI is distinct from trust in other technologies due to AI's unique characteristics, including its decision-making capabilities and its potential to surpass human performance in certain areas. Unlike static technologies, AI systems learn, adapt, and sometimes operate autonomously, which can raise concerns about reliability, transparency, and fairness.

Users need to trust that AI is not only designed to perform tasks effectively but also with their best interests in mind. Transparency in AI decision-making is crucial for building initial trust, as opaque processes can lead to suspicion or fear. This transparency should extend to both algorithmic processes, to a functional extent, and related work processes. While acknowledging the inherent complexity of algorithms, users need a basic understanding of AI concepts to foster trust, reinforced by the system's built-in checks and balances. This dual layer of trust – both in the AI system and the organization behind it – requires organizations to focus on transparency, explainability, and user engagement throughout the AI introduction and integration process.

16.5.2 The Multi-Layered Nature of Trust in AI

The research by Hoff and Bashir (2014) supports the view that trust is multilayered, involving several dimensions: dispositional trust (based on individual personality traits), situational trust (influenced by the specific context and environment), and learned trust (shaped by past experiences with AI systems). At the individual level, dispositional trust reflects a person's general openness to technology, which can play a significant role in determining how quickly and easily they adopt AI. For organizations, building and maintaining trust requires addressing all these layers, as AI systems become more embedded in workflows over time.

Overtrusting AI may lead to blind reliance on the system, potentially overlooking critical errors or flaws. Undertrusting AI can result in a reluctance to use the technology, stalling innovation and reducing efficiency. Therefore, instilling appropriate levels of trust in AI, through effective training, user feedback loops, and clear communication permeating the STS, is critical for ensuring safety and long-term adoption.

16.5.3 Trust in AI within the STS

Trust plays a pivotal role in shaping employees' attitudes toward AI in organizations. Trust in AI systems is closely linked to trust in the organization implementing the technology, creating a reciprocal relationship that influences cognitive, affective, and behavioral components of attitudes (Hoff & Bashir, 2014; Mayer et al., 1995). High levels of trust foster positive engagement with AI, while mistrust can result in resistance to its adoption (Jones & George, 1998). Trust is also dynamic and evolves based on interactions with AI systems, requiring organizations to continuously earn and maintain it through transparency and reliability (Rousseau et al., 1998). A key factor in building trust is ensuring that employees feel empowered and retain some level of control over processes that AI supports. When AI systems take on complex decision-making roles, maintaining a balance between human agency and AI automation is critical to avoid disempowerment (Bainbridge, 1983). Furthermore, trust can be strengthened by addressing potential biases in AI systems, ensuring that they are ethical and fair, particularly in sensitive applications such as recruitment or performance management (Floridi et al., 2018).

To build trust, organizations can focus on transparency, explainability, and providing hands-on opportunities for employees to interact with AI. Learning from other sectors – such as healthcare, financial services, and manufacturing – where trust in AI has been built through reliability and ethical use, can offer valuable strategies for successful AI integration (Floridi et al., 2018; Makarius et al., 2020).

16.6 Conclusion

Trust stands as the cornerstone of successful AI integration in STSs. It demands competence, benevolence, and integrity from AI systems and the organizations deploying them. Without trust, AI adoption falters, regardless of technological sophistication. Trust and attitudes form a mutually reinforcing core in successful AI integration within STSs and attitudes reciprocally influence each other, creating a complex feedback loop that significantly impacts the people within the socio-technical framework. Robust governance frameworks and ethical guidelines are non-negotiable for responsible AI implementation, ensuring that trust and positive attitudes are built on a foundation of integrity and societal benefit. As trust in AI systems grows, attitudes shift positively, fostering a more receptive environment for integration. Simultaneously, positive attitudes toward technology and innovation cultivate a greater propensity to trust AI systems.

The STS framework relies on harmonious balance between social and technical elements in AI integration. Organizational culture as part of the social subsystem, significantly, influences AI adoption attitudes. A culture of innovation and adaptability, reinforced by strong leadership, facilitates positive cognitive and affective responses to AI. Continuous assessment and adaptation, spanning all domains of the STS, are crucial for maintaining alignment between AI capabilities, user needs, and organizational goals in the long term. The dynamic interplay between organizational culture and AI adoption attitudes sets the stage for understanding the fluid continuum of AI integration across both individual and organizational dimensions.

AI adoption is a fluid continuum, not a series of lone stages. It spans both individual and organizational contexts within the STS. Individually, adoption can be viewed on a scale from initial high resistance to enthusiastic acceptance. Organizationally, it advances from no automation stage to full automation. This holistic approach of the STS framework provides an unparalleled lens for assessing AI integration potential. However, targeted solutions are equally vital. While the socio-technical perspective excels in capturing system-wide dynamics, narrow, focused interventions might hold the key to unlocking specific challenges within the system. Successful AI integration demands this dual approach: harnessing the comprehensive insights of the socio-technical framework and decisively deploying precision problem-solving for targeted issues.

References

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8.

Bakker, A. B., & Demerouti, E. (2007). The job demands-resources model: State of the art. *Journal of Managerial Psychology, 22*(3), 309–328. https://doi.org/10.1108/02683940710733115.

Brynjolfsson, E., & McAfee, A. (2017). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* W. W. Norton & Company.

Cherns, A. (1987). Principles of socio-technical design revisited. *Human Relations, 40*(3), 153–161. https://doi.org/10.1177/001872678704000301.

Clegg, C. W. (2000). Socio-technical principles for system design. *Applied Ergonomics, 31*(5), 463–477. https://doi.org/10.1016/S0003-6870(00)00009-0.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience.* Harper & Row.

Davenport, T. H., & Kirby, J. (2016). *Only humans need apply: Winners and losers in the age of smart machines.* Harper Business.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008.

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology, 86*(3), 499–512. https://doi.org/10.1037/0021-9010.86.3.499.

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2018). How to design AI for social good: Seven essential factors. *Science and Engineering Ethics, 24*(1), 19–31. https://doi.org/10.1007/s11948-017-9901-7.

Galbusera, F., Casaroli, G., & Bassani, T. (2019). Artificial intelligence and machine learning in spine research. *European Spine Journal, 28*(5), 913–919.

Hoff, K. A., & Bashir, M. (2014). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570.

Jones, G. R., & George, J. M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review, 23*(3), 531–546. https://doi.org/10.5465/amr.1998.926625.

Makarius, E. E., Mukherjee, D., Fox, J. M., & Fox, A. K. (2020). Rising with the machines: A socio-technical framework for bringing artificial intelligence into the organization. *Journal of Business Research, 120,* 262–273. https://doi.org/10.1016/j.jbusres.2020.07.045.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335.

Mumford, E. (1983). *Designing human systems for new technology: The ETHICS method.* Manchester Business School.

Nussbaum, M. C. (2011). *Creating capabilities: The human development approach.* Harvard University Press.

Organisation for Economic Co-operation and Development (OECD). (2019). OECD principles on AI. OECD Publishing. https://www.oecd.org/going-digital/ai/principles/.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.

Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In C. I. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1–14). Yale University Press.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393–404. https://doi.org/10.5465/amr.1998.926617.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68.

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *CUTTER Business Technology Journal, 31*(2), 47–53.

Tarafdar, M., Beath, C. M., & Ross, J. W. (2019). Using AI to enhance business operations. *MIT Sloan Management Review, 60*(4), 37–44.

Tarafdar, M., Beath, C. M., & Ross, J. W. (2019). Enterprise AI applications: A socio-technical systems perspective. *Journal of Information Technology, 34*(4), 328–343.

Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations, 4*(1), 3–38. https://doi.org/10.1177/001872675100400101.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science, 46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926.

Viskova-Robertson, A. (2023). The impact of artificial intelligence on organizational dynamics and employee well-being. *Journal of Organizational Change Management*, *36*(2), 210–225.

17 The Impact of Awareness and Risk-benefit Perceptions on Attitudes toward AI Adoption in Higher Education

Fatih Çetin, H. Nejat Basım and Joanna Paliszkiewicz

17.1 Introduction

The use of artificial intelligence (AI) in higher education has been receiving increased attention for its potential to enhance the quality and efficiency of educational and research practices (Aoun, 2018; Popenici & Kerr, 2017). However, the success of AI in higher education is not merely determined by its availability and capabilities of AI, but also by the attitudes and perceptions of faculty members toward AI adoption. Faculty members' decisions are influenced by cognitive and affective processes, making attitudes and perceptions crucial in determining their readiness to embrace AI technologies (Zawacki-Richter et al., 2019). In shaping attitudes and perceptions, the awareness and evaluation of the risks and benefits associated with AI technologies play significant roles in adoption and usage (Bousbahi & Alrazgan, 2015; Castelo et al., 2019; Huang et al., 2021).

Faculty members, as a responsible group for curriculum and research, are uniquely positioned to either facilitate or resist the integration of AI in higher education institutions. This chapter delves into the concepts of awareness, risk-benefit perception, and attitudes toward AI adoption, providing a theoretical foundation for understanding how these factors shape faculty members' decision-making processes regarding AI adoption in higher education. It encompasses the conceptual framework of the variables, the theoretical underpinnings of the interactions, the research model's purpose, methodology, research results, discussions of the results, and conclusions.

17.2 Conceptual framework

17.2.1 Awareness of AI technologies

Awareness of new systems and technology has a significant impact on attitude toward effective adoption (Porter & Graham, 2016). Awareness in the context of AI adoption refers to the extent to which individuals have accurate knowledge and understanding of AI technologies, their capabilities, and limitations.

DOI: 10.4324/9781003586944-20

According to cognitive psychology, awareness is not a static construct but exists along a continuum. At the low end of this continuum, individuals may have little to no understanding of AI, possibly having misconceptions or fears. On the high end, individuals possess a comprehensive understanding of AI's technical aspects, its potential for application, and the ethical and legal issues it raises (Zawacki-Richter et al., 2019).

In academic settings, faculty members' familiarity with AI varies based on their expertise, access to professional development resources, and prior experience with educational technology (Rahiman & Kodikal, 2023). The level of awareness significantly influences how faculty members perceive AI adoption. Faculty members who have a good understanding of AI applications are more likely to see the technology as a valuable tool for improving both teaching and research. On the other hand, those with less awareness may be more skeptical or fearful about AI, especially when it comes to worries about automation, surveillance, and job displacement (Celik et al., 2022).

Several studies have investigated the level of awareness among faculty members regarding AI technologies and their potential applications in education. For example, a study found that a significant proportion of faculty members had low awareness of AI and its applications in higher education, which negatively impacted their attitudes toward its adoption (Bousbahi & Alrazgan, 2015). The study also highlighted the role of professional development programs in increasing faculty awareness of AI, suggesting that institutions need to invest in ongoing training to ensure that faculty members are equipped to engage with these technologies. The results of a recent study showed that organizational policies, incentives, and the availability of professional development significantly influence the adoption of AI. Faculty awareness was found to be linked to understanding AI's pedagogical benefits and its usability in education, which encouraged a positive attitude toward AI implementation (Ofosu-Ampong, 2024).

Another study focused on ChatGPT and examined the role of trust, perceived ease of use, and awareness in influencing its adoption. It emphasized that increasing faculty awareness through educational initiatives and peer influence could enhance AI adoption intentions (Shahzad et al., 2024). Additionally, a needs assessment found that faculty members require further support in understanding generative AI's potential applications in instruction, suggesting that proper training and resources are essential to bridge the gap between awareness and actual usage (Mathew & Stefaniak, 2024). A systematic review highlighted that teachers' perceptions of AI are influenced by their level of understanding and experience with the technology. Those with limited knowledge often express concerns related to surveillance, loss of autonomy, and the risk of job displacement, especially in countries or regions with lower infrastructure preparedness for AI adoption (Almasri, 2024).

These studies highlight the varying levels of awareness and acceptance of AI in education, underscoring the need for targeted interventions to improve faculty members' understanding and use of these technologies in their teaching and research practices.

17.2.2 Risk-Benefit Perception

The process of risk-benefit evaluation is deeply rooted in psychological theories of decision-making, such as Prospect Theory, which suggests that individuals prioritize perceived gains (i.e., benefits) over perceived losses (i.e., risks) when making decisions (Tversky & Kahneman, 1974). This phenomenon, known as Loss Aversion Theory, indicates that individuals tend to choose options associated with potential gains when presented with two equal choices, one highlighting potential gain and the other potential losses. Accordingly, risk-benefit perception refers to the cognitive process by which individuals weigh the potential benefits of a new technology against the perceived risks it poses (Stolwijk et al., 1988).

The rise of AI presents a complex situation. On one hand, integrating AI into education offers a wide range of potential benefits for both teachers and students. However, it also introduces new challenges and the potential for misuse that demands careful consideration. One of the primary benefits of AI in education and research, as perceived by faculty members, is its ability to enhance the efficiency and productivity of various academic tasks (Owoc et al., 2021). A recent study highlighted the advantages of AI-powered tools for faculty members in research tasks. These tools can help with activities like literature review, data analysis, and even writing and editing research papers (BaHammam et al., 2023). This can greatly reduce the time and effort needed for these labor-intensive tasks, enabling faculty members to concentrate more on their primary responsibilities of teaching, research, and supervision.

A research showed that teachers with experience and knowledge of AI applications are more likely to understand the benefits of AI in enhancing personalized learning, improving student engagement, and supporting differentiated instruction. These teachers tend to have more trust in AI and appreciate its ability to assist with grading, lesson planning, and monitoring student progress (Almasri, 2024). Another study discovered that teachers' adoption of AI is influenced by various motivating and inhibiting factors. Motivating factors include the exploration of innovative educational technologies, personalized teaching and learning, time-saving, and professional development. On the other hand, inhibiting factors encompass concerns about reliability and accuracy, reduced human interaction, privacy and data security, lack of institutional support, and overreliance on AI (Al-Mughairi & Bhaskar, 2024).

In academic settings, some faculty members may be hesitant to adopt AI technologies due to perceived risks. These risks include the fear of losing control over the teaching process and concerns about the implications of data-driven decision-making, which may seem more significant than the potential benefits (Ertmer et al., 2012). Furthermore, risk perception is often influenced by emotions – particularly fear and anxiety – leading individuals to overestimate the likelihood or magnitude of negative outcomes (Slovic, 1987). This is particularly relevant in discussions of AI, which is often framed in popular media as a threat to human jobs or autonomy (Castelo et al., 2019).

With the increasing use of AI in education and research, primary risk concerns include the impact on decision-making and the potential for AI systems to introduce bias, lack transparency, and accountability, as well as promote laziness and insecurity (Ahmad et al., 2023; Gillani et al., 2023). A recent study discovered that faculty members perceive significant risks associated with AI, such as threats to academic integrity, concerns about job displacement, and ethical issues related to data privacy and surveillance (Abdelaal, & Al Sawy, 2024; Gustilo et al., 2024). In another study, faculty members have expressed significant concerns about the potential impact of AI-powered systems on the teacher-student relationship, the risk of dehumanizing the learning experience, algorithmic bias, and ethical issues such as privacy and equity (Gupta et al., 2024). In another study, educators who lack familiarity with AI tend to have concerns about the ethical implications of AI in education, such as privacy risks, biases in AI algorithms, and fears of losing jobs to automation (Celik et al., 2022).

These studies emphasize the benefits and risks of AI adoption in education, stressing the importance of evaluating them in relation to educational goals and research, while considering ethical implications.

17.2.3 Attitudes toward AI Adoption

Attitudes toward technology adoption are influenced by a combination of cognitive, affective, and behavioral factors (Ajzen, 1991). In the case of AI, faculty members' attitudes toward adoption are shaped by their perceived benefits, attitudes, behavioral intentions, and facilitating conditions (Rahiman & Kodikal, 2023). These attitudes, in turn, influence whether faculty are willing to incorporate AI into their teaching, research, and administrative practices. Faculty members show average readiness to integrate AI into teaching practices, with significant correlations between readiness and various factors (Alnasib, 2023).

Attitudes toward technology adoption are often explained through models such as the Technology Acceptance Model (TAM) (Davis, 1989) and the Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003). These models suggest that two key variables – perceived usefulness and perceived ease of use – are primary determinants of whether individuals will adopt new technologies. In the context of AI adoption, perceived usefulness refers to the extent to which faculty members believe that AI will enhance their teaching or research effectiveness. Perceived ease of use refers to how simple they believe it will be to integrate AI into their existing workflows. Faculty members who perceive AI as both useful and easy to use are more likely to adopt it, whereas those who view it as overly complex or marginally useful may resist its adoption (Zawacki-Richter et al., 2019).

17.2.4 Theoretical Underpinnings

TAM provides a framework for understanding the factors that influence individuals' decisions to adopt new technologies (Davis, 1989). The model posits that perceived usefulness and perceived ease of use are the primary predictors of technology adoption. Faculty members who perceive AI as highly useful for their teaching and research activities are more likely to develop positive attitudes toward its adoption. Similarly, those who find AI tools easy to learn and integrate into their workflows are more likely to adopt them. However, if faculty perceive AI as difficult to use or as offering limited benefits, their attitudes toward adoption are likely to be negative (Bousbahi & Alrazgan, 2015).

Theory of Planned Behavior (Ajzen, 1991) extends the TAM by incorporating the role of subjective norms and perceived behavioral control in the adoption process. In academic settings, faculty members' attitudes toward AI adoption are not only shaped by their own assessments of its usefulness and ease of use but also by the opinions of their colleagues and institutional leadership. If faculty perceive that their peers and supervisors expect them to adopt AI, this may positively influence their attitudes toward adoption. Additionally, perceived behavioral control – faculty members' beliefs about their ability to successfully implement AI – can also influence their adoption decisions (Ertmer et al., 2012).

Elaboration Likelihood Model (Petty & Cacioppo, 1986) is a dual-process theory of persuasion that explains how individuals process persuasive information through either the central route or the peripheral route. The central route involves careful and thoughtful consideration of the arguments presented, while the peripheral route involves less scrutiny and is more influenced by superficial cues. Faculty members with high levels of awareness about AI are more likely to engage in central route processing, evaluating the technology based on its merits and aligning their attitudes with their understanding of its functionality. On the other hand, those with limited awareness may rely on peripheral cues, such as media narratives or peer opinions, which may result in less favorable attitudes toward AI adoption (Tormala & Petty, 2007).

17.2.5 The purpose and hypotheses

Based on the discussions and theoretical explanations aforementioned above, this study aims to examine how faculty members' awareness, risk-benefit assessments, and perceptions of AI influence their attitudes toward its adoption. Our research hypotheses are listed below (also see in Figure 17.1);

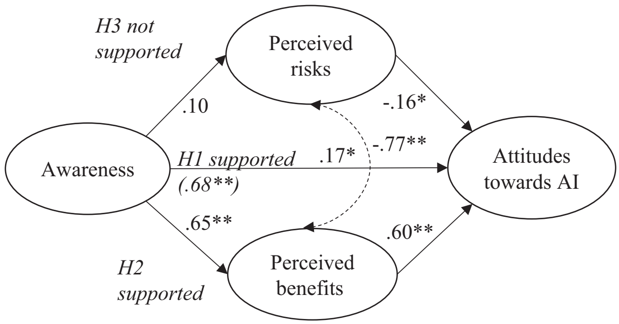

Figure 17.1 Research model and results

Source: Own elaboration.

H1: Awareness has a positive impact on attitudes toward AI adoption

H2: Perceived benefits partially mediate the influence of awareness on attitudes toward AI adoption

H3: Perceived risks partially mediate the influence of awareness on attitudes toward AI adoption

Drawing on both psychological theories and existing empirical research in higher education, this study provides insights into the cognitive and emotional processes underlying faculty members' acceptance of AI.

17.3 Method

17.3.1 Participants

Data were collected from 178 faculty members from different universities in Turkiye. The principles of voluntary participation and personal data privacy have been confirmed. An online questionnaire, including items from literature with the demographic details, has been sent to participants. Sample gender distribution is 51.6% female and 48.4% male with a mean age of 41.1 ($SD = 9.53$). The academic degree distribution is as follows: 12.4% professor, 25.8% associate professor, 31.2% assistant professor, 11.9% research assistant with PhD, and 18.7% research assistant.

17.3.2 Instruments

Risk-benefit perceptions. We have used 12-item (six items for each) risk-benefit assessment scale (Said et al., 2023). The sample items are "*When you think about use of AI, to what extent are you troubled*" (risk) and "*If you were to use of AI in your environment, how beneficial would the consequences be for you*" (benefit).

Awareness of AI. We have used four-item Cognitive Awareness Related to AI scale (Gaber et al., 2023). The sample item is "*I have sufficient knowledge of AI programs and applications.*"

Attitude toward AI adoption. We have used five-item Attitude Toward Artificial Intelligence scale (Sindermann et al., 2021). The sample item is "*I trust artificial intelligence.*"

17.3.3 Procedure

In order to ensure validity and reliability, a measurement model has been established, encompassing all items and related latent factors, which are interconnected. The Average Variance Extracted (AVE), Composite Reliability (CR), and Cronbach Alpha coefficients (α) have been calculated. Subsequently, a Structural Equation Model (SEM) was employed to test hypotheses and address the research question, using the following fit indices: X^2/df < 3, CFI and TLI >. 90, and RMSEA <.08. All scales rated on a 5-point Likert scale from strongly disagree (1) to strongly agree (5).

17.4 Results

17.4.1 Validity and Reliability

The results of the CFA fit indices for the scales and the measurement model (X^2/ df = from 1.32 to 2.74, CFI = from.90 to.98, TLI = from.89 to.98, RMSEA = from.043 to.071) confirmed the adequacy of the fit indices. The validity and reliability findings (AVE, CR, Cronbach Alpha, McDonalds' Omega) indicated acceptable values.

17.4.2 Hypotheses Testing

For the research hypotheses, we have constructed a SEM using the maximum likelihood estimation. In the SEM, the correlated variables of perceived risks, perceived benefits, and awareness have been entered as exogenous, and attitudes toward AI adoption as an endogenous variable. The structural model results have provided the confirm of the fit indices (X^2/df = 2.20, CFI = .95, TLI = .94, RMSEA = .058). In the first step, awareness has been entered as exogenous and attitudes toward AI adoption as an endogenous variable. The standardized influence (beta) from awareness to attitudes toward AI adoption is.68 (p<.01, *supporting H1*). In the second step, the correlated variables of perceived risks and perceived benefits have been added with awareness as exogenous and attitudes toward AI adoption as an endogenous variable. The standardized influence (betas) from awareness to attitudes toward AI adoption has been.17 (p<.05), from perceived risks to attitudes toward AI adoption has been -.16 (p<.05), and from perceived

Table 17.1 The validity and reliability statistics

		Factor loadings	*CFA Fit indices*				*Validity*		*Reliability*	
			χ^2/df	*CFI*	*TLI*	*RMSEA*	*AVE*	*CR*	α	ω
1. Awareness of AI		from.78 to.89	1.32	.98	.98	.043	.72	.91	.90	.90
2. Risk-benefit perceptions (two-factor solutions)	Risk	from.74 to.87	1.98	.93	.92	.061	.61	.86	.85	.85
	Benefit						.68	.88	.87	.87
3. Attitude toward AI adoption		from.71 to.85	2.74	.90	.89	.071	.59	.85	.85	.85
4. Measurement model		from.80 to.91	χ^2/df= 2.20, CFI=.95, TLI=.94, RMSEA=.058							
5. Single factor for testing Common Method Variance		from.51 to.88	χ^2/df= 4.47, CFI=.81, TLI=.79, RMSEA=.097							

Source: Author's research

benefits to attitudes toward AI adoption has been.60 ($p<.01$) in the model. The substantial standardized influences (betas) from awareness to perceived benefits were found to be highly significant (beta=0.65, $p<0.01$), indicating a strong link between awareness and perceived benefits. Moreover, the indirect effect of awareness on attitudes toward AI adoption, with the mediating role of perceived benefits, was also highly significant (indirect effect=0.39, $p<0.01$), *providing compelling support for H2*. However, the indirect effect of awareness on attitudes toward AI adoption, with the mediating role of perceived risks, was not found to be significant (indirect effect = -0.016, $p > 0.05$), thus *not supporting H3*.

17.5 Discussion

Enhanced awareness of AI has a positive impact on attitudes toward its implementation. However, deeper knowledge of AI can lead to an overestimation of benefits and an assumption of greater capability in managing risks. Therefore, it is crucial to assess both risks and benefits when shaping attitudes toward AI adoption. This study aims to examine how faculty members' awareness, risk-benefit assessments, and perceptions of AI influence their attitudes toward its adoption. Findings showed that being aware of AI has a positive impact on attitudes toward AI adoption, and perceived benefits play a mediating role in this relationship, outweighing the perceived risks.

The findings underscore the critical link between faculty members' awareness of AI and their attitudes toward AI adoption in higher education. It is clear that a deep understanding of AI leads to more favorable attitudes, highlighting the need for educational institutions to prioritize raising awareness about AI technologies among faculty. The research findings strongly support previous studies, indicating that a heightened awareness of AI technologies leads to overwhelmingly positive attitudes toward effective adoption (Porter & Graham, 2016), intention to adopt (Shahzad et al., 2024), and actual usage (Mathew & Stefaniak, 2024). Moreover, a comprehensive understanding of the technical aspects, potential applications, and limitations results in a clear grasp of the benefits and usability implementations (Ofosu-Ampong, 2024), alleviating concerns about automation, surveillance, and job displacement (Almasri, 2024; Celik et al., 2022).

The findings emphasize the pivotal role of perceived benefits in this relationship between awareness and positive attitudes toward AI adoption. Faculty members who recognize the vast potential benefits of AI, such as enhanced teaching efficiency, personalized learning experiences, and improved administrative processes, are more likely to embrace AI tools, even in the face of perceived risks. Parallel with the Prospect Theory (Tversky & Kahneman, 1974), faculty members tend to choose options associated with potential gains when presented with two equal choices, one highlighting the potential benefits of new technology and the other perceived risks (Stolwijk et al., 1988). The research findings clearly support the widespread preference among faculty members for utilizing AI to get educational and research-related benefits. This encompasses streamlining various academic tasks to boost efficiency and productivity (Owoc et al., 2021), including conducting literature reviews, data analysis, and crafting and refining research papers (BaHammam et al., 2023). Moreover, AI serves as a valuable assistant in grading, lesson planning, and monitoring student progress (Almasri, 2024). The adoption of AI in academia is driven by the appeal of innovative educational technologies, personalized teaching and learning, time-saving solutions, and professional development (Al-Mughairi & Bhaskar, 2024).

Based on these compelling results, it is imperative for educational institutions to introduce captivating professional development programs, workshops, or seminars that showcase the extraordinary capabilities and applications of AI in academic settings. This will provide faculty members with a firsthand understanding of how AI can revolutionize teaching and administrative practices. By effectively communicating these advantages, institutions can alleviate apprehensions and present a persuasive case for AI adoption. This dual approach of increasing awareness and articulating the benefits of AI adoption will significantly enhance faculty members' readiness to embrace AI technologies in their educational practices.

17.6 Conclusion

The findings of this study underscore the significance of awareness in shaping positive attitudes toward AI adoption among faculty members. It is evident that as individuals become more informed about AI technologies, their willingness to embrace these innovations increases. Furthermore, the perception of benefits derived from AI use serves as a critical mediator in this relationship, suggesting that individuals who recognize the advantages of AI are more likely to overcome their apprehensions and perceived risks. This highlights the importance of enhancing awareness and education about AI, as fostering a positive attitude toward its adoption can lead to more effective integration of AI technologies within educational environments. By focusing on the potential benefits and addressing concerns, institutions can facilitate a smoother transition into an AI-enhanced future.

The findings in the research also should be interpreted cautiously with the limitations of the cross-sectional research design, the survey methodology, the single-source data collecting, and the generalizability derived from a regional sample. Furthermore, the pervasive integration of AI within the academic community and the scarcity of applications such as *ethical AI* or *responsible AI* may shape participants' perceptions of the associated benefits and risks of AI.

Acknowledgment

This research is partly funded by the University of Economics Ho Chi Minh City, Vietnam.

References

Aoun, J. E. (2018). *Robot-proof: Higher education at the age of artificial intelligence*. MIT Press.

Abdelaal, N., & Al Sawy, I. (2024). Perceptions, challenges, and prospects: University professors' use of artificial intelligence in education. *Australian Journal of Applied Linguistics*, *7*(1), 1–24. https://doi.org/10.29140/ajal.v7n1.1309.

Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M. K., Irshad, M., Arraño-Muñoz, M., & Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities & social sciences communications*, *10*(1), 311. https://doi.org/10.1057/s41599-023-01787-8.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T.

Al-Mughairi, H., & Bhaskar, P. (2024). Exploring the factors affecting the adoption AI techniques in higher education: insights from teachers' perspectives on ChatGPT. *Journal of Research in Innovative Teaching & Learning*, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/JRIT-09-2023-0129.

Almasri, F. (2024). Exploring the impact of artificial intelligence in teaching and learning of science: A systematic review of empirical research. *Research in Science Education*, *54*, 977–997. https://doi.org/10.1007/s11165-024-10176-3.

Alnasib, B. N. (2023). Factors affecting faculty members' readiness to integrate artificial intelligence into their teaching practices: A study from the saudi higher education context. *International Journal of Learning, Teaching and Educational Research*, *22*(8), 465–491. https://doi.org/10.26803/ijlter.22.8.24.

BaHammam, A. S. (2023). Balancing innovation and integrity: The role of AI in research and scientific writing. *Nature and Science of Sleep, 15,* 1153–1156. https://doi.org/10.2147/NSS.S455765.

Bousbahi, F., & Alrazgan, M. S., (2015). Investigating IT faculty resistance to learning management system adoption using latent variables in an acceptance technology model, *The Scientific World Journal*, *ID* 375651, 1–11. https://doi.org/10.1155/2015/375651.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788.

Celik, I., Dindar, M., Muukkonen, H., & Sanna, J. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *Tech Trends, 66*, 616–630. https://doi.org/10.1007/s11528-022-00715-y.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008.

Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: A critical relationship. *Computers and Education*, *59*, 423–435. https://doi.org/10.1016/j.compedu.2012.02.001.

Gaber, S. A., Shahat, H. A., & Alkhateeb, I. A., et al., (2023). Faculty members' awareness of artificial intelligence and its relationship to technology acceptance and digital competencies at King Faisal University. *International Journal of Learning, Teaching and Educational Research, 22*, 473–496. https://doi.org/10.26803/ijlter.22.7.25.

Gillani, N., Eynon, R., Chiabaut, C., & Finkel, K. (2023). Unpacking the "Black Box" of AI in education. *Educational Technology & Society*, *26*(1), 99–111 https://doi.org/10.30191/ETS.202301_26(1).0008.

Gupta, G., Sreelatha, C., Shilpi Raj, A. L., & Singh. A. (2024). Navigating the future of education: The impact Oo artificial intelligence on teacher-student dynamics. *Educational Administration: Theory and Practice, 30*(4), 6006–6013. https://doi.org/10.53555/kuey.v30i4.2332.

Gustilo, L., Ong, E., & Lapinid, M.R. (2024). Algorithmically-driven writing and academic integrity: Exploring educators' practices, perceptions, and policies in AI era. *International Journal for Educational Integrity, 20*(3), 1–43. https://doi.org/10.1007/s40979-024-00153-8.

Huang, C. Y., Yang, M. C., & Huang, C. Y. (2021). An empirical study on factors influencing consumer adoption intention of an AI-powered chatbot for health and weight management. *International Journal of Performability Engineering*, *17*(5), 422. https://doi.org/10.23940/ijpe.21.05.p2.422432.

Mathew, R., & Stefaniak, J.E. (2024). A needs assessment to support faculty members' awareness of generative AI technologies to support instruction. *Tech Trends, 68*, 773–789. https://doi.org/10.1007/s11528-024-00964-z.

Ofosu-Ampong, K. (2024). Beyond the hype: exploring faculty perceptions and acceptability of AI in teaching practices. *Discover Education, 3,* 38. https://doi.org/10.1007/s44217-024-00128-4.

Owoc, M. L., Sawicka, A., & Weichbroth, P. (2021). Artificial intelligence technologies in education: Benefits, challenges and strategies of implementation. In M. L. Owoc, & M. Pondel (Eds.), *Artificial Intelligence for Knowledge Management* (pp. 37–58). Springer. https://doi.org/10.1007/978-3-030-85001-2_4.

Petty, R. E., & Cacioppo, J. T. (1986). *The Elaboration Likelihood Model of Persuasion. In: Communication and Persuasion. Springer Series in Social Psychology*. Springer. https://doi.org/10.1007/978-1-4612-4964-1_1.

Popenici, S.A.D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *RPTEL, 12*, 22. https://doi.org/10.1186/s41039-017-0062-8.

Porter, W. W., & Graham, C. R. (2016). Institutional drivers and barriers to faculty adoption of blended learning in higher education. *British Journal of Educational Technology*, *47*(4), 748–762. https://doi.org/10.1111/bjet.12269.

Rahiman, H. U., & Kodikal, R. (2023). Revolutionizing education: Artificial intelligence empowered learning in higher education. *Cogent Education, 11*(1), 1–24. https://doi.org/10.1080/2331186X.2023.2293431.

Said, N., Potinteu, A. E., Brich, I., Buder, J., Schumm, H., & Huff, M. (2023). An artificial intelligence perspective: How knowledge and confidence shape risk and benefit perception. *Computers in Human Behavior*, *149*, 1–25. https://doi.org/10.1016/j.chb.2023.107855.

Shahzad, M. F., Xu, S., & Javed, I. (2024). ChatGPT awareness, acceptance, and adoption in higher education: The role of trust as a cornerstone. *International Journal of Educational Technology in Higher Education, 21,* 46. https://doi.org/10.1186/s41239-024-00478-x.

Sindermann, C., Sha, P., & Zhou, M. et al., (2021). Assessing the attitude towards artificial intelligence: Introduction of a short measure in german, chinese, and english language. *Künstl Intell*, 35, 109–118. https://doi.org/10.1007/s13218-020-00689-0.

Slovic, P. (1987). Perception of risk. *Science, 236*(4799), 280–285. https://doi.org/10.1126/science.3563507.

Stolwijk, J. A. J., Doob, L. W., Tiemann, A. R., DeLuca, D. R., Gardner, G. T., & Gould, L. C. (1988). *Perceptions of Technological Risks and Benefits*. Russell Sage Foundation. https://muse.jhu.edu/book/38522.

Tormala, Z. L., & Petty, R. E. (2007). Contextual contrast and perceived knowledge: Exploring the implications for persuasion. *Journal of Experimental Social Psychology, 43*(1), 17–30. https://doi.org/10.1016/j.jesp.2005.11.007.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540.

Zawacki-Richter, O., Marín, V.I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators?. *International Journal of Educational Technology in Higher Education, 16,* 39. https://doi.org/10.1186/s41239-019-0171-0.

18 The Limits of Intuition in Artificial Intelligence

Edyta Skarzyńska, Jay Liebowitz, Giovanni Miragliotta and Joanna Paliszkiewicz

18.1 Introduction

Modern AI systems, such as generative AI models, base their predictions and decisions on large data sets (Skarzynska & Paliszkiewicz, 2025). However, for users, it is important that the systems work effectively and that their decisions seem natural, similar to humans. Intuition can be a key element in building credibility and Trust in AI (Liebowitz, 2024). Intuition plays a role in understanding humans' decisions based on implicit cognitive processes. In the case of AI, developing models that can mimic intuitive human processes to some extent is a technological, philosophical, and ethical challenge. Recent research has explored the integration of intuition into AI systems. Sun (2009) developed an artificially intuitive reasoner capable of making accurate predictions using sparse, low-quality data, mimicking human intuitive decision-making. Tsvetkova (2022) argues that intuition remains a competitive advantage for humans over AI, especially in information-deficit situations. However, Abbasi et al. (2022) suggest that intuition can be incorporated into AI-assisted decision-making through models of human-AI interaction or by replicating human intuition in AI systems. Trovati et al. (2022) propose embedding AI intuition into AI to enhance information processing and knowledge discovery.

While current AI systems excel at information overload, they struggle with information deficiency, where human intuition often proves valuable (Tsvetkova, 2022). These studies highlight the potential of integrating intuitive capabilities into AI systems to enhance decision-making capabilities in various domains. Human intuition is a product of evolution and involves complex cognitive processes such as emotional analysis, life experience, and subconscious analysis of the environment. It is fast, automatic, and often difficult to explain logically. AI algorithms are based on clearly defined mathematical and statistical rules (Korzynski et al., 2023). While human intuition operates under conditions of uncertainty, algorithms strive to make the most of available data. Intuition requires the ability to understand the context (Liebowitz, 2018). For AI, this means analyzing many variables in real time.

DOI: 10.4324/9781003586944-21

An example would be recommendation models that "predict" user preferences based on their previous choices and behaviors. Modeling intuition requires advanced data analysis techniques and access to diverse and rich data sets. Intuition can be key in developing AI systems that inspire Trust and provide natural interactions. However, its implementation requires consideration and a balance between efficiency and transparency. This chapter aims to evaluate approaches to the using and developing intuition in artificial intelligence (AI) systems. First, intuition and its importance in decision-making processes are defined, and the results of key studies are presented. This was followed by a discussion of the technical aspects of implementing intuition in AI and examples of algorithms inspired by human cognitive processes. Benefits, such as applications in diagnostics or management, and challenges, including the problem of data bias and the need for transparency, were considered. It then analyzed building trust in AI through intuitive mechanisms that can make understanding how algorithms work and reduce user uncertainty easier. The synergy between human intuition and AI capabilities is highlighted, leading to more effective decisions. This chapter concludes with predictions for the development of intuitive AI systems and the need for inclusive design that considers the diverse needs of users. Key findings are summarized, and recommendations are made for further research and practice.

18.2 The Importance of Intuition in Building Trust in AI

The development of AI technologies opens up unprecedented opportunities for humanity but, at the same time, raises numerous questions regarding the safety, ethics, and trust in these systems. One key challenge is building Trust between humans and AI systems. Intuition plays an important role in this context, as it significantly influences the perception of AI, its trustworthiness, and acceptance in everyday life and strategic decisions. Research suggests that intuition is essential for strategic decision-making and can be integrated with AI-assisted decision-making through models of human-AI interaction or by replicating human intuition in AI systems (Abbasi et al., 2022). Studies have identified three types of intuition involved in reasoning about AI predictions: intuition about task outcomes, characteristics, and limitations of AI (Chen et al., 2023). Trust is crucial for the recognition, progress, and development of AI in society, with factors such as transparency and lack of bias contributing to trust formation (Dashkov & Nesterova, 2021).

Trust models and reputation approaches using AI techniques have also been developed, addressing issues such as bootstrapping, trust propagation, and group trust modeling (Zhang et al., 2019). Understanding how users trust AI systems designed to act on their behalf is increasingly important for effective human-AI collaboration (Paliszkiewicz & Gołuchowski, 2024). Intuition is an integral part of the process of building Trust in AI. It is the first filter through which people evaluate AI; it influences their ability to accept the technology and how they understand its operation. At the same time, to fully exploit the potential of intuition in building trust, it is necessary to combine it with other elements such as explainability, transparency, and control mechanisms. As a result, AI systems can become tools that people trust but, at the same time, approach critically and responsibly.

While intuition is important, overreliance can lead to a false sense of security. People may intuitively trust AI systems that appear friendly when, in reality, they may be seriously flawed. This is especially true for systems that learn from historical data, which may contain biases and errors (Liebowitz, 2025).

In addition, intuition can be culturally conditioned and subjective, meaning that different groups may react differently to the same AI systems (Liebowitz et al., 2019). Therefore, technology designers need to consider the diversity of users and test their solutions in different contexts. User trust is a key element that determines the success of implementing AI systems in various areas of life. Without it, even the most advanced technologies can be rejected. Intuitive mechanisms in AI, such as transparency in decision-making or consistency of interactions, play an important role in building this Trust. Thanks to intuitive solutions, users can more easily understand how and why the algorithm made certain decisions, reducing uncertainty and increasing the sense of control. Interpretability is crucial for developing appropriate Trust because users must understand how AI systems make decisions (Bhatt et al., 2019). Trust in AI algorithms is built through collaborative validation processes, including negotiating criteria and comparison conditions (Winter & Carusi, 2022). Various computational trust models have been developed, but challenges remain in areas such as bootstrapping, trust propagation, and group trust modeling (Zhang et al., 2019). Researchers are exploring methods to assess user trustworthiness in the context of social media and applying trust filters to improve sentiment analysis for more accurate predictions in domains such as stock markets (Zhang et al., 2019). These studies emphasize the need for balanced Trust in AI systems and social media information. Combining human intuition with the power of AI algorithms leads to the creation of hybrid decision-making systems that can produce better results than either party acting alone.

Based on experience and the ability to think in context, human intuition perfectly complements the analytical precision of AI algorithms. The synergy between human intuition and AI computational capabilities increases efficiency and builds Trust, as users feel that decisions are made holistically. Intuition in AI system design carries significant ethical challenges. Although intuitive mechanisms can increase Trust, their overuse or improper implementation can lead to manipulation or misleading of users. (Sanders & Wood, 2020). Systems that seem to "understand" users' emotions but, in reality, only use advanced data analytics can create a false sense of security. The ethical boundary should always be ensuring fairness and transparency in the operation of systems. Developers are responsible for ensuring that intuitive mechanisms are consistent with the actual capabilities of the technology and support users and do not exploit their limited knowledge of how AI works (Sandler-Smith, 2023). An ethical approach also requires inclusiveness in the design of intuitive interfaces. AI systems must be intuitive for many users, regardless of age, education level, or technological experience. Only then can trust in AI be built sustainably and fairly.

Building Trust through intuition in AI is a complex process that requires advanced technologies and a deep understanding of human needs and boundaries. Intuitive systems support users in better understanding AI, connecting their abilities with algorithms' capabilities, and considering ethical responsibility, making interactions with AI more transparent and trustworthy (Vincent, 2021).

18.3 Intuition vs. Algorithms: Key Differences and Similarities

Intuition is considered unconscious intelligence based on experience, useful in uncertain situations where algorithms may falter (Gigerenzer, 2023). It involves a "yes or no" judgment, while insight refers to "what" the solution is (Zhang et al., 2016). The legal system generally favors algorithmic knowledge, but intuitive experts can contribute valuable insights to computational models (Solan, 2013). Intuition has historically been gendered as feminine and contrasted with logical rationality, leading to calls for its replacement by algorithms (Gigerenzer, 2023). However, intuition is essential to understanding our "algorithmic condition," in which machine learning technologies redistribute cognition between humans and machines (Pedwell, 2022). The relationship between intuition and algorithms has evolved through key historical moments, from the birth of AI to the rise of personal computers and our current algorithmic lives. Intuition and algorithms are distinct but often complementary approaches to decision-making and problem-solving. Intuition is a largely subjective process based on experiences and nonverbal pattern recognition, whereas algorithms are formal procedures based on precise rules that lead to a specific outcome. Analyzing these two approaches' main differences and similarities allows us to understand their potential applications and limitations.

One of the main differences between intuition and algorithms is the degree of transparency. Intuition, due to its subconscious nature, can be difficult to explain. Algorithms, however, are completely transparent – you can trace every rule and step of the process. For this reason, algorithms are widely used in situations where detailed justification of decisions is required, such as in science or engineering. Intuition is more adaptive and often copes better with unpredictable or novel situations without established rules (Demartini et al., 2024). Although effective in defined conditions, algorithms are limited to the scope for which they were designed. Only advanced machine learning algorithms begin to show the ability to adapt in some way, although they remain within the limits of the data on which they were trained. Intuition is based on cumulative experience and subconscious processing of information. This is often a fast process, but difficult to explain or objectively measure (Denford et al., 2024).

Conversely, algorithms are precisely defined sets of rules that lead step by step to solving a problem. Despite these differences, intuition and algorithms also have some common features. Both processes strive to solve problems based on detecting patterns – intuition through experience and algorithms through data analysis.

18.4 Development of Artificial Intuition

Although still developing, the concept of artificial intuition in computers is of growing interest to researchers and technology specialists. A growing body of evidence points to the possibility of implementing intuition – the ability to quickly recognize patterns and make decisions based on limited information – in AI systems. Intuition is key in decision-making processes, from management to medicine and the exact sciences. Research conducted by Johnny et al., (2019) and Trovati et al., (2023) indicates that the development of artificial intuition is possible, and its implementation can bring significant benefits in the automation of complex decision-making processes.

Traditional approaches to decision-making in computer systems require detailed analysis of all available data and scenarios. Although such methods are accurate, they can be time-consuming and require significant computational resources. In contrast, intuitive models allow for predicting outcomes based on a limited information set without fully understanding all parameters. This approach is particularly useful in situations of uncertainty when data is fragmentary or contradictory. Johnny et al. (2019) point out that intuition enables faster and more effective decision-making, which may be crucial in areas such as medical diagnostics or market analysis.

An important aspect of the development of artificial intuition is its programmability. Intuition, traditionally considered to be the domain of human experience and emotion, was long considered impossible to implement in computer systems. However, research such as that conducted by Van den Herik (2015) suggests that mental intuition can be programmable. This challenges traditional beliefs about its exclusively human nature. Van den Herik (2015) argues that appropriately designed algorithms can simulate intuitive processes and even generate decisions that resemble those made by humans.

One proposed approach to implementing artificial intuition is models based on semantic networks. Such models enable computers to recognize patterns and correlations in a way that resembles human intuitive thinking (Johanssen & Wang, 2021). For example, in data management systems, semantic networks can help identify key information in complex data sets, allowing for faster decision-making. Similar approaches have found applications in image analysis, where intuitive algorithms recognize anomalies in medical images.

Simon (1995) emphasized that AI has already reached a stage where it can simulate phenomena typical of the human mind, such as intuition, insight, and inspiration. Incorporating artificial intuition into AI systems allows for novel ways of identifying and processing information, leading to more efficient and effective decision-making processes. Trovati et al. (2022) suggest that intuitive systems play a key role in the future of technology, especially in areas requiring rapid response to changing conditions, such as crisis management or dynamic logistics planning.

The development of artificial intuition in computers opens up new possibilities for AI, especially in the areas of decision-making and data analysis (Diaz-Hernández, 2017). Although there are challenges related to its implementation and use, the potential benefits – such as increased efficiency and speed of operation – mean that artificial intuition could play a key role in the future of technology (Liebowitz, 2024). To achieve the full potential of this concept, further research is needed that considers both the technical and ethical aspects of its development.

18.5 Methodology

In April 2024, Liebowitz and Miragliotta at the Politechnic University of Milan (Polimi) conducted an initial survey using the Qualtrics platform. The study aimed to answer the question, "Can computers exhibit 'artificial intuition'?" The sample was later expanded in fall 2024 to 120 persons, through the help of Paliszkiewicz at the Warsaw University of Life Sciences. The results regarding the key questions asked in the study are presented below.

18.5.1 Results

The survey results provide interesting insights into expectations for the development of AI in various areas of functionality and capabilities. Respondents expressed optimism about AI's ability to simulate "intuitive hunches" (Figure 18.1). The largest group (43%) believes that AI will achieve this ability within one–five years (which may be optimistic), with another 16% predicting it will happen within six–ten years. At the same time, 12% of respondents believe that AI will never be able to fully emulate this ability, indicating a level of skepticism that still exists. Expectations of rapid development in this area suggest that intuitive decision-making is a key AI capability. Only 8% of respondents believe that AI can understand and express emotions, but as many as 47% predict that AI will achieve this within one–five years.

Nevertheless, 14% believe that AI will never acquire such skills. These results indicate high expectations for the development of emotional interaction in AI and concern about technology limitations in this area. In terms of biases, as many as 34% of respondents believe that AI will start to exhibit human-like biases resulting from excessive trust in data within one–five years. At the same time, 18% of respondents claim that AI will never be free from this problem. These results show an awareness of the risk of transferring human biases to AI algorithms, which requires further action to eliminate them. Most respondents (44%) believe that AI will be able to explain its decisions within the next one–five years, and 20% say that it can already do this. Only 6% of respondents believe that AI will never achieve this ability. These results reflect the growing demand for transparency in AI decisions, which is crucial in many applications. Respondents have different assessments of AI's ability to "create and intuit." Precisely, 38% of respondents believe that AI will achieve this ability within one–five years, but as many as 14% believe that AI will never equal humans in this area. Expectations for AI in this area are high, but there is also a significant level of skepticism.

Regarding AI self-awareness, only 9% of respondents believe that AI already has it, but as many as 38% predict that AI will achieve this ability within one–five years. However, 16% of respondents believe that AI will never be aware of its actions and thoughts. These results show that the ability to be self-aware is seen as a key but difficult-to-achieve stage of technological development. Nearly half of the respondents (49%) believe that a new field of "humanics," "combining people and AI," will begin to develop dynamically in the next one–five years. These expectations indicate the need to prepare people better to cooperate with intelligent systems. Precisely, 39% of respondents believe that organizational success depends primarily on people, while 29% indicate that AI already plays a significant role. These results underscore the continued importance of the human factor in business operations, even as technology continues to evolve. These results show optimism among respondents and an appreciation of the limitations of technology. As many as 66% of respondents believe intuition can be developed like other skills.

These results indicate a belief in improving both human and artificial intuition. So, 61% of respondents say that executive decisions in the future will be primarily data-driven, underlining the importance of analytics in decision-making. At the same time, 48% of respondents believe that AI can incorporate cultural elements into its intuitive decisions within one–five years. This result reflects AI's expectations to understand and integrate diverse cultural contexts better. The survey results show great optimism about the development of AI in many functional areas, such as intuition, creativity, and the ability to explain decisions. At the same time, there are significant areas of skepticism, especially regarding AI's full mastery of human capabilities. These results underscore the need for further research and development of the technology and efforts to increase its transparency, integration with culture, and understanding of human values.

The results in Figure 18.2 show what respondents believe about the capabilities of AI and its potential applications in the future. Most survey participants (63%) agree that a cluster of AI computers can develop the ability to "collective intuition" based on domain knowledge, experience, and cognitive abilities. Only 26% disagree, and 11% are undecided. There was similar optimism when asked

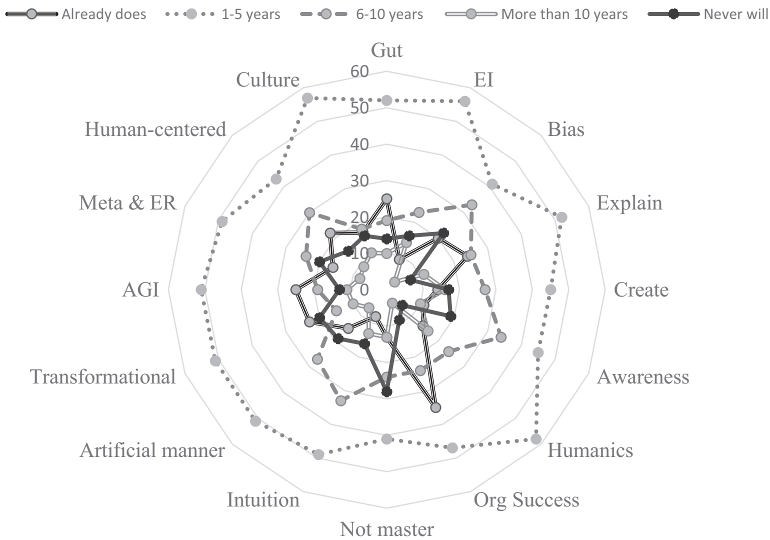

Figure 18.1 Expected development of AI capabilities in various functional areas

Source: Own study

about delegating decisions about the "relevance" of information to AI language models (LLMs). Consequently, 61% of respondents believe this is possible, 28% disagree, and 11% are undecided. More than half of respondents (58%) believe that intuitive and analytical approaches to decision-making can be combined into a unified rule-based model. However, just over a quarter (28%) believe this is impossible, and 14% are undecided.

When it comes to executive decisions, 61% of survey participants indicate that the future will belong to decisions based on evidence and data analysis rather than intuition. However, 30% believe that intuition will continue to play a key role in such processes, while 9% remain undecided. The most pronounced support (66%) came when asked whether intuition is a skill that can be improved, like any other trait. 28% disagreed, and only 6% had no opinion. The results indicate a prevailing optimism about the development of AI and its ability to integrate more advanced capabilities, such as collective intuition or data-driven decision-making. However, a noticeable minority of respondents expressed doubts, suggesting the need for further research and discussion on the future of AI and its impact on decision-making.

18.5.2 Discussion

The research results on the development of intuition in AI, considering other data and contexts, allow for a broad view of this technology's possibilities,

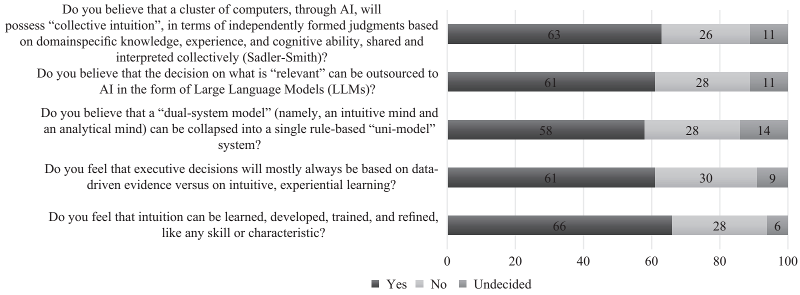

Figure 18.2 Prospects for the development of AI in the field of intuition, decision-making, and integration of AI models

Source: Own study

limitations, and potential development directions. The results presented in the research can be confronted with the results of other studies, which indicate challenges and different perspectives related to implementing intuitive mechanisms in AI systems. Tsvetkova's (2022) research indicates that intuition remains a competitive advantage for humans, especially when data is fragmented or access to information is limited. This contrasts with the optimistic results of the discussed study, in which the majority of respondents (43%) believe that AI will achieve the ability of intuition within one–five years. Tsvetkova's results suggest that even in the case of large technological advances, AI may encounter fundamental barriers related to the lack of contextual knowledge, which is the basis of human intuition. Similarly, the results of research conducted by Sun (2009), which concerned the development of "artificial intuitive reasoning," indicate difficulties related to the low quality of data and the need to model complex relationships between information. Although the Sun (2009) study provides evidence of the possibility of AI imitating human intuition in some limited scenarios, the results do not indicate a quick achievement of universal intuitive abilities. The discussed research, although optimistic, may need to pay more attention to this aspect. It is also worth paying attention to the research results by Abassi et al., (2022), which indicate the effectiveness of integrating human intuition and AI algorithms in managerial decision-making. This study suggests that a synergistic approach, in which human intuition complements the analytical abilities of AI, can yield significantly better results than AI alone. These results complement the conclusions of the discussed study, in which as many as 66% of respondents believe that intuition can be developed and combined with technology, which indicates the wide possibilities of hybrid decision models. Another important aspect is the issue of transparency of AI operations, which is crucial for building trust in these systems. The results of the research by Bhatt et al., (2019) emphasize that the interpretability of AI systems is fundamental for users. However, the discussed study suggests that only 20% of respondents believe that AI can already explain its decisions. These results indicate a discrepancy between user expectations and the actual state of technology development. In turn, the research by Dashkov and Nesterova (2021) emphasizes that trust in AI depends not only on transparency but also on the ability to have predictable and consistent interactions, which should be a priority in the further development of the technology. An interesting contrast can also be seen in cultural intuition. The results of Johanssen and Wang (2021) suggest that intuition in AI that takes into account cultural differences may be difficult to implement due to the lack of universal patterns. The results of the discussed study, in which 48% of respondents believe that AI can integrate diverse cultural contexts in its decisions, can therefore be seen as optimistic but also require deeper analysis. Winter and Carusi's (2022) research points to the potential of hybrid models that combine human intuition with AI's analytical capabilities, which is also reflected in the current study. The results emphasize that trust in AI can be built when human decisions are supported by technologies that are both precise and tailored to user needs. However, these studies also show that overreliance on AI can lead to a false sense of security, which requires a balance in the design of such systems. The comparison of the results of the discussed study with those of other scientific works indicates that there are many challenges ahead of the development of intuition in AI. Although the respondents expressed great optimism about the possibility of AI achieving intuitive abilities, the scientific literature suggests that full implementation of this ability requires a much more complex approach, considering technological, cultural, and ethical constraints. Combining human intuition with the precision of AI algorithms may be a key development direction. However, advanced research and careful design are required to avoid potential pitfalls and risks.

18.6 Conclusion

The development of artificial intuition in AI systems opens up new possibilities but requires a multidimensional approach that considers both technical, social, and ethical aspects. Intuition can be useful in designing AI systems, but its implementation risks introducing biases or difficulties in interpreting the results. Combining human intuition with AI analytical capabilities offers the best results, increasing efficiency and user trust in technology. The development of artificial intuition is an area that combines technological, social, and ethical challenges. The right approach can significantly contribute to increasing the efficiency of AI systems and building lasting user trust. However, achieving this goal requires technological innovation and a deep understanding of human needs and limitations.

The study presented in this chapter has several significant limitations. The study results are based on subjective assessments of respondents, whose perspectives may be strongly influenced by their professional experience and level of technological knowledge. Experts may be more aware of technical limitations, while management may focus on strategic opportunities. The rapid development of AI technology means that predictions about its capabilities can quickly become outdated. The study results should be interpreted taking into account potential changes in the short term. The number of respondents is relatively small, allowing the study to expand to include other professional or social groups.

To better understand and leverage the potential of AI, future research should focus on developing intuitive models that can operate in real time, taking into account the context of decisions made. It is also worth investigating how different cultural groups perceive AI intuition and what are the ethical boundaries of its implementation. Finally, the development of AI should consider users' diversity and the need to ensure inclusiveness and fairness when using new technologies.

Acknowledgment

This research is partly funded by the University of Economics Ho Chi Minh City, Vietnam.

References

Abbasi, M. F., Bilal, M., & Rasheed, K. (2022). The role of human intuition in AI-Aided managerial decision making: A review. *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, 713–718. IEEE.

Bhatt, U., Ravikumar, P., & Moura, J. M. (2019). Building human-machine trust via interpretability. *AAAI Conference on Artificial Intelligence*, *33*(1), 9919–9920.

Chen, V., Liao, Q. V., Vaughan, J. W., & Bansal, G. (2023). Understanding the role of human intuition on reliance in human-AI decision-making with explanations. *Proceedings of the ACM on Human-Computer Interaction, 7*, 1–32.

Dashkov, A. A., & Nesterova, Y. (2021). Building trust when using artificial intelligence. *E-Management*, *4*(2), 28–35.

Demartini, G. et al. (2024). Opinion: Who determines what is relevant? humans or AI? why not both?. *Communications of the ACM*, *67*(4), 31–34.

Denford, J., Dawson, G, & Desouza, K. (2024). National AI strategic plans for the public versus private sectors: A cross-cultural configurational analysis. *Proceedings of the 57th Hawaii International Conference on System Sciences,* Hawaii, 1840–1860.

Diaz-Hernández, O. (2017). Analysis of human intuition towards artificial intuition synthesis for robotics. *Libraries & Information Technology Journal*, *1*(1), 1–17.

Gigerenzer, G. (2023). *The intelligence of intuition*. Cambridge University Press.

Johanssen, J., & Wang, X. (2021). Artificial intuition in tech journalism on AI: Imagining the human subject. *Human-Machine Communication*, *2*, 173–190.

Johnny, O., Trovati, M., & Ray, J. (2019). Towards a computational model of artificial intuition and decision making. In L. Barolli, H. Nishino, & H. Miwa (Eds.), *International Workshop on Intelligent Networking and Collaborative Systems*, 463–472. Springer.

Korzynski, P., Mazurek, G., Altmann, A., Ejdys, J., Kazlauskaite, R., Paliszkiewicz, J., ... & Ziemba, E. (2023). Generative artificial intelligence as a new context for management theories: analysis of chatGPT. *Central European Management Journal*, *31*(1), 3–13.

Liebowitz, J. (2018). Using your intuition. In J. Liebowitz (Ed.), *How well do executives trust their intuition* (pp. 3–26). Taylor & Francis.

Liebowitz, J. (2024) (ed.). *Developing the intuitive executive: Using analytics and intuition for success*. Taylor & Francis.

Liebowitz, J. (2025) (ed.). *Regulating hate speech created by generative AI.* Taylor & Francis.

Liebowitz, J., Chan, Y., Jenkin, T., Spicker, D., Paliszkiewicz J., & Babiloni, F. (2019). If numbers could 'feel': How well do executives trust their intuition?. *VINE Journal of Information and Knowledge Management Systems*, *49*(3), 531–545.

Paliszkiewicz, J., & Gołuchowski, J. (2024). *Trust in artificial intelligence directions. Trust and artificial intelligence: Development and application of AI technology.* Taylor & Francis.

Pedwell, C. (2022). Speculative machines and us: More-than-human intuition and the algorithmic condition. *Cultural Studies*, *38*(2), 188–218.

Sadler-Smith, E. (2023). *Intuition in business.* Oxford University Press.

Sanders, N., & Wood J. (2020). *The humachine: Humankind, machines, and the future of enterprise*. Routledge.

Simon, H. A. (1995). Explaining the ineffable: AI on the topics of intuition, insight and inspiration. *International Joint Conference on Artificial Intelligence*, *1*, 939–948.

Skarzynska, E., & Paliszkiewicz, J. (2025). The use of generative artificial intelligence (GenAI) capabilities for early detection of threats in the digital environment: The good side of GenAI. In: Liebowitz (ed.) *Regulating hate speech created by generative AI*, Taylor & Francis, 91–104.

Solan, L. M. (2013). Intuition versus algorithm: The case of forensic authorship attribution. *Journal of Law and Policy*, *21*, 551.

Sun, Y. C. (2009). *The development of an artificially intuitive reasoner*. McGill University.

Trovati, M., Johnny, O., Xu, X., & Polatidis, N. (2022). A new model for artificial intuition. In: *International Conference on Artificial Neural Networks,* 454–465. Springer International Publishing.

Trovati, M., Teli, K., Polatidis, N., Cullen, UA., & Bolton, S. (2023). Artificial intuition for automated decision-making. *Applied Artificial Intelligence*, *37*(1), 2230749.

Tsvetkova, A. D. (2022). Investigator's Intuition: Conditions of Formation and the Possibility of Implementing Artificial Intelligence. *Полицейская и следственая деятельность, 6*, 23–46.

Van dhe Herik, H. (2015). Computers and intuition. *International Computer Games Association Journal*, *38*, 195–208.

Vincent, V. U. (2021). Integrating intuition and artificial intelligence in organizational decision-making. *Business Horizons*, *64*(4), 425–438.

Winter, P., & Carusi, A. (2022). 'If you are going to trust the machine, then that trust has got to be based on something'. *Science & Technology Studies*, *35*(4), 58–77.

Zhang, J., Bentahar, J., Falcone, R., Norman, T. J., & Sensoy, M. (2019). Introduction to the special section on trust and AI. *ACM Transactions on Internet Technology (TOIT)*, *19*, 1–3.

Zhang, Z., Lei, Y., & Li, H. (2016). Approaching the distinction between intuition and insight. *Frontiers in Psychology*, *7*, 1195.

19 To Trust or Not to Trust

Is Generative Artificial Intelligence Taking Sides in Higher Education?

Cezar SCARLAT and Alexandra IOANID

19.1 Introduction

On 6 September 2024, the news agencies announced that the United States, EU, and United Kingdom have signed, the previous day, the Council of Europe's convention on AI – which was the *first legally binding international treaty on the use of AI*.

This convention was "drafted over two years by more than 50 countries, including Canada, Israel, Japan and Australia," while other countries are continuing to sign the pact; it requires signatories "to be accountable for any harmful and discriminatory outcomes of AI systems" and that "outputs respect equality and privacy rights" (Murgia and Espinoza, 2024). Notably, this treaty "comes as governments develop a host of new regulations and agreements to oversee AI software" – including *European Artificial Intelligence Act* (European Parliament, 2024), *G7 Leaders' Statement* on AI process (European Commission, October 2023), and *Bletchley Declaration* (AI Safety Summit, November 2023), which was signed by 29 participant countries. The G7 Leaders' Statement refers particularly to international *Guiding Principles* for all AI actors in the AI eco-system and a voluntary *Code of Conduct* for organizations developing advances AI systems.

In education, AI is being used more and more, whether we are aware of it or not. The reluctance of teachers is justified mainly due to the excessive use of ChatGPT-type resources by students, who no longer go through the learning process. Thus, students submit assignments that do not belong to them, or the assignments contain errors that the AI systems generate (Baidoo-Anu et al., 2023).

Many prestigious universities are concerned with regulating the use of ChatGPT or similar tools by students. For example, the University of Cambridge (2024) publishes the following recommendations on its website:

> Artificial intelligence tools can be used by students to enhance their personal study, research and formative work… they are advised to discuss this topic with the professor coordinator to understand how I can do this without violating the ethics and academic integrity regulations.

DOI: 10.4324/9781003586944-22

Kumar looked at the usefulness of artificial intelligence (AI) in academic writing and observed that although the text is generated in less than two minutes and the content is systematic, it still lacks academic rigor (Kumar, 2023).

The time for discussions about softer than software, and intangible issues related to AI development – as AI ethics and AI trust – has arrived. Is university ready for it?

19.2 Scope of Work

The focus of this chapter is on the influence of the GenAI technologies (in particular ChatGPT) on the relationship teacher-student in the context of higher education. In other words, trust-related GenAI-mediated relationships are under scrutiny, at individual and organizational levels, in light of the *triadic and multi-triad models* (Scarlat, 2021a), and considering the trust triad relationships (Scarlat & Ioanid, 2023) in particular (Figure 19.1).

The trust relationship between students and professors with regards to using GenAI or not has more importance on the professor's side as they need to evaluate students, for example to check if the homework or report was really done by the student or not (it could be generated automatically, or even written by someone else). Also, the students need to trust the professors and to admit when using GenAI and in which phases of their work.

At the same time, trust relationship exists between students that access the AI technologies and the technology itself that generates more or less accurate information.

The next sections explore generative AI emergence as well as influences exercised by GenAI on main actors active in higher education (professors and students) and how the GenAI-related technologies impact (either help or hinder) the trust relationships among these actors. Discussions on a pilot survey, conclusions, and further research paths close this chapter.

19.3 AI Emergence and Its Impact

In general, the issue of reciprocal trust between students and their mentors might be extended at higher education organization-level in the sense of dynamic force of academics *versus* academia as set of relatively rigid regulations.

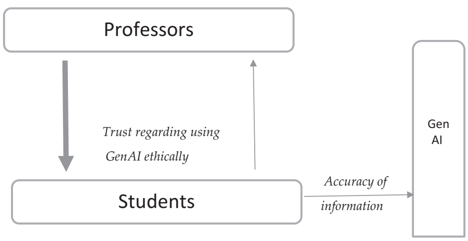

Figure 19.1 The scope of work in the context of higher education relationship, in which students and professors need to trust each other.

Source: Authors

What role will GenAI play as an intermediary in this relationship? Would it "take sides"? What advantages and disadvantages occur for each side? These all are legitimate questions. Although they are posed for the sake of research, amid specific interests of each side, both sides ultimately have a common goal: better educated graduates.

19.3.1 ChatGPT Emergence

AI emergence coincided with the retrogression of coronavirus pandemic – from January 2020 to May 2022 (Rigby & Satija, 2023) – and, probably, its emergence was accelerated by this pandemic (Scarlat & Stănciulescu, 2021; Scarlat et al., 2022; Scarlat, 2023a). Scarlat (2023b) also mentions the transformation process of work settings and habits as post-pandemic effects, via emergence of AI technologies (specifically the increasing use of ChatGPT).

After ChatGPT "dazzled the world last November" [2022], *The Economist* (2023) echoed the peoples' first interest and main concern: "whether AI will obliterate entire categories of jobs"; and concluded that

> evidence so far hints instead at another, more hopeful possibility: by augmenting workers, rather than replacing them altogether, generative AI could lead both to better jobs and better experiences for customers. After years of frustration and rage, that would come as a relief to people on both sides of the customer-service line.
>
> (2023, p.58).

As survey of 100,000 Danish workers (Humlum & Vestergaard, 2024) shows that collaboration with a virtual assistant like ChatGPT can halve time spent on about a third of work tasks. "AI will transform the global economy without booting people out of jobs" (Economist, 2024a, p. 58).

Amid general concern connected to GenAI-related business effects–productivity and jobs–the academic world has its specific, supplementary worry: *What impact would the use of GenAI-related technologies have on higher education processes and relationships?*

19.3.2 Can We Trust AI?

This question has no definite answer, as the opinions are shared. *The* Economist (2024b, p. 10), citing Adam and Carter (2023), doubted:

> Existing LLMs could not be trusted to produce finished intelligence reports, which require lateral thinking and counter-factual ('what if') reasoning. New hybrid models would be needed for that, such as neurosymbolic networks, which combine the statistical approach of neural networks with old-fashioned logic-based ('if this, then that') AI. Until then, the LLMs were best confined to early stages of drafting 'an extremely junior analyst'.

On the other side, the same publication in a balanced report cites Schoenegger et al. (2024) with a counterargument: "Volunteers given access to LLMs made forecasts that were 23% more accurate than a control group."

Cao et al. (2024), exploring a vast literature, discuss the future of soft robots with AI, since "a diversity of functions like real-time object classification, gesture estimation, and touch modality recognition have already been achieved by the soft robots with the assist of ML [machine learning]." (2024*,* p.214).

As a research report produced by Microsoft and LinkedIn has shown that "75% of global 'knowledge workers' (folk who sit in front of a computer all day) use it" (Economist, 2024a, p. 57), which demonstrated that large majority of computer users (three quarters) "are, by such accounts, already in an AI world." However, according to the same source, America's Census Bureau asked AI-related questions to firms "in a wider range of industries than Microsoft and LinkedIn" and found that

> only 5% of businesses have used AI in the past fortnight. […] It is a similar story elsewhere. According to official Canadian numbers, 6% of the country firms used AI to make goods and provide services in the past 12 months.

Also, "British surveys suggest use there is higher – at 20% of all businesses in March [2024] – though the questions are asked differently. And even in Britain use is growing slowly."(Economist, 2024a*,* p. 58).

As far as areas where GenAI is applied, *The Economist* (2024a, p. 58) shows that

> companies that are going beyond experimentation are using generative AI for a narrow range of tasks. Streamlining customer service is perhaps most common. ADP, a payroll firm, boasts of 'a new feature that enables our small-business clients to…leverage gen AI to answer questions and better understand how to initiate an HR action'. Others use the tech for marketing. Verizon, a telecoms firm, says it employs AI to create a better 'personalised plan recommendation' for its customers; Starbucks, a coffee chain, uses it to make 'more personalised customer offers'.

Overall, even if expected higher productivity increase is moderate, the investment in GenAI continues to be impressive (Economist, 2024a, p. 57). According to the same source, the GenAI impact on labor market is far to be destructive as it was feared; on the contrary, the reality displays a collection of positive results (Economist, 2024a, p. 58):

- Unemployment across the rich world is below 5%, close to an all-time low
- The share of rich world workers in a job is near an all-time high
- Wage growth remains strong (all these contrary to the IMF gloomy previsions on the labor market)
- Workers are not moving between companies faster than usual

- The share of employment in white-collar professions (ranging from back-office support to copywriters) is a percentage point higher than before the pandemic.

However, *the study of GenAI-enhanced trust-related relationships among the main academy actors*, within and without university, individual and organizational – in the spirit of trust triad (Scarlat & Ioanid, 2023) – remains quite unchartered territory. Does the higher education sector follow the general trend?

19.4 A Pilot Study on the Use of AI in Higher Education

The higher education environment deserves attention in two main respects:

– University is the tipping point in which young students turn into a highly qualified labor force; and

– Academia is the place where many new ideas are debated to be refined and eventually hammered into new technologies to serve the business community, industry and, ultimately, drive the advance of society.

The novel GenAI is among these new technologies. How does it serve the higher education itself?

19.4.1 AI in Higher Education

School was always well-regarded as a trusted institution – since education has its roots intertwined with religion (Namdeo, 2024, p. 325; Scarlat, 2021b, p. 267). The oldest schools in antiquity (Marrou, 1948) appeared in Mesopotamia by religious temples (to train the priests and scribes), China (as early as the 3rd millennium BC), and Egypt by the pharaoh court (for the royal family and administrators) – in general, limited for the use of ruling families and religious elites (Lyons, 2013). Later, during the Middle Ages, the precursors of modern universities from the 6th century (Riché, 1978) originated in religious establishments – either cathedral or monastic schools.

As the time has passed and technology has developed, Namdeo (2024, p. 325) notes: "the spread of educational technology is changing the nature of learning, having effect on classrooms, schools, online platforms, and even schooling at home" and has the potential to revolutionize the entire educational system. Significantly, the impact of AI technologies on higher education teaching and learning was observed and analyzed by scholars (Almaraz-Menéndez et al., 2022; Verma & Tomar, 2021) even *before the novel GenAI emerged*.

Amid AI-driven adaptive learning systems, which customize educational content and streamline administrative tasks for educators, Kocoglu (2024, p. xiii) singled out some challenges introduced by AI: "ethical concerns and privacy issues that necessitate thorough scrutiny and careful implementation," emphasizing "the importance of comprehensive teacher training and ongoing ethical evaluation to ensure responsible use of AI in education."

Based on the findings from the *Gordon Commission on the Future of Assessment* (Gordon, 2020), which supports the ideas that educational assessments can and should be used to inform and improve teaching and learning processes and outcomes, Namdeo (2024, p. 326) is pessimistic about the negative impact that Covid-19 pandemic had on education in India: "[it] caused a global learning crisis that is the biggest in modern history and has set back the education industry by three decades."

Research conducted by Grigorescu-Pirvu and Scarlat (2023) supports the pandemic's negative impact on a limited sample in Romanian higher education environment: a master program, 2020–2022 (2023, pp. 70–72). However, the *effect of pandemic as technology accelerator is undeniable* (Scarlat & Stănciulescu, 2021; Scarlat et al., 2022; Scarlat, 2023a) – so that the combined effect of coronavirus pandemic on education system is not necessarily controversial, but yet to be seen in time.

Overall, the emergence of GenAI technology in November 2022 (Economist, 2023) was the definitely strong *answer to the accelerated technology development to the challenge of the paradigm shift*[1] *in the educator's role when the critical point in education was reached* (Scarlat, 2021b, pp. 268–272).

The extant literature on new education technologies (Aslam & Nisar, 2023; Asian Development Bank, 2023; Hai-Jew, 2024) is optimistic and emphasizes their advantages. Reckoning that ChatGPT "could revolutionize personalized adaptive learning" a report of Asian Development Bank (2023, p. 1) emphasized the advantages in an optimistic note:

> to enable personalized learning at scale, helping higher education institutions improve course success and reduce student withdrawal rates. In sum, adaptive and immersive learning technologies are transforming teaching and learning, making truly personalized learning a reality. These technologies facilitate student-centered learning, promoting lifelong learning, and help expand opportunities across the globe.

Cooper (2023) provided a documented example of how ChatGTP answered questions related to science education itself, while Le-Nguyen and Tran (2024) provided case studies and ethical frameworks for responsible AI integration.

The students' traditional trust in traditional education system and teachers (i.e. before coronavirus pandemic and ChatGPT emergence) is gone. The "dual" education system that was carved by the post-Covid emergence of GenAI (Grigorescu-Pîrvu & Scarlat, 2023) is going to take new shapes and colors, but not much of it is addressing the hot and "softer than software" subjects (as trust and ethics) in relation to the use of AI in higher education:

- Are faculty members ready to accept and become familiar with the current use of GenAI-enabled education technologies (i.e. are they aware of the paradigm shift in education)?

- Do the users of new technologies, both students and teachers, consider GenAI as trustworthy?
- Is there a reciprocal trust-based relationship between students and their mentors (both parties being well-aware about responsive and ethical use of GenAI technologies)?

This study is an attempt to contribute to the general picture of using GenAI in higher education with a flash on the Romanian university life.

19.4.2 Use of AI in Education. A Pilot Survey in Romania

The authors conducted a pilot study using a cross-sectional survey method, targeting both students and professors at the National University of Science and Technology Politehnica Bucharest, Romania. A total of 109 students from various educational levels (bachelor, master, and PhD) and 12 professors participated in the study, which took place between July and September 2024. The data was collected using a quiz instrument administered during dedicated meetings, followed by semi-structured interviews with participants who provided additional details, blending both quantitative (questionnaire) and qualitative (interviews) techniques.

All students who participated in the research declared that they had used technologies based on AI at least once in the completion of the assignments received during their studies. On the other hand, only 25% of the faculty members who participated in the survey declared they used AI platforms at least once, and others prefer not to answer this question or declare they never used AI technologies.

Faculty members may not yet be fully prepared to embrace or adapt to the current use of GenAI-enabled educational technologies. With only 25% of those surveyed reporting prior use of AI platforms, while others either declined to respond or stated they had never used AI tools, a gap in acceptance of AI-driven changes in education is clear. This highlights the need for increased awareness, training, and openness to GenAI, both among faculty members and within the institution.

19.4.2.1 Purpose of Using GenAI

When asked about the purpose of using AI, 83% of students reported using ChatGPT to complete their homework, 77% for writing research reports, 41% to find correct answers during exams, 70% to search for general, personal information (such as tourist recommendations or restaurants), and 22% to assist with tasks at work.

In contrast, when professors who had used AI were asked the same question, approximately 70% said they tried AI out of curiosity without relying on it for creating reports or documents, while the remaining 30% indicated they used AI to generate documents or reports.

Recognizing that ChatGPT is primarily used to generate homework and research reports, the authors explored during interviews which phases of the research process AI is most often applied to. Student responses indicated that 70% use AI to generate ideas, which might be acceptable if limited to this stage. However, 40% of students admitted using AI, particularly ChatGPT, to write introductions, describe methodologies, interpret data, and even draft conclusions.

In terms of ethical considerations, when asked if they disclose their use of ChatGPT for generating reports and homework, 73% of respondents stated they are transparent about it, while 27% avoid informing their professors. During interviews, students who chose not to disclose their use of AI explained that they feared how professors might react and were concerned about receiving lower grades (79%), or, in some cases, they were solely focused on passing the course without regard for the professors' opinions (13%).

Student responses reveal a lack of trust in their professors, leading many to avoid admitting that they use AI to complete homework and research reports. The significant proportion of students in this study who withhold their use of GenAI supports the perception among professors that students are increasingly relying on AI while investing less effort in their studies. The absence of reliable technologies to accurately detect AI-generated content further complicates professors' ability to distinguish between students' original work and AI-generated text.

The authors also examined *students' trust in AI technology* overall. When asked whether they trust AI to generate homework and reports, 90% of the students indicated that they do. In contrast, 10% expressed doubts about the accuracy of the information and mentioned that they seek alternative sources to validate the automatically generated content. Based on the findings, the authors noticed that *students seem to trust the technology itself,* but they are reluctant about how their professors might perceive its use.

19.4.2.2 Reciprocal Trust-Based Relationship Between Students and Their Professors

It does not appear to be a reciprocal trust-based relationship between students and their professors regarding the use of GenAI technologies.

While 90% of students trust AI for generating homework and reports, many hesitate to disclose their use of AI to their professors due to concerns about how it might be perceived, fearing negative reactions and, possibly, lower grades.

In case of faculty members, the *gap between high familiarity and low acceptance* – i.e. limited engagement with AI (only 25% reporting prior use) – suggests that *mentors may not be fully aware of the ethical implications or benefits of GenAI, further complicating the trust dynamic.*

For a truly *reciprocal trust-based relationship, both mentors and disciples would need to have a common understanding* on what would be a responsible and ethical use of AI in higher education. Flexible but clear guidelines and not rigid regulations regarding the use of GenAI would help it.

19.5 Conclusion, Limitation, and Further Studies

The results of this pilot study complete the image of AI use in higher education, bridging the gap related to its use in Romanian environment.

The research brings to light an issue that stakeholders tend to avoid discussing, even though many students use GenAI for their homework and reports, and most professors are aware of this but lack the tools to control it.

In other words, the university (as organization) should take further steps to implement the EU Regulation 2024/1689 of the European Parliament (2024) – in order to be prepared to responsibly and ethically administer the structure and processes GenAI-enabled within the university.

As it is clear that *GenAI is here to stay*, the trust relationship between professors and students needs to evolve in response to these new technologies. Students are encouraged to be more transparent and ethical in their use of GenAI, while professors should adapt their evaluation methods so that the use of AI becomes less of a concern.

One limitation of the study is that it was conducted exclusively at a technical university in Romania; however, the findings align with global trends regarding students' ethical use of AI solutions.

A path for further studies is to analyze the *honesty of the higher education actors* (any of them), using *sentiment analysis* (Chang & Ma, 2024; Mahalakshmi et al., 2024).

In addition, since the use of AI robots is developing, and integrated chatbots will assist higher education (Shen, 2024), the next question to answer would be: How fair would be their use?

A possible scenario is imagined by the contemporary American novelist Hellen Phillips: a big city of the future, darken by smoke and chemicals, where people are out of jobs because of new generation of AI robots, called *hums*. In that distant future, the main character, May Webb, was teaching communication skills to AI. Eventually she is out her job simply because the AI network became able of self-teaching (Phillips, 2024).

May Webb's job is teaching communication skills to AI …

Note

1 The *paradigm shift* in higher education regarding the educator's role as formulated by Scarlat (2020, 2021b) should not be confused with the *paradigm shift* in the Indian

education system: "transforming Ancient India into Digital India" (Jatwani et al., 2021).

References

Adam, C., & Carter, R. (2023). *Large language models and intelligence analysis. Expert analysis*. The Alan Turing Institute. Centre for Emerging Technology and Security – CETaS.

AI Safety Summit. (2023). *The bletchley declaration by countries attending the AI safety summit, 1–2 November 2023*. Policy paper. Published 1 November 2023. Official website of the UK Government: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.

Almaraz-Menéndez, F., Maz-Machado, A., López-Esteban, C., & Almaraz-López, C. (Eds.), (2022). *Strategy, Policy, Practice, and Governance for AI in Higher Education Institutions.* IGI Global. https://doi.org/10.4018/978-1-7998–9247-2.

Asian Development Bank. (2023). *Reimagine tech-inclusive education*: *Evidence, practices, and road map.* Asian Development Bank. https://dx.doi.org/10.22617/TCS230233.

Aslam, M. S., & Nisar, S. (2023). *Artificial intelligence applications using chatGPT in education: case studies and practices*. IGI Global. https://doi.org/10.4018/978-1-6684–9300–7.

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of chatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52–62.

Cao, Y., Xu, B., Li, B., & Fu, H. (2024). Advanced design of soft robots with artificial intelligence. *Nano-Micro Letters* (pp. 213–260). Published online: 13 June 2024. Shanghai Jiao Tong University Press and Springer. https://link.springer.com/content/pdf/10.1007/s40820-024-01423-3.pdf.

Chang, W., & Ma, S. (2024). A sentiment analysis model based on attention map convolutional network. *International Journal of Information Technologies and Systems Approach, 17*(1), 1–14, January-December 2024. https://doi.org/10.4018/IJITSA.348658.

Cooper, G. (2023). Examining science education in chatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology, 32,* 444–452. https://doi.org/10.1007/s10956-023-10039-y.

Economist. (2023). Your call is important: Customer service is getting worse. What could fix it? *The Economist*, September 30th, 2023, p. 58.

Economist. (2024a). Technology and the economy: A sequence of zeros. *The Economist,* July 6th, 2024, 57–58.

Economist. (2024b). Artificial intelligence. SpookGPT. *The Economist*, *Technology Quarterly*, July 6th, 2024, 9–10.

European Commission. (2023). *G7 Leaders' Statement on the Hiroshima AI Process, 30 October 2023*. European Commission. https://digital-strategy.ec.europa.eu/en/library/g7-leaders-statement-hiroshima-ai-process.

European Parliament. (2024). Regulation (EU) 2024/1689 of the European parliament and of the council of 13 June 2024. *Official Journal of the European Union*, 12 July 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/0j.

Gordon, E. W. (2020). Toward assessment in the service of learning. *Educational Measurement: Issues and Practice, 39*(3), 72–78. https://doi.org/10.1111/emip.12370.

Grigorescu-Pîrvu, A. F., & Scarlat, C. (2023). Online technologies in teaching and learning. lessons learnt while teaching during COVID-19 pandemic in Romania: Towards a "dual" education system. In M. Carmo (Ed.), *Education applications & developments – EAD VIII. advances in education and educational trends* series (pp. 63–80). World Institute for Advanced Research and Science – WIARS, InScience. https://insciencepress.org/education-applications-developments-viii/.

Hai-Jew, S. (Ed.), (2024). *Generative AI in teaching and learning*. IGI Global. https://doi.org/10.4018/979-8-3693-0074-9.

Humlum, A., & Vestergaard, E. (2024). *The adoption of chatGPT.* Discussion paper series IZA DP No. 16992. May 2024. IZA institute of labor economics. Retrieved July 25, 2024 from https://www.econstor.eu/bitstream/10419/299920/1/dp16992.pdf.

Jatwani, P., Tomar, P., & Dhingra, V. (2021). Technology-enabled education: Paradigm shift in higher education. In S. Verma, & P. Tomar (Eds.), *Impact of AI technologies on teaching, learning, and research in higher education* (pp. 49–61). IGI Global. https://doi.org/10.4018/978-1-7998-2.ch003.

Kocoglu, I. (2024). Revolutionizing learning: The impact of artificial intelligence on educational paradigms. Keynote lecture. In M. Carmo (Ed.), *Education and new developments 2024* (*Vol 2,* p. xiii). World Institute for Advanced Research and Science (WIARS).

Kumar, A. H. (2023). Analysis of chatGPT tool to assess the potential of its utility for academic writing in biomedical domain. *Biology, Engineering, Medicine and Science Reports, 9*(1), 24–30.

Le-Nguyen, H. -T., & Tran, T. T. (2024). Generative AI in terms of its ethical problems for both teachers and learners: Striking a balance. In S. Hai-Jew (Ed.), *Generative AI in teaching and learning* (pp. 144–173). IGI Global. https://doi.org/10.4018/979-8-3693-0074-9.ch006.

Lyons, M. (2013). *Books. A Living History.* Thames & Hudson.

Mahalakshmi, T., Zulaikha Beevi S., Navaneethakrishnan, M., Puppala, R., & Prasad Kumar, S. N. (2024). Optimized attention-driven bidirectional convolutional neural network: Recurrent neural network for facebook sentiment classification. *International Journal of Business Data Communications and Networking*, *19*(1), January-December 2024, 1–20. https://doi.org/10.4018/IJBDCN.349572.

Marrou, H. I. (1948). *Histoire de l'education dans l'antiquité.* Editions du Seuil.

Murgia, M., & Espinoza, J. (2024). Groundbreaking treaty agreed on use of AI. *Financial Times,* Friday, 6 September 2024, p. 4.

Namdeo, A. (2024). Digital education: Plea for the golden fleece. In A. Anand, A. Madaan, & A. Danielsson (Eds.), *Intersections between rights and technology* (pp. 325–342). IGI Global. https://doi.org/10.4018/979-8-3693-1127-1.ch017.

Phillips, H. (2024). *Hum*. Simon & Schuster.

Riché, P. (1978). *Education and culture in the barbarian west: From the sixth through the eighth century.* University of South Carolina Press.

Rigby, J., & Satija, B. (2023). WHO declares end to COVID global health emergency. *Reuters news,* published online May 8, 2023. Retrieved August 12, 2024 from https://www.reuters.com/business/healthcare-pharmaceuticals/covid-is-no-longer-global-health-emergency-who-2023-05-05/.

Scarlat, C. (2020). Today's higher education at a crossroads: Is the educator's role undergoing a paradigm shift?. In M. Carmo (Ed.), *Education and new developments 2020* (pp. 276–280). InScience Press.

Scarlat, C. (2021a). The triadic and multi-triad models: Typology and potential for entrepreneurship and business management studies. *International Conference on Management and Industrial Engineering*, *10,* 86–95.

Scarlat, C. (2021b). Today's higher education at a crossroads: The *critical point* and *paradigm shift* in the educator's role. In M. Carmo (Ed.), *Education applications & developments VI. advances in education and educational trends series* (pp. 265–277). InScience Press.

Scarlat, C. (2023a). Post-pandemic business effects as evolution of strategic thinking: Strategic networking and digital acceleration. *International Conference on Management and Industrial Engineering*, *11*, 417–424.

Scarlat, C. (2023b). Changes in business organizations: Transformation of work settings and habits as post-pandemic effects. *International Conference on Management and Industrial Engineering*, *11*, 439–446.

Scarlat, C., & Ioanid, A. (2023). The triad of digital trust: Organizational trust, reputation, and ethics in cyberspace. In J. Paliszkiewicz, J. L. Guerrero Cusumano, & J. Gołuchowski (Eds.), *Trust, digital business and technology. Issues and challenges* (pp. 14–32). Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781003266495-3.

Scarlat, C., & Stănciulescu, G. D. (2021). Covid-19 pandemic: Threat or accelerator? The case of Romanian publishing industry. In M. Al Haziazi, A. Al Hajri, K. P. Subramanian, & S. Muthuraman (Eds.)*, Co-creating the post covid-19 world. Exploring sustainable paths* (pp. 83–98). Arab Open University.

Scarlat, C., Stănciulescu, G. D., & Panduru, D. A. (2022). Covid-19 pandemic as accelerator: Opportunity for digital acceleration. *Journal of Internet and e-Business Studies, 2022,* ID 296375, 1–14. https://doi.org/10.5171/2022.296375.

Schoenegger, P., Greenberg, S., Grishin, A., Lewis, J., & Caviola, L. (2024). *Can AI understand human personality? -- Comparing human experts and AI systems at predicting personality correlations.* arXiv preprint, arXiv:2406.08170v1. https://doi.org/10.48550/arXiv.2406.08170. Retrieved July 24, 2024 from https://arxiv.org/pdf/2406.08170.

Shen, I. (2024). The robot wrote my college papers: Integrating chatbots to assist higher education. In S. Hai-Jew (Ed.), *Generative AI in teaching and learning* (pp. 311–327). IGI Global. https://doi.org/10.4018/979-8-3693-0074-9.ch013.

University of Cambridge. (2024). *Plagiarism and academic misconduct.* University of Cambridge. Retrieved September 20, 2024 from https://www.plagiarism.admin.cam.ac.uk/what-academic-misconduct/artificial-intelligence.

Verma, S., & Tomar, P. (Eds.) (2021). *Impact of AI technologies on teaching, learning, and research in higher education*. IGI Global. https://doi.org/10.4018/978-1-7998-4763-2.

20 To Trust or Not to Trust Generative AI-Supported Academic Scientific Research Publications as a Matter of Trust and Ethics

Cezar SCARLAT and
Alexandru-Ionuţ PETRIŞOR

20.1 Introduction

The dynamic of scientific publications follows the trend of technology development (Scarlat, 2024). *The Economist* (2024a, p. 62) shows that higher-education institutions across the world employ 15 million researchers who "produce five times the number of papers each year. […] In theory, therefore, universities should be an excellent source of productivity growth. In practice, however, the great expansion of higher education has coincided with a productivity slowdown."

Arora et al., (2023, p. 38) argue that "sluggish growth in productivity over the last three decades or more in the face of sustained growth in scientific output" happened because abstract ideas (i.e. generated by PhD research in universities) are difficult to use. Then "the firms, especially those not on the technological frontier, appear to lack the absorptive capacity to use externally supplied ideas unless they are embodied in human capital and inventions" (Scarlat, 2024, p. 39).

However, amid the general trend that features booming volume of publications, (Economist, 2023a) highlights that *publications related to significant scientific advances are not progressing at similar pace; moreover, their proportion is going down*. Even more alarmingly, Park, Leahey and Funk (2023, p. 138) declare: "slowing rates of disruption may reflect a fundamental shift in the nature of science and technology."

This state of facts and affaires (lagging productivity and decreasing proportion of *quality publications* despite booming volume of publications) *invites to careful examination* (Scarlat, 2024, p. 294).

This chapter explores just one of the possible reasons the applied research and associated productivity do not progress as rapidly as the number of publications, namely the uneven influence of novel technologies (i.e. generative artificial intelligence, GenAI) on reciprocal trust among the main actors involved in the process of academic research and publication of their scientific research results (i.e. authors, editors, and reviewers). In addition, the role of the scientific research fraud is emphasized.

DOI: 10.4324/9781003586944-23

It is suggested that *these phenomena* (scientific research fraud and sluggish growth in productivity of quality papers published, emergence of GenAI, and reciprocal trust among actors involved in scientific research and publication) *might be intertwined,* making things even more complicated.

20.2 The Scientific Research Frauds and Its Lessons

Frauds always happened, and not only in recent scientific research (Chevassus-au-Louis, 2019; Evans, 2020; Judson, 2004; Park, 2000). Just think about costly art frauds!

Like in any other sector, scientific research fraud seems to be driven by material interest: then, it is not surprising that sound frauds happened in the highly rewarding business sectors, specifically pharmaceuticals and health services: just remember cases of Erich Poehlman (obesity treatment) and Dong-Pyou Han (HIV vaccine) (Evans, 2020), and the huge fraud engineered be Elizabeth Holmes who has fallen from $9 billion to zero valuation of her blood-testing company, Theranos (Carreyrou, 2018; Senior, 2019), and currently serving a criminal penalty of 11.25 years in prison.

The next case is illustrative not only for the sector where it happened but also for the impact produced, hot debates as well as trust and ethical issues it has raised.

20.2.1 A Case: Trust But Check! [1] *(or Jan Hendrik Schön and His Fabricated Data)*

In the late '90s, the German physicist Jan Hendrik Schön was working at Bell Laboratories in New Jersey, United States. By the turn of millennia, he claimed contributions to build the world's first organic electrical laser, construction of the first ever light-emitting transistor, and "he even claimed to have built the world's smallest transistor by wiring up a single molecule" (Reich, 2009a, 2009b). While conducting experiments on organic nanomaterials, he also reported major progress on their super-conductibility (Evans, 2020).

Between 2000 and 2001, Schön published eight articles in two prestigious journals (*Nature* and *Science*); as a result, he has received accolades and was awarded several prizes, being even enrolled in the race for the Nobel Prize (Evans, 2020). In the spring of 2002, the 31-year-old physicist was considered "the world's most productive scientist [… as he] was emerging with breathtaking speed as a star researcher in physics, materials science and nanotechnology" (Reich, 2009b).

It was one the Schön's papers which described "the construction of molecular transistors that tripped the first domino when two fellow physicians attempted to patent research that showed that soft lithography could be used to make softer and gentle contact with organic molecules" (Physics, 2009): Julia Hsu and Lynn

Loo "accidentally stumbled across duplicated data and raised an alarm bell that led eventually to Schön's dismissal." In May 2002, Bell Laboratories (Schön's employer) established a committee chaired by Malcolm Beasley from Stanford University to investigate the matter.

In September 2002, after thorough investigation, a report was produced (Beasley et al., 2002), presenting details of 24 allegations of Schön's misconduct: at least 16 found evidence of misconduct (the whole data sets had been reused for a number of different experiments). Plainly, the same diagram was used to illustrate results of different experiments. *All co-authors were exonerated of scientific misconduct*, which generated widespread debates in the scientific community (Norman, 2009).

Following to *Beasley Report*, Bell Laboratories fired Schön (Evans, 2020).

Snowball effects followed: Schön's PhD degree was revoked in 2004 and remained definitive after a ten-year-legal battle; between October 2002 and May 2003, a total number of 28 articles (co)authored by Schön were withdrawn by the journals that published them between 1998 and 2002 (*Science, Nature, Physical Review, Applied Physics Letters,* and *Advanced Materials*). All prizes and awards that Schön had received before he was exposed were later withdrawn (Labini, 2016, p. 120).

Currently, ex-Doctor Schön is absent from public view.

20.2.2 Lessons Learnt: Trust and Ethics are Centerpieces

This case deserves several comments in the area of *lessons learnt*, since the right question to ask is *how was this case possible to happen*?

Schön was found guilty of scientific misconduct; therefore, he was punished administratively, losing scientific merits. Post-factum reaction looks right. However, wasn't anybody else wrong? Was anything else right?

- *A matter of peer-review quality*. A legitimate question concerns the quality of peer-reviewing of fraudulent scientific production that was published, mainly top-level journals (such as *Nature, Science,* and alike). Either quality of the peer reviewers or quality of review procedures should be questioned. Implications are both in trust and ethics areas.
- *A matter of trust*. It is quite difficult to accept that co-authors (i.e. research collaborators) did not notice anything wrong during experiments, while reporting experiment results and developing the corresponding papers. Or, equally wrong, they did not care about the work done (then: why did they accept the co-authorship?). A more rational explanation is the (blind) *trust they have had in their colleague*; trust that was not reciprocally honored (their trust was misplaced actually).
- *A matter of abnormal scientific super-production*. Referring to the articles published by Schön between 1999 and 2001, Labini (2016, p. 120) observed

that "Schön wrote, on average, a scientific paper every eight days when considering also less visible journals […] This production is really alarming for an experimental scientist, since even a theoretician, what does not need instruments or data, rarely writes more than seven articles in a whole year (although there are cases of super-production, and no one is shocked or at least worried)." Whose responsibility is to be the whistle-blower of abnormal scientific super-productions, when the promotion system in academia is, sometimes, flawed?

- *A matter of ethics*. Exoneration of co-authors caused hot debates in academic community: how the blame of misconduct should be distributed among co-authors, particularly when they share a significant part of the credit (Beasley et al., 2002).
- *Ethics of the whistle-blower*. Professor Yueh-Lin (Lynn) Loo and her colleague Julia Hsu, just happened to work at the Bell Laboratories during Schön's fake experiments and uncovered his fraud accidentally (phase one: mix of coincidence, luck, and professional acumen). Phase two: it was their genuine professional ethics that made the fraud public, and their action that singled them out.
- *Is science self-correcting?* According to Reich (2009a, p. 7), a senior manager and a press officer from Bell Laboratories "expressed the view that all was well because the fraud has come to light" (i.e. "if someone is determined to be unethical, it is not easy to detect"!) Consequently, the organization procedures were fine ("it was inevitable that he [Schön] would get found out") and "it would be unfair to question the integrity of others without proof."

As opposed to the official position of organization, Eugenie Samuel Reich, an attorney representing whistleblowers of fraud in science and technology, science author and editor, has a different opinion, around the same idea that *science is self-correcting*. Even Beasley's report concluded that "Schön's collaborators might ideally have acted as the first line of defense against fraud," despite all of them were cleared of misconduct (Reich, 2009a, pp. 7–8). Hence, there are some lessons to learn (Reich, 2009a, p. 8):

- Strengthen the monitoring function exercised by research managers and research team leaders
- *Better the scientific journals' peer-review system*[2]
- Improve the research record-keeping practices in this digital age.

Yet, there are some points to justify the inclination for publishing unverified scientific research results, even not qualified as fraud (Reich, 2009a, p. 9):

- *Placing excessive (and un-checked) trust in research colleagues*
- Researchers' desire for quicker recognition of self-merits

- Scholars' desire for rapid advance in academic hierarchy
- Organization pressure[3] for reporting research results quicker, and publishing them faster, in better ranked publications (i.e. common in most universities, for better place in international ranking)
- Increasing pressure exercised by the accelerated pace of technology advance.

Ultimately, the case presented and analyzed is an argument for a *thorough re-examination of the peer-review process, mainly for the journals that feature "fast-forward" publishing policies.*

20.2.3 Data Fabrication and Paper Withdrawing

Recent research demonstrates that *scientific research frauds are not accidents*; on the contrary, it looks like *their frequency has a positive slope*. In any case, the publication in academic journals is an important part of scientific research process, which plays a key role in fraud detection.

20.2.3.1 Data Fabrication (Doctored Data)

Evans (2020) reports revealing responses following research conducted in 2009 at the University of Edinburg:

- 2% of (scientist) respondents admitted data fabrication and/or falsification, at least once in their careers.
- 14% knew another scientist who fabricated and/or falsified research data.
- 34% admitted doubtful practices (rejecting data in contradiction to results of own previous studies, and/or rejecting data which appeared as incorrect).
- 70% knew another scientist who practiced such doubtful practices.

20.2.3.2 Bypassing Scientific Journals and Withdrawal.

There are two more specific paths, at limits of scientific research ethics: formally legal, but not entirely observing the ethics standards of fair scientific research.

- Bypassing the top-level academic journals while announcing the uncertain results of their experiments. For example, in 1989 Pons and Fleischmann announced positive results of their "cold fusion experiment" in a press conference and avoided the publication in highly regarded journals such as *Nature* or *Science*. The case was analyzed by Ackermann (2013) in *Scientometrics.*
- Publication of results, followed by retraction (paper withdrawal) for understandable reasons (not validated results, by similar experiments).

The percentage of withdrawn papers out of total number of published papers might be an indicator of scientific fairness.

In this respect, Evans (2020) mentions an index of 0.04% in 2018 (4 out of 10,000), which is a reasonably less significant value, also considering that only a part of retractions (about 40%) is caused by frauds.

The purpose of this essay is not to develop a typology of scientific fraud (even referring only to frauds like doctoring or manipulation of scientific data and experiments, therefore reporting flawed results) neither to calculate how frequent (even only these types of frauds) are.

Despite the fact that each of them might be an attractive research path to follow, our discourse continues by exploring the views of the main actors involved in the scientific research and publication process, and what role might GenAI eventually play in this respect.

20.2.4 AI Emergence and Higher Education Research

The emergence of AI coincided with the retrogression of coronavirus pandemic (Rigby & Satija, 2023) and, probably, its emergence was accelerated by this pandemic (Scarlat, 2023a). Scarlat (2023b) also mentions the transformation of work settings and habits as post-pandemic effects, via emergence of AI technologies (specifically the increased use of ChatGPT).

Regardless of the concerns and beyond the specific doubts of the scientific world, the AI investments continue to increase. Recently, *The Economist* has announced that "big five" tech firms – Alphabet, Amazon, Apple, Meta and, Microsoft – are investing vast sums: "This year [2024] they are budgeting an estimated $400bn for capital expenditures, mostly on AI-related hardware, and for research and development." (Economist, 2024b, p. 57).

While the GenAI impact on the research process is undeniable (e.g., Economist, 2023b, 2024c), the focus of this essay is on the GenAI impact *on the publication-side of the scientific research process* only, which is relatively less discussed.

20.3 Main Actors Involved in Scientific Publication, Their Roles and Views

Each professional body has its code of ethics, and each university has its own standards and regulations. Specific standards of *academic integrity* are also developing under pressure of rapid GenAI development (Mahmud, 2024).

The main actors (article author/s, journal editor/s, and article reviewer/s) and the main flows are schematized in Figure 20.1. The use of AI-powered communication interface between author/s and editors (on one side) and between editor/s and reviewer/s (on the other side) is also suggested (GenAI?).

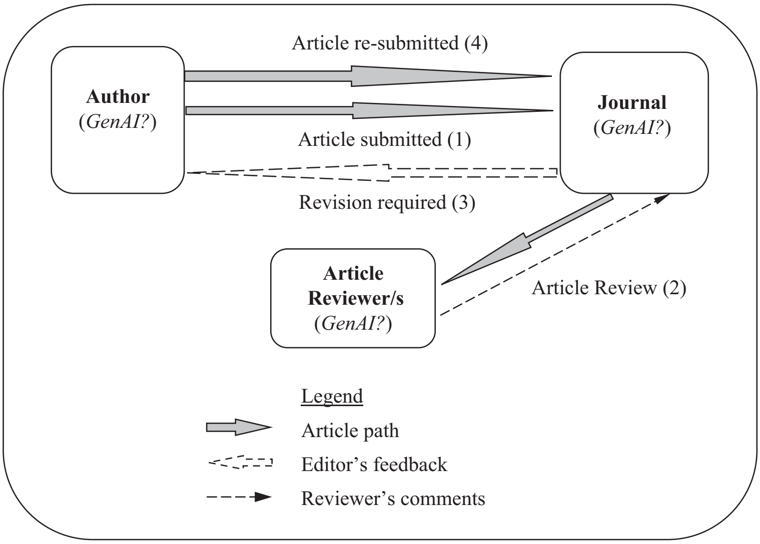

Figure 20.1 A scheme of main actors and their relationships involved in the process of scientific research publication. The numbers between brackets indicate the succession order of the flow exchange (the cycle author–editor–reviewer–editor–author can be iterative).

Source: Authors' research

20.3.1 Editors' Role.

The traditional role of editors' (i.e. journals'), as analyzed by Marusic et al., (2007) was the detection and prevention of scientific misconduct. After AI emergence, the editors' role does not change fundamentally, but it becomes more complex, twofold:

i detection of fraud (in that sense of GenAI usage by authors); and
ii detection of unfair use of AI by reviewers. Ideally, the activity of editors should be supported by more powerful AI-based tools, able to detect the use of AI by either authors or reviewers.

As a principle, the prevention is realized by administrative means; the reviewers are clearly asked by editors: "*Please confirm that you did not use any generative AI or AI-assisted technological methods to prepare this reviewer report*" (author's experience). It is highly debatable to what extent this firm interdiction (either to reviewers or authors) of using AI-assisted methods or tools is / is not productive, ethical or fair. In any case, Gen AI could be used to detect AI fakery (Economist, 2024d, pp. 70–71).

20.3.2 Authors' Perspective.

Undoubtedly, the authors in their research role gain advantage, since the scientific work in top research areas is facilitated by the AI use (Economist, 2023b, p. 65): from research methods (quantifying uncertainty); to neuroscience (understanding the brain), fundamental physics (discovery of Higgs boson) and pure mathematics (uncovering hidden patterns in network knots); to apparently remote areas as linguistics (decoding whale sounds); to wildlife conservation (counting endangered animal species). According to the same source, which cited the Australia's Science Agency (CSIRO), in 2023 "with the rise of deep learning, more than 99% of research fields were producing AI-related results" (Economist, 2023b*,* p. 63).

20.3.3 AI-Based Scientific Predictions

Each of the AI-related discoveries – to count only new materials and new drugs – is a success story as well as an excellent case study. Spectacular results were achieved in medicine – just to mention new antibiotics such as *abaucin* (Liu et al., 2023) and *halicin* (Stokes et al., 2020) – and in materials science, as new materials for building new generation batteries (Moses et al., 2022).

CSIRO, cited by *The Economist* (2023b, p. 64) shows that AI-related publications (as percentage of total publications) are dominated by physical sciences (more than 10%), followed by social sciences and humanities, life sciences, and health sciences (around 4% each), which means that in 2022, *AI-related publications counted for about a quarter* of total scholarly publications.

Grossmann et al. (2023) came with a fresh perspective on the transformational role the AI may have in social science research, in social scientists' work (as authors). As "social sciences rely on a range of methods, including questionnaires, behavioural tests, mixed-method analyses of semi-structured responses, agent-based modelling, observational studies, and experiments […] *LLMs* (large language models) *can more accurately simulate human behavioural responses in social science research,"* making large language models (LLMs) useful (among others) in high-risk projects where "traditional data collection is impractical, allowing for the testing of interventions in simulated populations *before real-world implementation*." (2023*,* p. 1).

The Economist (2023b, p. 66) reported that back in 2019, Vahe Tshitoyan and colleagues from Lawrence Berkeley National Laboratory "used an AI technique called unsupervised learning to analyse the abstracts of materials-science papers." After retraining their system, the researchers demonstrated that "an unsupervised method can recommend materials for functional applications several years before their discovery" which "suggests that latent knowledge regarding future discoveries is to a large extent embedded in past publications" (Tshitoyan et al., 2019, p. 96). Sourati and Evans (2023) extended Tshitoyan's research based on LBD (literature-based discovery) systems, observing that "AI models trained on published scientific findings […] succeed by predicting human predictions and the scientists who will make them."

Sourati and Evans (2023) show that "incorporating the distribution of human expertise by training unsupervised models on simulated inferences that are cognitively accessible to experts dramatically improves (by up to 400%) AI prediction of future discoveries beyond models focused on research content alone, especially when relevant literature is sparse."

The concept of "robot scientist" (a technology that combines robotics and AI) is discussed by King et al., (2023). They identify key trends and make recommendations for "continued investment in the development of both AI and robotics and their interface across the medium and long term." Dr. King is also a strong supporter of the *Stockholm declaration on AI ethics* (King et al., 2024).

20.3.4 Reproducibility Crisis

The "reproducibility crisis" in science is the situation of uncertainty created when, for various reasons, the result of a certain scientific experiment cannot be reproduced (or replicated) is a topic of concern for many researchers (Korbmacher et al., 2023; Romero, 2019; Roper et al., 2022). Notably, Gundersen (2023) have shown that *AI-supported reproducibility* of scientific research contributes not only to increased productivity but also to *higher level of trust.*

20.3.5 Dealing with Predatory Journals

Both authors and editors are concerned (even if for different reasons) by the negative effects of predatory journals. There is no universally agreed definition as the defining border is fuzzy and criteria are rather difficult to assess. However, according to *InterAcademy Partnership,* there are some features for considering a journal as "predatory": (i) solicit articles from researchers through practices that exploit the pressure on researchers to publish (e.g., rapid pay-to-publish models); (ii) less rigorous peer-review; (iii) fake editorial boards (e.g., falsely listing respected scientists); (iv) fraudulent impact factors; (v) journal titles that are deceptively similar to those of legitimate journals; (vi) paid review articles that promote fake science; (vii) aggressive spam invitations to submit articles (IAP, 2022, p. 13). Moreover, practicing the "author-pays" model of open access for financial gain is common practice for predatory journals.

While launching a predatory journal means *lack of ethics*, the growing number of predatory journals and conferences leads to *dented trust in publications in general*.

Amid China's global scientific rise (academic journal included), Teixeira de Silva (2023) published meaningful reflections related to "predatory" journals and references (Teixeira da Silva et al., 2024).

Significantly, journals themselves (institutionally) act along professional associations and authors against predatory journals as a means of protection, warning both authors and reviewers.

20.3.6 Reviewers' Perspective

One of the authors, in his capacity of reviewer for a prominent journal belonging to a large publishing group, repeatedly experienced 7 (seven) days to review a proposed paper at first reading, and 3 (three) days for reviewing the revised version of that paper, rather quickly (in a week or such). To make the already tense review atmosphere uncomfortable too, the "dialog" reviewer-editor was mediated by an AI-supported robot. In other words, speeding up the publication (which is not necessarily bad) at any rate.

Additionally, besides the issues raised by the excessively blitz-publication durations (although understandable from the standpoint of rapid dissemination of scientific knowledge in this age of accelerated pace of technology progress), the quality of the review process entirely, the quality of the reviewers able to perform in such conditions as well as the quality of the authored paper to be revised (often significantly) in such short time are issues of concern. Moreover, all these are very permissive filters – even "facilitators" – for allowing fraudulent research publications.

As the focus of this piece of work is on *trust relationships*, both individual and organizational trust relationships are mentioned, at intra- and extra-organizational levels. Nevertheless, in the light of the *trust triad model* (Scarlat, 2021), the *trust relationships are closely interlinked with ethics* (Scarlat & Ioanid, 2022).

While Gaggioli (2023) and Spirling (2023) discuss the ethical way to use GenAI models in scientific papers, Evans (2020) provides a good anti-example (both trust and ethics lacking): Woo Su Hwang, a Korean researcher, found guilty in October 2009 for misuse of research funds (illegal purchase of human ovules for his experiments) conjugated with withdrawn of two articles (previously published in *Science* in 2004 and 2005).

20.4 The Ethical Side While Using GenAI in Academic Research

Along with the widespread use of GenAI software, its area of application expanded and included research. Many researchers found GenAI useful especially for writing literature reviews, but, while the software evolved, other uses drew their attention too. The academic community concerned with research ethics reacted immediately. Like the general public, the reaction was negative, dominated by fear; the general feeling was that something must be wrong with using GenAI. But what was actually wrong remains the subject of controversy, because GenAI does not match any of the existing categories addressed by academic ethics. Certainly, a researcher using GenAI may have some benefits due to its capabilities, but how wrong is, after all, making use of these capabilities?

The first reaction was banning its use, but those attempting to put the interdiction into practice failed soon, because the use of GenAI is not traceable.

Therefore, one must rely on the honesty of people who should declare whether they made use of GenAI. The next step was, consequently, asking authors to acknowledge the use of GenAI in their research, especially in writing articles. But how should be acknowledged the use of GenAI? The first way was to list the most used GenAI software, namely ChatGPT, as an author. As a result, ChatGPT had in 2023, according to Stokel-Walker (2023), at least four Web of Science publications in its portfolio. By June 26, 2024, the number had already increased to ten indexed by the Web of Science (including a highly cited article) and three indexed by Scopus. This figures place ChatGPT in the series of fictional authors with an internationally recognized author profile (e.g., Web of Science or Scopus), together with the non-existing Stronzo Bestiale (Penders & Shaw, 2020), the dog Galadriel Mirkwood, and many other animals (Lăzăroiu, 2020).

However, the *solution of listing GenAI as author is not correct*. According to the Copyright Predatory Journals and Conferences website (Anonymous, 2023), GenAI does not meet two authorship conditions, i.e., being responsible for what it "writes," and holding the copyright of the text it produces.

Similarly, according to the "Scientist Sees Squirrel" website, one cannot thank GenAI for "writing" or rephrasing a text, as it would be possible when the same is done by a colleague, or as authors thank those who funded their research, to give only a few examples (Heard, 2023).

The Committee of Publication Ethics (2023) requires authors to mention the use of GenAI in the methodological section, but, according to the "Scientist Sees Squirrel" (Heard, 2023), GenAI cannot be considered a "method," at least for the fact that its output cannot be replicated. However, the use of GenAI can be part of a method: "*The doctoral research applies a meta-heuristic approach, similar to the heuristic method of "*incompetence*" [...], consisting in an interview with a subject not specialized in the investigated subject, i.e., ChatGPT*" (Gârjoabă, 2023). Eliminating the possibilities of assimilating GenAI to an author, collaborator, or method, "Scientist Sees Squirrel" concludes that it should be regarded as a research tool (Heard, 2023).

If GenAI is a tool, the ethical questions are reduced to those defining its fair use. Of course, as stated before, GenAI has multiple facets, and defining the acceptable uses requires fine-tuning. For example, one of its functions is to rephrase a text. A researcher whose native language is not English may find frustrating to have a quality article rejected because it cannot be understood due to the language barrier. This condition explains, among others, why such authors eventually get their work published in predatory journals, which do not require a high level of English (Soler, 2020). Rephrasing the text using GenAI becomes an appealing option; is it something necessarily "bad"? After all, is it worse in nature than using specialized proofreading services for the same purpose? Of course, this facet has a simpler answer, unlike others, such as using GenAI for writing up a literature review.

In a similar way, journals turned to the reviewers and editors with respect to using GenAI. In this case, it seems that the attitude toward using GenAI evolves passing through the same phases as in the case of authors, with a certain lag. Currently, reviewers are not allowed to use GenAI for preparing their reports. An exploration of Elsevier's policies explains the three reasons beyond the decision.

First, "the reviewer is responsible and accountable for the content of the review" (Elsevier, 2004b). This statement can easily be related to the explanation for GenAI not to be considered an author, as it cannot be held responsible for the text it produces.

Second, "critical thinking and assessment required for peer-review are outside the scope of generative AI and AI-assisted technologies, and there is a risk that the technology will generate incorrect, incomplete or biased conclusions" (Elsevier, 2004b). While this idea is not found explicitly among the "fears" for the authors using GenAI, it could explain while authors must acknowledge the use of GenAI , beyond receiving some 'help' in writing up their research.

However, the Committee of Publication Ethics has a different position: "AI and automation tools have demonstrated success in assisting faster and accurate peer review" (Committee of Publication Ethics, 2021).

The most interesting point made by Elsevier is the third and last one, banning the use of GenAI in writing a review report "even if it is just for the purpose of improving language and readability" (Elsevier, 2004b).

Similarly, Elsevier (2024a) states that "Generative AI should not be used to assist in the review, evaluation or decision-making process of a manuscript." Therefore, similarly to the reviewers, editors are not allowed to use GenAI in order to carry out their role in the peer review process.

In a nutshell, there are substantial differences between the authors, on the one side, and reviewers and editors, on the other side, with respect to using GenAI.

At this moment, authors can use GenAI as long as they acknowledge it, but reviewers and editors cannot. However, the attitude toward using GenAI changed in time in the case of authors; it might change for reviewers too. Also, as it could be seen, sometimes there are contrasting viewpoints, such as the discussion on whether the use of GenAI by reviewers has the potential to affect the quality of review reports.

20.5 Conclusion

This essay is a result of the authors' continuous endeavor to identify the best (i.e., trustful and ethical) ways to conduct their research teams. In addition, it continues its trust-related studies in new circumstances (GenAI emergence) and less investigated areas (higher education scientific research) from multiple perspectives.

Not only authors' but also editors' and reviewers' standpoints are worth-to-explore how to increase the reciprocal trust to effectively cope with fraud, in the new GenAI era. Hence, several areas to further investigate: GenAI-mediated relationship, GenAI trust triad, author-editor reciprocal trust-based relationship.

In addition, the closer examination of the author-editor-reviewer reciprocal trust (e.g., is it a linear or network-type relationship?) There would be implications for all actors (and respective institutions, too) involved in the academic research process.

Lack of in-depth, quantitative research on the above issues is a limitation, but a natural stage of study. One challenging path to follow would be assessing the level of trustfulness by the use of sentiment analysis (Chang and Ma, 2024; Mahalakshmi et al., 2024).

The timing of this type of study is just-in. On 6 September 2024, the news agencies announced that the United States, EU, and United Kingdom have signed, the previous day, the Council of Europe's convention on AI that was the *first legally binding international treaty on the use of AI* and followed soon after the European Union Regulation EU 2024/1689 or the *Artificial Intelligence Act* (European Parliament, 2024). Related to this event, a good counterexample is suitable. Anecdotally, by the time, this Act was published (along with a wealth of training programs), an outlet in a certain member state had launched its own webinar targeted to HR managers. The online invitation contained "catchy and stimulative" messages (e.g., obligations, interdictions to use AI, and alike), all under bolded warning:

> Attention!
>
> Failure to observe the 'new Regulation' may have consequences: fines of 7% applied to annual turnover (for breaking interdictions to use AI) or 3% for other cases of breaking the law.

Highly "attractive message" for promoting the use of AI technologies!

20.5.1 Instead of Concluding Remarks

It is impossible to stop thinking about a stunningly analogous situation that happened more than a decade ago. Facing the reality of significantly increased banks' accumulated loss caused by e-fraud at the global level, a meeting of the world's top bankers was called in quasi-secrecy to decide urgent measures against e-fraud (complete cancellation of e-banking among options). Decision: just continue to develop e-banking services. Reason: profits were considerably higher than losses to be covered.

Beyond the strange similitude, the decision pattern is neatly valid: frauds cannot stop technology progress, but fighting fraud should continue.

Notes

1 Old saying (in Latin: *Confido sed cognoscere*). Also known as *Trust but verify*!
2 Reich (2009a) also notes that improperly conducted peer-review process might turn the journals' review into opportunities for additional fabrication.
3 This was exactly the case of Bell Laboratories, owned by Lucent Technologies during that period, which suffered significant drop of shares price. Currently (August 2024) Bell Labs are owned by Nokia as part of its acquisition of Alcatel-Lucent.

References

Ackermann, E. (2006). Indicators of failed information epidemics in the scientific journal literature: A publication analysis of polywater and cold nuclear fusion. *Scientometrics, 66*(2006), 451–466. Published by Springer, 2013. https://doi.org/10.1007/s11192-006-0033-0.

Anonymous. (2023). *If you use chatGPT, you must acknowledge it*. Retrieved 26 June 2024 from https://predatory-publishing.com/if-you-use-chatgpt-you-must-acknowledge-it/.

Arora, A., Belenzon, S., Cioaca, L., Sheer, L., & Zhang, H. (2023). The effect of public science on corporate R&D. Working Paper 31899, *NBER Working Paper Series*. National Bureau of Economic Research (NBER). Retrieved 13 February 2024 from https://www.nber.org/papers/w31899.

Beasley, M. R., Kroemer, H., Kogelnik, H., Monroe, D., & Datta, S. (2002). *Report of the investigation committee on the possibility of scientific misconduct in the work of hendrik schön and coauthors*. Property of lucent technologies. Distributed by the American physical society with permission of lucent technologies. September 2002. Retrieved 5 August 2024 from https://media-bell-labs-com.s3.amazonaws.com/pages/20170403_1709/misconduct-revew-report-lucent.pdf.

Carreyrou, J. (2018). *Bad blood: Secrets and lies in a silicon valley startup*. Alfred A. Knopf.

Chang, W., & Ma, S. (2024). A sentiment analysis model based on attention map convolutional network. *International Journal of Information Technologies and Systems Approach, 17*(1), January-December 2024. https://doi.org/10.4018/IJITSA.348658.

Chevassus-au-Louis, N. (2019). *Fraud in the lab: The high stakes of scientific research.* Harvard University Press.

Committee of Publication Ethics. (2021). *Artificial intelligence (AI) in decision making*. https://doi.org/10.24318/9kvAgrnJ. Retrieved 26 June 2024 from https://publicationethics.org/resources/discussion-documents/ai-artifical-intelligence-decision-making.

Committee of Publication Ethics. (2023). *Authorship and AI tools*. Retrieved 26 June 2024 from https://publicationethics.org/cope-position-statements/ai-author.

Economist. (2023a). The changing nature of science: Ripe for disruption? *The Economist*, January 7th, 2023, 59.

Economist. (2023b). AI in Science (1. Faster, better, more productive. 2. I, robot scientist). *The Economist,* September 16th, 2023, 63–67.

Economist. (2024a). Ivory Sour. How universities contribute to slow economic growth. *The Economist*, February 10th, 2024, 62.

Economist. (2024b). Technology and the economy: A sequence of zeros. *The Economist,* July 6th, 2024, 57–58.

Economist. (2024c). Artificial intelligence (1. Meet your new copilot. 2. The meaning of Mistral). *The Economist,* March 2nd, 2024, 50–53.

Economist. (2024d). Detecting AI fakery: True or false? *The Economist*, January 20th, 2024, 70–71.

Elsevier. (2024a). *How to conduct a review*. Retrieved 26 June 2024 from https://www.elsevier.com/reviewer/how-to-review.

Elsevier. (2024b). *The use of generative AI and AI-assisted technologies in the review process for Elsevier.* Retrieved 26 June 2024 from https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-the-review-process.

European Parliament. (2024). Regulation (EU) 2024/1689 of the European parliament and of the council of 13 June 2024. *Official Journal of the European Union*, 12 July 2024, 2024/1689, L series. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

Evans, J. (2020). *The big ideas in science. A complete introduction*. Chapter 21: Fraud, fakery and fantasy. Revised and updated edition. Teach Yourself, an imprint of John Murray Press.

Gaggioli, A. (2023). Ethics: disclose use of AI in scientific manuscripts. *Nature,* 2023 February, *614*(7948), 413. https://doi.org/10.1038/d41586-023-00381-x.

Gârjoabă, A.-I. (2023). *Natural protected areas of the urban environment* (in Romanian), Doctoral Dissertation. 'Ion Mincu' University of Architecture and Urbanism.

Grossmann, I., Feinberg, M., Parker, D. C., Christakis, N., Tetlock, P. E., & Cunningham, W. A. (2023). AI and the transformation of social science research. Careful bias management and data fidelity are key. *Science, XXX*, 24 May 2023, 1108–1109.

Judson, H. F. (2004). *The Great Betrayal: Fraud in Science.* Harcourt.

Gundersen, O. E. (2023). Improving reproducibility of artificial intelligence research to increase trust and productivity. In OECD (Ed.), *Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research,* OECD Report (pp. 262–270). OECD Publishing.

Heard, S. (2023). ChatGPT: author, acknowledgement, method, or tool? Retrieved June, 26 2024 from https://scientistseessquirrel.wordpress.com/2023/05/23/chatgpt-author-acknowledgement-method-or-tool/.

IAP. (2022). *Combatting predatory academic journals and conferences.* Full report. IAP – The InterAcademy Partnership. Retrieved August, 12 2024 from https://www.interacademies.org/publication/predatory-practices-report-English.

Judson, H. F. (2004). *The great betrayal: Fraud in science.* Harcourt.

King, R., Peter, O., & Courtney, P. (2023). Robot scientists: from Adam to Eve to Genesis. In OECD (Ed.), *Artificial intelligence in science: Challenges, opportunities and the future of research,* OECD Report (pp. 129–139). OECD Publishing. https://doi.org/10.1787/d89b39b6-en.

King, R. D., Scassa, T., Kramer, S., & Kitano, H. (2024). Stockholm declaration on AI ethics: why others should sign. *Nature,* 2024 February; *626*(8000), 716. https://doi.org/10.1038/d41586-024-00517-7.

Korbmacher, M., Azevedo, F., Pennington, C. R., …, Lecuona, O., Micheli, L., and Evans, T. (2023). The replication crisis has led to positive structural, procedural, and community changes. *Communication Psychology, 1*(3), 1–13. https://doi.org/10.1038/s44271-023-00003-2.

Labini, F. S. (2016). *Science and the economic crisis: Impact on science, lessons from science.* Springer.

Lăzăroiu, G. (2020). Whose scientific work is it anyway? Knowledge production in the socially constructed fuzzy authorship. *Educational Philosophy and Theory, 54*(9), 1290–1295.

Liu, G., Catacutan, D. B., Rathod, K, …, Barzilay, R., Collins, J. J., & Stokes, J. M. (2023). Deep learning-guided discovery of an antibiotic targeting *Acinetobacter baumannii*. *Nature Chemical Biology, 19*, 25 May 2023, 1342–1350. https://doi.org/10.1038/s41589-023-01349-8.

Mahalakshmi, T., Zulaikha Beevi, S., Navaneethakrishnan, M., Puppala, R., & Prasad Kumar, S.N. (2024). Optimized attention-driven bidirectional convolutional neural network: Recurrent neural network for facebook sentiment classification. *International Journal of Business Data Communications and Networking*, *19*(1), January-December 2024, 1–20. https://doi.org/10.4018/IJBDCN.349572.

Mahmud, S. (2024). Fostering academic integrity in the age of artificial intelligence. In S. Mahmud (Ed.), *Academic integrity in the age of artificial intelligence* (pp. 1–20). IGI Global. https://doi.org/10.4018/979-8-3693-0240-8.

Marusic, A., Katavic, V., & Marusic, M. (2007). Role of editors and journals in detecting and preventing scientific misconduct: Strengths, weaknesses, opportunities, and threats. *Medicine and Law, 26,* 545–566.

Moses, I. A., Barone, V., & Peralta, J. E. (2022). Accelerating the discovery of battery electrode materials through data mining and deep learning models. *Journal of Power Sources, 546,* 30 October 2022, 231977. https://doi.org/10.1016/j.jpowsour.2022.231977.

Norman, M. R. (2009). Cargo-cult science Redux. *Nature Physics, 5*(7), 451–452. https://doi.org/10.1038/nphys1316.

Park, M., Leahey, E., & Funk, R. (2023). Papers and patents are becoming less disruptive over time. *Nature, 613,* 138–144.

Park, R. (2000). *Voodoo science: The road from foolishness to fraud*. Oxford University Press.

Penders, B., & Shaw, D. M. (2020). Civil disobedience in scientific authorship: Resistance and insubordination in science. *Accountability in Research, 27*(6), 347–371.

Physics (2009). Lessons from Schon – the worst physics fraudster? *Physics* online. Posted on May 5, 2009. Institute of Physics. Retrieved 5 August 2024 from https://phys.org/news/2009-05-lessons-schon-worst-physics.html.

Reich, S. E. (2009a). *Plastic fantastic. How the biggest fraud in physics shook the scientific world.* St. Martin's Press.

Reich, S. E. (2009b). The rise and fall of a physics fraudster. *Physics World, 22*(05), 24. https://doi.org/10.1088/2058-7058/22/05/37.

Rigby, J., & Satija, B. (2023). WHO declares end to COVID global health emergency. *Reuters News, published online* May 8, 2023. Retrieved August, 12 2024 from https://www.reuters.com/business/healthcare-pharmaceuticals/covid-is-no-longer-global-health-emergency-who-2023-05-05/.

Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass, 14*(e12633), 1–14. Published by John Wiley & Sons. DOI: https://doi.org/10.1177/10892680211066468.

Roper, K., Abdel-Rehim, A., Hubbard, S., Carpenter, M., Rzhetsky, A., Soldatova, L., & King, R. D. (2022). Testing the reproducibility and robustness of the cancer biology literature by robot. *Journal of The Royal Society Interface, 19*, 20210821. The Royal Society Publishing. https://doi.org/10.1098/rsif.2021.0821.

Scarlat, C. (2021). The triadic and multi-triad models: Typology and potential for entrepreneurship and business management studies. *International conference on management and industrial engineering*, *10,* 86–95.

Scarlat, C. (2023a). Post-pandemic business effects as evolution of strategic thinking: Strategic networking and digital acceleration. *International Conference on Management and Industrial Engineering*, *11*, 417–424.

Scarlat, C. (2023b). Changes in business organizations: Transformation of work settings and habits as post-pandemic effects. *International Conference on Management and Industrial Engineering*, *11*, 439–446.

Scarlat, C. (2024). *Facilis descensus averno* or interactions of doctoral research projects with university's other projects. In M. Carmo (Ed.), *Education and new developments 2024,* (Vol. 2 pp. 293–297). World Institute for Advanced Research and Science (WIARS).

Scarlat, C., & Ioanid, A. (2022). The triad of digital trust: Organizational trust, reputation, and ethics in cyberspace. In J. Paliszkiewicz, J. L. Guerrero Cusumano, & J. Gołuchowski (Eds.), *Trust, digital business and technology. Issues and challenges* (pp. 14–32). Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781003266495-3.

Senior, P.C. (2019). *Billion dollar façade: The rise and fall of Theranos and Elizabeth Holmes*. Independently Published.

Soler, J. (2020). Linguistic injustice and global english: Some notes from its role in academic publishing. *Nordic Journal of English Studies, 19*(3), 35–46.

Sourati, J., & Evans, J. A. (2023). Accelerating science with human-aware artificial intelligence. *Nature Human Behaviour, 7*, 1682–1696.

Spirling, A. (2023). Why open-source generative AI models are an ethical way forward for science. *Nature,* 2023 April, *616*(7957), 413. https://doi.org/10.1038/d41586-023-01295-4.

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: many scientists disapprove. *Nature*, *613*, 620–621.

Stokes, J. M., Yang, K., Swanson, K., …, Jaakola, T. S., Barzilay, R., & Collins, J. J. (2020). A deep learning approach to antibiotic discovery. *Cell, 180*(4), February 20, 2020, 688–702. https://doi.org/10.1016/j.cell.2020.01.021.

Teixeira de Silva, J. A. (2023). A reflection on springer nature's current editorial policies related to "predatory" journals and references. *Science Editor and Publisher, 8*(2), 110–123.

Teixeira da Silva, J. A., Nazarovets, S., Daly, T., & Kendall, G. (2024). The chinese early warning journal List: Strengths, weaknesses and solutions in the light of China's global scientific rise. *The Journal of Academic Librarians, 50*(4), 102898 https://doi.org/10.1016/j.acalib.2024.102898.

Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K. A., Ceder, G., & Jain, A. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature, 571,* 95–98.

21 Generative AI Use Among Non-financial Companies – Trust Perspective

Magdalena Mądra-Sawicka and Marzena Ganc

21.1 Introduction

Trust is fundamental in shaping a company's internal dynamics and external relationships. Internally, trust fosters employee engagement, enhances collaboration, and drives productivity while strengthening partnerships, customer loyalty, and brand reputation externally. The division of trust between internal (employees, management) and external (clients, investors) stakeholders highlights the need for tailored strategies to cultivate and maintain it. Research shows that organizations with high levels of trust outperform competitors by promoting innovation and resilience (Mayer et al., 1995). Building and sustaining trust becomes crucial to long-term success and sustainable growth as businesses navigate increasingly complex markets.

Trust among internal stakeholders should prioritize building trust by promoting transparency, delivering personalized experiences, fostering emotional connections, and proactively addressing privacy concerns (Kim et al., 2024). Building trust with Generative AI (GenAI) technology requires involving employees in the AI integration process, offering training, and fostering a collaborative environment where human oversight plays a central role. By addressing these challenges, companies can ensure smoother AI adoption, improve project outcomes, and strengthen team cohesion (Paparic & Bodea, 2024).

Managers increasingly recognize GenAI as a valuable tool for enhancing company productivity. Its capacity to automate routine tasks and support complex decision-making is reshaping business functions and boosting efficiency. Moreover, as AI models become more sophisticated, they enable deeper personalization and innovation potential for companies. Thus, the growing adoption of GenAI reflects a broader trend toward digital transformation, with companies leveraging AI to cut costs and unlock new revenue streams and market opportunities.

This interactivity makes GenAI easier for non-experts to use while also serving as an intuitive and personalized tool for managing complex tasks.

DOI: 10.4324/9781003586944-24

Companies can lower operational expenses and improve efficiency by automating routine tasks and optimizing workflows. GenAI's ability to analyze vast datasets allows businesses to recognize patterns and trends, enabling more informed decisions and revealing new growth opportunities. Moreover, GenAI-driven personalized customer experiences foster stronger engagement and loyalty, leading to increased revenue. Non-financial firms can harness GenAI to drive innovation, create new products, and streamline supply chain processes, further enhancing financial performance and strengthening their competitive edge.

The chapter aims to present the generative artificial interface engine in companies from various industries from a trust perspective among stakeholders. The literature review led to the GenAI implementation in companies. The methodology part describes the external sources of data used in the analysis. The research results present the main conclusion concerning GenAI implementation and companies' motivation to build trust. The discussion part includes an analysis of the benefits and challenges that the company faces. The last part of the chapter presents the conclusion, the study's limitations, and the future direction of further research.

21.2 Literature Review

Trust is a cornerstone of organizational success, influencing relationships at every level of a company's ecosystem. Understanding how trust operates across different contexts allows businesses to build stronger connections, drive engagement, and sustain long-term growth. Trust among companies can be perceived through a few theories: stakeholder theory, cognitive and affective trust, organizational justice, and international trust framework. Stakeholder theory suggests that the nature and level of trust vary depending on the depth of the relationship (whether it is close or superficial) and the type of stakeholder (internal or external) (Freeman, 2010). This theory highlights the importance of adapting trust-building strategies to different groups (Pirson & Malhotra, 2011). Internal stakeholders, such as employees, often require more profound personal trust, while external stakeholders, like partners or clients, may base their trust on surface-level credibility. In modern businesses, this aligns with a growing emphasis on employee engagement and transparent stakeholder communication, ensuring each relationship has specific needs. Cognitive and affective trust is shaped by both rational evaluations (such as assessing competence and fairness) and emotional connections (like forming emotional bonds) (McAllister, 1995; Yan-Hong & Cai-Zhuan, 2011). This dual perspective emphasizes the need for organizations to balance performance and empathy. It joins with demonstrating fairness and competence to foster emotional loyalty among employees and partners. Leaders who combine logical decision-making with empathy are more likely to inspire long-term trust and cooperation, reinforcing the value of emotional intelligence in leadership roles (Koohang et al., 2017).

Furthermore, organizational justice highlights that trust within a company is strongly influenced by employee perceptions of fairness in outcomes (distributive justice), processes (procedural justice), and interpersonal treatment (interactional justice) (Colquitt et al., 2001; Greenberg, 1987; Malla & Malla, 2023; Mukherjee & Bhattacharya, 2013). Employees are more likely to trust leadership when they see equitable distribution of rewards, transparent procedures, and respectful communication. This theory underscores the importance of addressing biases, promoting equal opportunities, and ensuring employees feel heard and valued, contributing to a more harmonious and productive work environment. Interorganizational trust focuses on relationships between organizations, suggesting that trust can be measured on a single scale, while distrust operates on a more complex, two-dimensional model. This perspective reflects the reality of business partnerships, where companies may collaborate while remaining cautious (Lewicka & Zakrzewska-Bielawska, 2022). The absence of trust does not automatically imply distrust but rather a neutral, guarded state. This approach encourages organizations to balance cooperation and risk management, carefully fostering trust while protecting their interests. By integrating these frameworks, companies can develop more effective trust-building strategies that enhance internal cohesion, stakeholder confidence, and sustainable external partnerships, ultimately strengthening organizational resilience and performance.

21.3 Implementing GenAI in Companies

Trust is key to implementing GenAI in enterprises, impacting employee adoption and stakeholder confidence – transparent guidelines and proven reliability foster trust and accelerate integration. Companies use GenAI to automate tasks, boost efficiency, cut costs, improve data analysis, and enhance decision-making. It drives innovation, improves customer experiences, and ensures competitiveness in fast-changing markets. Furthermore, companies benefit from improving employee performance. Companies are incorporating GenAI into their operations to support research, idea generation, drafting business communications, and summarizing or refining text (Cardon et al., 2024).

The financial and operating activity of companies could be supported by GenAI, which could be widely used in cases of financial reporting, accountancy, data analysis, financial performance management, innovation, investment and assets management, supply chain management, and working capital management, or improvement in cybersecurity and employment performance.

GenAI enhances financial processes and decision-making by automating tasks, delivering precise analyses, and identifying patterns in large datasets for comprehensive reporting (Pandey & Rana, 2024; Peng et al., 2023; Shakdwipee et al., 2023). GenAI improves financial statement analysis, enabling better investment decisions and operational efficiency through automated reconciliation and intelligent reporting (Guo & Polak, 2024; Yang, 2022). It forecasts and manages financial risks, detects supplier distress, and recommends mitigation strategies, supporting managerial accounting (Hong, 2020; Rajagopal et al., 2023).

AI advances Environmental, Social, and Governance (ESG) reporting by enhancing internal controls, aligning sustainability goals, and ensuring regulatory compliance (Barde & Kulkarni, 2023; Chen & Zhang, 2024; Zhang & Yang, 2024). AI-driven ESG platforms automate data collection, improving report quality and stakeholder engagement (Markova-Karpuzova et al., 2024). Effectiveness relies on technological capabilities, compliance, and cultural factors (Chen et al., 2024).

In accounting, AI adoption necessitates skills in data analytics and AI technologies (Ballantine et al., 2024). Major firms leverage AI to enhance audits, risk assessments, and data analysis, reducing junior staff needs and improving service quality (Boritz & Stratopoulos, 2023; Kokina & Davenport, 2017; Munoko et al., 2020). From the other perspective, trust in GenAI fosters employee engagement and performance, boosting efficiency by automating tasks and enabling strategic focus (Barney & Reeves, 2024; Łapińska et al., 2021; Manresa et al., 2024). Early adoption correlates with increased trust and productivity (Semujanga & Mikalef, 2024).

AI optimizes financial performance by automating invoicing, payroll, and expense management, reallocating resources to strategic initiatives, and minimizing errors (Artene et al., 2024).

Robotic Process Automation (RPA) lowers operational risk, while explainable AI (XAI) fosters trust by clarifying AI decisions (Falatouri et al., 2023). AI enhances credit risk analysis, fraud detection, and compliance, improving loan approvals and risk evaluations (El-Qadi et al., 2023).

Predictive analytics driven by AI optimize resource allocation and financial planning, enhancing profitability and pricing strategies (Ismanov et al., 2024; Wang, 2024).

AI streamlines supply chain and working capital management by optimizing inventory, renegotiating supplier agreements, and predicting demand fluctuations (Erdoğan, 2019; Rajagopal et al., 2023). GenAI enhances liquidity, cash flow, and financial resilience (Manresa et al., 2024). GenAI improves cash flow forecasting and scenario analysis, refining business models and driving innovation (Chen, 2024; Mądra-Sawicka, 2021, 2022).

GenAI accelerates innovation by automating design and prototyping, fostering product development, and strengthening competitiveness (Iuga & Brad, 2025; Solaiman, 2023). Furthermore, AI refines investment strategies through portfolio optimization and asset valuation, integrating AI-driven stock selection with quantitative models (Romanko et al., 2023).

21.4 Methodology

The study is based on a comprehensive analysis of multiple reputable sources, each providing insights into the adoption, application, and impact of GenAI across various industries and functions. By leveraging data from leading global firms and research organizations, the study ensures a robust and multidimensional perspective on the evolving role of GenAI in the business landscape. The methodology adopts a mixed-methods approach, combining quantitative data from large-scale surveys and qualitative insights from industry reports. The study identifies patterns, trends, and discrepancies in GenAI adoption by analyzing cross-industry data, allowing for a comprehensive assessment of AI's transformative potential.

The selected sources represent the following statistics and surveys:

1. Microsoft, User rate of generative artificial intelligence (AI) in the workplace globally in 2024
2. Microsoft Advertising, & IAB Europe, Factors necessary to create trust towards generative artificial intelligence according to advertising industry professionals in Europe as of April 2024
3. EY, Reimagining Industry Futures Study 2024, What are your organization's most important GenAI priorities in the future?
4. EY, Reimagining Industry Futures Study 2024, Most relevant artificial intelligence (AI) and generative artificial intelligence (GenAI) application types for companies worldwide in 2023
5. Deloitte 2024a, Deloitte's State of Generative AI in the Enterprise, Share of professionals using generative artificial intelligence (AI) in organizations worldwide as of 2024, by function and expertise
6. Deloitte 2024b, Deloitte's State of Generative AI in the Enterprise, Key benefits global organizations hope to achieve with the adoption of generative artificial intelligence (AI) as of 2024
7. Thomson Reuters, Industries where generative artificial intelligence (GenAI) can or cannot be applied in professional services worldwide in 2024
8. McKinsey & Company, Potential impact of generative artificial intelligence (AI) on productivity worldwide in 2023, by business functions (in billion US dollars)

Analyzed report results reflect micro and macro-level dynamics, providing a holistic understanding of GenAI's impact on productivity, trust, and organizational strategy.

21.5 Report Results Concerning GenAI in Companies

A 2024 Microsoft survey by LinkedIn found that 75% of employees use AI at work, with 46% adopting it in the past six months (Microsoft, 2024). It shows rapid AI integration, though nearly half are early adopters. The trend reflects the growing importance of AI, with companies investing to boost efficiency and innovation. However, ongoing training is crucial for maximizing AI's potential, potentially giving businesses a competitive edge.

A 2024 Microsoft Advertising and IAB Europe report found that 78% of professionals see policies and regulations as key to responsible GenAI use (Microsoft Advertising & IAB Europe, 2024). Human moderation is also vital, with 67% emphasizing oversight to maintain accountability – additionally, 47% highlight technical skills as crucial for trust and effective GenAI application. Only 14% cited other factors, reinforcing the focus on governance, oversight, and expertise as essential for successful GenAI integration.

A 2023 EY Reimagining Industry Futures Study survey found that 46% of global organizations prioritize improving data governance to address AI risks like accuracy and ethics (EY, 2024). Integrating GenAI into tech strategies follows 41%, 40% focus on understanding GenAI, and 38% prioritize selecting use cases for deployment. Other goals include evaluating productivity (36%), engaging vendors (35%), and addressing skills gaps (32%). A 2023 EY survey also presents the most relevant AI and GenAI applications. According to this study, employee training and collaboration are the top AI/GenAI applications (36%), highlighting AI's role in upskilling staff and boosting efficiency. Customer sales and support follow at 35%, with software development and testing at 34%. Applications like product design, supply chain, security, and content creation each account for 33%, showing AI's broad use. Personalized services, predictive maintenance (32%), and infrastructure monitoring (31%) are also key, while legal and financial services lag at 25%. The data focuses on immediate operational gains, with gradual AI expansion into complex areas.

The 2024 Deloitte study highlights GenAI adoption across global organizations (Deloitte, 2024a, 2024b), with the highest rates in marketing, sales, R&D, and customer service (73%), followed by IT and cybersecurity (71%). Finance, strategy, and supply chain report moderate adoption (61-64%), while legal and compliance have the lowest levels (60%). Adoption is most advanced in customer-facing and innovation-driven areas, while legal sectors progress more slowly due to regulatory scrutiny. The top benefits of GenAI adoption include improving efficiency and productivity (56%), reducing costs (35%), and driving innovation (29%). Organizations see GenAI as a tool for automating tasks, enhancing operations, and boosting growth. Regarding expertise, 45% of organizations report "some expertise," while 35% demonstrate high proficiency. Only 9% claim very high expertise, with 10% acknowledging limited knowledge, reflecting broad adoption but few industry leaders.

A 2024 Thomson Reuters study shows that 81% of professionals believe GenAI can be applied in legal, corporate risk, government, and tax advisory sectors. Law firms and corporate legal services show the highest confidence (85%), followed by corporate risk (82%) and tax/government services (77%)

(Thomson Reuters, 2024). Only 5% think GenAI cannot be applied, while 14% are uncertain. Uncertainty is highest in the tax sector (19%) and government services (15%), indicating that while most professionals see potential for GenAI, some remain unsure or need more information.

The survey of McKinsey & Company concerns the potential impact of GenAI on productivity worldwide in 2023 by business functions conducted worldwide (McKinsey & Company, 2024). The results were demonstrated by an analysis of 63 use cases in 2023. Marketing and sales are expected to experience the largest gains, with an estimated value increase ranging from \$ 760 billion to \$ 1.2 trillion. Similarly, software engineering is projected to see substantial value growth, highlighting AI's transformative role in these areas. The findings highlight GenAI's role in boosting growth, especially in data-driven areas. AI enhances marketing by improving engagement and streamlining campaigns, while in software engineering, it speeds up development and improves code quality. Companies investing in GenAI can expect significant productivity and financial gains.

21.6 Discussion

Trust in GenAI emerges as a critical factor influencing its adoption and effectiveness across various domains, including financial reporting, accounting, supply chain management, and innovation. In financial reporting, trust ensures the credibility of automated disclosures, promoting compliance and investor confidence. In accounting, trusted AI systems can reliably handle audits and reduce errors, streamlining operations. Within supply chain management, confidence in AI predictions supports better inventory control and risk mitigation. For innovation, trust in GenAI accelerates product development and market responsiveness, driving competitive advantages. In financial reporting and ESG performance, GenAI automates complex tasks and enhances data accuracy, contributing to more reliable disclosures and regulatory compliance. However, the degree to which stakeholders trust AI-driven reports affects their engagement, and the perceived integrity of these reports varies depending on factors such as data quality, model transparency, and previous experiences with AI systems. It reflects broader industry trends, as highlighted by Microsoft's 2024 survey, which shows widespread AI adoption yet emphasizes the need for transparency and ongoing training (Microsoft, 2024). Transparent AI models foster confidence and improve the acceptance of automated reporting systems, while a lack of explainability may undermine trust and limit GenAI's broader impact (Chen et al., 2024; Markova-Karpuzova et al., 2024). Similarly, GenAI streamlines audits and financial analysis in accountancy, reducing operational costs and reallocating tasks traditionally managed by junior staff. However, workforce trust remains pivotal for smooth adoption. Employees who trust GenAI are more likely to embrace AI tools, boosting productivity and minimizing resistance to technological change (Barney & Reeves, 2024; Manresa et al., 2024).

The Deloitte 2024 survey highlights the highest AI adoption rates in functions where trust and expertise align. Sectors like legal and compliance show slower adoption due to lower trust levels and regulatory constraints (Deloitte, 2024a, 2024b). It reinforces that trust-building measures, such as XAI and comprehensive training, are essential for fostering a collaborative AI-human work environment. GenAI plays a key role in financial stability and planning, automating processes, managing risks, and improving decision-making. However, transparency in AI-generated outcomes remains a cornerstone for cultivating trust. XAI provides insights into complex decision-making algorithms, enabling users to understand and trust the rationale behind AI-driven recommendations (Falatouri et al., 2023). It aligns with the EY 2024 report, highlighting the focus on improving data governance and integrating GenAI into broader technology strategies to address AI risks (EY, 2024). This transparency reduces hesitation in adopting AI for credit risk assessments and fraud detection, reinforcing AI's role as a trusted partner in financial management. By predicting demand fluctuations, optimizing inventory, and renegotiating supplier contracts, GenAI enhances efficiency and cash flow.

Confidence in AI-driven adjustments is necessary to fully realize these benefits (Erdoğan, 2019; Rajagopal et al., 2023). Trust can be perceived through consistent, accurate performance, demonstrating GenAI's reliability in mitigating risks and aligning operational changes with broader financial strategies (Manresa et al., 2024). Finally, in driving innovation, GenAI automates design and accelerates product development, fostering competitive advantages and reshaping business models. The extent to which organizations invest in GenAI's innovative potential often depends on internal trust and acceptance. Leaders who perceive GenAI as a valuable tool for growth are more likely to prioritize its integration. The Deloitte 2024 study highlights efficiency and productivity gains as key drivers of GenAI adoption, with innovation and growth also ranking as significant factors (Deloitte, 2024a, 2004b). Involving diverse perspectives during deployment ensures balanced risk management and enhances team trust (Solaiman, 2023). The cyclical relationship between trust, adoption, and performance underscores GenAI's transformative potential across industries, with transparency and organizational engagement as essential pillars for success.

21.7 Conclusion

Companies that successfully integrate GenAI into their operations can significantly boost their competitiveness by increasing efficiency, lowering costs, and accelerating innovation. AI enhances risk evaluation and helps companies mitigate potential risks. Additionally, GenAI automates compliance checks, reducing the likelihood of legal issues and fines. However, research studies indicated that integrating GenAI into existing financial and supply chain management systems can be challenging, necessitating thorough planning and precise execution.

The widespread adoption of GenAI across industries is expected to significantly impact the global economy, fueling growth and reshaping the competitive industry environment.

Key findings suggest organizations integrating GenAI experience measurable productivity and operational efficiency improvements. Enhanced collaboration between AI systems and human teams creates opportunities for innovation and improved strategic decision-making. The data highlights that companies that proactively address employee concerns regarding AI, invest in skills development, and maintain ongoing oversight build more resilient AI ecosystems.

GenAI aligns closely with trust, emphasizing ability, integrity, and benevolence as key trust drivers. By enhancing risk evaluation and ensuring regulatory compliance, GenAI demonstrates high competence (ability) and adherence to established norms (integrity), fostering trust in automated processes. The complexities and uncertainties surrounding GenAI adoption underscore the importance of perceived fairness and transparency. In GenAI, leadership is crucial in balancing innovation with risk management, ultimately shaping a culture of trust that will support long-term AI-driven growth and competitiveness.

This research focuses on non-financial sectors, which may limit the generalizability of findings to financial industries. Additionally, the reliance on secondary data from industry reports may introduce bias or incomplete representations of GenAI adoption trends. Future studies should investigate the longitudinal impact of GenAI on trust dynamics within organizations, focusing on cross-sectoral comparisons. Exploring how industries address trust-building strategies and ethical AI governance could provide deeper insights.

References

Artene, A. E., Domil, A. E., & Ivascu, L. (2024). Unlocking business value: Integrating AI-Ddiven decision-making in financial reporting systems. *Electronics*, *13*(15), 3069. https://doi.org/10.3390/electronics13153069.

Ballantine, J., Boyce, G., & Stoner, G. (2024). A critical review of AI in accounting education: Threat and opportunity. *Critical Perspectives on Accounting*, *99,* 102711. https://doi.org/10.1016/j.cpa.2024.102711.

Barde, K., & Kulkarni, P. A. (2023). Applications of generative AI in fintech. *ACM International Conference Proceeding Series,* (pp. 1–5). https://doi.org/10.1145/3639856.3639893.

Barney, J. B., & Reeves, M. (2024). AI Won't give you a new sustainable Advantage. *Harvard Business Review*, September–October 2024 issue.

Boritz, J. E., & Stratopoulos, T. C. (2023). AI and the accounting profession: Views from industry and academia. *Journal of Information Systems*, *37*(3), 1–9. https://doi.org/10.2308/ISYS-2023-054.

Cardon, P., Fleischmann, C., Logemann, M., Heidewald, J., Aritz, J., & Swartz, S. (2024). Competencies needed by business professionals in the AI Age: Character and communication lead the way. *Business and Professional Communication Quarterly*, *87*(2), 223–246. https://doi.org/10.1177/23294906231208166.

Chen, H. (2024). The applied research on internet of things economic behavior of generative AI and its social and economic impacts. *2024 7th International Conference on Artificial Intelligence and Big Data, ICAIB*D 2024, 555–560. https://doi.org/10.1109/ICAIBD62003.2024.10604626.

Chen, P., Chu, Z., & Zhao, M. (2024). The Road to corporate sustainability: The importance of artificial intelligence. *Technology in Society*, *76,* 102440. https://doi.org/10.1016/j.techsoc.2023.102440.

Chen, R., & Zhang, T. (2024). Artificial intelligence applications implication for ESG performance: can digital transformation of enterprises promote sustainable development?. *Chinese Management Studies*, https://doi.org/10.1108/CMS-11-2023-0653.

Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizatioanl justice research. *Journal of Applied Psychology*, *86*(3), 424.

Deloitte. (January 15, 2024a). Share of professionals using generative artificial intelligence (AI) in organizations worldwide as of 2024, by function and expertise [Graph]. In *Statista*. Retrieved December 29, 2024, from https://www.statista.com/statistics/1545783/ai-adoption-among-organizations-worldwide/.

Deloitte. (January 15, 2024b). Key benefits global organizations hope to achieve with the adoption of generative artificial intelligence (AI) as of 2024 [Graph]. In *Statista*. Retrieved December 29, 2024, from https://www.statista.com/statistics/1450727/key-benefits-genai-organizations-worldwide/.

El-Qadi, A., Trocan, M., Conde-Cespedes, P., & Frossard, T. (2023). Predicting corporate solvency using sentiment analysis of risk analyst textual assessments. *Colloquium in Information Science and Technology, CIST*, 277–281. https://doi.org/10.1109/CiSt56084.2023.10409941.

Erdoğan, M. (2019). Working capital management and firm performance of Turkish companies. *In Studies on Balkan and Near Eastern Social Sciences, 3,* 119–130.

EY. (February 8, 2024). What are your organization's most important GenAI priorities in the future? [Graph] and Most relevant artificial intelligence (AI) and generative artificial intelligence (GenAI) application types for companies worldwide in 2023 [Graph]. In *Statista*. Retrieved December 29, 2024, from https://www.statista.com/study/50485/in-depth-report-artificial-intelligence/.

Falatouri, T., Nasseri, M., Brandtner, P., & Darbanian, F. (2023). Shedding Light on the Black Box: Explainable AI for Predicting Household Appliance Failures. In *International Conference on Human-Computer Interaction* (pp. 69–83). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-48057-7_5.

Freeman, R. E. (2010). *Strategic management: A stakeholder approach*. Cambridge University Press.

Greenberg, J. (1987). A taxonomy of organizational justice theories. *Academy of Management Review*, *12*(1), 9–22.

Guo, H., & Polak, P. (2024). Finance centralization—research on enterprise intelligence. *Humanities and Social Sciences Communications*, *11*(1), 1–9. https://doi.org/10.1057/s41599-024-04098-8.

Hong, Y. (2020). Intelligent financial development based on artificial intelligence. *Proceedings* - 2020 *2nd International Conference on Applied Machine Learning, ICAML 2020*, 195–199. https://doi.org/10.1109/ICAML51583.2020.00047.

Ismanov, I., Qayumov, N., Mukhamadjonova, D., & Akhmadaliyev, B. (2024). AI and cost management: Strategies for reducing expenses and improving profit margins in business. In *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (Vol. 1, pp. 1–7). IEEE. https://doi.org/10.1109/ICKECS61492.2024.10616761.

Iuga, M., & Brad, S. (2025). Integrating generative AI with TRIZ for evolutionary product design. In D. Cavallucci, S. Brad, & P. Livotov (Eds.), *World conference of AI-powered innovation and inventive design. TFC 2024. IFIP advances in information and communication technology* (Vol. 735, pp. 32–49). Springer. https://doi.org/10.1007/978-3-031-75919-2_3.

Kim, D., Kim, S., Kim, S., & Lee, B. H. (2024). Generative AI characteristics, user motivations, and usage intention. *Journal of Computer Information Systems,* 1–16. https://doi.org/10.1080/08874417.2024.2442438.

Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. *Journal of Emerging Technologies in Accounting*, *14*(1), 115–122. https://doi.org/10.2308/jeta-51730.

Koohang, A., Paliszkiewicz, J., & Goluchowski, J. (2017). The impact of leadership on trust, knowledge management, and organizational performance: A research model. *Industrial Management & Data Systems*, *117*(3), 521–537.

Łapińska, J., Escher, I., Górka, J., Sudolska, A., & Brzustewicz, P. (2021). Employees' trust in artificial intelligence in companies: The case of energy and chemical industries in Poland. *Energies*, *14*(7), 1942. https://doi.org/10.3390/en14071942.

Lewicka, D., & Zakrzewska-Bielawska, A. F. (2022). Trust and distrust in inter-organisational relations-scale development. *PLoS ONE*, *17*(12), 0279231. https://doi.org/10.1371/journal.pone.0279231.

Malla, S. S., & Malla, S. (2023). Does the perception of organizational justice determine employees' affective commitment? The mediating role of organizational trust. *Benchmarking*, *30*(2), 603–627. https://doi.org/10.1108/BIJ-07-2021-0408.

Mądra-Sawicka, M. (2021). Trust building strategy among food listed companies in the digital economy era. In J. Paliszkiewicz & K. Chen (Eds.), *Trust, organizations and the digital economy* (pp. 245–257). Routledge.

Mądra-Sawicka, M. (2022). Digital finance technology in trust creation among Customers. In J. Paliszkiewicz, K. Chen, & M. Launer (Eds.), *Trust and digital business* (pp. 216–228). Routledge.

Manresa, A., Sammour, A., Mas-Machuca, M., Chen, W., & Botchie, D. (2024). Humanizing GenAI at work: Bridging the gap between technological innovation and employee engagement. *Journal of Managerial Psychology*, https://doi.org/10.1108/JMP-05-2024-0356.

Markova-Karpuzova, M., Marinov, E., & Kotzev, N. (2024). Sustainable solutions: Advancing in Tech-based ESG reporting platforms. *Vide. Tehnologija. Resursi - Environment, Technology, Resources*, *1*, 235–241. https://doi.org/10.17770/etr2024vol1.7947.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734.

McAllister, D. J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24–59.

McKinsey & Company. (July 6, 2023). *Potential impact of generative artificial intelligence (AI) on productivity worldwide in 2023, by business functions (in billion U.S.*

dollars) [Graph]. In *Statista*. Retrieved December 29, 2024, from https://www.statista.com/statistics/1415436/generative-ai-adoption-commerce-use-worldwide/

Microsoft. (May 7, 2024). *User rate of generative artificial intelligence (AI) in the workplace globally in 2024* [Graph]. In *Statista*. Retrieved December 29, 2024, from https://www.statista.com/statistics/1482102/rate-of-generative-ai-utilization-globally/.

Microsoft Advertising, & IAB Europe. (April 30, 2024). *Factors necessary to create trust towards generative artificial intelligence (gen AI) according to advertising industry professionals in Europe as of April 2024* [Graph]. In *Statista*. Retrieved December 29, 2024, from https://www.statista.com/statistics/1470083/trust-factors-for-gen-ai-europe/.

Mukherjee, K., & Bhattacharya, R. (2013). Exploring the mediating effect of organizational trust between organizational justice dimensions and affective commitment. *Management and Labour Studies*, *38*(1–2), 63–79. https://doi.org/10.1177/0258042X13491363.

Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The ethical implications of using artificial intelligence in auditing. *Journal of Business Ethics*, *167*(2), 209–234. https://doi.org/10.1007/s10551-019-04407-1.

Pandey, S. K., & Rana, N. S. (2024). Automatic financial report using accounting software based on artificial intelligence. *Journal of Lifestyle and SDG'S Review*, *4*(4), e03619–e03619. https://doi.org/10.47172/2965-730X.SDGsReview.v4.n04.pe03619

Paparic, M., & Bodea, C.-N. (2024). Building trust through responsible usage of generative artificial intelligence in projects: A case study. *Issues in Information Systems*, *25*(4), 143–157. https://doi.org/10.48009/4_iis_2024_112.

Peng, Y., Ahmad, S. F., Ahmad, A. Y. A. B., Al Shaikh, M. S., Daoud, M. K., & Alhamdi, F. M. H. (2023). Riding the waves of artificial intelligence in advancing accounting and Its implications for sustainable development goals. *Sustainability*, *15*(19), 14165. https://doi.org/10.3390/su151914165.

Pirson, M., & Malhotra, D. (2011). Foundations of organizational trust: What matters to different stakeholders?. *Organization Science*, *22*(4), 1087–1104. https://doi.org/10.1287/orsc.1100.0581.

Rajagopal, M., Nayak, K. M., Balasubramanian, K., Shaikh, I. A. K., Adhav, S., & |Gupta, M. (2023). Application of Artificial Intelligence in the Supply Chain Finance. In *2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICONSTEM56934.2023.10142286

Romanko, O., Narayan, A., & Kwon, R. H. (2023). Chatgpt-based investment portfolio selection. *Operations Research Forum*, *4*(4), 91.

Semujanga, B., & Mikalef, P. (2024). Exploring the productivity impacts of generative AI in organizations. In R. van de Wetering, et al. (Eds.), *Disruptive innovation in a digitally connected healthy world. I3E 2024. Lecture notes in computer science* (Vol. 14907, pp. 103–111). Springer. https://doi.org/10.1007/978-3-031-72234-9_9.

Shakdwipee, P., Agarwal, K., Kunwar, H., & Singh, S. (2023, August). Artificial intelligence in finance and accounting: Opportunities and challenges. In *International Conference on ICT for Sustainable Development* (pp. 165–177).Springer Nature Singapore. https://doi.org/10.1007/978-981-99-5652-4_17.

Solaiman, I. (2023). The gradient of generative AI release: Methods and considerations. *ACM International Conference Proceeding Series*, 111–122. https://doi.org/10.1145/3593013.3593981.

Thomson Reuters. (April 11, 2024). Industries where generative artificial intelligence (GenAI) can or cannot be applied in professional services worldwide in 2024 [Graph]. In Statista. Retrieved December 29, 2024, from https://www.statista.com/statistics/1477928/professional-services-genai-application/.

Wang, M. (2024). Artificial intelligence empowers the construction of first-class financial management system. *Applied Mathematics and Nonlinear Sciences*, *9*(1), 1–15. https://doi.org/10.2478/amns-2024-0518.

Yang, N. (2022). Financial big data management and control and artificial intelligence analysis method based on data mining technology. *Wireless Communications and Mobile Computing, 2022,* 1–13, 7596094. https://doi.org/10.1155/2022/7596094.

Yan-Hong, Y., & Cai-Zhuan, L. (2011). Organization trust of employees and its measurement. *International Conference on Management Science and Engineering - Annual Conference Proceedings*, 416–420. https://doi.org/10.1109/ICMSE.2011.6069995

Zhang, C., & Yang, J. (2024). Artificial intelligence and corporate ESG performance. *International Review of Economics and Finance*, *96,* 103713. https://doi.org/10.1016/j.iref.2024.103713.

Index