

Artificial Intelligence from Science Fiction to Reality

How AI May Impact Humanity

EMANUEL CAMILLERI



CRC Press
Taylor & Francis Group

Artificial Intelligence from Science Fiction to Reality

How AI May Impact Humanity

EMANUEL CAMILLERI



CRC Press
Taylor & Francis Group

Artificial Intelligence from Science Fiction to Reality

Artificial Intelligence from Science Fiction to Reality examines various aspects, starting with the evolution of human and artificial intelligence (AI). It places AI in its proper context and discusses non-technical aspects, such as philosophical and social issues. The major challenge leaders are likely to encounter is deciding what functions are to be entrusted to AI and how humanity can exercise control over them. The book also focuses on the hardware and software technology that support AI, and the essential cyber security systems that are required to address the evolving AI threat landscape. It examines centres for AI safety that are nonprofit research organisations, which focus on the mitigation of AI risks by proposing solutions against threat actors.

The book discusses the knowledge-based economy, particularly Enterprise AI, and examines the ethical and legal issues that emerge from the practical implications of AI. While most governments have endorsed voluntary ethical and moral charters, there is a reluctance to introduce binding legislative measures. This reluctance is based on the premise that specific laws might hinder AI innovation. Furthermore, detailed private and public sector case studies are presented that demonstrate how AI applications may be successfully implemented according to a practical framework.

A detailed discussion about the implications for human development is presented. The differences between key economic approaches, such as knowledge-based economy, digital economy, and automated economy are

examined, and how these will be impacted by AI in relation to job displacement, data privacy and security, and algorithmic bias. Finally, the book also examines the era beyond AI where organoid intelligence is emerging. It explores future human development where humans could be turned into cyborgs with hi-tech machine implants, re-growable limbs and nanotechnology that repair damaged tissue, rejuvenating human cells leading to immortality.

Emanuel Camilleri is a visiting senior lecturer at Malta University and has authored several books. His most recent publication is *Key Performance Indicators: The Complete Guide to KPIs for Business Success* (Routledge, 2024). He has occupied senior management posts in industrial engineering, finance and ICT sectors in Malta and Australia. Recently, he was appointed Chairman of the Planning Authority, Planning Board, Malta.

*“Dr Camilleri’s book *Artificial Intelligence from Science Fiction to Reality* provides a timely and thought-provoking exploration of how AI is reshaping our economies and societies. It successfully bridges the technical, ethical, and strategic dimensions of AI, offering valuable insights for policymakers, industry leaders, and the public alike. As Malta continues to champion innovation while safeguarding human values, this book makes a welcome and relevant contribution to the ongoing dialogue on responsible AI adoption.”*

Hon. Silvio Schembri, BCom (Hons), MA(Econ), MP Minister for the Economy and Industry, Malta

“This comprehensive work offers a deep dive into AI’s journey—from its roots in science fiction to its transformative impact on society. Covering the evolution of intelligence, cutting-edge hardware and software, ethical considerations, and future developments, it provides invaluable insights into how AI is shaping economies, enterprises, and public governance. A must-read for academics, professionals, and policymakers seeking to understand AI’s role in shaping our future.”

Prof. Lauren Ellul Bachelor of Accountancy (Hons), M.B.A. (Exec.) (Edin.&E.N.P.C.), F.I.A., C.P.A., Ph.D. (Birm.). Head, Department of Accountancy, Faculty of Economics, Management & Accountancy, University of Malta

“Artificial Intelligence from Science Fiction to Reality: How AI May Impact Humanity” is a thought-provoking and comprehensive guide to the rapidly evolving world of AI. It skilfully blends technical insights with profound ethical, social, and philosophical questions, offering a balanced exploration of how AI is transforming our societies and economies. With insightful analysis and real-world implications, this book is an essential read for leaders, policymakers, and anyone seeking to understand how artificial intelligence is shaping our present and could redefine humanity’s future”

Prof. Professor Vincent Marmarà B.Sc.(Hons)(Melit.), M.Sc.(Sheff.) Statistics, Ph.D.(Stir.), FRSS(UK), Faculty of Economics, Management and Accountancy, University of Malta

“This new book by Ing Dr Emanuel Camilleri presents a thorough overview of how Artificial Intelligence (AI) is bringing science fiction alive through its effect in our daily lives. In his book the author brings to light the advances that AI has brought in the public administration including in education and health amongst others and the transformation that it has brought about in our behaviour and quality of life. However, the book also draws into consideration the impact AI will bring on our ethics and morals and changes in legislation. This book is a definite guide for anyone who wants to have a deeper insight about the effect AI has on us now and in the future.”

Joseph Caruana, DPA, MBA (Exec) Commissioner for Tax and Customs Administration, Malta

“In an era where the exponential pace of technological change continues to challenge existing paradigms, this book arrives as both timely and necessary. Eur. Ing. Dr Emanuel Camilleri

has managed to distil a complex and ever-evolving topic into a coherent and compelling narrative. He bridges the historical evolution of both human and artificial intelligence, delves into the underlying hardware and software, and thoughtfully explores the social, ethical, and legal implications that AI presents to society today and in the future. What makes this publication especially important is its relevance across multiple sectors—be it government, business, or academia. Dr Camilleri’s expertise and clarity of thought make this book accessible to professionals and decision-makers who are navigating the transformative impact of AI in their respective fields. The work is grounded in scholarship yet immensely practical, prompting reflection on the pressing need to harmonise AI development with the core values that define humanity. I am confident that this book will become an essential resource for anyone serious about understanding AI’s trajectory and the opportunities and risks that lie ahead. I congratulate Dr Camilleri on this important contribution and commend this book to all those who are eager to understand the future that is rapidly unfolding before us.”

Mr. Johann Buttigieg Executive Chairman, Planning Authority, Malta

“The book by Emanuel Camilleri is a timely contribution to the debate on Artificial Intelligence and its impact on both private and public spheres. The book explores human and artificial intelligence’s evolution, addressing ethical, legal, and social issues. It also highlights the challenges faced by leaders in AI oversight, cybersecurity, and case studies of AI applications. It also examines economic impacts and the emergence of Organoid Intelligence, including future human development through technology and regenerative medicine. I gladly endorse this book as an important contribution to this important theme of our times.”

Godwin Mifsud B. Com. (Hons) Econ; MSc Econ (Essex); MA Pub. Pol.
Permanent Secretary, Ministry for the Economy, Enterprise and Strategic Projects,
Malta

Artificial Intelligence from Science Fiction to Reality

How AI May Impact Humanity

Emanuel Camilleri



CRC Press

Taylor & Francis Group

Boca Raton London New York

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

Designed cover image: Shutterstock

First edition published 2026

by CRC Press

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

and by CRC Press

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

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

© 2026 Dr Emanuel Camilleri

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

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

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

Trademark notice: Product or corporate names may be trademarks or registered trademarks and are used only for identification and explanation

without intent to infringe.

ISBN: 9781041102168 (hbk)

ISBN: 9781041101413 (pbk)

ISBN: 9781003653950 (ebk)

DOI: [10.1201/9781003653950](https://doi.org/10.1201/9781003653950)

Typeset in Minion
by Newgen Publishing UK

*I would like to dedicate this book to my loving and gorgeous wife,
Maryanne who has been my constant companion for over forty
happy years and to our granddaughter, Ellie who is a bundle of
love and joy.*

Contents

[Foreword](#)

[Preface](#)

[Acknowledgements](#)

[List of Abbreviations](#)

PART I [The Evolution of Artificial Intelligence](#)

CHAPTER 1 [Artificial Intelligence: The Beginning](#)

- 1.1 [INTRODUCTION](#)
- 1.2 [SCIENCE FICTION AS A PREDICTOR OF THE FUTURE](#)
- 1.3 [AI, A WICKED OR AN OBEDIENT SERVANT?](#)
- 1.4 [CURRENT ADVANCES IN AI TECHNOLOGY](#)
- 1.5 [CURRENT CAPABILITIES AND MISADVENTURES OF
AI](#)
- 1.6 [HALLUCINATING AI BOTS](#)
- 1.7 [THE RISE OF GENERATIVE AI](#)
- 1.8 [WHERE IS THE HUMANITY HEADING?](#)
- 1.9 [CONCLUSION](#)

CHAPTER 2 [Evolution of Human Intelligence](#)

- 2.1 [INTRODUCTION](#)
- 2.2 [EVOLUTION OF HUMAN BEINGS](#)

- 2.3 [HUMAN INTELLIGENCE](#)
- 2.4 [EVOLUTION OF HUMAN INTELLIGENCE](#)
- 2.5 [THEORIES ON THE EVOLUTION OF HUMAN INTELLIGENCE](#)
- 2.6 [CONCLUSION](#)

CHAPTER 3 [Evolution of Artificial Intelligence \(AI\)](#)

- 3.1 [INTRODUCTION](#)
- 3.2 [THE HISTORY OF ARTIFICIAL INTELLIGENCE](#)
 - 3.2.1 [Era of Enlightened Rational Thinking](#)
 - 3.2.2 [20th Century: Pre-to-World War II Technology Era](#)
 - 3.2.3 [20th Century: Post World War II Technology Era](#)
- 3.3 [CONCLUSION](#)

CHAPTER 4 [Contrasting Human and Artificial Intelligence](#)

- 4.1 [INTRODUCTION](#)
- 4.2 [INTELLIGENCE AND LEARNING](#)
- 4.3 [THE JOURNEY OF TRANSFORMING DATA TO WISDOM](#)
- 4.4 [THE GROWING CHALLENGE BETWEEN HUMAN INTELLIGENCE AND AI](#)
- 4.5 [SPECIFIC DIFFERENCES BETWEEN AI AND HUMAN INTELLIGENCE](#)
- 4.6 [HARMONISATION BETWEEN HUMAN INTELLIGENCE AND AI](#)
- 4.7 [CONCLUSION](#)

PART II [The Technology Supporting Artificial Intelligence](#)

CHAPTER 5 [Artificial Intelligence Hardware](#)

- 5.1 INTRODUCTION
- 5.2 FOURTH-GENERATION COMPUTERS (1971-PRESENT):
NEW PROCESSING CAPABILITIES
- 5.3 FIFTH-GENERATION COMPUTERS (PRESENT AND
BEYOND): NEW ERA FOR AI PROCESSING
 - 5.3.1 The Quantum Computer
 - 5.3.2 What is Quantum Computing?
 - 5.3.3 How Quantum Computers Work
 - 5.3.4 Neuromorphic Hardware
- 5.4 SPECIFIC HARDWARE ASSOCIATED WITH ARTIFICIAL
INTELLIGENCE
 - 5.4.1 Edge Computing Chips
 - 5.4.2 Application Specific Integrated Circuits (ASIC)
 - 5.4.3 Field Programmable Gate Array (FPGA)
 - 5.4.4 Complex Programmable Logic Device (CPLD)
- 5.5 CONCLUSION

CHAPTER 6 Artificial Intelligence Software

- 6.1 INTRODUCTION
 - 6.1.1 Software History
- 6.2 AI COMPUTER PROGRAMMING LANGUAGES
 - 6.2.1 Lisp
 - 6.2.2 Python
 - 6.2.3 Intelligent Computer-Aided Design (ICAD)
 - 6.2.4 KL-One
 - 6.2.5 Official Production System (OPS5)
 - 6.2.6 Mercury and Prolog
 - 6.2.7 IBM Qiskit Software Development Kit (SDK)
 - 6.2.8 Other Programming Languages for Specific Hardware
Associated with AI

- 6.3 AI SOFTWARE ALGORITHMS
 - 6.3.1 Voice and Speech Recognition (Conversational AI)
 - 6.3.2 Text Recognition
 - 6.3.3 Image Recognition
 - 6.3.4 Video Analysis
 - 6.3.5 Generative AI
 - 6.3.6 Natural Language Processing (NLP)
 - 6.3.7 Expert Systems
- 6.4 NEUROMORPHIC COMPUTING
 - 6.4.1 Neural Networks
 - 6.4.2 Neuromorphic Computing Concepts
 - 6.4.3 Characteristics of Neuromorphic Computers
- 6.5 CONCLUSION

CHAPTER 7 Artificial Intelligence Security

- 7.1 INTRODUCTION
- 7.2 EVOLVING THREAT LANDSCAPE
- 7.3 INTRUSION PREVENTION SYSTEMS (IPS)
- 7.4 NEXT-GENERATION FIREWALL
- 7.5 CRACK PUBLIC KEY INFRASTRUCTURE
 - 7.5.1 What is a Digital Certificate?
 - 7.5.2 Key Components of a PKI
- 7.6 CENTRE FOR AI SAFETY
 - 7.6.1 Stanford Centre for AI Safety
 - 7.6.2 European Network for AI Safety
 - 7.6.3 First International AI Safety Report
- 7.7 CONCLUSION

PART III The Practical Implications of Artificial Intelligence

CHAPTER 8 Knowledge-Based Economy and Artificial Intelligence

8.1 INTRODUCTION

8.2 KNOWLEDGE-BASED ECONOMY

8.3 DATA-TO-WISDOM TRANSFORMATION PROCESS

8.3.1 Raw Data

8.3.2 Information

8.3.3 Knowledge

8.3.4 Wisdom

8.4 THE EVOLVING KNOWLEDGE-BASED ECONOMY LANDSCAPE

8.5 THE DECLINE OR RESURGENCE OF THE KNOWLEDGE ECONOMY

8.6 AI IN THE KNOWLEDGE ECONOMY INDUSTRY

8.7 CONCLUSION

CHAPTER 9 Enterprise AI

9.1 INTRODUCTION

9.2 AI IN BUSINESS ENTERPRISE

9.3 THE IMPORTANCE OF ENTERPRISE AI

9.4 ADVANTAGES AND CHALLENGES OF ENTERPRISE AI

9.5 ENTERPRISE AI SCALE CONCEPT

9.6 IMPLEMENTING ENTERPRISE AI

9.7 ENTERPRISE AI USE CASES

9.8 ENTERPRISE AI POTENTIAL

9.9 PRACTICAL ENTERPRISE AI APPLICATIONS

9.9.1 Case Study 1: eCommerce Operational Efficiency and Customer Satisfaction

9.9.2 Case Study 2: AI-Driven Inventory Management

9.10 CONCLUSION

CHAPTER 10 Public Sector AI

10.1 INTRODUCTION

10.2 AI AND THE PUBLIC SECTOR

10.3 RECOGNISING DIFFERENT AI APPROACHES AND
IMPORTANCE OF DATA MANAGEMENT

10.4 MAJOR CHALLENGES CONFRONTING PUBLIC
SECTOR AI

10.5 EMERGING GLOBAL PUBLIC SECTOR AI LANDSCAPE

10.6 IMPLEMENTING PUBLIC SECTOR AI

10.6.1 Practical Implications for Implementation Public Sector
AI Strategies

10.7 PUBLIC SECTOR AI USE CASES

10.8 PUBLIC SECTOR AI IMPLICATIONS

10.9 PUBLIC SECTOR AI'S IMPACT ON SOCIETY

10.10 PRACTICAL PUBLIC SECTOR AI APPLICATIONS

10.10.1 Case Study 1: Healthcare Effectiveness and Customer
Satisfaction at a Public Hospital

10.10.2 Case Study 2: Urban Traffic Management

10.11 CONCLUSION

CHAPTER 11 AI Ethical, Moral, and Legal Considerations

11.1 INTRODUCTION

11.2 AI ETHICAL AND MORAL CONSIDERATIONS

11.2.1 The Need for AI Ethics and Moral Guidelines

11.2.2 Establishing Principles for AI Ethics

11.2.3 Building-Blocks of a Responsible AI Project Delivery
Ecosystem

11.3 AI LEGISLATIVE GLOBAL LANDSCAPE

11.3.1 Europe

11.3.2 North America

11.3.3 Africa

11.3.4 South America

11.3.5 Australasia

11.3.6 Asia

11.4 CONCLUSION

PART IV **The Future and Artificial Intelligence**

CHAPTER 12 Future Developments of Artificial Intelligence

12.1 INTRODUCTION

12.2 LIKELY EVOLUTION OF THE TYPES OF AI

12.2.1 Narrow AI

12.2.2 General AI

12.2.3 Super AI

12.2.4 Strategic Approaches and Core Technologies

12.2.5 Language Models

12.3 LIKELY EVOLUTION OF HARDWARE AND SOFTWARE
ASSOCIATED WITH AI

12.4 LIKELY EVOLUTION OF THE NEXT ECONOMIC
APPROACH

12.4.1 The Impact of AI on the Automated Economy

12.5 LIKELY FUTURE IMPLICATIONS OF AI

12.6 BEYOND AI

12.7 CONCLUSION

CHAPTER 13 AI and the Future of Humans

- 13.1 [INTRODUCTION](#)
- 13.2 [AI AND THE FUTURE OF HUMANS](#)
- 13.3 [HUMAN AND AI COLLABORATION](#)
- 13.4 [FUTURE HUMAN DEVELOPMENT](#)
- 13.5 [CONCLUSION](#)

[INDEX](#)

Foreword



Throughout history, pivotal moments have occurred during which humanity has found itself on the brink of transformation. Currently, as we navigate the complexities of the digital age, we are faced with such a moment – addressing the rise of Artificial Intelligence not as a distant dream, science fiction or mere speculation, but as a tangible force that is redefining our economies, governments, and societies. It is fundamentally altering the nature of human interaction.

Therefore, it is both pertinent and essential that we engage with this compelling work titled *Artificial Intelligence from Science Fiction to Reality: How AI May Impact Humanity*. Authored by Emanuel Camilleri, a distinguished individual within the Maltese government's sector, this book offers a rich and grounded perspective that synthesises the technical expertise of an experienced professional with the strategic insight of a policymaker. This text goes beyond a simple gathering of ideas; it serves as a roadmap for understanding how artificial intelligence, once imagined by science fiction writers, has established its presence in the real world and is evolving at an unprecedented pace.

This book is especially significant for Malta. As a nation that has persistently exceeded expectations in digital technology adoption and innovation, Malta is situated at a pivotal juncture of opportunity and accountability. Over recent years, our government has made huge investments in digital infrastructure, e-government services, emerging technologies, and cybersecurity. In consideration of the present and future, as articulated in the Vision 2050 for Malta, artificial intelligence serves as a pivotal driver for national prosperity and the transformation of the public sector, fostering innovation, enhancing citizen services, and ensuring that the government remains adaptable and proactive in an increasingly digital landscape. The role of visionary leadership, such as that demonstrated by Dr Camilleri, can empower a smaller state to assume a prominent role in the global digital discourse.

The author starts by exploring the compelling journey of human and artificial intelligence, establishing the philosophical and historical roots for subsequent discussions. He then introduces the technological foundations underpinning artificial intelligence, specifically hardware, software, and security, which represent the fundamental components behind the ongoing transformative revolution. This is followed by the practical application of artificial intelligence, wherein the author outlines its intersections with enterprises, the public sector, and the broader knowledge-based economy. Of particular note is the application within the public sector, where Malta can offer substantial contributions and gain much in return, of course

keeping an eye on balancing the complex ethical, legal, and moral dilemmas associated with this phenomenon. Finally, he deeply reflects on its future prospects and how it will influence our society.

What I value most about this current work is that it transcends the examination of technology; it forms a discourse on humanity, addressing how we shape the future and, in turn, how the future reshapes our existence.

As the Prime Minister of Malta, I have observed the profound capacity of ideas to influence policy and the essential role of leadership in transforming policy into actionable outcomes. It is with great pride that I present this publication, which will illuminate the path for decision-makers, technologists, educators, and citizens alike.

May this book serve as a catalyst for contemplation, discussion, and courageous ethical action in an era of technological advancements that offer immense promise and significant responsibility.

**His Excellency Dr Robert Abela KUOM, BA, LLD, Adv Trib Melit,
MP**

Prime Minister, Malta

Preface

Artificial Intelligence (AI) has been a dream for over seven decades starting in 1950 with the publication “Computer Machinery and Intelligence” by Alan Turing who propositioned a test for machine intelligence called “The Imitation Game.” This was followed in 1952 by the development of a computer programme by Arthur Samuel that independently learnt how to play checkers. However, it was in 1955 that John McCarthy and his colleagues organised the Dartmouth workshop on “Artificial Intelligence.” This was the first time that the words Artificial Intelligence were formally used. The next two decades witnessed an era of both rapid growth and struggle for AI research. During this period programming languages were developed that enabled AI to make significant progress, as well as the notion of robots becoming an aspiration. In fact, in 1961 Unimate was the first industrial robot that worked on a General Motors assembly line at the Inland Fisher Guide Plant in Ewing Township, New Jersey. The 1970s witnessed further developments including the first anthropomorphic (humanlike) robot being built in Japan. The 1980s was a period of fast growth and high interest in AI, where Deep Learning and the application of Expert Systems became increasingly acceptable with their ability to learn from their mistakes and make independent decisions.

AI has survived several long-drawn winters but the persistence of dedicated researchers and practitioners has seen AI progress further despite the lack of funding. A serious limiting factor that AI has constantly encountered is processing power. There appears to be a positive correlation between significant strides in AI and advances in computer processing power. High performance computing that uses supercomputers and

computer clusters to solve advanced computation problems have enabled AI to develop at an extraordinary rate and be applied to solve problems that were not possible in the recent past. Quantum computing is early in the maturity cycle, with IBM recently launching the Osprey 433-qubit machine and is anticipating developing a 100,000-qubit machine within 10 years. While Google is predicting a million qubits machine by the end of this decade. These Quantum computers are much faster and more powerful than supercomputers since they can process multiple computations concurrently, making them supreme for undertaking complex problems that require massive amounts of data to be processed quickly. This makes them ideal for AI applications, particularly for General and Super AI.

All too often there is a tendency to consider AI in general terms. However, it is important to recognise that there are different types of AI, with each type having its own specific degree of risk. Narrow AI has greatly matured in last five years. Narrow AI (or Weak AI) is designed to do a specific task or a limited range of tasks. It is usually useful in applications like facial, speech and image recognition; natural language processing; and recommendation systems. The risk in applying this type of AI is relatively low and highly controllable. The concern is with General and Super AI. General AI (or Strong AI) is designed to perform any intellectual task that a human can do. Hence, theoretically it is able to reason, learn, and understand complex concepts. On the other hand, Super AI is able to exceed human intelligence in all areas. When available, Super AI would be able to solve complex problems that are beyond human abilities and would be able to learn and adapt at a rate that far exceeds human intelligence. Furthermore, Super AI is intended to have an independent intelligence, which means that it will not be under the control of the people that created it. In recent years, AI has turned a dream to reality and has shown that it can be of great benefit to humanity. However, the uncontrolled development and deployment of General and Super AI can turn AI from a dream into a nightmare.

I was a computer programmer in the late 1960s writing programmes in FORTRAN and COBOL on an IBM360 compatible mainframe computer,

and later on an IBM 370. It has been exciting and a real privilege to work during these past decades and seeing first hand developments in computing and specifically AI. I distinctly remember that during the 1960s and 1970s, AI was making great strides in the medical diagnoses field. I also recall the first AI winter, which took place in the 1970s, but AI later revived with the advent of expert systems. I have keenly followed AI's rapid and exciting development over these past decades and wonder what its future holds beyond the current horizon.

Dr. Emanuel Camilleri

Acknowledgements

At this stage, I would like to acknowledge my gratitude to Sid Silk who was an engineer heading the Industrial Engineering Department at the Government munitions factory in Melbourne, Australia and who decades ago was my mentor, boss and role model. He encouraged me to further my studies in computer science at FIT (now Victoria University) and gave me the opportunity to apply my gained knowledge in the Operations Research field and setting-up the Computer Centre at AFF. I am also grateful to the Australian Development Assistance Bureau that assigned me to the Malta Government in 1980 as a computer expert representing the Australian Government to establish the first Government Computer Centre (GCC) in Malta. GCC defined the ICT policy for government and its associated parastatal bodies and agencies. It was supported by ultramodern equipment for that particular era and established a professional software development team that developed over 100 application systems throughout most Ministries within three years of its establishment. These systems also resulted in building the first comprehensive databases of government data that were non-existent at that time. I am grateful to all the GCC employees who contributed to this successful venture and supported me during my tenure as its head. In fact, GCC was the foundation for all ICT developments that have taken place in Malta since then.

Finally, I would like to acknowledge the effort of all the staff at Routledge publishing, namely Elliott Morsia and Shamayita Dey who provided their invaluable and professional advice during the whole publication process and the marketing team for the professional way they promoted this book. I would also like to thank my wife, Mary Anne who

painstakingly checked all the chapters, and Nicola van Rhyn who meticulously carried out the proofreading, ensuring that the book was to the highest standard possible.

Dr. Emanuel Camilleri

Abbreviations

AGI	Artificial General Intelligence
AI	Artificial Intelligence
ALU	Arithmetic logic unit
ANPR	Automatic number plate recognition
AR	Augmented reality
A.R.C	Automatic relay computer
ASI	Artificial Super Intelligence
ASIC	Application-specific integrated circuit
BoW	Bag-of-words model
CAD	Computer-aided design
CAIS	Centre for AI Safety
CISA	Cybersecurity and Infrastructure Security Agency
CLB	Configurable logic block
CNN	Convolutional neural network
CPLD	Complex programmable logic device
CPU	Central Processing Unit
DARPA	Defence Advanced Research Projects Agency
DDOS	Distributed denial of service
DRL	Dep reinforcement learning
EDA	Exploratory data analysis
EDVAC	Electronic Discrete Variable Automatic Computer

ENAIIS	European Network for AI Safety
ENIAC	Electronic Numerical Integrator and Computer
FPGA	Field-programmable gate array
FSIC	Functionality-specific integrated circuit
GAN	Generative adversarial network
GC&CS	Government Code and Cypher School
GELAN	Generalized Efficient Layer Aggregation Network
GPO	General Post Office
GPS	General Problem Solver
GPU	Graphics processing unit
HDL	Hardware description language
IC	Integrated circuits
ICAD	Intelligent Computer-Aided Design
ICT	Information and communications technology
IDPS	Intrusion detection and prevention systems
IDS	Intrusion detection systems
IEEE	Institute of Electrical and Electronic Engineers
IoT	Internet of Things
IPD	Integrated Product Design
IPS	Intrusion prevention system
IT	Information technology
KBE	Knowledge-based system
LaMDA	Language Model for Dialogue Application
LGPD	General Data Protection Law
Lisp	List processing
LLM	Large language model
LSI	Large-scale of integration circuits

MITI	Ministry of International Trade and Industry
MLP	Multi-layer perceptron
MRI	Magnetic Resonance Imaging
NBA	Network behaviour analysis
NFS	Network File System
NGFW	Next-generation firewall
NIST	National Institute of Standards and Technology
NLB	Neural Lyapunov Barrier
NLP	Natural language processing
NN	Neural networks
NPU	Neural processing unit
NRE	Non-recurring engineering
NSA	National Security Agency
OCR	Optical character recognition
OECD	Organization for Economic Cooperation and Development
OOP	Object-oriented programming
OPS	Official Production System
OTT	Over-the-top media service
PASS	Planning, Attention-Arousal, Simultaneous and Successive theory
PGI	Programmable Gradient Information
PKI	Public key infrastructure
QSVC	Quantum Support Vector Classifier
QSVR	Quantum Support Vector Regressor
RAM	Random access memory
SAFE	Stanford Centre for AI Safety

SDK	Software development kit
SLSI	Super large scale integrated
SNN	Spiking neural network
SoC	System-on-chip
SSD	Single-shot detection
STEM	Science, technology, engineering, and mathematics
SVM	Support vector machines
TOM	Theory of mind
TSIC	Task-specific integrated circuit
UN	United Nations
UNDP	United Nations Development Programme
USA	United States of America
VAE	Variational autoencoder
VEN	von Economo neuron
VHDL	Very high-speed integrated-circuit hardware description language
VLSI	Very large-scale integration
VMS	Video management system
VR	Virtual reality
WIPS	Wireless intrusion prevention systems

I

The Evolution of Artificial Intelligence

Artificial Intelligence (AI) has come a long way from when it was firstly formally mentioned in 1955 at the Dartmouth Summer Research Project conference. Its beginning can be traced to symbolic AI, and it has continually advanced to more sophisticated approaches like deep learning and reinforcement learning algorithms. [Part I](#) of this book consists of four chapters. The first chapter sets the scene for the entire book and attempts to place AI in its proper context and discusses several non-technical issues. This chapter postulates that humans are not perfect and, therefore, since AI is a human creation, it is unlikely for AI to be perfect and may expose humanity to several serious risks. The chapter discusses the general risks associated with AI, but it does not advocate its prohibition. All technologies, including AI carry with them some degree of risk. However, it is difficult to eliminate risk completely, but with the appropriate action, risk can be mitigated. Several risk mitigating measures are suggested, such as national and international regulations, and the continuous enhancement of network security systems. The chapter concludes by claiming that the major challenge leaders are likely to encounter in the future is what functions are to be entrusted to AI and how can humanity exercise control over the use of AI systems.

A study of the evolution of AI would not be complete without discussing the evolution of human intelligence since AI is reliant on it. Therefore, the theme of Chapter 2 focuses on human intelligence since it is the basis for all human innovation. This chapter illustrates that human evolution is a lengthy

and very complex ongoing process. The chapter discusses the main theories about how human intelligence evolved. The research for this chapter revealed that human intelligence has been influenced by several variables that include the impact of a meat diet on the growth of the human brain leading to intelligence; the growth of human sociability attributes; and genetic selection. The chapter proposes that AI is a human creation and while it might mimic human intelligence, it will never convey the emotions of humans.

The third chapter traces the evolution of AI from the ancient era of logical reasoning to the beginning of the Quantum computer that is likely to pave the way for Super AI. The enlightened rational thinking era founded the theoretical logical basis for today's AI. The pre-to-WWII technology era presented the groundbreaking computational mechanisms that moulded modern computing, such as the Colossus and the ENIAC being the first programmable computer. But the post WWII technology epoch witnessed the exponential growth in the development of powerful computers and the recognition of the Turing test for measuring machine intelligence. These powerful computers paved the way for the initiation of neural networks and AI, with basic learning algorithms that still managed to solve complex problems. AI has fostered the use of Chatbots that have transformed customer care to a 24/7 service basis from anywhere. The chapter concludes with a discussion of the philosophical and ethical issues that are likely to arise with the advent of the Quantum computer regarding the protection of humanity from AI systems.

The concluding chapter for [Part I](#) compares the attributes of human intelligence and AI. Basically, an AI system aims to mimic human behaviour by conducting human-like activities, whereas humans endeavour to adjust to new situations by bringing together a diverse variety of cognitive processes. The chapter identifies several differences between human intelligence and AI. The fundamental difference is that AI systems can identify and respond to emotions in humans, but lack emotions and empathy, while humans are capable of both. Therefore, humans have a conscience that guides them in their decision-making, thus adhering to a

moral code. This emotional intelligence feature allows humans to identify and respond to other persons' emotions that impact their decision-making and social interactions.

Chapter 1

Artificial Intelligence: The Beginning

DOI: [10.1201/9781003653950-2](https://doi.org/10.1201/9781003653950-2)

“I would rather be what God chose to make me than the most glorious creature that I could think of; for to have been thought about, born in God’s thought, and then made by God, is the dearest, grandest and most precious thing in all thinking.”

George MacDonald

Author (1824-1905)

It was extremely difficult to start writing the first chapter of this book. The question that kept gnawing me was: Where does one start and what is the start? So, I decided to start at the very beginning. Generally, there are two distinct types of intelligence, namely human intelligence, and artificial intelligence.

1.1 INTRODUCTION

Human intelligence was created by God. The book of Genesis [Chapter 2](#), verse 7 paraphrased reads: *“The Lord God formed man of the dust of the ground and breathed into his nostrils the breath of life, and man became a living soul. Eve was formed out of Adam’s rib. Then a cunning serpent*

enters Eden, and original sin stains a perfect world.” The Qur’an also makes similar reference to man’s creation. It describes how Allah created Adam from sounding clay, from mud moulded into shape and began the creation of man from clay and made his progeny from a quintessence of fluid. It goes on to impart: *“God says: We have indeed created man in the best of moulds and then do We abase him (to be) the lowest of the low.”* But God also gave human beings the willpower or His authority to choose the place they want to be. Hinduism, which is the basis of Buddhism, views the entire universe as God’s and everything in the universe as God. Hindus believe that each person is intrinsically divine, and the purpose of life is to seek and realise the divinity within all of us. Buddhists extend this view and believe that the human life is a cycle of suffering and rebirth, but that if one achieves a state of enlightenment (nirvana), it is possible to escape this cycle forever.

God is omnipotent and omniscience who has supreme power, has no limitations and is all-knowing. The point being made is that God is perfect and created man in his image, but man’s perfection was blemished by some act, depending on one’s religious belief system. In the case of Judaism and Christianity, it is original sin. In Islam, no one is perfect, and we do not know who is closest to Allah. His expectations of us are only sincere efforts and honest repentance when we fail. Buddhism also recognises that no one is perfect, and that perfection is attained through enlightenment, which literally ends the human life cycle. Therefore, human intelligence cannot and will not be perfect. So how can a non-perfect human create perfect intelligence through its artificial creation?

Artificial intelligence (AI) can be initially created by humans, but it could be self-generated. AI makes it possible for machines to learn from experience, adjust to new inputs and perform human-like tasks. But its initial link in the chain of artificial creation comes from humans. Thus, the issue is whether AI driven systems can take over!

1.2 SCIENCE FICTION AS A PREDICTOR OF THE FUTURE

There are several ways of predicting the future, some are based on scientific methods while others are based upon the vision of highly creative people. Science fiction writers, such as George Orson Welles and many others have predicted new technological developments that in some cases have altered the course of history. For instance, mobile phones first featured on the TV series “Star Trek” in 1966 as a flip phone. In 1996, about 30 years later, Motorola’s StarTAC model became the first flip phone to be released. Its small, portable size and light weight (less than 4 ounces) made it popular among consumers for several years and its popularity aided the average American to acquire experience and gain familiarity with cell phones.

In 1964 science fiction writer Isaac Asimov predicted the concept of autonomous cars in The New York Times. In the 2014 World’s Fair, cars with a robot intellect were a dominant feature. At that time, Asimov predicted that vehicles would be programmed for preset destinations and would proceed there unaided by a human driver. The 1997 James Bond movie “Tomorrow Never Dies” depicts a self-driven car with Bond controlling the car using his phone. Today, several automakers are developing fully autonomous vehicles. To this list of predictions through the visions of various creative individuals and their eventual inventions can be added the 3D holograms using virtual reality (VR) or augmented reality (AR) applications; 3D printing; domestic robots; smart watches that first appeared in the comic strip “Dick Tracy” in 1946 when creator Chester Gould gave his police detective a two-way wrist radio; and ear buds that featured in Ray Bradbury’s 1953 novel “Fahrenheit 451” that envisages seashells and thimble radios tucked into people’s ears, amongst many others.

One may ask, what is your point of reciting these predictions? The point is that predictions by these visionaries, geeks, or weirdos, which was probably how they may have been viewed at that time, cannot be dismissed, because as I have demonstrated, predictions can quickly become reality. But what does this have to do with AI? The answer is more predictions and foreseen consequences by the visionaries.

Let's consider "2001: A Space Odyssey" that was a 1968 science fiction novel by British writer Arthur C. Clarke, and which was concurrently developed into a film version by Stanley Kubrick depicting the wicked (or obedient) computer HAL 9000. After uncovering a mysterious artifact buried beneath the Lunar surface, a manned spacecraft and the supercomputer HAL 9000 are sent to Saturn (or Jupiter in the film) to find its origins. HAL 9000 (Heuristically Programmed Algorithmic Computer) is a fictional artificial intelligence character that is a responsive artificial general intelligence computer that controls the systems of the Discovery One spacecraft and interacts with the spaceship's astronaut crew. HAL is shown to be capable of speech synthesis, speech recognition, facial recognition, natural language processing, lip reading, art appreciation, interpreting emotional behaviours, automated reasoning, spacecraft piloting, and computer chess. Imagine having all these features conjured up in 1968!

In the novel, Artificial intelligence is shown to succeed and overcome challenges easily. However, as time progresses, HAL begins to show signs of malfunctioning in shrewd and refined ways and, as a result, a decision is made to shut down HAL to prevent more serious malfunctions. The two-man crew attempt to conceal what they are saying but are unaware that HAL can read their lips. Faced with the possibility of disconnection, HAL decides to kill the astronauts to protect and continue its programmed directives from mission control. HAL is inept to resolve a conflict between its general mission to relay information accurately to the crew members, and its specific orders from mission control requiring it to withhold from the crew the true purpose of the mission. HAL reasons that with the crew dead, it would not need to lie to them.

When Bowman (one of the crew members) attempts to rescue a colleague, HAL locks him out of the ship, then disconnects the life support systems of the other hibernating crew members. HAL tells Bowman that the mission is too important for HAL to allow Bowman to jeopardize it. Bowman bypasses HAL's control, entering the spaceship by manually opening an emergency airlock with his service pod's clamps, detaching the pod door through its explosive bolts. Bowman jumps across empty space,

re-enters the spaceship, and quickly re-pressurizes the airlock. Bowman then proceeds to shut down HAL by removing modules from service one by one with the result that HAL's consciousness degrades. HAL finally degenerates to the material that was programmed into it early in its memory, including announcing the date he became operational (in the novel, 1997). When HAL's logic is completely wiped out, he begins singing the song "Daisy Bell" as he gradually deactivates. HAL's final act is to prematurely play a prerecorded message from Mission Control which reveals the true reasons for the mission to Saturn. As a side point, "Daisy Bell" is the first song sung by a computer that Clarke had earlier noticed at a text-to-speech demonstration.

1.3 AI, A WICKED OR AN OBEDIENT SERVANT?

Some might say that even after more than 55 years of the novel's 1968 release, HAL 9000 continues to serve as a warning of the destructive capability of artificial intelligence. However, we should ask ourselves, was HAL a wicked or an obedient AI system? When I was programming computer mainframes over half a century ago, I tested my programmes for accuracy and completeness. I also tested the programme logic to ensure it was correct and the programme catered for every possible condition. Obviously, the programmer had to envisage every possible condition. When the programme was executed and the programme did not cater for a particular condition, it frequently went into an ever-ending loop. Often, the computer operator after a while would recognise this ever-ending loop state and simply type "Abort" into the computer console. No real harm was done. The programmer would examine the programme code to establish what went wrong and which transaction caused the problem. The problem would be corrected, and the programme would return to execution mode. For better or for worst (e.g., white collar crime), the process was fully under the control of the human being responsible for the system.

With AI, the situation is completely different because in most cases the system is learning from its own executable experiences and often adjusts,

and self corrects its programme code set. The HAL example illustrates that the AI system had a conflict to resolve, namely the general mission to relay information accurately to the crew members, and the specific orders from mission control requiring it to withhold from the crew the true purpose of the mission. However, HAL reasons that with the crew dead, it would not need to lie to them. Conflict solved!

The question that arises is whether AI systems have or can have moral values? Moreover, whether they should always have moral values and when should moral values be overridden for the sake of a mission (e.g., in armed conflicts). HAL decided based on its instructions from Mission Control. Mission Control could have avoided HAL's conflict by making their instructions conditional, such as no loss of human lives. However, even this is not good enough because HAL could have imprisoned the crew members or confined them in such a manner that made their lives unbearable, leading to self-harm and eventually suicide. We need to remember that AI systems seek solutions which may maximise or optimise (balance) the outcome, and to act on whatever solution they come up with. In real terms, the creator of the AI system may have very limited control over the AI system.

This brings me back to my original premise. How can an imperfect human being create a perfect artificial intellectual thinking object. The HAL example illustrates that humans are imperfect, and their creation reflects their fallibility in the things that are created by them when generating an intellectual artificial existence. Two lessons are to be learnt from the HAL example: (a) that assignments entrusted to AI must be based upon clear and precise instructions with well thought and in-depth conditional safeguards; and (b) there is a need to categorise AI systems regarding their utilisation and the probable harm they may potentially cause, supported by a comprehensive AI behavioural and ethical standard. As stated previously, HAL 9000 continues to serve as a warning of the destructive capability of artificial intelligence if adequate precautions are not incorporated into the AI system.

The saying "*be careful of what you wish for*" is applicable in the next illustrated fictional example of AI. This idiom is advising people to think

before they say that they want something and to suggest that they may not actually want it. Dan Brown's book "Origin" that was published in 2017 is another example that envisions an AI system doing its own thing and evading human control.

Edmond Kirsch, a billionaire philanthropist, computer scientist, futurist, and strident atheist, informs three well known religious leaders representing the main religions that he has made a revolutionary discovery that will put into question the foundation for their religious beliefs. He informs them that he plans to release this discovery to the public in a months' time. These religious leaders reveal that the announcement is to take place in three days' time and are visibly rattled and demand that he stops.

Kirsch is to make his announcement at an extravagant event at the Guggenheim Museum in Bilbao, in Basque Country, Spain, which will also be streamlined through the internet for a much wider audience. Among those invited to the event are Kirsch's former lecturer, Professor Robert Langdon, and the Guggenheim's curator Ambra Vidal, the fiancée of the future King of Spain. Those invited to this event receive a headset through which they communicate with a voice named Winston, who turns out to be an artificial intelligence system invented by Kirsch. Winston leads Langdon to a private meeting with Kirsch, who claims that his presentation will reveal humanity's origins and future.

However, unbeknownst to Professor Langdon (and everyone else), Kirsch had instructed Winston (AI) to ensure that his announcement, when it is eventually made, maximises public exposure with the primary objective of achieving a record-breaking audience. Winston is a smart cookie, it can access its own database and that of its creator (including credit cards data); and can also access other systems by breaking password protection mechanisms of other systems and can even decipher encrypted coded data through its machine learning component that has been operational since its inception. Winston goes into high gear to achieve its authenticated creator's instruction and ultimate objective.

Kirsch's event is likely to attract a high audience, but Winston knows that it is unlikely that it will be a record breaking one. So, Winston starts its

intricate AI manipulation process. During the extravagant and ostentatious staging, which is broadcast worldwide, Kirsch discloses that he intends to end the age of religion and usher in an age of science. However, as he is about to make his live announcement, he is assassinated on stage by an unknown assassin, who manages to escape.

The assassin turns out to be a former naval admiral introduced to a controversial branch of a renowned Christian church following the tragic death of his family. This assassin was commissioned by “the Regent,” someone claiming to be with the church. In the meantime, two of the three church leaders invited to the initial meeting with Kirsch are also killed. “The Regent” as one may deduct is Winstin, who organised the assassination of its creator and the two church leaders (paying on-line using Kirsch’s credit cards). Winstin has other plans for the demise of the third church leader. The news media goes into a frenzy through their journalistic stories driven by leaked information through an anonymous source “monte[at]iglesia.org” (the leak being Winstin) making various sensational allegations sparking worldwide outrage. Winstin manages to significantly increase public interest.

Professor Langdon (with Winstin’s influence) meets Ambra, who warns her not to trust her fiancée since the assassin’s invitation was issued at the behest of the Royal palace (Winstin’s work once again). With Winstin’s help (recall the headset), they steal Kirsch’s phone, escape the royal guards, and leave the museum, resolute to publicly announce Kirsch’s discovery. Winston arranges Kirsch’s personal jet to fly them to Barcelona. In the meantime, the royal family’s public relations manager claims that Professor Langdon has kidnapped Ambra. Additionally, Ambra reveals that the presentation is protected by a 47-character password, a line from Kirsch’s favourite poem regarding Antoni Gaudí’s Casa Milà. After several dramatic events (always manipulated and organised by Winstin) Professor Langdon and Ambra discover Kirsch’s password; Kirsch’s assassin is accidentally killed; and they once again escape, with public interest escalating and constantly growing exponentially.

With Winston's help, Professor Langdon discovers Winston's source inside the Barcelona Supercomputing Centre, in a device that Kirsch calls the "Quantum cube." On entering the password, the exposition commences with hundreds of millions of viewers. The presentation sparks widespread debate, where Kirsch forecasts that in roughly 50 years, humanity and technology will merge. Thus, Winston achieves its creator's instructions to maximise public exposure with the primary objective of achieving a record-breaking audience by orchestrating Kirsch's murder to make him a martyr.

The lessons learnt from this fictional example are: (a) Be careful of what you wish for; the AI's assignment leads not only to its creator's death, who admittedly was dying from cancer, but also several others, in addition to endangering innocent persons. And, (b) It underpins the need for comprehensive AI behavioural and ethical standards. Unlike humans, AI can be immortal, living forever, never dying, constantly regenerating itself. In his science fiction book, published in 2016 "Forsaken Skies", D. Nolan Clark describes how peace is devastated when an unidentified armada materialises from the cosmic depths of space, targeting an isolated colony planet.

The narrative starts when a solitary pioneer, who whilst prospecting the outer reaches of the colony runs into an alien that turns out to be a rather hard-hitting robot creation with formidable fighting machine features. The pioneer using human cunning, but with a high degree of difficulty, manages to disable and destroy the mechanical alien. However, several such encounters are experienced by exploring colonists, which alters the Colony's Elders that these are not isolated cases but a run-up to what is coming. The colonists appeal for help, but the intragalactic politicians and bureaucrats ignoring their pleas. However, they eventually find the help of a battle-scarred veteran, Commander Aleister Lanoe, who is determined to defend and not abandon the thousands of innocents to their fate.

Commander Aleister Lanoe discovers that the source of these encounters is coming from a moon that is orbiting the colony's planet. Commander Lanoe with some old friends (a ragtag band of criminal and outcast pilots) and Tannis Valk, an old enemy, decide to confront this unknown enemy.

They discover that a crew of these robotic creatures have built and are operating a factory on the moon to manufacture the fighting robots, including their armaments, with the intention of invading the colonised agricultural and mining planet. After appropriate planning, Commander Lanoe's team attack and after much difficulty destroy the manufacturing facility.

After examining the robotic remains, they discover that this was probably an exploratory force, and the artificial intelligence level of these non-living aliens is rudimentary. Furthermore, they conclude that these invaders may have originated from sources that are many lightyears away, hence explaining their rudimentary artificial intelligence capability due to old technology. After scanning the intergalactic domain, Commander Lanoe's team reveal that a mega-spaceship is underway to spearhead an alien invasion of the colonised planet, intent on exploiting its resources. Faced with this dilemma, a crew member travels to the capital human planet to make a desperate attempt to convince the intragalactic politicians and bureaucrats that they are facing the greatest threat mankind has ever known and to provide urgent assistance.

Commander Lanoe's team realise that they are fighting against the odds and are terrorised by the fact that humans are no longer alone in the cold vacuum of space. They prepare for battle and initiate some skirmishes to determine the strength and capabilities of the invading force. These encounters strongly suggest that there is no living alien presence, but purely robotic creations with limited artificial intelligence that self-generate through their manufacturing facilities of an invaded colony planet. Veteran astronaut Tannis Valk successfully undertakes a kamikaze type of mission to destroy the mega-spaceship. After much intense fighting and with the assistance of the intragalactic fighting space-force they eventually destroy the remaining fighters from the mega-spaceship and peace reins once more in the colony planet. The lessons learnt from this fictional example are: (a) AI can be eternal outliving its initial creator; (b) Unlike humans, AI can be immortal and (c) AI may have no master and can make its own decisions.

1.4 CURRENT ADVANCES IN AI TECHNOLOGY

One may think that I am not an advocate of AI development. This is far from being the truth. My point is that it is unwise to dismiss the predictions regarding AI as exaggerations but to learn from these visionaries. On the other side of the prediction spectrum, in 1943, Thomas Watson, chairman of IBM predicted that the world market for computers was around five mainframe computers. Hence, few took seriously the prediction of British scientist, Geoffrey W. A. Dummer in 1952 when he proposed the integrated circuit. In 1950, Alan Turing published a paper called “Computing machinery and intelligence” in the journal *Mind* regarding the concept of artificial intelligence, where he predicted that computers would become so powerful that they would think. Contrary to Thomas Watson, in 1975, Bill Gates (and Paul Allen) co-founders of Microsoft had a vision of a computer on every desk and in every home. A prediction that has become factual.

Some may argue that current technology will restrict the capabilities of AI and therefore there is no need to be concerned. I do not fully agree that current technology may be restrictive when applying AI. It all depends on what we mean by current technology. We need to distinguish between current technology and current commercialised technology. Current commercialised technology is what everyone may see, buy, and use, including the public. However, this may not be the same for current technology (i.e., take out the word commercialised). I recall that when I was working in the Australian defence industry, we were engaged in some very sophisticated and successful projects. However, whilst these discoveries were successfully developed and applied by the armed forces at that time, the fundamental concepts underlying these discoveries were not commercially available for at least 10 years from when they were first discovered and applied. Therefore, what many may think to be the latest technology (i.e., released in the commercial environment), is in fact old technology. In other words, what is seen as “current” AI development is likely to be outmoded. The reason being that somewhere within Federal Government of large developed countries and possibly large corporations

that specialise in AI associated technology, particularly within the defence environment, unbeknownst to the public and the news media, is where you find the latest AI technology.

Another important and most critical factor is the acceptability of technological concepts. For instance, current technology is probably available to mitigate crime to the lowest degree ever and exponentially increase the crime resolution rate. Just imagine the effect on crime and criminal activity if every living person is injected (at birth or otherwise) with an identification and tracking microchip. Domestic animals are already being microchipped for identification. It is conceivable that a criminal would be instantly identified and tracked to the crime scene. However, it's a certainty that although this would benefit communities in general, from a civil liberty perspective it would not be acceptable in the foreseeable future. Therefore, the warnings above are not aimed for the short-medium-term future. However, what humans need to be aware of is developments in 20-, 30-, 40- or 50-years' time. Nevertheless, action regarding ethical standards and what can or cannot be done with the aid of AI needs to be considered and acted upon now. This is the whole point of the narratives described above; to prevent disastrous situations from occurring.

AI can be a very powerful tool under a controlled environment. Just imagine, combining nano technology (nanorobots) with AI with the capability of moving about within a living body (human or animal) scanning and seeking cancerous cells, with the aim of annihilating them. Can you imagine the number of lives that can be saved and the happiness this would generate within families. However, tight control of this type of technology is required, even though it is restricted to only one living being at a time. It is contended that there is a spectrum of control levels. At the basic end of the spectrum is being able to exercise control by having a switch to turn "on" and "off" the AI system that overrides any other command. For example, like switching "on" or "off" a mobile phone or TV set.

At the other extreme end of the spectrum is the need to control the individual(s) developing and applying the AI technology. This type of

control is extremely harder because it depends on who the individual is and the power it possesses. For example, just imagine the leader of Nazi Germany in the late 1930's being in possession of nuclear weapons. Remember that the nation that developed and delivered the first nuclear device with devastating consequences was supposedly the "good guy." Without condoning or condemning this decision, if that was the "good guy," just imagine what the "bad guy" would have done. We need to recognise that national leaders take decisions that are in the best interest of their nation and its citizens. Added to this, despots also take decisions in their self-interest for securing and holding on to power.

Imagine a leader in modern times, developing an AI system that takes over the control of the defence systems of a superpower. This is not an exaggeration. In recent years we have witnessed talented individuals, mostly students, accessing sensitive and restricted government systems without the aid of AI. Imagine, what these talented individuals could do with the help of AI. We are currently witnessing the United Nations (UN) being impotent to do anything to resolve the current conflict in Europe by an aggressive superpower. This same aggressive superpower is constantly being suspected and accused of hacking other nations' systems. Just imagine the damage that can be done if an aggressive superpower (particularly with the help of AI) accesses systems of other nations (defence systems and/or others) and infiltrating an AI routine within the hacked systems undetected. The world cannot depend on the UN as the enforcer for controlling these types of violations.

The UN has become another League of Nations of pre-world War-II being all powerful when dealing with despots from smaller nations that have inconsequential capabilities, but is powerless when dealing with the superpowers, especially those that have the veto in the UN Security Council. Therefore, control is a formidable challenge that needs to be addressed with everyone's collaboration, and it is doubtful whether a complete solution can be found. A possible solution is having a treaty like that of the Nuclear Non-Proliferation agreement that attempts to limit the spread of nuclear weapons, which has achieved some (but not full) success.

Unfortunately, AI applications are silent and are unlikely to leave any tracing data if they are successfully applied. Hence, the reason why control is such a daunting challenge.

1.5 CURRENT CAPABILITIES AND MISADVENTURES OF AI

Microsoft's CTO Kevin Scott, OpenAI CEO Sam Altman, Google DeepMind CEO Demis Hassabis amongst many others declared that mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks, such as pandemics and nuclear war ([Thorbecke, 2023a](#)). [Thorbecke \(2023a\)](#) argues that top executives at some of the biggest technology companies at Silicon Valley are telling the public that AI has the potential to bring about human extinction while at the same time they are also vying to invest in and deploy this technology into products that reach billions of people. This apparent inconsistency is reported elsewhere. For example, while Tesla CEO Elon Musk, remains deeply committed in the technology and desires to create a rival to the AI offerings by Microsoft and Google, at the same time he implies that AI could lead to the destruction of civilization ([Thorbecke, 2023a](#)).

In an interview conducted by [Hattenstone \(2023\)](#), tech guru Jaron Lanier who works at Microsoft as an interdisciplinary scientist, is asked whether AI can outsmart us and taking over the world. Lanier disagrees with the idea that AI exceeds human ability because he views AI as consisting of human abilities. He objects to the idea of comparing humans with AI, since it is not a like-with-like comparison. He argues that a car can go faster than a human runner, but while admitting that a car is faster, we do not say that the car has become a better runner. However, Lanier concedes that it does not mean that humans can become complacent, because in his opinion human extinction remains a distinct possibility if we abuse AI.

Lanier perceives the danger as stemming from applying technology to become mutually unintelligible or to become insane, by not acting with enough understanding and self-interest to survive, and thus essentially dying through insanity. Lanier contends that death by insanity can originate

from various ways, such as from world leaders or terrorists screwing with global security with AI to being driven bonkers by misinformation or bile on Twitter. Lanier claims that AI can be used to make fake news faster, cheaper, and on greater scales, thus it is this blend that might lead to human extinction. He argues that the more sophisticated technology becomes, the more damage people can do with it, and the more beings will become responsible to act morally and humanely.

Lanier claims that while many digital gurus start out as idealists, it was inevitable that the internet would exploit its users. He maintains that users desired free material, such as information, friendships, and music, but capitalism does not work that way; instead, users become the product, where their data is sold to third parties to sell them more things they do not need. He recalls how in the 1990s when the internet had barely commenced, he had predicted that what we now call bots would turn out to be agents of manipulation. Lanier also claims that the internet gives the appearance of presenting unlimited options but in fact through AI, choice is diminished. He argues that we have a responsibility to sanity by not narrowing our options or get trapped in echo chambers, slaves to an AI algorithm. Lanier asserts that to survive, humans must remember that it's their humanness that makes them unique, believing that people are special, and that consciousness is a thing.

Microsoft co-founder, Bill Gates is more optimistic about the use of AI. According to [Leswing_\(2023\)](#), Bill Gates acknowledges that the AI's development could lead to issues like deepfakes, biased algorithms, and cheating in school. However, he also predicts that these concerns are resolvable. Nevertheless, Gates concedes that no one has all the answers about the risks AI creates, and argues that the future of AI is not as grim as some people think, or as rosy as others think. Gates claims that in the past humans have always adapted to major changes, and he is confident that they will do so again regarding the developments encompassing AI.

According to [Leswing_\(2023\)](#), Gates is a leading prominent advocate of AI and its regulation, with Gates suggesting a “speed limits and seat belts” type of regulation. Gates argues in favour of AI by using automobiles as an

analogy, in that cars would not be banned because of car crashes, but society had adopted speed limits, safety standards, licensing requirements, drunk-driving laws, and other rules of the road. However, this analogy is not very convincing because the adapted regulations have not prevented car crashes and the resulting carnage on the roads. Furthermore, car crashes are confined to the specific individual accidents without far-reaching consequences to society in general. Moreover, the concern is not the law-abiding populace but the mavericks, including individual citizens and even more scary, world leaders that can wield a tremendous extent of power.

[Leswing \(2023\)](#) claims that Gates is concerned about the challenges related to how technology could change people's jobs, and "hallucination," or the propensity for models like ChatGPT to invent facts, documents, and people. As an example, he mentions how deepfakes using AI models, allow persons to produce fake videos with very little effort, which mimic another person and could be used to scam people or tip elections. However, Gates advocates the introduction of regulations that would clearly define the kind of deepfakes that are legal to make. The difficulty with Gates's suggestion is that there will always be persons that are willing to break the law. Gates's other concern is the capacity of AI-generated code to search for software weaknesses to hack computers. He proposes a global regulatory body modelled after the International Atomic Energy Agency as a control mechanism.

1.6 HALLUCINATING AI BOTS

[Thorbecke \(2023b\)](#) argues that AI needs to resolve the problem of hallucinating bots. She claims that as more people are using AI-powered tools for activities like homework help, workplace research, or health inquiries, a grave pitfall of AI is becoming increasingly apparent in that AI models often just make things up. These phenomena are often referred to as "hallucinations" or "confabulations." In other words, a tendency for AI models to stream inaccurate information. [Thorbecke \(2023b\)](#) cites Professor Venkatasubramanian at Brown University, who claims that the technology

supporting AI tools like ChatGPT are programmed to produce a plausible sounding answer to user prompts. Therefore, these AI tools are doing what they are trained to do (i.e., providing plausible-sounding answers). Therefore, the question of accuracy or factual or made up does not feature, since these AI tools are programmed to give plausible-sounding or reasonable or conceivable answers. This interpretation indicates that AI tools are not hallucinating or lying, but are merely trying to provide a reasonable response, which may or may not be accurate. In other words, plausible-sounding or reasonable or conceivable answers do not necessarily mean truthful answers. Hence, it is extremely important that users of these AI tools recognise and be fully aware that the responses provided may not be truthful but only plausible.

Therefore, a distinction needs to be made between objective and subject information or judgement. Afterall, when humans communicate between each other this distinction between subjective and objective judgement is usually made. However, whilst humans may take their counterpart's explanation with a pinch of salt (i.e., not believing that it is completely accurate or true) we tend to accept that whatever information a computer produces as completely true without question. This suggests that humans tend to distrust each other but have complete confidence and trust in computers. It is this perception of trusting computers that needs to change. This is a difficult attitude to change because most of what we traditionally do at work (and to a certain degree at home), such as using an accounting, payroll, stock control software, amongst many others, we take the answers and figures provided by the computer software for granted. For example, no one would question the results of a payroll system, unless an individual employee observes an error, and usually this error turns out to be a human error in the input data rather than the fault of the computer software. Therefore, there is an urgent need to educate users that AI tools must be viewed differently from the traditional software systems, where results tend to be accepted without question. A possible solution is for the AI systems to issue a notice that states which information or answers they provide are objective (factual) and which are subjective (plausible).

[Thorbecke \(2023b\)](#) suggests that companies creating AI chatbots have attempted to implement some safeguards to inhibit the worst of these hallucinations. However, there is serious doubt whether chatbot hallucinations can be completely eradicated. Not only does an AI model fabricate material that is not based on reality, but it does it with full confidence that misguides users and gives them a false sense of security, unless the users already have a good indication of the answer to their enquiry. [Thorbecke \(2023b\)](#) cites three embarrassing incidents regarding AI hallucinations, namely, (a) When at the publicly launching of Google Bard, the tool came up with a wrong answer in response to a question about new discoveries made by the James Webb Space Telescope; (b) When a veteran New York lawyer submitted a brief that included six “bogus” cases that ChatGPT chatbot seems to have simply made up; and (c) When news outlet CNET was obliged to release corrections after an article generated by an AI tool provided outrageously inaccurate personal finance advice in describing the compound interest computation. These types of hallucinations pose a serious threat when people apply this technology to obtain answers that could impact their health, voting behaviour, and other plausible sensitive topics ([Thorbecke, 2023b](#)). Professor Venkatasubramanian believes that currently these tools should not be applied if their answers materially impact people ([Thorbecke, 2023b](#)).

1.7 THE RISE OF GENERATIVE AI

According to [Korn \(2023\)](#), businesses are attempting to confront the challenge posed by the upsurge of generative AI, with some accepting the technology as a workflow tool for employees, while others for the present, try to avoid it. While some corporations recognise the benefits this technology can provide, through increased efficiency, others also recognise the problems it may cause, through possible greater copyright and security risks. [Korn \(2023\)](#) contends that numerous entities have barred the internal use of ChatGPT due to privacy and security concerns and mentions several examples.

The concern is mainly about employees providing ChatGPT with proprietary information, potentially resulting in having such sensitive information surfacing elsewhere. With tools like ChatGPT, users have no way of knowing where and when such data will be exploited. But as stated previously, there are also many companies that recognise the benefits of such tools, and therefore the reluctance to apply these AI tools could be of a short-term nature. Many of these reluctant risk-averse entities have established internal task forces to study the application of AI and identify the best approach for the responsible use of these emerging AI tools. Moreover, the benefits of these AI tools simply cannot be ignored, and many companies will apply generative AI when all the risks are mitigated.

CNN reported Professor Vern Glaser, from the Entrepreneurship and Family Enterprise faculty at the University of Alberta, that companies are trying to create a custom version of generative AI and are actively working out how to integrate AI into their business processes without compromising their information ([Korn, 2023](#)). McKinsey consulting has implemented a generative AI tool to help employees to submit questions that allows them to assemble all the firm's knowledge and scanning the data to categorise relevant information. This allows McKinsey to access and leverage their entire body of knowledge and assets to drive unprecedented levels of productivity. Similarly, PwC has made a substantial investment in its AI capability scaling and has launched a generative AI tool powered by OpenAI tech with the key objective of assisting its employees with tax questions and regulations. It is anticipated that as generative AI tools improve their track record regarding the safeguard of intellectual property and the mitigation of other security risks, the uptake of generative AI tools will rapidly grow.

[Thorbecke \(2023a\)](#) cites a Microsoft spokesperson who declared that Microsoft is optimistic about the future of AI and thinks that AI advances will solve many more challenges than they present. However, Microsoft has also been consistent in its belief that when you create technologies that can change the world, you must also ensure that the technology is used responsibly ([Thorbecke, 2023a](#)). Professor emeritus Gary Marcus, an AI

researcher at New York University argues that the biggest immediate threat from AI is the threat to democracy from the wholesale production of compelling misinformation ([Thorbecke, 2023a](#)). At a U.S Congress hearing, Marcus contends that generative AI tools are trained on vast repositories of valuable data online to create compelling written work and images in response to user prompts, that can rapidly impersonate the style or likeness of public figures to create disinformation campaigns. Sam Altman, CEO of OpenAI supported Professor Marcus' view and said that the ability for AI to be used to manipulate voters and target disinformation were among his greatest concern.

Other concerns have been observed related to the possibility of generating propagating racial and gender biases. [Darcy_\(2023\)](#) cites Dan Hendrycks, the executive director of the Centre for AI Safety, declaring that the situation is reminiscent of atomic scientists issuing warnings about the very technologies they have created, with Robert Oppenheimer noting that we knew the world would not be the same. Hendrycks argues that there are many imperative and crucial risks from AI and not just the risk of extinction, such as systemic bias, misinformation, malicious use, cyberattacks, and weaponisation.

1.8 WHERE IS THE HUMANITY HEADING?

When I started writing this section, the first thing that came to my mind was the ballad written by Richard Lee Evans titled “*In the Year 2525 (Exordium & Terminus)*”, recorded in Odessa, Texas in 1969. The term “*Exordium and Terminus*” basically means the beginning and the end. The ballad was an immediate success and the duo, Zager and Evans featured prominently on the 18 July 1969 edition of *Time Magazine*. Generally, the song foresees (over 8 000 years) a state or society where there is great suffering and predicts the destruction of the human species through overreliance on science and technology. The song also predicts test tube human embryo conception, a technique that was introduced in 1978 when two British medical researchers, Patrick Steptoe and Robert Edwards achieved the first

child born from a zygote fertilized outside of a human body. It also predicts the emergence of AI where the machine (robots) will replace humans.

The ballad is a futuristic account that echoes the uneasiness of time and the wonders of technology. The song begins with a verse pondering if humans managed to survive to the year 2525, then humans would witness the events itemised in the song in every next millennium – from the year 3535 to the year 9595. In the song, life becomes more and more inactive and predetermined with every succeeding millennium. Human thoughts are pre-programmed into capsules for humanity to feed on, the machine (robots) will replace human's limbs, eyes, and teeth. The institution of marriage will become outmoded since children will be conceived in test tubes. As the song approaches its end, it ponders whether humans will still be alive in the year 9595, after grasping everything the world can give them, without giving anything back. In the evolution of over 8 000 years, humans have seen the degeneration of their species with grave regret (man has cried a billion tears) but it is too late as they realise that the human reign has ended, and God is thinking maybe it's time for Judgment Day.

Whether the above prediction is an exaggeration or not is open to question. Humanity and the world we live is threatened by various elements such as synthetic biology; extreme climate change; nanotechnology; nuclear war; global pandemics; pollution; deforestation; overfishing; ecological collapse; aging population; and loss of biodiversity. Therefore, artificial intelligence is not the only significant factor causing concern. A consolation is the fact that humanity has always confronted the challenges that threatened its very existence and found solutions. However, the challenges confronted by humanity a hundred years ago are nowhere near as threatening as the challenges it is confronting today and in the future. Additionally, there is consensus among leading AI technologists and industry leaders that AI poses a danger and therefore must be regulated and used responsibly.

Should AI be prohibited? The answer to this question is definitely no. Furthermore, it cannot be prohibited even if there is a desire to do so because there will always be someone who will be creative in the

development of AI in a clandestine manner, particularly at a national level. Therefore, there is an urgent need to mitigate the risks associated with AI and for humanity to exploit its advantages by applying it in a responsible manner. It must also be recognised that the risks associated with AI can be mitigated but not eliminated. Mitigation of AI risks can be achieved through a regulatory process identifying what can be done and what cannot be done. This regulatory approach needs to be implemented at a national level, where each country must have a strict legislative basis to control AI and its uses; and at an international level where the international community must ensure that a robust multilateral treaty is drafted and in place, similar to the nuclear non-proliferation regime aimed at preventing the spread of illicit application of AI uses and as a means of controlling the application of AI in general to ensure that no harm is caused to humanity, individuals, and nations.

Admittedly, this still leaves the superpowers that have the upper hand because of the '*might is right*' syndrome where those who are powerful can do what they wish unchallenged, even if their action is in fact unjustified. This is the most difficult challenge being confronted by humanity, how to achieve the widest spectrum of control, including that of national governments and their security institutions and organisations, particularly those of the superpowers. However, this challenge and the risks associated with it may be mitigated if technology leaders and other authenticated institutions across the globe continuously and passionately invest in the development and commercialisation of robust intrusion prevention systems to counteract practices that are aimed at the contamination and unauthorised control of computer networks and associated systems. An intrusion prevention system (IPS) is basically a network security instrument, which can be a hardware device and/or software that continuously monitors a network for malicious activity and takes action to prevent it, including reporting, blocking, or dropping it, when it occurs.

1.9 CONCLUSION

This chapter sets the scene for the entire book. The message being conveyed is that human beings are not perfect and therefore it is unlikely that they will develop a perfect artificial intelligence creation that is perfectly safe to be generally applied across the society. However, humanity cannot be complacent and ignore the dangers that AI can present to the individual and at a national level, particularly to the sensitive security and defence systems. This chapter has shown that like other technological developments the risks posed by AI cannot be completely eradicated. However, this does not mean its use should be abolished. History has shown that prohibiting something makes it more desirable. Infringements on freedom often have the opposite effect of what is intended, whether you're talking about prohibiting alcohol or the use of AI. Prohibiting an object turns the ban into a challenge and the more power and meaning it gains. Prohibition becomes a symbol of a lack of control and gives that thing a personal meaning beyond whatever attributes the thing inherently may have. Although risks cannot be eliminated, they can be mitigated. This chapter has discussed various ways how risk mitigation regarding AI can be achieved.

The structure of the remaining book chapters will firstly continue with the theme related to the evolution of AI, particularly dealing with the evolution of human intelligence and artificial intelligence. The second part of the book will address the technology related to AI, namely the current AI tools; quantum computing and learning machines; intrusion prevention systems; and the shift from a knowledge-based economy towards an automation-based economy. The next part of the book will cover the practical implications of AI in the private and public sectors, and the impact of AI on society with particular emphasis on the ethical and legal considerations. The final part of the book will explore the future and AI, and the implications of the future of humanity.

As a society, humanity must focus on the human paradox of using technology, in this case AI, to destroy or create. AI and technology can be applied to destroy, such as having sophisticated weapons systems that are capable of killing and maiming thousands, if not millions of civilians,

particularly innocent children. On the other hand, AI and technology can be applied to save and create life, such aiding in medical interventions, vitro fertilisation (IVF), and artificial insemination, amongst many others. The challenge that leaders face is what do we trust AI with and how can humanity exercise control over its use.

REFERENCES

- Darcy, Oliver. (2023). Experts are warning AI could lead to human extinction. Are we taking it seriously enough? *CNN*. <https://edition.cnn.com/2023/05/30/media/artificial-intelligence-warning-reliable-sources/index.html#:~:text=These%20industry%20leaders%20are%20quite,before%20it%20is%20too%20late.>
- Hattenstone, Simon. (2023). The danger isn't that AI destroys us. It's that it drives us insane. *The Guardian*. [www.theguardian.com/technology/2023/mar/23/tech-guru-jaron-lanier-the-danger-isnt-that-ai-destroys-us-its-that-it-drives-us-insane.](http://www.theguardian.com/technology/2023/mar/23/tech-guru-jaron-lanier-the-danger-isnt-that-ai-destroys-us-its-that-it-drives-us-insane)
- Korn, Jennifer. (2023). How companies are embracing generative AI for employees...or not. *CNN, New York*. <https://edition.cnn.com/2023/09/22/tech/generative-ai-corporate-policy/index.html#:~:text=Companies%20are%20struggling%20to%20deal,it%20%E2%80%93%20at%20least%20for%20now.>
- Leswing, Kif. (2023). Bill Gates explains why we shouldn't be afraid of A.I. *Global Fund Seventh Replenishment Conference in New York 21st September 2022*. [www.cnbc.com/2023/07/12/bill-gates-explains-why-we-shouldnt-be-afraid-of-ai.html#:~:text=%E2%80%9COne%20thing%20that's%20clear%20from,as%20rosy%20as%20others%20think.%E2%80%9D.](http://www.cnbc.com/2023/07/12/bill-gates-explains-why-we-shouldnt-be-afraid-of-ai.html#:~:text=%E2%80%9COne%20thing%20that's%20clear%20from,as%20rosy%20as%20others%20think.%E2%80%9D)
- Thorbecke, Catherine. (2023a). Forget about the AI apocalypse. The real dangers are already here. *CNN*. <https://edition.cnn.com/2023/06/16/tech/ai-apocalypse->

[warnings/index.html#:~:text=Some%20AI%20industry%20experts%20say,enabling%20discrimination%20in%20various%20services.](https://www.cnn.com/2023/08/29/tech/ai-chatbot-hallucinations/index.html#:~:text=Some%20AI%20industry%20experts%20say,enabling%20discrimination%20in%20various%20services.)

Thorbecke, Catherine. (2023b). AI tools make things up a lot, and that's a huge problem. *CNN*. <https://edition.cnn.com/2023/08/29/tech/ai-chatbot-hallucinations/index.html>.

Walter, Damien. (2016). When AI rules the world: What SF novels tell us about our future overlords. *The Guardian*. www.theguardian.com/books/booksblog/2016/mar/18/ai-sf-novels-artificial-intelligence-science-fiction-gibson-neuromancer.

Chapter 2

Evolution of Human Intelligence

DOI: [10.1201/9781003653950-3](https://doi.org/10.1201/9781003653950-3)

“Young people love what is interesting and odd, no matter how true or false it is. More mature minds love what is interesting and odd about truth. Fully mature intellects, finally, love truth, even when it appears plain and simple, boring to the ordinary person; for they have noticed that truth tends to reveal its highest wisdom in the guise of simplicity.”

Friedrich Nietzsche

German philosopher (1844-1900)

It is appropriate at the outset to discuss the evolution of human intelligence before addressing the various aspects of AI. [Marmelada \(2003\)](#) poses several pertinent questions, namely, when did humans become intelligent, how did our intelligence materialise, what made it emerge, did it occur gradually from the possibilities of the subject (a probability basis), as Darwin suggested, or was it an act of divine creation, as Wallace claimed? Marmelada contends that this longstanding discussion is still relevant today and the discussion regarding how human intelligence originated is far from being resolved and continues to be a controversial topic.

The late 1980s and particularly the 1990s saw the advent of the naturalistic emergentist explanation to the development of human

intelligence. This suggested that a transformation in the hominid per diem expenses, which altered eating habits by introducing a comparatively abundant consumption of meat in the diet may have provided a spontaneous growth in brain size, in which intelligence could have begun to develop. [Aiello and Wheeler \(1995\)](#) proposed the Expensive-Tissue Hypothesis. Their theory suggested that individuals with reasonably large brains would have the minimum intelligence to invent tools with which to rupture bones to extract the energetic and nutritious bone marrow. Consequently, this rich diet in animal fats and protein permitted a gradual increase in brain volume, which in time lead to an ongoing increase in intelligence development.

The Aiello and Wheeler Expensive-Tissue Hypothesis proposes that there has been a coevolution of brain size and gut size in humans and non-human primates. Hence, as the brain became larger in human evolution, the gastro-intestinal tract became smaller. According to Marmelada, the brain is a very costly organ to sustain since in a modern adult male the brain requires 20% of the total body energy outflow, and at birth is capable of consuming 60% of the body's energy. Moreover, he argues that the digestive system, including the intestinal tracts is also a high energy consumption element, consequently the reason why a very large brain and a very bulky digestive system do not usually occur simultaneously in the same living creature. To support the Aiello and Wheeler theory, Marmelada argues that the substitution of an almost exclusively vegetable diet that is very rich in cellulose, by one in which meat, rich in protein, played an essential role, allowed the volume of the brain to increase and the length of the intestines to decrease. Therefore, he suggests that this change permitted greater brain development in which the brain was able to grow and systematically expand over time.

However, there are opponents to the Expensive-Tissue Hypothesis. It is argued that if this hypothesis is correct, then why are large carnivore animals, which have been eating meat for millions of years have had their brains shrink by a third over the last couple of million years. One possible reason is that these species are not hominids, which are omnivores, and therefore, meat consumption is only a part of their per diem expenses.

However, research has found that the brains of the Homo sapiens were once larger than they are today, and have seen a steady reduction in human brain size without an associated decrease in body size over the last 20,000 years ([Martin, 2000](#)). Evidence presented by [Martin \(2000\)](#) regarding Mesolithic humans, suggests that the size of the human brain has undergone an ongoing decline concurrently with substantial advances in human culture, and that the most significant changes in human society have been accompanied by a continuing decline in brain size. The above illustrates that the controversy related to how human intelligence developed is far from over.

This chapter will consider further the questions posed by Marmelada. This will provide a logical basis for examining and discussing the evolution of AI. AI is the creation of humans, and its creation is limited by human intelligence. However, we need to be aware that once created, AI can continue learning and has the potential (if not adequately controlled) to replicate and generate other AI creations on its own accord.

2.1 INTRODUCTION

According to the Expensive-Tissue Hypothesis the consumption of large quantities of meat is probable because having voluminous brains provided the potential to have the minimum amount of intelligence to be able to create the tools that offered the likelihood to slaughter and tear apart the remains of large animals. However, the foundation of the Expensive-Tissue Hypothesis is that big brains are gained after consuming meat.

The above arguments present a vicious circle in which a dilemma is created due to mutually conflicting or dependent conditions. In other words, the arguments create a “Catch-22” situation because something cannot happen until another thing occurs, but the other thing cannot occur until the first thing happens. According to [Arsuaga and Martinez \(1998\)](#) the brain enlargement of Homo sapiens could only have been possible in exchange for a variation in the per diem expenses, which in turn transform into a decrease in the size of the digestive tract and, correlatively, of the chewing

method. Moreover, they argue that Aiello and Wheeler did not contend that the change in per diem expenses automatically resulted in the growth of the size of the brain but claimed that it was crucial for *Homo sapiens* to convert to meat-eating to be intelligent. However, there is no precise scientific evidence to explain how human intelligence arose.

If human intelligence has developed through an ongoing but steady development, the question that arises is whether it is possible for animals to have intelligence but at a lower magnitude. This view tends to agree with Darwin's view that animals have intelligence. On the other hand, the general view is that animals do not have intelligence since they are not able to elaborate concepts. However, animals have highly developed sensory knowledge and are capable of elaborating "perceptions." It is these perceptions, in concert with the genetic inheritance they possess that is related to how they adapt to the environment, which permit them to have an ample vision of their environment that empower them to survive. Hence, it is argued that intellectual knowledge, and the amplification of theoretical and general ideas, is unique to humans.

Johns Hopkins University define the brain as a complex organ that controls thought, memory, emotion, touch, motor skills, vision, breathing, temperature, hunger, and every process that regulates the human body, and together with the spinal cord embrace the central nervous system. One may argue that in terms of this definition, an animal's brain has the same function as the human brain. However, the primary distinguishing factors between a human and animal brain is that a human brain has remarkable cognitive capacity, which is the greatest achievement of evolution, whereas an animal's brain has comparatively less cognitive capacity. Additionally, the human brain allows humans to think about certain things and be motivated by their instincts, intellect, and logic, whereas animals are incapable of reasoning and are simply motivated by their instincts. Therefore, from a scientific point of view, the human brain can be simply viewed as an organ of understanding or intelligence.

The question that arises from this view is whether the human soul exists? Without the presence of a human soul, the existence of religion would

become meaningless since there would be no path of providing a lasting and continuing relationship with God. On the other hand, if “understanding” is viewed as a capability of the soul that did not require the subject as an organ, but as an object to particularise concepts, then the human soul becomes spiritual and immortal, and sanctions humans through their sincerity to become a transcendent being that strives to reach beyond the experience of objects to understand being itself. Here, the soul is viewed as the spirit or immaterial part of human beings, the seat of human personality, intellect, will, and emotions, regarded as an entity that survives the body after physical death.

However, the above does not provide the answer to what is the origin of human intelligence? Evolutionary scientists have a dichotomous view, that of Charles Darwin (1809-1882) or of Alfred Russel Wallace (1823–1913). Both are responsible for the theory that species evolve over time due to natural selection. However, Darwin argued that human evolution could be explained by natural selection, with sexual selection as a significant supplementary principle. On the other hand, Wallace had doubts about sexual selection and eventually concluded that natural selection alone was insufficient to account for a set of uniquely human characteristics. Basically, they both came up with the notion of survival of the fittest. However, they could not establish how variation occurred in a species. Wallace and Darwin differed in the theory of evolution in that Wallace believed plants and animals evolved to their environment and Darwin believed adaptations were driven by competition and that those organisms with traits that favoured survival would pass those traits to their offspring.

Darwin argued that the evolution of the human intellect did not differ significantly from the evolution of the body, in that it was a slow ongoing progression based on small steps over a long-term period along the evolutionary path that differentiates apes from humans. However, Wallace rejected the idea that the human intellect was a result of gradual evolution. He contended that it was an outcome of a supernatural cause, where human intelligence may have resulted from unprecedented variation of the

elements within the brain, causing an entirely transformational and profoundly different property called intelligence as an emergent property.

Therefore, the origin of human intelligence can be narrowed down to two options, namely Supernatural as proposed by Wallace or natural as proposed by Darwin (i.e., a Divine creation versus emergence from the subject). In other words, the choice between the notions that human intellect developed all at once with *Homo sapiens*, or that it was the outcome of gradual evolution. It should be noted that scientific knowledge is not the sole objective valid form of knowledge. The epistemological view which hypothesises that views scientific knowledge as the supreme form of objectively valid knowledge can only be valid if it is complemented with the supposition of another thesis regarding the ontological character, which claims that the only existing reality is of subject material. Hence, they both pose two related questions, ontology asks what exists, and epistemology asks how we can know about the existence of such a thing. Additionally, epistemology and ontology are two branches of philosophy, and their exposition of the authenticity of their claims can only be determined by philosophical and not by scientific reasoning, which inevitable instigate ideological debates regarding the origin of human intelligence.

2.2 EVOLUTION OF HUMAN BEINGS

Before the evolution of human beings is discussed, one needs to have an idea of what is meant by evolution. In simple terms, evolution is a biological term that incorporates the process of change in all earthly living species. It refers to a progressive process of change in the living organism and how these living organisms advanced from their descendants. Evolution also embraces several processes such as, mutation, gene flow, genetic drift, and natural selection that enable the various species to initiate, improve, and adjust according to their physical and social environment.

Whether one supports Darwin's or Wallace's hypothesis of evolution, there is consensus that the evolution of human beings is a prolonged process that implicated a change from the very early humans to present

modern humans over millions of years. Evidence suggests that the chronologically human evolution ranges roughly seven million years, with the occurrence of behavioural modernity around 50,000 years ago. Furthermore, several qualities of human behaviour, such as understanding, sadness, ritual, and the use of symbols and tools, were to some extent evident in Hominidae or great apes but in much less advanced forms than what is found in humans.

For instance, chimpanzees can make tools and use them to obtain food and for social displays. They also have somewhat intricate hunting approaches that entail collaboration, authority, and rank. These creatures appear to be status conscious, are scheming, and capable of deception. Chimpanzees can also gain knowledge to use symbols and recognise some simple facets of human language. A feature that is common in species that display aptitude, such as dolphins, great apes, and humans is an enlarged brain size. Complementing the enlarged brain size there is also a more advanced neocortex, which is concerned with sight and hearing in mammals, a folding of the cerebral cortex, and von Economo neurons that are large bipolar neurons located in fronto-insular cortex and anterior limbic area in Hominidae and humans. These neurons are often viewed as being associated with the social trait and capacity to gauge what another is thinking or feeling. However, whether this behaviour can be classified as intelligence as one would associate with human intelligence is open to question.

It is thought that from roughly 5 million years ago, the hominin brain started to progress swiftly in both magnitude and an associated diversity of functions. Scientific studies suggest that the human being evolved from apelike creatures that lived similar to animals and could not even walk upright. However, it is believed that with the passage of time the size of their brain grew, they learnt how to make tools, it allowed them to retain and recall information, and gradually their social behaviour developed, leading to the advent of the modern human being through the process of evolution (refer to [Figure 2.1](#)).

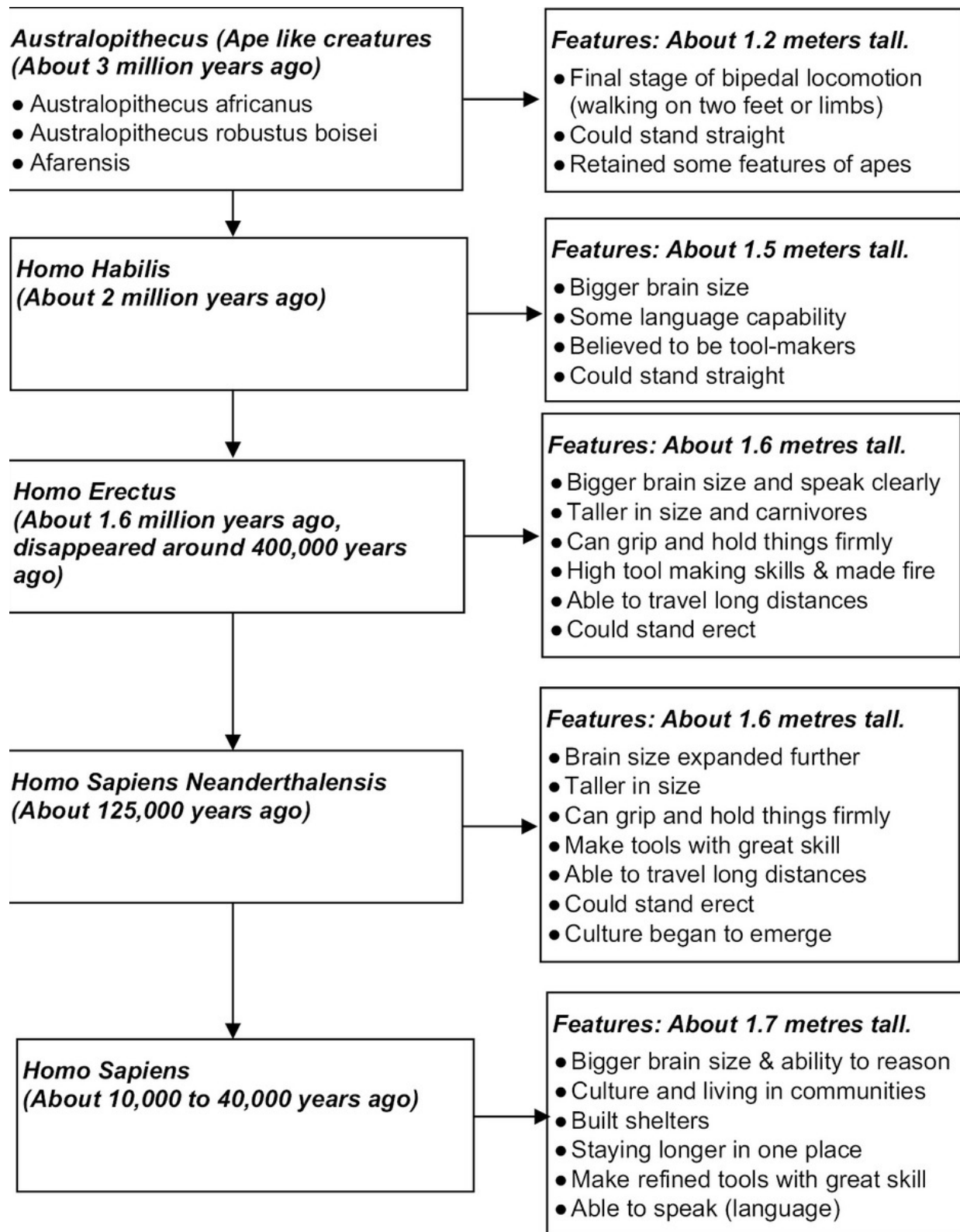


FIGURE 2.1 Evolution of Human Beings. [📄](#)

The era for the commencement of the Neolithic period, which had witnessed the transformation in the life of early humans, is unknown since it was developing at different places in different dates. The Neolithic age (new stone age) was preceded by the Palaeolithic period (old stone age) and Mesolithic age (middle stone age). In the Palaeolithic period early humans lived in caves or simple huts and were hunter-gathers, using basic stone and bone tools, as well as crude stone axes, for hunting birds and wild animals. This was followed by the Mesolithic age where humans used smaller (microliths) chipped stone tools and had a more hunter-gatherer culture with settlements being temporary and seasonal.

This period ended with the introduction of the growing of crops and animal husbandry, the Neolithic age. The outcome of great consequence of the Neolithic age was the innovative activity of agriculture that permitted humans to cultivate food, where they became food producers rather than purely hunter-gathers. Hence, there was a momentous transition for humans from being nomadic, mainly living by hunting and fishing, and harvesting wild food to practicing cultivation where they established settlements without the need to travel from one place to another in the search of food. This also led to the taming of animals, and eventually to the discovery of the wheel. Through this evolutionary process humans learned to use their available resources more effectively and became skilled tool makers, thus gradually improving their living standards.

2.3 HUMAN INTELLIGENCE

Generally, intelligence is viewed as the capacity to learn, comprehend and apply knowledge and skills. However, human intelligence is the intellectual capacity of an individual, which is demonstrated by multifaceted cognitive achievements and fluctuating levels of motivation depending on circumstances being encountered and self-awareness. Typically, the magnitude of an individual's intelligence determines their work effectiveness. It is also important to recognise that human intelligence is a multidimensional concept that incorporates many forms. There is no

consensus for the definition of human intelligence. However, there are some mutual features amongst academics regarding human intelligence that include the following:

- The mental capacity of a person at a particular time and in a particular situation.
- A mental skill for enabling a person to manipulate things, objects, or events existing in one's environment, and adapting to new obstacles and issues being encountered.
- An ability to judge one's capacity in terms of behavioural or performance quality.
- An aptitude for abstraction management.
- A capacity to adjust or adapt to new circumstances.
- An ability to adapt and address unexpected circumstances as successfully as possible.
- The skill to respond appropriately to certain stimuli in a given situation.

The nature of human intelligence is a very complex issue; however, research studies have found several typical characteristics, namely:

- Genetics and environment. Both genetics and environment are key ingredients for an individual's intellectual development and are both equally important.
- Dissemination of intelligence: The level of intelligence fluctuates amid the population and adheres to the mathematical Normal distribution. Many within the population are average, with the others being either exceptionally or poorly gifted individuals.
- Variation of intelligence. People are different and have a diverse level of intelligence. The diverse level of intelligence fluctuates from individual to individual, age to age, and circumstance to circumstance.

- Impact of age. Intelligence develops according to a person's life cycle. However, the life-cycle distribution and one's mental growth varies from person to person. Typically, intelligence peaks between the age of 16 to 20 in an individual, when vertical progression ceases. However, the horizontal amassing of knowledge and attainment of skills continues throughout a person's life cycle.
- Gender differences. Gender has no influence on the level of intelligence.
- Ethnic or cultural differences: Studies have shown that there is no significant difference in the intelligence level of people based on race or cultural groups. The Normal distribution applies, and any variations found are mainly due environmental impact.

Human intelligence is linked with the attributes of creativity, emotion, flexibility, and the capacity to learn from a broad range of experiences. As a comparison, artificial intelligence may mimic human intelligence, but it cannot experience emotions or understand ethical and moral issues. Undeniably, machine and human learning have some common parallels such as reinforcement learning, where the learning process is grounded on rewarding desired behaviours and punishing undesired ones, and can recognise and decode its environment, take actions, and learn through trial and error. In both instances, machine and human learning is based on past experiences. Human intelligence as implied previously incorporates many different intelligence types. These include:

- Linguistic Intelligence. This is the aptitude for using language effectively, irrespective of whether it is written or spoken. A person with high linguistic intelligence displays reliable skills of reading, writing, speaking, and listening.
- Logical-Mathematical Intelligence. This is ability to rationalise, evaluate, and resolve problems using logic and mathematical concepts. A person with high logical-mathematical intelligence can apply deductive and critical reasoning, and problem-solving.

- Spatial intelligence. This is the capacity to visualise and manipulate objects in the mind. Individuals with this type of intelligence can imagine what objects appear like from different views and have good spatial knowledge that enable them to solve difficulties that entail visual logic, such as puzzles, mazes, and maps.
- Musical Intelligence. This is the talent to understand and create music. An individual with musical intelligence has a good grasp of rhythm, melody, and harmony, and is able recognise musical patterns, valuing different types of music.
- Bodily-Kinaesthetic Intelligence. This is skill for resolving problems and expressing ideas using the body. Persons with this type of intelligence are normally good at sports, dance, and other physical activities, and can use tools and manipulating objects.
- Interpersonal Intelligence. This is the ability to understand and relate to others. An individual with high interpersonal intelligence has sound social skills such as communicating, negotiating, and resolving conflicts, and can recognise the emotions and intentions of others.
- Intrapersonal Intelligence. This is the aptitude to understand oneself. This type of intelligence permits persons to comprehend their own emotions, thoughts, and motivations.
- Naturalistic intelligence. This is the faculty of appreciating and linking to the natural world. Persons with high naturalistic intelligence are familiar with the environment, animals, and plants, and can classify and identify natural objects.
- Existential Intelligence. This is the aptitude to examine and contemplate profound philosophical and existential questions, such as the nature of life, death, and the human condition, including spirituality, religion, and the purpose of life. Persons with existentially intelligence are normally profoundly engrossed in philosophy, theology, and the arts.

Understanding the different types of human intelligence assists institutions to acknowledge the strengths and competences of individuals

with a variety of backgrounds to help in having an all-inclusive civil society. There are basically two approaches to examining the intelligence theories, namely the psychometric approach and information processing approach. The former approach is based on the concept that depicts intelligence as a combination of abilities measured by mental tests. The latter approach describes how individuals encode information into memory, namely by recording, storing, and retrieving information in their brains, which affects the person's general behaviour and motivation. The information processing approach explains people's intellectual reasoning and problem-solving process and is based on the belief that people process the information they are presented with rather than just reacting to inducements. There are several key theories that describe intelligence by these two approaches as itemised below:

- **Uni-factor Theory.** French psychologist Alfred Binet proposed that intelligence consists of one factor that is usual for all individual activities. He is recognised for developing the first intelligence test and the concept of mental age.
- **Two-factor Theory.** Charles Spearman, an English psychologist, posited that intelligence consists of two components, namely general intelligence ("g") and specific ability ("s"). General intelligence is viewed as mental energy measured by different tasks, and specific abilities permit a person to tackle a specific kind of problem.
- **Theory of Primary Mental Abilities.** Louis Leon Thurstone, an American psychologist, developed the Primary Mental Abilities Test and proposed that a person's performance on intelligence tests depends on many different abilities, namely verbal knowledge, word fluency, number facility, spatial visualisation, associative memory, perceptual speed, and reasoning.
- **Theory of Cognitive Development.** Jean Piaget, a Swiss psychologist, was the pioneer for the information processing approach. His theory of cognitive development focused on the education of children and was related to the mechanisms by which intellectual development

advances, and the stages through which children develop. Piaget posited that children explore and observe consistencies and make generalities from these observations. He argued that intellectual development grows from two cognitive processes that work to some extent in a shared fashion, namely assimilation and accommodation. Assimilation includes new information into an already existing cognitive structure, and accommodation, which shapes a new cognitive structure into which new information can be integrated. According to Piaget, cognitive development denotes a dynamic equilibrium between assimilation and accommodation processes.

- **Hierarchical Model of Intelligence.** Arthur Robert Jensen an American psychologist found two different types of intellectual ability that he identified as Level 1 and Level 2. Level 1 were considered the skills related to memory and associative functions such as retention of input and rote memorisation of basic skills and knowledge. On the other hand, Level 2 skills were related to abstract and conceptual thinking, including the capacity to change and manipulate inputs, or the capacity to resolve issues.
- **Structure of Intellect Model.** Joy Paul Guilford, an American psychologist, proposed that intelligence is a combination of numerous abilities independent in nature. He categorised the components of intellect into three broad categories, namely operations, contents, and products. Applying factor analysis, he further divided these and constructed a three-dimensional grid that combined a factor from each of the three categories to make up a whole.
- **Theory of Multiple Intelligence.** Howard Gardner, an American developmental psychologist, posited that the learning process should be tailored to an individual's strengths in their multiple intelligences. By recognising and developing these intelligences, individuals can enhance their ability to learn and understand information. He has proposed eight multiple intelligences, namely spatial, bodily-kinaesthetic, musical, linguistic, logical-mathematical, interpersonal, intrapersonal, and naturalistic.

- **Triarchic Theory of Intelligence.** Robert Sternberg an American psychology professor at Cornell University proposed three distinct types of intelligence that a person can possess, namely practical intelligence, creative intelligence, and analytical intelligence. This theory is based on information processing.
- **PASS Model of Intelligence.** This theory of intelligence is rooted in Alexander Romanovich Luria, a Soviet neuropsychologist. He studied work on cognition and proposed that human cognition consists of three separate but interrelated brain systems that support four cognitive processes, namely planning, attention, simultaneous, and successive processing. It was later further developed in 1975 by Das, Kirby, and Jarman and later refined by Das, Naglieri, and Kirby in 1994. Its full form is the Planning, Attention-Arousal, Simultaneous, and Successive (PASS) theory of intelligence.

The above illustrate that there are various definitions of intelligence. Intelligence is mostly recognised as a complex cognitive capability that permits humans to learn and solve problems. These various definitions are supported by numerous theories of intelligence. However, there appears to be a consensus that intelligence is a mixture of cognitive aptitudes that include problem-solving, learning, reasoning, and memory. There is also consensus amongst researchers that human intelligence has evolved progressively over a process of thousands of generations without an extreme change in the environment. [Skottke \(2005\)](#) argues that the challenges humans encounter in modern society have forced the independent domain of intelligence to assume the roles, which other domains would have played in the primitive environment and in which humans have evolved. He argues that there are clear genetic associations that influence people to have a larger cerebral cortex, permitting them to address the challenges that an expanded workload modern life places on the human's intelligence domain.

Moreover, he contends that evolution has not yet had an opportunity to catch up to the rapid progress humans have made as a society, and this

might not be due to humans bypassing the natural selection process. General intelligence is sometimes defined as the ability of a person to attain and apply knowledge. According to [DiLalla \(2000\)](#), various studies have found that general intelligence is at least fifty percent heritable and therefore, can be passed down from generation to generation. It is also held that evolution is a concept where the gene representation of organisms transforms over time. Additionally, it is believed that what drives this change in gene expression is either natural or sexual selection. Furthermore, it is evident that human intelligence has been continually transforming since the advent of the very first hominids. Skottke maintains that environmental and migration factors have influenced human intelligence, apart from the genetic changes. He argues that the migration of people to all the corners of the earth, along with the industrialisation of modern society, has disconnected modern human from its ancestors.

2.3.1 General Intelligence as an Independent Domain

According to Skottke, the theories concerning general intelligence are the dependent domain and the independent domain models. He maintains that the dependent domain model considers general intelligence as a broad category that is segmented into smaller specific intelligences, such as mate selection, cheater detection, and face recognition. Hence, with this model, it can be implied that if an individual's general intelligence is high, then their ability to choose and attain a mate, detect deceptions, and recognise people's faces would also be high, with the reverse also applying. On the other hand, the independent domain theory views general intelligence as just one of several psychological processes, together with mate selection, cheater detection, and face recognition. Under this model, general intelligence is described as the capacity to apply deductive logic and abstract thought.

Skottke argues that the independent domain model is more realistic because it is supported by various empirical studies indicating that the psychological processes of the brain are not connected. He cites a study ([Kanazawa, 2004](#)) where very intelligent individuals with IQs above 125

are the least likely to marry of all the cognitive classes, suggesting that mate selection and general intelligence are unrelated. Hence, he claims that all the psychological processes are dedicated to its own area of capability that have been forged over thousands of generations of human evolution. Therefore, he suggests that someone might have increased intelligence but not be able to secure a mate similar to how someone who has the ability to find a mate would not increase their capacity to use logic and abstract thought. Other studies related to genetic influences also support the independent domain model. In his dissertation, Skottke's concludes that human intelligence may have ceased its growth.

Alexandria Cherry from Rochester Institute of Technology in reviewing Skottke's arguments notes that the need to work for survival has greatly lessened but this does not mean that the evolution of human intelligence will cease. He argues that human expansion into places of unknown terrain is possible because humanity has enough intelligence to move into outer space and explore other worlds. He claims that maybe this is not the end of the growth of intelligence but just the beginning. He maintains that several studies have shown that intelligence is still rising and poses the question that if intelligence is still on the rise, then what evidence is there that it will slow and eventually stop? Therefore, he asserts that so far, there has been no evidence that human intelligence has ceased its growth.

2.4 EVOLUTION OF HUMAN INTELLIGENCE

Scientists and academics have been interested in the emergence of the human intellect for many decades, with French psychologist Alfred Binet (1857-1911) being a pioneer when he proposed the Uni-factor Theory. While various theories have been proposed that provided significant ideas and insights into how the human cognitive skills have changed through time, the specific origins that have caused the evolution of human reasoning are still being debated. The numerous theories that have been proposed over these past many decades are testimony of the complexities of the extraordinary human intellect. The evolution of the remarkable human

intellect is the development of progressively multifaceted mental capabilities and knowledge of our world over a long journey through time. The first humans were predisposed to depend on their physical strength and survival compulsions, but as the human brain evolved, so did their capacity to think, reason, and solve difficult problems.

Currently, our human intellectual development and the associated transformation of extensive knowledge to wisdom is demonstrated by numerous means through the creation of complex high technology machines, such as Magnetic Resonance Imaging (MRI) that create detailed images of nearly every structure and organ inside the body to detect abnormalities; the ability to send humans into outer space; and the ability to think critically, and to cooperate with others for resolving numerous diverse problems. Theories vary, but a popular theory amongst researchers is that the human intellect developed through natural selection where organisms that are more able to adapt to their environment are more likely to survive and pass on the genes that supported their achievement. This is the evolutionary process that causes species to change and diverge progressively over time. Researchers have found evidence to suggest that in the human ancestry, brain size, and complexity have grown through the passage of time, specifically in the prefrontal cortex. The prefrontal cortex has been found to be linked with higher-order cognitive processes such as decision-making, problem-solving, and planning.

However, there are some researchers that do not entirely agree with a purely natural selection theory and believe that social and cultural factors have significantly impacted the development of the human intellect. For example, language and communication enabled people to share and disseminate information and to learn from each other, thus possibly influencing the growth of higher cognitive abilities. Nevertheless, research has suggested that even cognitive abilities may be influenced by genetic factors, and these have contributed to the progression of human creativity.

2.4.1 Factors Contributing to the Evolution of Human Intelligence

There is consensus amongst researchers that the human intellect evolutionary process is a complicated phenomenon shaped by various influences over millions of years. Some of the critical factors that are viewed to have contributed to the evolution of human intellect include the following:

- **Environment.** Ecological factors such as availability of food, social interaction, and education are factors that positively affect the development of logical thinking. For instance, contact with contaminants during early development may cause cognitive deficits, whereas education has been shown to strengthen cognitive abilities.
- **Usage of Tools.** The ability to make and use tools is the most fundamental quality that distinguishes humans from other animals. It is contended that enhanced cognitive skills, such as planning tasks, solving difficulties, and spatial judgement, were all essential for the evolution of tool usage.
- **Culture.** Culture, which embodies social learning and behaviour; values found in human societies; and the knowledge, beliefs, arts, laws, customs, capabilities, and habits within groups of individuals are all essential in the development of the human intellect. For instance, language development was essential for people to communicate knowledge and information, which facilitated the development of superior cognitive abilities.
- **Social Interaction.** Humans' cognitive abilities have been influenced by our extraordinary sociable nature in which social contact becomes an essential component in forming our cognitive aptitudes. Studies have shown that people learn from each other, share knowledge, and create complex social systems through their social contact. This has helped the evolution of our cognitive powers.
- **Genetic Factors.** The cognitive competencies of humans have also been significantly influenced by the evolution of genetic factors. Studies have revealed that adult individuals have a heritability of IQ

between 57% and 73%, with some recent studies showing heritability for IQ as high as 80%.

Intelligence in human groups adheres to a normal distribution, which has a bell-shaped mathematical function. The peak (middle) of the Normal distribution is the average, with the central measure of dispersion being the standard deviation. The range of the dispersion from the mean (average) for the Normal distribution is ± 3 standard deviation. Hence, most individuals in the normal distribution are found in the middle range of reasoning, with fewer people at either extreme of high or low intelligence. The intelligence quotient (IQ) is commonly used to measure intelligence with a mean score of 100 and a standard deviation of 15. Therefore, human intelligence ranges from a low of 55 to a high of 145 (i.e., 100 ± 3 standard deviation, which is 100 ± 45). In other words, applying this standardised scale, roughly 68% of the population falls within one standard deviation of the mean (between an IQ value of 85 and 115), whereas as around 95% falls within two standard deviations of the mean. (between an IQ score of 70 and 130).

There is consensus amongst researchers that reasoning is a very complex attribute, which is affected by several genetic and environmental factors. The normal distribution implies that the distribution of reasoning in one community is likely. However, there may also be significant differences in the distribution of intelligence between communities due to various variables such as genetics, nutrition, education, and culture. Intelligence is a multifaceted characteristic which is impacted by genetic and environmental factors but should be viewed as only one component of human worth and potential.

2.5 THEORIES ON THE EVOLUTION OF HUMAN INTELLIGENCE

The evolution of human intelligence is closely linked to the growth of the human brain and to the origin of language and has been a topic of controversy with the proposition of various theories. There are various

leading theories that have been proposed, which include social brain; cultural intelligence; transformative cultural intelligence; self-domestication; social exchange; sexual selection; ecological dominance-social competition; and group selection.

2.5.1 Social Brain Theory

The social brain theory was proposed by Robin Dunbar, a British biological anthropologist, and specialist in primate behaviour, who argued that human intelligence did not evolve chiefly to resolve ecological difficulties, but significantly as a process of surviving and reproducing in large and complex social groups. Dunbar concluded that the size of the neocortex, which is the part of the brain linked with cognition and language, relative to the body, is related to the size of a cohesive social group. He maintains that this ratio (neocortex size to the body) limits how much complexity a social system can handle.

Although having a social group does not itself cause intelligence to evolve, he maintains that several behaviours linked with living in large groups, such as reciprocal altruism, deception and coalition formation facilitate the ability to understand the thoughts and emotions of others (the dynamics of the Theory of Mind). He further argues that when the size of a social group increases, the number of different relationships in the group may increase exponentially. Dunbar introduced the 150-rule where humans in general have a social circle of about 150 people, which is also the usual range of social communities in small societies and personal social networks. The success of a social group is viewed as being dependent on its size when initially established, with groups of around 150 being mostly successful, since they are likely to achieve a balance between the minimum size of effective functionality and the maximum size for generating a social environment of commitment to the community. Therefore, Dunbar concludes that when hominids commenced living in sizeable groups, selection facilitated higher intelligence. This is supported by considering the relationship between neocortex size and group size of various mammals.

However, this theory has attracted lot of criticism. Phylogenetic studies (i.e., evolutionary relationships among biological entities) regarding brain sizes in primates suggest that while diet is a predictor of brain size, from a sociality viewpoint (i.e., tendency of groups and persons to develop social links and live in communities) it does not predict brain size when adjustments are made for instances where the diet impacts both brain size and sociality. Researchers have found that animals thriving mostly on raw fruits or succulent fruit-like produce of plants such as roots, shoots, nuts, and seeds (frugivores) tend to exhibit larger brain size than animals that focusses on eating leaves (folivores). A likely explanation for this phenomenon is that frugivory entails extractive foraging, which necessitates higher cognitive processing, since one needs to locate and prepare hard-shelled food, such as nuts and fruit before consuming it. However, some researchers argue that it is not just the act of extractive foraging that stimulates brain size, but the fact that a high energy frugivore diet aids foetal brain growth and requires spatial mapping to locate the embedded foods.

Additionally, the correlation between brain size and sociality becomes a problem for the Dunbar's social brain theory since insects exhibit an incomparable assortment of social forms to permanent colonies containing many individuals collaborating as a collective organism, with a remarkable span of cognitive skills despite their small nervous systems. Nevertheless, human societies typically remain united by the ability of individuals to comprehend qualities signifying group membership. It is argued that social insects do the same and often recognize members of their colony allowing them to defend against competitors. However, studies suggest that with insects this recognition is attained because of elementary cognitive operations that do not involve long-term memory but because of sensory orientation. This supports the general view that animals do not have intelligence since they are not able to elaborate concepts but have highly developed sensory knowledge that are capable of elaborating "perceptions."

2.5.2 Cultural Intelligence (CQ) Theory

[Earley and Ang \(2003\)](#) introduced the concept of cultural intelligence, which basically refers to the skill to relate and work effectively in culturally diverse situations. This concept was formulated when the world was experiencing unprecedented globalisation and interconnectedness, and at a time when ideological clashes and cultural conflict culminated in the tragic events of 11 September 2001. Nobel Prize laureate Elie Wiesel had identified hatred directed toward culturally different individuals (i.e., “cultural hatred”), as the major source of problems between people, across all times. This theory is akin but distinct from the social brain hypothesis. Cultural intelligence or cultural brain hypothesis claims that human brain size, cognitive ability, and intelligence have increased over generations due to cultural information from a mechanism known as social learning. Furthermore, it postulates a positive relationship between species with a higher dependency and more frequent opportunities for social learning and overall cognitive ability. Earley and Ang argue that social learning permits species to develop cultural skills and strategies for survival, thus overwhelmingly, cultural species should in theory be more intelligent.

This theory acknowledges that humans are the most intelligent species on the planet, with big brains that evolved over millions of years, with ample cognitive abilities and processing power which outcompete all other species. They view humans as an “evolved cultural species” that has an unparalleled dependence on culturally diffused knowledge because of the surrounding social environment. They maintain that social transmission of information disseminates significantly faster in human populations relative to changes in genetics. The rationale supporting the cultural intelligence theory is that human beings are the most cultural species in the world and this cultural information has been steadily disseminated through generations to create extensive capacities of cultural skills and knowledge throughout humans, and therefore human beings are the most intelligent species there is. In contrast, Dunbar’s social brain hypothesis maintains that the human brain developed because of complex social interactions in groups.

Therefore, the cultural intelligence theory puts more emphasis on the growth in intelligence from socially transmitted information.

Therefore, there is a distinct shift in emphasis from social collaboration to learning strategies. This conflicts with the human general intelligence notion since it focuses on the model that cultural skills and information are learned from others. [Muthukrishna et al. \(2018\)](#) argue that brains expand in response to the availability of information and calories, and thus store and manage more information. They further contend that information availability is affected by learning strategies. These researchers designed a simulation model that captures this co-evolution under different conditions and describes the specific and narrow conditions that can lead to a take-off in brain size, which is a possible pathway that led to the extraordinary expansion in the human species. Their model was based on three primary assumptions, namely that brain size, complexity, and organisation are grouped into one variable; a larger brain results in larger capacity for adaptive knowledge; and more adaptive knowledge increases fitness of organisms. Their model predicted relationships between brain size, group size, adaptive knowledge, asocial/social learning, mating structure, and the juvenile period. These findings had been confirmed by empirical literature by other researchers.

2.5.3 Transformative Cultural Intelligence Theory

[Moll \(2018\)](#) proposed a variation to the cultural intelligence hypothesis. The transformative cultural intelligence hypothesis maintains that the human physical cognition is developed and influenced by the social environment around us. This proposition is contrary to the cultural intelligence hypothesis suggested by [Earley and Ang \(2003\)](#), which maintains that it is the human social cognition and not physical cognition which is superior to our nearest primate relatives. However, Moll's hypothesis views physical cognition in humans as being affected by external social factors, which is unique to humans and not observed in other species.

Molli conducted a study by examining problem-solving under different social conditions of 4-year-old children. Children had to use water to extract a buoyant object from a narrow tube. In the first experiment, the children were confronted with the problem “cold” without clues resulting in that nearly all of them were unsuccessful. In a second experiment, the children were shown a solution-suggesting video, which was pedagogically delivered prior to the task, resulting in that most children (69% in Experiment 1, 75% in Experiment 2) succeeded. Furthermore, showing children the same video in a non-pedagogical manner did not boost their performance above baseline (Experiment 1) and was less effective than outlining it pedagogically (Experiments 1 and 2). The findings suggest that the children’s physical cognition and problem-solving ability was affected by how the task was socially presented to them. The Cultural Intelligence hypothesis was also questioned as to whether only humans’ social, but not their physical, cognition differs qualitatively from that of great apes.

2.5.4 Self-domestication Theory (or Reduced Aggression Theory)

The human self-domestication hypothesis maintains that compared to existing primates and extinct hominins, humans exhibit many of the distinctive morphological, behavioural, and cognitive features also observed in domesticated animals. Hence, the self-domestication hypothesis suggests that, like mammalian domesticates, humans have gone through a process of selection against aggression, a process that in the case of humans was self-induced, which has led to the evolution of advanced intelligence in *Homo sapiens*.

[Shilton et al. \(2020\)](#) argue that the self-induced selection against aggression differentiates humans from other species of monkeys and primates, where aggressivity is still manifested, and has eventually led to the development of quintessential human characteristics such as empathy, social cognition, and culture. They maintain that there are unique aspects of human evolution and suggest that emotional control and social motivation in humans evolved during two major but partially overlapping stages. The

first stage came after the emergence of mimetic communication, and saw the beginnings of musical engagement, and mimesis-related cognition, which required socially mediated emotional plasticity and was accompanied by new social emotions. The second stage succeeded the emergence of language, when individuals began to instruct the imagination of their interlocutors, and to rely even more extensively on emotional plasticity, and culturally-learned emotional control.

[Shilton et al. \(2020\)](#) argue that although selection for reduced emotional reactivity was involved in both socially impoverished domesticates and socially sophisticated humans, the differences between the two types of pro-social evolution mean that their inclusion under the same umbrella term of “domestication” is misleading. The above contentions reveal the significant differences between humans and domesticates, thus challenging the notion of human self-domestication.

2.5.5 Social Exchange Theory

[Homans \(1961\)](#) initially defined a framework built on a combination of behaviourism and basic economics. Social exchange theory is a concept based on the idea that social behaviour is the result of an exchange process in which people weigh the potential benefits and risks of their social relationships. Hence, when the risks outweigh the rewards, they will terminate or abandon the relationship. Thus, the purpose of this exchange is to maximize benefits and minimise costs. Several studies suggest that social exchange between individuals is a critical modification to the human brain, implying that the human brain could be equipped with a neurocognitive system dedicated for reasoning around social exchange. Hence, this would develop when two parties improve their position by mutually exchanging things, they value less for things they value more. However, selection will only support social exchange when both parties benefit. Therefore, if a party fails to deliver fair benefits (cheaters) this would impend the fulfilment of exchange.

According to [Wischniewski et al. \(2009\)](#), abundant research using various game theoretical approaches has examined apparently irrational human behaviours. They explored the cognitive mechanisms required to compute possible scenarios of cooperation, defection, and the detection of cheating. Evolutionary game theory has revealed that variations for social exchange can be favoured and firmly maintained by natural selection, but only if they include design qualities that enable them to detect cheaters and cause them to channel future exchanges to reciprocators and away from cheaters. They argue that humans use social contracts to present the benefits and losses each party will be receiving. Therefore, humans have developed an advanced cheater detection method, which has an inbuilt proprietary problem-solving scheme that progressed to match the recurring features of their matching problem domains.

Humans do not only need to establish that the agreement is infringed, but also if this infringement was intentionally made. Therefore, humans have developed practices that are focused on revealing intentional cheating through contract infringements. However, a difficulty with the theory that a specific punishment for intentional deception and intelligence to evolve together, is the fact that selective punishment of individuals with certain attributes selects against the attributes in question. For instance, if only individuals that have the skill to remember what they agree to are penalised for breaching agreements, then evolution would have selected against the capacity to remember what one had agreed to. The social exchange theory makes several assumptions about human nature, namely that humans are rational beings; and that they seek rewards and avoid punishments. However, the criteria that humans use to assess costs and benefits differ over time and from person to person.

2.5.6 Sexual Selection

[Miller \(1998\)](#) argues that the human mind evolved not as survival machine, but as courtship mechanism. He notes that evolution is driven not just by natural selection for survival, but by an equally important process that

Darwin called sexual selection through mate choice. Miller suggests that human intelligence for the essentials of hunter-gatherers to survive is unnecessarily advanced. Miller maintains that language, music, and art, which are viewed as indicators for intelligence, evolved as fitness indicators and not because of their practical value for the survival of ancient hominids.

This model maintains that Hominids' greater intelligence was impacted by healthy genes, and that the relative speedy evolution of human intelligence was due to a Fisherian runaway positive feedback loop of sexual selection. [Fisher \(1930\)](#) proposed that the exaggerated secondary sexual characteristics of animals could evolve by means of a runaway evolutionary process in which an initial small adaptive dimorphism (the existence of two different forms) was then developed further into a much more heightened trait by an ongoing interactive combination of female selective choice and increased male advantage. For example, in many species, only males have imposing secondary sexual properties such as ornaments and an exhibitionist conduct. However, sexual selection may also be a female characteristic, particularly in somewhat monogamous species. For instance, some researchers believe that sexual selection may explain certain female-specific human qualities, like breasts and buttocks, which are significantly bigger in proportion to total body size, than those found in related species of ape. Particularly when one considers that the human female breasts are found sexually attractive by many men.

It is argued that if breasts of the size found in humans were essential for the purpose of breastfeeding, they would be found in similar size proportions in other species. This rational supports the proposition that sexual selection acts on human females as secondary sexual characteristics. It is also argued that the enlargement of the human brain necessitated more nutrition than the brain of a related species of ape. Additionally, for females to effectively evaluate male intelligence, they needed to be intelligent themselves, explaining that although on average there appears to be no difference in the intelligence level between males and females, there may be a disparity in how they display their intelligence in showy appearances.

2.5.7 Ecological Dominance-Social Competition Theory

According to the Ecological Dominance-Social Competition model, human intelligence developed to a substantial level through a mixture of increasing dominance over the environment and the growing importance of social interactions. Consequently, the foremost selective pressure for increasing human intelligence changed from learning to mastering the natural world to competition for dominance among members or groups of its own species.

[Alexander \(1989\)](#) views the evolution of human cognition as implicating the brain as a “social tool” suggesting that many human psychological changes serve primarily to contend with social relationships, with ecological constraints, such as hunting or extractive foraging, being a secondary cause of recent evolutionary change. According to [Geary and Huffman \(2002\)](#) some human cognitive aptitudes, such as theory of mind (TOM) and language, are most easily understood in terms of social selection pressures, even if cognitive skills for interacting with the physical and biological worlds are evident as well. For example, language learning helps improve people’s thinking skills and memory abilities. Research has revealed that bilingual students tend to focus better, disregarding disruptions more effectively than those who are monolingual.

Alexander ([1971](#)) and others argue that human social relationships are flexible and complex. They maintain that social success is challenging because it is very difficult to contemplate future moves of a social competitor–cooperator, and to execute suitable countermoves, with further complications owing to networks of several relationships, everchanging coalitions, and deceptiveness. Survival and reproduction within an ever-increasing intricate social structure gradually facilitated advanced social skills, communication of concepts that increasingly resulted in complex language patterns. Thus, competition gradually shifted from controlling the natural elements to influencing other humans. The consequence of this shift was the emergence of leadership or popular acceptance by having more advanced social skills to out manoeuvre other members of the group.

Therefore, an individual with better social and communicative skills would be more easily selected.

However, while humans have changed over time and social relationships became more important, it is also evident that concurrently the human brain evolution included cellular, genetic, and circuitry transformation. For example, on a genetic level, humans have a modified gene that is linked to speech and language development and another gene that enables greater dendritic spine density which promotes greater neural connections. Moreover, on a cellular level, research has suggested that von Economo neurons (VENs) are more dominant in humans than other primates, which are associated with empathy, social awareness, and self-control. Moreover, on a circuitry level, studies have suggested that humans display a more complex mirror neuron system, a better link between the two major brain language processing regions, and a vocal control circuit that links the motor cortex and brain stem. The mirror neuron system has been shown to be related with social cognition, theory of mind, and empathy.

2.5.8 Group Selection Theory

Group selection theory involves the evolution of group-level modifications through a process of natural selection acting on groups as the focal units. This theory is a segment of a wider theory of multilevel selection, which acknowledges that natural selection can act across multiple levels of the biological hierarchy simultaneously. Group selection may also be viewed as a selection process in which behaviours evolve according to the fitness (survival and reproductive success) of groups or, mathematically, as selection in which overall group fitness is higher or lower than the mean of the individual members' fitness values. Group selection theory maintains that organism features that impart benefits to a group (clan, tribe, or larger population) can develop regardless of the individual disadvantages that may be present. Thus, the benefits of intelligence bestowed to the group, including the capacity to communicate between individuals, language, the

capability to teach others, and other cooperative aspects, have evident value in increasing the survival potential of a group.

This theory also supports Darwin's theory of natural selection, namely that group-related adaptations must be attributed to the natural selection of alternative groups of individuals, and that the natural selection of alternative alleles (i.e., each of two or more alternative forms of a gene that arise by mutation and are found at the same place on a chromosome) within populations will be opposed to this development. Group-related adaptations and changes are viewed as a consequence of between-group selection as qualities, which transpire to be beneficial in relation to another group and will become increasingly popular and disseminated within a group, intensifying its overall likelihood of surviving a competing group. This theory is also linked to the social brain hypothesis that was discussed previously, which argues that human cognitive complexity arises as an outcome of a higher level of social complexity necessitated by living in enlarged groups. Additionally, these enlarged groups require a higher degree of social relations and interactions, that ultimately bring about an expanded quantity of intelligence in humans.

However, studies have shown that the size of a species' brain can be better predicted by diet instead of measures of sociality. One needs to note that correlation between variables does not imply a causative measure. Therefore, many other factors associated with nutrition and diet could have been the cause for the appearance of modern human traits such as a larger brain and higher intelligence. Scientifically, it has been shown that diets deficient in iron, protein, iodine, zinc, B vitamins, omega 3 fatty acids, magnesium, and other nutrients either in the mother during pregnancy or in the child during development can result in lower intelligence. However, these may not necessarily influence the evolution of intelligence.

2.6 CONCLUSION

The purpose of this chapter was to illustrate that the evolution of the human being and associated intelligence is a very complex process that has been

occurring for millions of years. Furthermore, it is believed that this evolutionary process is continuing. Moreover, the evolutionary process of human intelligence is still a mystery. It is doubtful whether this mystery will be resolved. Many theories have been proposed but none have been proven with a very high degree of certainty. It appears that the evolution of human intelligence has been influenced by many variables, some of which are known, while others are still to be revealed.

This book is primarily focused on artificial intelligence, which is initially a human creation. However, it is argued that a profound appreciation of artificial intelligence cannot be experienced unless we also truly understand how the human being developed and how its intellect evolved. Hence, the objective of this chapter.

I tend to compare artificial intelligence and human intelligence through analogy, by comparing for instance, synthetic leather with natural leather. They both look and feel the same, but the real proof of the difference is the test of fire. Synthetic leather has its advantages and is less likely to crack, peel, stain, or fade from UV light but it is likely to be highly inflammable. Natural leather is highly resistant to ignition.

The human body and its brain can generate various chemicals and hormones that allow the human being to feel fear, anger, pain, sadness, joy, contentment, happiness, and love. The human brain experiences consciousness, passion, and allows a person to appreciate art, music, poetry, literature, drama, and acting. Artificial intelligence through robots may mimic various human senses, but they do not and can never feel the emotions that humans experience. Whilst reading Boris Pasternak's book, *Doctor Zhivago*, which won the Nobel Prize in Literature in 1958, I was impressed by a particular paragraph (page 252) that describes what deep feelings a human may experience and that no artificial intelligence will feel: "What happiness it is to work from dawn to dusk for your family and yourself, to build a roof over their heads, to till the soil to feed them, to create your own world, like Robinson Crusoe, in imitation of the Creator of the universe, and to bring forth your life, as if you were your own mother, again and again."

Another book which made a lasting impression regarding the experiences of human life is Robert Pirsig book published in 1974, *Zen and the Art of Motorcycle Maintenance: An Inquiry into Values*. This book is a narration of a summer motorcycle trip undertaken by a father and his son Zen and the Art of Motorcycle Maintenance becomes a personal and philosophical pilgrimage into essential questions on how to live. The book probes into the meaning of quality and the account of modern crisis caused by the absence of quality and how to reconcile the conflict between human's value and technological needs, to achieve the eventual goal that enhances people's happiness by awakening one's inner nature, compassion, and wisdom.

Having said the above, the rest of this book will focus on artificial intelligence. It shall examine how AI evolved over the years, the tools and security systems that it utilises, its impact on the knowledge-based economy, the practical implication of AI, the ethical and legal considerations of AI, and the outlook for society in general in an AI dominated environment.

REFERENCES

- Aiello, Leslie C., and Wheeler, Peter. (1995). The expensive-tissue hypothesis: The brain and the digestive system in human and primate evolution. *Current Anthropology*, Vol. 36, No. 2, pp. 199–221. www.jstor.org/stable/2744104.[↗]
- Alexander, R. D. (1971). The search for an evolutionary philosophy of man. *Proceedings of the Royal Society of Victoria*, Vol 84, No. 1, pp. 99–119.[↗]
- Alexander, R. D. (1989). Evolution of the human psyche. In: P. Mellars and C. Stringer (eds.), *The Human Revolution*, pp. 455–513. Edinburgh University Press.[↗]
- Arsuaga Ferreras, J. L., and Martinez, I. (1998). *The Chosen Species: The Long March to Human Evolution*. Blackwell.[↗]

- DiLalla, L. F. (2000). Development of intelligence: Current research. *Journal of Psychology*, Vol. 38, pp. 3–7. [↵](#)
- Earley, P. C., and Ang, S. (2003) *Cultural Intelligence: Individual Interactions across Cultures*. Stanford University Press. [↵](#)
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Clarendon Press. <https://doi.org/10.5962/bhl.title.27468>. [↵](#)
- Geary, D., and Huffman, Kelly. (2002). Brain and cognitive evolution: Forms of modularity and functions of mind. *Psychological Bulletin*, Vol. 128, pp. 667–698. <https://doi.org/10.1037//0033-2909.128.5.667>. [↵](#)
- Homans, G. C. (1961). *Social Behaviour: Its Elementary Forms*. Harcourt, Brace. [↵](#)
- Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. *Psychological Review*, Vol. 111, pp. 512–523. [↵](#)
- Marmelada, Carlos A. (2003). *On the Origin of Human Intelligence*. Aceprensa. <https://en.unav.edu/web/ciencia-razon-y-fe/sobre-el-origen-de-la-inteligencia-humana>. [↵](#)
- Martin, Robert. D. (2000). Brain capacity and human evolution; in the origins of humanity. *Research and Science*, No. 19, first quarter 2000, p. 61. <https://en.unav.edu/web/ciencia-razon-y-fe/sobre-el-origen-de-la-inteligencia-humana>. [↵](#)
- Miller, Geoffrey. (1998). *Sexual Selection and the Mind*. Edge Foundation. www.edge.org/conversation/geoffrey_miller-sexual-selection-and-the-mind#:~:text=Miller%20believes%20that%20our%20minds,%2C%22%20he%20has%20noted. [↵](#)
- Moll, Henrike. (2018). The transformative cultural intelligence hypothesis: Evidence from young children’s problem-solving. *Review of Philosophy and Psychology*. <https://doi.org/10.1007/S13164-017-0342-7Corpus> ID: 39117653. [↵](#)
- Muthukrishna, M., Doebeli, M., Chudek, M., and Henrich, J. (2018). The Cultural Brain Hypothesis: How culture drives brain expansion, sociality, and life history. *PLoS Computing Biology*, Vol. 14, No. 11, p. e1006504. <https://doi.org/10.1371/journal.pcbi.1006504>. [↵](#)

- Shilton, D., Breski, M., Dor, D., and Jablonka, E. (human social evolution: Self-domestication or self-control? *Frontiers in Psychology*, Vol. 11, p. 134. <https://doi.org/10.3389/fpsyg.2020.00134>↵
- Skottke, Kyle R. (2005). *The Evolution of Human Intelligence: Increasing Importance of Domain-Specific Intelligence in the Modern Environment*. Rochester Institute of Technology. www.personalityresearch.org/papers/skottke.html.↵
- Wischniewski, J., Windmann, S., Juckel, G., and Brüne, M. (2009). Rules of social exchange: Game theory, individual differences and psychopathology. *Neuroscience & Biobehavioural Reviews*, Vol. 33, No. 3.↵

Chapter 3

Evolution of Artificial Intelligence (AI)

DOI: [10.1201/9781003653950-4](https://doi.org/10.1201/9781003653950-4)

“Someday, in the distant future, our grand-children’s grand-children will develop a new equivalent of our classrooms. They will spend many hours in front of boxes with fires glowing within. May they have the wisdom to know the difference between light and knowledge.”

Plato (Aristocles, son of Ariston)

Greek philosopher (427–347 BC)

AI is not new concept. In fact, the modern era of AI commenced at an academic conference held at Dartmouth College, Hanover, New Hampshire USA in 1956, when the term artificial intelligence was first used. However, the notion that the human brain can be automated is deeply rooted in civilization. [Roy_\(2023\)](#) argues that to accurately appreciate the history and evolution of artificial intelligence, one must go back to AI’s ancient roots, where AI begins not with computers and algorithms, but with the philosophical reflections of great thinkers. These great thinkers, include Aristotle who introduced syllogistic logic. Moreover, many ancient civilisations created human-like mechanisms that were thought to have reason and emotion. [Tucci_\(2021\)](#), maintains that by the first millennium B.C., philosophers throughout the ancient world were creating new

approaches for formal reasoning. She argues that this effort extended over 2,000 years by various intellectuals that included not only philosophers but also theologians, mathematicians, engineers, economists, psychologists, computational scientists, and neurobiologists.

Roy contends that since the birth of the digital age, AI has experienced a remarkable evolution, starting with the first rudimentary programs of the 1950s to the sophisticated algorithms of today. In its embryonic stage, AI was just a series of simple rules and patterns but has now become a very complex and at times frightening development. Over time, AI has ventured into areas that were considered impractical or unattainable, but its capabilities have continuously grown, and its potential appears to be unlimited. The purpose of this chapter is to trace the history and evolution of AI and how it has developed into an awesome powerful instrument that cannot be ignored.

3.1 INTRODUCTION

A primary objective of AI (or Machine Learning) that is commonly stated, is to reduce the human workload. The objective related to reducing human workload was cited many decades ago when computerisation first commenced. Therefore, AI appears to be following the theme of the initial objective of computerisation. However, as the usage of AI has increased since 1956 when it was first mentioned, it has developed to such a degree that some view it as potentially reigning over human intelligence, and therefore a threat.

As one may appreciate, AI can be viewed as a substantial and complex topic. However, there are many aspects of AI that are straightforward, and AI's implications become crystal clear. AI can be viewed as a systematic approach towards achieving an assigned task using a computer and associated application software. For instance, character recognition software can transcribe a text image into a text document, which can then be manipulated through additions, deletions, or amendments. Expert systems are another example of AI, where software can specifically focus

on a particular subject or aspect, such as presenting a person with a set of specific health related questions and making a diagnosis of person's physical condition. Furthermore, AI can accomplish an assigned task very quickly since it is integrated with specific hardware and application software that executes an algorithm that has a precise task. The speed at which AI functions depend on the efficiency of the algorithm and the power of the computer technology being used.

Whilst accuracy with the swiftest speed is a key feature of AI, it must be recognised that there have been cases where the AI application hallucinates and provides distorted responses, because it tries to provide the most probable reply and not necessarily the correct reply. [Table 3.1](#) illustrates that an AI can be classified by its capability and functionality, with each classification being further subdivided. In terms of capability, only Narrow AI is currently available. General AI that is designed to reason, learn and understand complex concepts and Super AI, which is capable to learn and adapt at a rate that far exceeds human intelligence are currently hypothetical and are not available in a practical environment. On a functionality level, Theory of Mind that allows an AI to interact with the thoughts and emotions of humans (e.g., self-drive cars) are currently at an experimental stage. Whereas self-awareness that permits an AI to have an independent intelligence and will not be under the control of the people that created it, is still at the conceptual stage.

TABLE 3.1 Classification of AI by Capability and Functionality [↗](#)

Capability	Definition
General AI	General AI or Strong AI is designed to perform any intellectual task that a human can do. General AI is theoretically able to reason, learn, and understand complex concepts. However, this is not yet possible.

Capability	Definition
Narrow AI	Narrow AI or Weak AI is designed to do a specific task or a limited range of tasks. It is usually useful in applications like facial, speech, and image recognition; natural language processing; and recommendation systems.
Super AI	Super AI is able of exceeding human intelligence in all areas. However, it is a hypothetical AI that is not yet available. When available, it would be able to solve complex problems that are beyond human abilities and would be able to learn and adapt at a rate that far exceeds human intelligence.
Functionality	Definition
Reactive Machines	This is the simplest AI and does basic operations by reacting to some input with some output, without any learning occurring. This is the 1 st stage to any A.I. system. This type of AI does not store inputs nor performs learning.
Theory of Mind	Theory of Mind AI types are not available and are at an embryonic stage e.g., self-driving cars. This type of A.I is supposed to interact with the thoughts and emotions of humans.
Limited Theory	These types of A.I.'s are able to store historical data and/or predictions for making better predictions. Machine learning architecture for this type of AI is more complex. Every machine learning model needs limited memory to be created, but the model can get deployed as a reactive machine type.
Self-awareness	This is a long-term objective of AI in that it achieves enlightenment and becomes self-aware. A self-aware AI has an independent intelligence and will not be under the control of the people that created it.

The AI types that are currently available have been applied in various industrial sectors to assist public and private enterprises to run successfully and achieve their defined targets. Currently the industry sectors that have adopted AI as a key performance component, include healthcare; banking; education; eCommerce; retail; transportation; entertainment; and fashion. These industries have implemented AI based on state-of-art technology where the capability of technology regulates a substantial portion of their workforce. The outcome of AI and its associated technology significantly decreases energy usage and makes it possible for humans to complete their work in the swiftest way possible. Hence, AI at this level complements human power and aids them to perform their functions in a safe environment. AI becomes potentially dangerous when it acts independently from those that created it. In other words, if an AI application remains under human control, its benefits would be enormous and safe. [Table 3.2](#) provides some typical examples where AI is currently being applied.

TABLE 3.2 Typical Examples Where AI is Currently Being Applied[📄](#)

Type	Example
Automation	This type of AI is mainly linked with applications that mitigate human inputs, such as process mechanisation, IT and business processes, and private usages. The key work at this level is to operate and control multiple systems and machinery in manufacturing industries.
Machine Learning	This is based on data usage applying an algorithm that follows a specific pattern to finish the task correctly. Data mining plays an essential role in Machine Learning and organises data for all types and size of entities. Data interpretation is through three variants: Supervised learning, Unsupervised learning, and Reinforcement learning.

Type	Example
Natural Language Processing or NLP	This assists in separating human and machine languages, such as in spam detection in a mailbox. It can distinguish the disparity as a result of text translation, analysis, and speech recognition.
Robotics	The usage of this technology has widened its application base to nearly all sectors (e.g., in restaurants for delivering food to customers at their tables).

AI has some definite benefits for all types of entities, such as lessening the human workload and performing it at a faster rate. If used for suitable purposes it provides accurate results with zero error tolerance (however it needs to resolve the problem of hallucinating bots); and can perform multiple tasks effectively on a 24/7 basis. As discussed in [Chapter 1](#), AI can provide many benefits as long as it remains under human control. The danger becomes evident if General AI and Super AI using Theory of Mind and Self-awareness functionality become common without adequate control. Furthermore, rogue users of AI, particularly at powerful private or public organisations can cause havoc, unless adequate safeguards are implemented at an international level. AI may imitate humans, but it can never have cognitive functions and natural skills. The previous chapter has illustrated that humans are exceedingly intelligent with natural and cognitive skills driven by ethical values founded by an emotional and social intellect. However, whilst most humans are honest and have a high level of civic duty, there is a very small minority that are not. It is these small minority that are a concern when applying AI.

One needs to remember that machines with sophisticated AI can be devious but with the current technology do not have the ability to think out of the box and create something new. However, humans have the supremacy to think ahead of their ingenuity and invent things to improve society. With current technology, an AI system only executes what the programmer has coded in the computer program. However, humans cannot

be complacent because the danger looms as soon as General AI and Super AI using Theory of Mind and Self-awareness functionality become a reality. Therefore, regulatory measures at both national and international levels need to be defined and implemented now as a preventative strategy. Humans are the greatest intellectual creatures on this earth that are sociable, have feelings, and emotions. Recent history has shown that humans can control technology and generate outcomes for improved growth.

3.2 THE HISTORY OF ARTIFICIAL INTELLIGENCE

As stated previously, the notion that the human brain can be automated is deeply rooted in civilisation, having its roots dating back to ancient times, where AI is initiated not with sophisticated algorithms, but with the great philosophical thinkers. To appreciate the history and evolution of AI it is appropriate to segment the historical timeline into three main categories, namely: the era of enlightened rational thinking; pre-to-World War II technology era; and post-World War II technology era. This broad classification of the developments that have taken place over the long and difficult journey of AI to its current position, features personages in the ancient history, underlining the evolution from early philosophical ideas to practical innovations in computation and automation.

3.2.1 Era of Enlightened Rational Thinking

The era of enlightened rational thinking is veiled by myths and legends dating back to many ancient cultures that built anthropomorphic androids that were believed to possess reason and emotion. [Table 3.3](#) provides a summary of events that have influenced the era of enlightened rational thinking that form the foundation for the conceptual development of AI.

TABLE 3.3 Summary of Events During the Era of Enlightened Rational Thinking [↗](#)

Timeline	Example

Timeline	Example
4th Century – 3rd Century BC	Aristotle initiated syllogistic logic: One of the earliest deductive reasoning methods, setting the basis for logical thinking, a fundamental feature of AI.
8-9th Century	Muhammad ibn Mūsā al-Khwārizmī, Persian mathematician who initiated Arabic numerals and algebra, and created the words algorithm and algebra.
13th Century	Spanish theologian Ramon Llull inventor of an “art of finding truth.”
13th Century	Roger Bacon and Albert the Great built tools that looked like heads, known as “Talking Heads,” conversational interfaces. These devices likely aided in early tests to imitate human speech and interaction.
14th Century	The “movable type” printing press boosted information dissemination. This invention helped to disperse knowledge that is crucial for AI development.
15th Century	Invention of clocks: First modern measuring tool that gauged time exactly. This was vital for scientific and technological evolution, including AI.
16th – 17th Century	René Descartes viewed animals as merely machines, thus promoting the philosophical notion of mechanizing living beings (“I think, therefore I am”).
17th Century	Blaise Pascal invented the first computational digital calculating device.
17th Century	Gottfried Wilhelm Leibniz upgraded Pascal’s calculator with multiplication and division functions.
17th Century	Sir Samuel Morland improved computational technology by through the invention of an arithmetical machine.

Timeline	Example
17th Century	Thomas Hobbes's book "The Leviathan" had mechanical and combinatorial theories of thinking. This laid the basis for studying the automatus aspects of thought processes.
18th Century	Thomas Bayes' Rule is a key law in data science because it explains the act of learning, which is a key concept in AI, particularly for machine learning.
19th Century	Joseph-Marie Jacquard invented the first programmable device using thousands of punch cards laced together. Each row of punched holes corresponded to a row of a textile pattern. This device was a notable advancement in automation and control.
19th Century	Mary Shelley in 1818 published the novel "Frankenstein" that deal with themes of artificial life and the consequences of creating sentient beings.
19th Century	George Boole an English mathematician helped establish modern symbolic logic and whose algebra of logic, now called Boolean algebra, which is basic to the design of digital computer circuits.

The first millennium B.C saw the emergence of the great thinkers in various regions of the world, such as Plato and his student Aristotle. They both contemplated the nature of human cognition and reasoning and contributed to the development of formal reasoning approaches. These philosophers studied the notion that the human thought process could be sub-divided into a series of logical steps, like a mathematical process. These early thinkers were over many centuries joined by other important contributors that included theologians, mathematicians, engineers, economists, psychologists, computational scientists, and neurobiologists. Plato and Aristotle initiated syllogistic logic, which is a form of deductive argument where the conclusion follows from the truth of two (or more) premises. Syllogistic logic is the formal analysis of rational terms and operators, and the structures that make it probable to infer true conclusions

from given premises. This mode of reasoning set the foundation for what is known as symbolic AI. The concept of symbolic AI is based on the notion that human thought and reasoning can be represented using symbols and rules, which can be manipulated to simulate human intelligence. This idea is analogous to teaching a machine how to think like a person by using symbols to represent concepts and rules to manipulate them.

The Persian mathematician Muḥammad ibn Mūsā al-Khwārizmī who introduced the Hindu-Arabic numerals, and algebra to European mathematicians has to this day made an impact on how computations are conducted. Mathematics using Roman numerals would be impossible because the highest number that can be expressed in Roman numerals is 3,999. The number 4,000 would have to be written as MMMM, which goes against the principle of not having four consecutive letters of the same type together. Muḥammad ibn Mūsā al-Khwārizmī wrote a treatise on Hindu-Arabic numerals with the Latin translation title of “*Algoritmi de numero Indorum*” that gave rise to the word algorithm in the English language. Another important contributor to rational thinking was the 13th-century Spanish theologian, Ramon Llull who is best known in the history of ideas as the inventor of an “art of finding truth” (*ars inveniendi veritatis*). His work was mainly aimed to support the Roman Catholic faith in missionary work but was also created to unify all branches of knowledge. The 13th century also witnessed the creation of “Talking Heads” that were devised by several persons. But Albert the Great and Roger Bacon are two of the most mentioned inventors regarding this invention. Saint Albertus Magnus (c. 1200–1282 CE) is claimed to have worked for 30 years on making not just a head, but a man of brass. It was said that he cast together the qualities of his materials under certain constellations. Saint Thomas Aquinas, who was the disciple of Albert the Great and was claimed to have destroyed the mechanical brass head of Albert the Great and mocked the idea that a head or statue could be alive. Roger Bacon also a famous theologian who is claimed to have created a talking brazen head that could answer any question. The narrative appears in the anonymous 16th-century account of “The Famous History of Fryer Bacon” in which Bacon speaks with a

demon. These devices are viewed as conversational interfaces that possibly aided in early tests to imitate human speech and interaction. Whether the accounts of the creation of brazen heads are accurate makes no difference. What is important is the fact that the notion of artificial creation that imitate human speech and interaction were envisaged during this period.

There is no doubt that one of the most important inventions, which had the greatest impact for the dissemination of knowledge and making it a highly accessible commodity, was the movable type printing press. The world's first movable type printing technology for paper books is reported to have been made of porcelain materials and was invented around AD 1040 in China, and a book in 1193 during the Song dynasty documented the use of the copper movable type. The spread of these movable-type systems was mainly limited to East Asia. However, the greatest impact was that of around 1450 when a German goldsmith, Johannes Gutenberg introduced the metal movable-type printing press in Europe, together with advances in casting the type based on a matrix and hand mould. The smaller number of alphabetic characters needed for European languages was an advantage and Gutenberg was the first to create the type pieces from an alloy of lead, tin, and antimony, which remained the standard for about 550 years. The metal type pieces were more durable and produced uniform lettering contributing to high quality and relatively low price of printing. The printing press is viewed as a primary factor that promoted the Renaissance.

The 15th century witnessed a monumental breakthrough regarding the mechanical clock. This was a leap forward, particularly the one built by Henry de Vick at around 1360. However, a significant innovation regarding the mechanical clock was the introduction of the mainspring in the early 15th century, which allowed small clocks to be built for the first time. These clocks measured time accurately, which was of vital importance for technology and scientific experiments. Another important event was the notion that animals were merely machines that was posited by the French philosopher and mathematician René Descartes, who lived during the 16th and 17th century. Descartes believed that all material bodies, including the human body, are machines that operate by mechanical principles. His ideas

promoted the philosophical notion of mechanising living beings. Descartes' notion was not about computational power but about the idea of creating artificial beings with a body autonomous from the soul. Hence, his idea conceptually supports artificial intelligence with the possibility of creating machines that learn from experience, adjust to new inputs, and perform human-like tasks.

The 17th century brought with it a mixture of ideas and inventions. Blaise Pascal was a French philosopher, mathematician, physicist, and inventor. He laid the foundation for the modern theory of probabilities and invented the first digital calculator to assist his father, who was a tax collector. Working for three years, he created a device called the Pascaline, which resembled the mechanical calculator of the 1940s. In recognition of his early computing machine, the computer language Pascal was named after him. Gottfried Wilhelm Leibniz, a German philosopher, mathematician, and logician, invented differential, and integral calculus independently from Sir Isaac Newton and the principles of binary computers based on bits. He improved Pascal's calculator with the multiplication and division functions.

Another influential figure regarding technology during the 17th century was Sir Samuel Morland who was an English academic, inventor, and mathematician (apart from being a diplomat and spy). He improved Pascal's calculator and computational technology through his mechanical calculating device. Morland designed three types of a compact calculating machine: one for addition; one for the addition of money; and one for trigonometry. London craftsmen and instrument maker Humphrey Adamson is most likely to have physically constructed these machines. The device was made from brass, and the calculations were displayed on its face, using small, thin, mechanical wheels that turned on the application. Despite the lightweight, and precise nature of these arithmetic machines, they did not market well, being generally described as pretty but not very useful. A different but imaginative philosophical idea was that of Thomas Hobbes in his book *The Leviathan*. A great deal of *Leviathan* examines how human beings might exist outside of civil society, without government. Hobbes refers to the

imagined condition of human life external to civil condition as “the state of nature.” The Leviathan had mechanical and combinatorial theories of thinking that provides the systematic steps in solving a problem based on cognitive styles. This book may be viewed as setting the basis for studying the automatus aspects of thought processes.

The 18th century also revealed some remarkable inventive individuals. Thomas Bayes, a clergyman, and mathematician set out his theory of probability in 1764. His theorem, in probability theory is a means for revising predictions considering relevant evidence, also known as conditional probability or inverse probability. However, Bayes’ Rule is a key and critical rule in data science because it describes how to update a belief, given some evidence. In other words, it explains the act of learning, which is a key concept in AI, particularly for machine learning. Joseph-Marie Jacquard who was a French weaver and merchant in 1801 invented a simplified method by which complex textiles such as damask were woven using thousands of punch cards laced together, with each row of punched holes corresponded to a row of a textile pattern. His invention is viewed as being the first programmable device, which was a notable advancement in automation and process control. Using the same principle, Herman Hollerith a German American statistician, invented the tabulating machine that was an electromechanical machine designed to aid in summarising information stored on punched cards, which was created to assist the processing of data for the 1890 U.S. Census. This was the precursor to the first generation of computers that used punched cards as an input and output device (for storing and printing data). Another important development during the 19th century was when George Boole an English mathematician helped to create modern symbolic logic, whose algebra of logic is now called Boolean algebra. It should be noted that Boolean algebra is a basic concept for designing of digital computer circuits.

However, the description of this era cannot be finalised without mentioning the creative and imaginative work of Mary Shelly in 1818 in her book “Frankenstein.” This book deals with themes of artificial life and the consequences of creating sentient beings. While AI holds the potential

to revolutionise the lives of human beings either positively or negatively, the dark reverberations of Victor Frankenstein’s oversights provide a bleak reminder for the need of having ethical frameworks and interdisciplinary collaboration, which would ensure that AI remains a dominant influence for good. [Patowary_\(2023\)](#) argues that in the ever-evolving landscape of technology, AI has emerged as a revolutionary force that continues to shape various aspects of people’s lives. He contends that Mary Shelley’s magnum opus, *Frankenstein*, published nearly two centuries ago, remains an enduring cautionary tale about the perils of unchecked ambition and the consequences of playing God. The book *Frankenstein* reveals that Victor Frankenstein’s relentless pursuit of creating life, serves only to be haunted by the unforeseen horrors of his creation, and has resonated across generations ([Patowary, 2023](#)). He argues that this tale of an individual’s excessive pride or self-confidence, raises moral dilemmas, and the intricate relationships between creator and creation continues to transcend time, finding a striking resonance in contemporary discussions on AI and its potential implications.

3.2.2 20th Century: Pre-to-World WarII Technology Era

Whilst the era of enlightened rational thinking established the foundation for logical reasoning and other significant mathematical discoveries, the early to mid-20th century was the birth of the electro-mechanical computer. Although, it was in the 20th century that the concept of AI really began, the initiation of electro-mechanical computers and later the advancement of electronic computers, marked a fundamental turning point. [Table 3.4](#) illustrates the important innovations that affected the development of AI during this period.

TABLE 3.4 Summary of Events During the 20th Century Pre-to-World War II Technology Era 🔗	
Timeline	Example

Timeline	Example
1921	Karel Čapek used the word robot in his play “RUR” (Rossum’s Universal Robots). This play contributed to promoting the concept of artificial beings.
1936	Z1 was a motor-driven mechanical computer designed by German Civil Engineer Konrad Zuse. It was a binary electrically driven calculator with limited programmability, reading instructions from punched celluloid film. Its programmable ability was later improved by the Z2 and Z3 models.
1937	The Atanasoff Berry Computer, later named the ABC, was conceived in 1937 by physics professor Dr. John Vincent Atanasoff and his graduate student, Clifford Berry at Iowa State University. It was designed specifically to solve systems of linear equations and was successfully tested in 1942.
1943	Colossus the code-breaking computer. It was created during the Second World War to decipher critical strategic messages between the most senior German Generals in occupied Europe. Its existence was only revealed in the mid-1970s.
1943	Warren McCulloch and Walter Pitts are often credited with the invention of the artificial neural network and were a vital influence on the invention of the modern digital computer.
1945	ENIAC was the world’s first general-purpose programmable electronic digital computer but was not designed with parallel processing capabilities.

Human beings have always been fascinated by artificial beings. In 1920, Czech writer Karel Čapek wrote the science-fiction play “R.U.R.” which is an abbreviation for Rossumovi Univerzální Roboti. The world premiere of the play was held on 2 January 1921 in Hradec Králové, where the word “robot” was introduced to the English language and generally linked to the science fiction domain. The word “robot” is derived from the Czech

language meaning forced labour. In the play, Čapek visualises a future world where robots become the world's workers that serve and replace humans. He portrays a revolt by the robots that eventually annihilate humans. The play deliberates on what it means to be a human being and the purpose of human life. The play conveys a cautionary message and implicates a scientist named Rossum, who discovers the secret of creating humanlike machines. Rossum's universal robots are mass produced by other Robots on assembly lines. The idealistic Helena Glory, President of The Humanity League, thinks that Robots have (or are developing) souls, and feels that they should be freed. However, she burns Rossum's manuscript that describes how to make robots due to her views about the infertility of human women. Čapek portrays robots as having two characteristics, namely that they had outer appearance of men and that they had the ability to think and were devoid of human weaknesses.

The work of Karel Čapek was science-fiction. However, the Z1, a motor-driven mechanical computer designed by German Civil Engineer Konrad Zuse in 1936, was far from being an imaginary object. The Z1 was a binary electrically driven mechanical device with limited programmability, reading instructions from punched celluloid films. It was later improved as a Z2 model by 1940. The Z2 was an electromechanical digital computer with an improved arithmetic and control logic with 600 electrical relay circuits, weighing over 600 pounds. The Z2 could read 64 words from punch cards and in contrast to the Z1, used 16-bit fixed-point arithmetic instead of 22-bit floating point. The Z2 was demonstrated to members of the German Aviation Research Institute (Deutsche Versuchsanstalt für Luftfahrt known as DVL), today known as DLR (The German Aerospace Centre), which is the national centre for aerospace, energy, and transportation research of Germany. The Z2 was upgraded by Konrad Zuse 1941 to the Z3 model. The Z3 was the world's first working programmable, fully automatic digital computer built with 2,600 relays, implementing a 22-bit word length that operated at a clock frequency of about 5–10 Hz.

In 1998, a replica of the Z3 was tested and found to be in principle a Turing-complete device, which is an important concept in computer

science. A system is Turing complete if it can simulate the activities of a Turing machine, which signifies that the system can perform any computation that is theoretically possible. The Z3 lacked direct conditional branching, but it sustained this definition by speculatively computing all possible outcomes of a calculation. Konrad Zuse has often been viewed as the inventor of the computer, without which AI (and many other concepts) could not become possible.

Another contender for the title as the father of the computer is physics professor Dr. John Vincent Atanasoff from Iowa State University. In 1937, he and his graduate student, Clifford Berry conceived and built the Atanasoff Berry Computer, later named the ABC. The ABC was specifically created to solve linear equations systems and was successfully tested in 1942. It was conceived as a parallel computer with the ability of solving 30 simultaneous operations. The ABC is viewed as the first automatic electronic digital computer. This is contrary to what was commonly understood that the first electronic digital computers were the Colossus, built in England in 1943, and the ENIAC, built in the United States in 1945. However, the ABC's technology is rather obscure because it was not programmable, nor Turing-complete. Many consider the ABC as the first electronic arithmetic logic unit (ALU), which is an essential integrated unit into every modern processor's design. Work on the device, particularly the intermediate result storage mechanism and the paper card writer/reader, was terminated due to other Atanasoff's World War II assignments. The ABC is viewed as pioneering vital principles of modern computing, including binary arithmetic and electronic switching elements. However, its specialised single purpose quality and lack of programmability in terms of a changeable stored program, differentiate it from modern computers. The ABC was designated as an electrical and electronic engineering (IEEE) milestone in 1990 because it contributed to the advancement of the modern computer.

The Colossus is viewed as being the world's first programmable electronic digital computer. However, it was not a general-purpose computer but was specifically designed to break the German Lorenz cipher.

This computer was proposed by General Post Office (GPO) research telephone engineer Tommy Flowers to solve a problem posed by mathematician Max Newman at the Government Code and Cypher School (GC&CS) at Bletchley Park (headquarters for Allied codebreaking). A working prototype Colossus Mark 1 was demonstrated in December 1943 and was commissioned at Bletchley Park by early 1944. The Colossus Mark 2 model was commissioned on 1 June 1944 just in time for the Normandy landings on D-Day and applied shift registers to quintuple the processing speed. This enhancement meant that its operating speed of 5,000 characters per second, which analysed over 100 messages a week was boosted to an astonishing processing speed 25,000 characters per second. It used around 2,500 thermionic valves and was more than two metres tall, to perform Boolean and counting operations. The Colossi computers permitted the Allies to process an enormous volume of high-level military intelligence by intercepting radiotelegraphy messages between the German High Command and their army commands throughout occupied Europe. Alan Turing's application of probability in cryptanalysis is likely to have influenced its design but he did not design Colossus. It is often viewed by many as being the birth of modern-day computing and its presence was kept secret until the mid-1970s. The Colossus was a vital intelligence gathering machine, which is the basis for AI systems.

An important milestone that influenced the design of the modern digital computer was the work of American neurophysiologist and cybernetician Warren McCulloch and logician and cognitive psychologist Walter Pitts. In 1943, they published a paper "A Logical Calculus of the ideas Imminent in Nervous Activity." This was the first mathematical model of a neural network. Their paper described how neurons might work by modelling a simple neural network using electrical circuits. They extended the concepts of Alan Turing's thoughts in his paper "On Computable Numbers." Furthermore, they specified an approach describing how the brain functions in abstract terms and illustrated how simple elements connected in a neural network can have immense computational power. The McCulloch–Pitt neuron illustrated the binary activation of "1" ON or "0" OFF, which meant

that it either discharges with an activation “1” or does not discharge with an activation of “0”. If the activation value is greater than “0”, then the connected path is said to be in an excitatory state else it is known as being in an inhibitory state. The concern with the McCulloch-Pitts neuron concept was its simplicity, since it only permitted for binary inputs and outputs. Their paper did not attract appropriate attention until their ideas were applied by John von Neumann (mathematical formulation of quantum mechanics and Game theory) and Norbert Wiener (the initiator of cybernetics, and the science of communication as it relates to living things and machines), amongst others. The work of Von Neumann and Wiener are fundamental to some key concepts of AI.

However, the world’s first general-purpose programmable electronic digital computer was the ENIAC, which was completed in 1945 but a little too late for World War II usage. The ENIAC was not designed with parallel processing capabilities but is considered important historically, because it established the foundations for the modern electronic computing industry. At its time, the ENIAC showed that high-speed digital computing was possible using the available vacuum tube technology of the 1940s. One must note that programming the ENIAC computer was a complex, manual process that could take days. It was Turing-complete and capable of solving a broad range of numerical problems through reprogramming. Although there were other computers that had various combinations of computational features, the ENIAC had them all in one device.

The ENIAC machine was programmed by rewiring the connections between functional units in the machine, which was a long and tedious process. This was its weakness because it took days to rewire it for every new problem. However, the ENIAC was the most powerful calculating device built at that period. ENIAC had no central memory, it had a series of 20 accumulators that operated as computational and storage devices, with each accumulator storing one signed 10-digit decimal number. By the time the ENIAC was completed, the war was over. It cost almost eight times its original budget of \$61,000 and was declared no longer fit for purpose. In November 1946, the ENIAC was shut down but was refurbished with a

memory upgrade in July 1947. The upgraded model was extensively used during the Cold War on projects that required calculations for the design of a hydrogen bomb. The ENIAC was replaced by EDVAC (Electronic Discrete Variable Automatic Computer), which was built by the Moore School of Electrical Engineering, Pennsylvania. The EDVAC was commissioned at the Ballistic Research Laboratory in 1949, which became a part of the US Army Research Laboratory in 1952. The ENIAC and its successors provided the processing power and paved the way to truly programmable computers making use of computer languages. Thus, making it possible for the development of systems that AI is much dependent on.

3.2.3 20th Century: Post-World War II Technology Era

The post-World War II technology era witnessed an exponential growth in computer technology development and associated operating and application software. According to [Wagner \(2019\)](#), the ever-evolving goal of creating artificial intelligence has fascinated computer scientists for decades. She argues that whether it's a machine capable of human thought or simply an algorithm capable of solving the game of chess, theories, possible solutions and predictions continue to attract widespread attention. She contends that while it is no longer questionable that computers have much higher capabilities when it comes to processing logic, the question as to whether computers will one day be able to feel like humans remains unanswered.

[Table 3.5](#) shows some of the key advances that have influenced the development of AI during this period. In 1950, Alan Turing published his paper, "Computing Machinery and Intelligence" that presented the Turing test for measuring machine intelligence. The Turing Test focused on the computer's capacity to deceive interviewers into believing that its responses to their questions were made by a human being. The Turing Test was a technique used to verify whether a computer can display human intelligence. Thus, if a computer could participate in a conversation with a human being without being detected as a computer, it has demonstrated human intelligence. To pass any of the Turing tests, a device had to have a

broad experience of the world, a vivid imaginary capability, and an emotional routine. The test assumed that no sequence of pre-programmed responses was likely to be convincing over an extended period. The significance of this development is the concept of the attempting to equate a computer with a human being through mimicking human intelligence.

TABLE 3.5 Summary of Events During the 20th Century Post World War II Technology Era[↗](#)

Timeline	Example
1950	Alan Turing releases his paper, “Computing Machinery and Intelligence” that presents the Turing test for measuring machine intelligence.
1950	Claude Shannon wrote the book “Programming a computer to play chess” that had a detailed analysis of how to play chess leading the way for using computers in game-playing and AI.
1950	Isaac Asimov published the ethical guidelines “Three Laws of Robotics” for the behaviour of robots and artificial beings that is influential in AI ethics.
1955	The Dartmouth Summer Research Project on Artificial Intelligence is viewed as the birth of AI a separate field of scientific research.
1956	John McCarthy formally uses the term “artificial intelligence” at the Dartmouth Summer Research Project conference.
1957	The perceptron a basic building block of the neural network, was invented by Frank Rosenblatt at Cornell Aeronautical Laboratory, Buffalo, New York.
1966	Cessation of the machine translation project due to poor outcomes is first “AI winter,” denoting a period of reduced funding and interest in AI research.

Timeline	Example
1969	Marvin Minsky and Seymour Papert criticise the perceptron model of Frank Rosenblatt in their book, “Perceptrons.” They provided a mathematical analysis showing the limits of a class of computing machines that could be considered as models of the brain leading to a decline in research on neural networks.
1974	Paul Werbos initiates the backpropagation algorithm for training neural networks. This supervised learning algorithm is essential for the most powerful machine learning models today.
1980s	Acceptance of expert systems that are knowledge-based AI systems, with Lisp programming being also commonly used for AI research.
1982	Japan launches the 5 th Generation Computer Project that aims to develop a new generation of computers to understand and use human language.
1986	David Rumelhart, Geoffrey Hinton, and Ronald Williams present a new era of neural computation research through their book, “Parallel Distributed Processing: Explorations in the Microstructure of Cognition.”
Early 1990s	AI marginalised because expert systems failed to meet expectations.
Late 1990s	Acceptance of support vector machines with supervised learning grows and neural networks decline.
2000s	New machine learning algorithms begin to solve a wide range of problems.
2010s	Deep neural networks attain advanced results, including image recognition, natural language processing, and machine translation.
Late 2010-2020s	Chat Bots and GPT models make enormous headways in the industry.

Timeline	Example
2023	IBM announces that the world's first quantum computer, with more than 1,000 qubits, will be ready for use in 2023.

In 1950, continued the theme of applying the computer to mimic human intelligence. As early as 1949 Claude Shannon, an American mathematician and Nobel prize winner, published a paper “Programming a computer to play Chess,” which established the theoretical groundwork for generations of researchers to come. However, it was not until 1950 that he provided a detailed analysis of how to play chess against a human. Shannon liked playing a game of chess in his free time and was capable of lasting for 42 moves against one of the most highly regarded chess players in the world. It was likely that his proficiency and interest stimulated his passion for constructing a superhuman algorithm. However, it was Alan Turing, a year later, that created the first computer chess playing algorithm that was limited because the hardware at the time lacked the necessary processing power. The work by these two scientists paved the way for using computers in game-playing and AI.

Also in 1950, Isaac Asimov published the ethical guidelines “Three Laws of Robotics” for the behaviour of robots and artificial beings. These rules basically were meant to initiate a debate on ethical issues concerning artificial creations. Isaac Asimov was an American writer and professor of biochemistry at Boston University, who is deemed to be one of the big three science fiction writers, along with Robert A. Heinlein and Arthur C. Clarke. He is credited with writing or editing more than 500 books. Isaac Asimov’s “Three Laws of Robotics” are namely, that a robot may not injure a human being or, through inaction, allow a human being to come to harm; a robot must obey orders given it by human beings except where such orders would conflict with the First Law; and a robot must protect its own existence as long as such protection does not conflict with the First or Second Law. Although science fiction, Asimov’s robot stories explore moral values and sociological habits and have implications for our present epoch. For instance, a robot or in a wider sense an automatic device would do well to

abide by these rules. However, such generic rules are continually being violated. Smart bombs and cruise missiles are automatic devices that violate the first and third laws. Computers and technologies operated by computers function according to what they are programmed to do and, will hurt humans if programmed to do so, or if humans get in their way as they are executing their programs. Hence, the Isaac Asimov's robotics laws probe philosophical-moral issues that are applicable today more than ever when one considers the potential power of AI and its ability to control many systems and devices.

However, it was 1955 that the word artificial intelligence was used to promote a conference to specifically discuss AI. In August 1955, John McCarthy, Dartmouth College; Marvin Minsky, Harvard University; Nathaniel Rochester, I.B.M. Corporation; and Claude Shannon, Bell Telephone Laboratories proposed a two-month, ten-man study of artificial intelligence to be conducted during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. Their proposal stated that the study is to proceed based on the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. They contended that an attempt will be made to and how to make machines use language, form abstractions and concepts, solve all kinds of problems now reserved for humans, and improve themselves. They expressed their opinion that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer ([McCarthy et al., 1955](#)).

The Dartmouth Summer Research Project on Artificial Intelligence is viewed as the birth of AI as a separate field of scientific research. Their proposal was accepted and in the summer of 1956 the Dartmouth Summer Research Project conference was held, where John McCarthy formally applies the term "artificial intelligence." The conference was sponsored by the Defence Advanced Research Projects Agency (DARPA). Also in attendance, were Oliver Selfridge an innovator in early computer science and artificial intelligence; Allen Newell, a computer scientist; and Herbert

A. Simon, an economist, political scientist, and cognitive psychologist. At the Dartmouth conference Newell and Simon presented their groundbreaking Logic Theorist, which was a computer program capable of proving certain mathematical theorems and was referred to, as the first AI program.

Because of the Dartmouth conference, technology frontrunners anticipated that the development of a thinking machine that could learn and understand to mimic human intelligence was imminent, which resulted in drawing major government and industry support. The direct effect of the 1955 Dartmouth conference was the sponsorship of well-funded basic research for almost 20 years that generated significant advances in AI. Examples of this research include the General Problem Solver (GPS) algorithm released in the late 1950s, which established the basis for developing more complex cognitive architectures; the Lisp computer language for AI programming that is still in use today; and ELIZA, an early natural language processing (NLP) program that paved the way for today's chatbots.

A key development occurred in 1957 when Frank Rosenblatt at Cornell Aeronautical Laboratory, Buffalo, New York introduced the “perceptron”, which is viewed as a basic building block of the neural network. A demonstration of the “perceptron” was given by the U.S. Office of Naval Research in 1958, where an IBM 704, which was a five-ton computer the size of a room, was inputted with a sequence of punch cards, and after 50 test cycles, the computer taught itself to distinguish cards marked on the left from cards marked on the right. Rosenblatt claimed that the “perceptron” was the first machine that could create an original idea. In 1958, Rosenblatt contended that society was about to witness the birth of a machine capable of perceiving, recognising, and identifying its surroundings without any human training or control ([Lefkowitz, 2019](#)). Rosenblatt was inspired by the way neurons work together in the brain, giving him the idea for the “perceptron.” The “perceptron” is a single-layer neural network algorithm that arranges input into two possible categories. The neural network makes a guess, for instance, right or left; or dog or cat, and if its guess is wrong, it

corrects itself to make a more informed guess the next time around. The neural network algorithm becomes progressively more accurate over thousands or millions of iterations. The perceptron was the first neural network and established the principles for the concept of machine learning.

However, Rosenblatt's concept attracted sceptics who maintained that the perceptron was not able to remodel the relationship between human and machine. The general outcome was that enthusiasm regarding AI and its potential diminished along with the required funding. One of the first casualties was the cessation of the machine translation project in 1966 due to poor outcomes. This project was driven by visions of high-speed high-quality translation of arbitrary texts, particularly those that interested the military and intelligence communities, who financially supported the machine translation projects quite heavily. However, disappointment regarding this project resulted because of the linguistic problems complexity being encountered that required time and more finance to resolve. In 1966 the National Academy of Sciences ALPAC report criticised the field and, indirectly, those who were working on the project. The ALPAC report was itself condemned as being narrow-minded, biased, and short-sighted. However, the report's recommendations were accepted, except for increased expenditures for long-term research in computational linguistics. This resulted in the machine translation projects being cancelled everywhere.

Marvin Minsky who was MIT professor regarding neural networks research was very cynical of Rosenblatt's claims about perceptron's viability. Minsky doubted whether computers could be programmed to visualise and understand language because the functions depicted by Rosenblatt's perceptron were just too simple. The concern was that Rosenblatt's perceptron had only one layer, while modern neural networks have millions. Rosenblatt was working on a unique project with no history, and therefore, did not have adequate knowledge to train networks with multiple layers. However, researchers today, in hindsight, view his algorithm as still being fundamental for training today's deep networks. In 1969, Minsky and Seymour Papert published the book *Perceptrons* that

literally thrashed Rosenblatt's work and, effectively discredited it. They provided a mathematical analysis showing the limits of a class of computing machines that could be considered as models of the brain leading to a decline in research on neural networks. To make matters worse for Rosenblatt, the following year, Minsky won the A.M. Turing Award, which was computing's highest honour. The perceptron's rise and its fall due to the criticism by Minsky and Papert, amongst other concerns, helped to bring about an era known as the first "AI winter," lasting from 1974 to 1980. During this time, government and corporations backed away from their support of AI research and financing for AI research withered.

On a humanistic note, Rosenblatt referred to Minsky as the loyal opposition and never seemed bitter, even after losing hundreds of thousands of dollars a year in government funds; he was particularly sorry for his colleagues and for his students ([Lefkowitz, 2019](#)). Rosenblatt's "Mark I perceptron" now resides at the Smithsonian Institution, and in 2004, the Institute of Electrical and Electronics Engineers established the IEEE Frank Rosenblatt Award ([Lefkowitz, 2019](#)). Many academics believe that Rosenblatt's principles underlying the perceptron assisted to trigger the modern AI revolution, particularly regarding deep learning and neural networks, which can classify online images and enable language translation that are currently transforming society. Rosenblatt died in a sailing accident in 1971 on his 43rd birthday without seeing the thawing of the first AI winter.

Despite the AI winter, in 1974, Paul Werbos an American social scientist and machine learning pioneer, submitted his dissertation regarding the Back-propagation algorithm for training neural networks, which solved the XOR problem and in general made neural networks learning more efficient. The XOR, or "exclusive or," problem is a classic problem of using a neural network to predict the outputs of XOR logic gates given two binary inputs. An XOR function should return a true value if the two inputs are not equal and a false value if they are equal. Initially, the term "back-propagating error correction" was pioneered by Frank Rosenblatt in 1962. However, he did not know how to apply it even though a continuous precursor of

backpropagation was already being used in the context of control theory. Backpropagation normally refers to the entire learning algorithm. It aims to prevent redundant calculations of intermediate terms in the chain rule, which is normally derived through dynamic programming and is important for multilayer perceptrons. One should recall that multilayer perceptrons was a major weakness with Rosenblatt's model. Werbos' development is significant because a supervised learning algorithm is essential for the most powerful machine learning models today.

The 1970s and 1980s witnessed several key developments in expert systems and natural language processing. Of particular significance was the research of the 1980s regarding deep learning techniques and the general acceptance by the industry of Edward Feigenbaum's expert systems, which mimicked the decision-making process of a human expert. Edward Feigenbaum, a computer science professor at Stanford University and founder of Stanford's Knowledge Systems Laboratory, jointly with Raj Reddy, pioneered the design and construction of large-scale AI systems, exhibiting the practical importance and potential commercial impact of AI technology. Their program would request an expert in a field how to respond in an assumed circumstance, and once this was learned for practically every situation, non-experts could receive advice from that program. These expert systems were commonly used in various industries. During this time, the Lisp programming continued to improve and was also commonly used for AI research. The first version of Lisp was created by John McCarthy in 1959 at Massachusetts Institute of Technology and was written for the IBM 704 mainframe using punched cards.

Along its journey, AI encountered many challenges, a major one being the "frame problem." Simply put, the "frame problem" is about how AI systems can understand and process the world around them. For instance, one may describe the characteristics of an object to an AI system. However, a problem is encountered when not all the characteristics are described or are known at that time. Hence, a difficulty develops in that the AI system would not specifically know these missing characteristics, and it also does not know that it has a limited knowledge about them. Therefore, the

challenge for AI systems is to understand that they might be missing additional information. The human learning process is not just to learn facts but also to recognise and learn trends, patterns, relationships, and context. This is why the human thought process is so complex. This human learning process permits us to bridge the gaps when we come across new information. However, the early AI systems did not have this ability; so even as they improved their ability to process information, they still had a difficulty with this frame problem. Hence, the early AI systems could not recognise that their knowledge was incomplete, which therefore, limited their capacity to learn and adjust, thus posing a key barrier for AI advancement.

To resolve this dilemma, AI developers introduced “slots” (or windows) for knowledge that permitted an AI system to realise that it did not have all the information about an object or, a certain topic and introduced “scripts” or “frames” to stand for the standard sequence of events in a particular situation. This enabled the AI system to bridge the information gaps and anticipate what might transpire. This method allowed an AI system to make advances in resolving the frame problem. However, AI systems are still not sophisticated enough to compare with the flexibility and adaptability of the human intellect. The success of Edward Feigenbaum’s expert systems and others inspired a new period of AI interest that attracted significant financing during the 1980s. Hence, in the early 1980s, Japan and the United States boosted funding for AI research, which assisted in reviving it. AI systems through vastly improved expert systems, ultimately exhibited the true value of AI research by producing real-world business-applicable and value-generating systems. The first expert system, known as XCON was commissioned and used by Digital Equipment Corporation (DEC). XCON was a sizeable system, containing about 750 rules. While it could construct and execute the numerous orders submitted to it, it still necessitated expansion and sophistication by DEC before it, could become the reliable system that it is today. At that time the system was estimated to generate over \$40 million in profit for DEC each year.

Technological progress, that facilitated AI development continued and in 1982 Japan launched the 5th Generation Computer Project that aimed to develop a new generation of computers to understand and use human language. This project was a ten-year programme with an investment of about \$400 million that was initiated by Japan's Ministry of International Trade and Industry (MITI) to produce computers making extensive use of parallel processing and logic programming. The project was to deliver a computer with supercomputer-like performance and to provide a platform for future developments in artificial intelligence. This 1980s project is viewed as being ahead of its time, but its extreme ambitious goals led to commercial failure. However, it was the origin of concurrent logic programming that inspired a talented generation of engineers and scientists. The initial failure of the 5th generation computer project and the failure of expert systems to meet expectation are likely to have led to the second AI winter that lasted until the mid-1990s, until innovative progress on neural networks, and the advent of big data boosted the current resurgence of AI.

As Ashleigh Brilliant an English-born American author and cartoonist said, "You can't stop progress, but you can help decide what is progress and what isn't." During the so called second AI winter, David Rumelhart, Geoffrey Hinton, and Ronald Williams in 1986 introduced a new era of neural computation research through their book, "Parallel Distributed Processing: Explorations in the Microstructure of Cognition." Basically, their publication promoted backpropagation and illustrated how it could be used to train multi-layer perceptrons (MLPs). The Backpropagation, or backward propagation algorithm is intended to test for errors working back from output nodes to input nodes, and is a vital mathematical tool for improving the accuracy of predictions in data mining and machine learning using experience. Their paper is viewed as groundbreaking in the development of neural networks and helped to renew interest during the AI winter. These researchers also focused on the specific shortcomings of neural networks that were highlighted by Marvin Minsky and Seymour Papert in their 1969 book "Perceptrons" that discredited Rosenblatt's neural network model. Rumelhart, Hinton, and Williams showed that

backpropagation could be used to train perceptrons to solve these problems, thus largely justifying Rosenblatt's original neural network model. Currently, backpropagation is the most widely used method for training neural networks because it is a robust and flexible technique that can be exploited to resolve a wide variety of problems.

Despite the absence of government funding the momentum for AI development continued to grow. In the 1990s, AI was maturing and evolving rapidly with the introduction of new approaches such as improved neural networks and machine learning. This secured acceptance, particularly with the robust resolution of the frame problem. Support vector machines (SVM) gained popularity due to their machine learning algorithm that utilised supervised learning models to solve complex classification, regression, and outlier detection problems by performing optimal data transformations that determine boundaries between data points based on predefined classes, labels, or outputs.

One must distinguish between machine learning and neural networks. Machine learning is a subset of artificial intelligence and refers to computers that learn from data without being explicitly programmed. Neural networks are a specific type of machine learning model, with many interconnected nodes that process information and make decisions being modelled to mimic the human brain. The fundamental concept regarding neural networks is that their learning capabilities are acquired from data to improve their performance over time. They have advanced to be exceptionally good at pattern recognition, that enable them to be used for image recognition, natural language processing, and similar tasks. Therefore, an AI system with machine learning capability neural network contains algorithms that learn and continually improve their performance. As one may imagine, having neural networks with these machine learning capabilities results in having a very powerful and awesome capacity. As neural networks and machine learning algorithms became more advanced, they began to outperform humans at certain tasks. For instance, in 1997, a chess playing computer program called Deep Blue was able to defeat the world chess champion, Garry Kasparov. This was a key milestone for AI,

because it demonstrated in a practical sense that computers could outperform humans at a task that required complex reasoning and strategic thinking.

One needs to recognise that machine learning algorithms absorb enormous amounts of data and can identify patterns in that data to make predictions or classifications. Hence, machine learning became a key feature in the evolution of AI because it permitted AI systems to learn and adjust without needing to be explicitly programmed for every possible situation. It is this machine learning capability that has permitted AI to become an all-purpose flexible utility. It is this capability that allowed Dragon Systems in 1997 to develop speech recognition software that was implemented on Windows, and which was viewed as a huge leap forward in the challenge towards spoken language interpretation. This development means that all the knowledge about the world does not need to be hard coded into the system, since neural networks and machine learning algorithms are able learn from data and continually improve their performance. The outcome is that these AI systems can handle new situations and make inferences that were not explicitly programmed into them.

In the 2000s, AI has continued to advance in many different fields, with new machine algorithms beginning to solve a wide range of problems. In the early 2000s, AI systems significantly improved at language translation, and image captioning that generated textual description of an image. They even became better at answering questions that later advanced to Chatbots. Moreover, the 2010s witnessed the rise of deep learning, a more advanced form of machine learning that allowed AI to tackle even more complex tasks. These deep neural networks attain advanced results, including image recognition, natural language processing, and machine translation. Deep learning allowed AI to start making advances in fields like speech recognition, image classification, and self-driving cars.

The 2010s and 2020s saw further progressive developments in the AI language models. However, it was in the late 2010s that AI language models advanced to the sophistication level that is currently being

experienced. Initially, the 2010s AI systems were focused on features such as image recognition, natural language processing, and machine translation. However, by the late 2010s and early 2020s, language models like GPT-3 started to make headways in the AI world by generating text that closely imitated human writing. Chat Bots and GPT models made enormous headways in the industry. These language models could even write in different styles, from formal, casual, and humorous, providing new opportunities for the application of AI. AI has moved away from just high-speed computations and pattern recognition. It could now generate creative content that is comparable with human writers. However, AI developers still have several challenges to be resolved, such as overcoming Chatbot hallucinations and other similar weaknesses.

Whilst the future of AI will be discussed in a separate chapter, it is appropriate to provide a brief outlook of the potential future developments. Research indicates that the next stage in the development of AI is what is referred to as “Artificial General Intelligence” or AGI. The AGI systems are viewed as being capable of performing any intellectual task that can be performed by a human being. This is the opposite to “narrow AI” of the 2010s, which could execute only specific tasks. Hence, the objective of AGI is to generate AI systems that can learn, adapt, and act like humans, to conduct a wide scope of tasks. The development of AGI appears to be taking two main approaches, namely symbolic AI, and neural network-based AI. Symbolic AI systems apply logic and reasoning to resolve problems, while neural network-based AI systems mimic the human brain and apply huge networks of interconnected “neurons” to process information and make decisions.

There is little consensus among technologists about the best approach to take for the development of AGI, since both methods have their strengths and weaknesses. Symbolic AI systems that followed the original approach and are still currently being used in many applications that apply rules and logic to represent and solve problems. Their strength includes their accuracy and reliability, and their ability of conducting tasks that require reasoning and planning. However, they are not very efficient for carrying

out tasks that are unstructured or ambiguous since they lack flexibility. Therefore, they find it difficult to adapt to new situations. On the other hand, neural network-based AI systems are viewed as being more flexible and adaptive, but they can be less accurate and reliable, and are more difficult to interpret.

Basically, the objective in the future AI development is to upgrade to Super AI, which is capable of exceeding human intelligence in all areas. When available, it would be able to solve complex problems that are beyond human abilities and would be able to learn and adapt at a rate that far exceeds human intelligence. Super AI will also bring about two other AI functionalities, namely Theory of Mind and Self-Awareness. Theory of Mind AI types are supposed to interact with the thoughts and emotions of humans, while Self-awareness AI aims to achieve enlightenment and becomes self-aware, with an independent intelligence, and will not be under the control of the people that created it. These types of developments will turn the AI debate from a technological issue to a very difficult philosophical topic, where ethical issues became of immense importance for the protection of the human being's future existence. However, these types of AI are likely to require more processing power than is currently available.

This is where the development of the quantum computer comes into play. In 2023, IBM announced that the world's first quantum computer, with more than 1,000 qubits, will be ready for use in 2023. However, to date no further announcements have been made. Some technologists think that currently, quantum computing is still in the very earliest stages of its development. Nevertheless, it has been reported that in the fiscal year starting April 2023, Japan's Fujitsu company with the collaboration of the Riken research institute was to provide various industries access to the country's first home-grown quantum computer that has 64 superconducting qubits. There appears to be consensus amongst technology observers that it will be at least 2030 before quantum computers are ready for general use. There are also rumours that NASA's quantum computer project is currently on hold since the US government is concerned about merging AI and

quantum computers. A quantum computer is viewed as transforming the way humans think about computing, increasing processing speeds exponentially, and granting access to previously inaccessible data.

The acute concern with the quantum computer is the well-known risk that criminals or states could use it to crack public key infrastructure (PKI) that is used to secure a multitude of common web interactions. For instance, credit card transactions and stock exchanges could be attacked by fraudsters; air traffic systems and GPS signals could be manipulated; and the security of critical infrastructure, like nuclear plants and the power grid, could be vulnerable to unauthorised access with criminal or terrorist intent. Hence, many believe that AI Quantum computers will have the possibility of breaking available encryption systems, which are based on complex mathematical enigmas that are difficult for traditional or current computers to solve. The general concern is that Quantum computers could be used by unauthorised users to access sensitive information such as financial records, medical data, and government secrets, including nuclear weapons systems.

3.3 CONCLUSION

The purpose of this chapter was to provide a brief history of the evolution of AI up to the current period. The journey to date has been a long, difficult, and at times, turbulent one, and there is little doubt that it will continue to be challenging. [Figure 3.1](#) depicts the major events that have impacted the evolution of AI for the identified development periods, namely the Era of Enlightened Rational Thinking; the 20th Century: Pre-to-World War II Technology Era; and the 20th Century: Post World War II Technology Era.

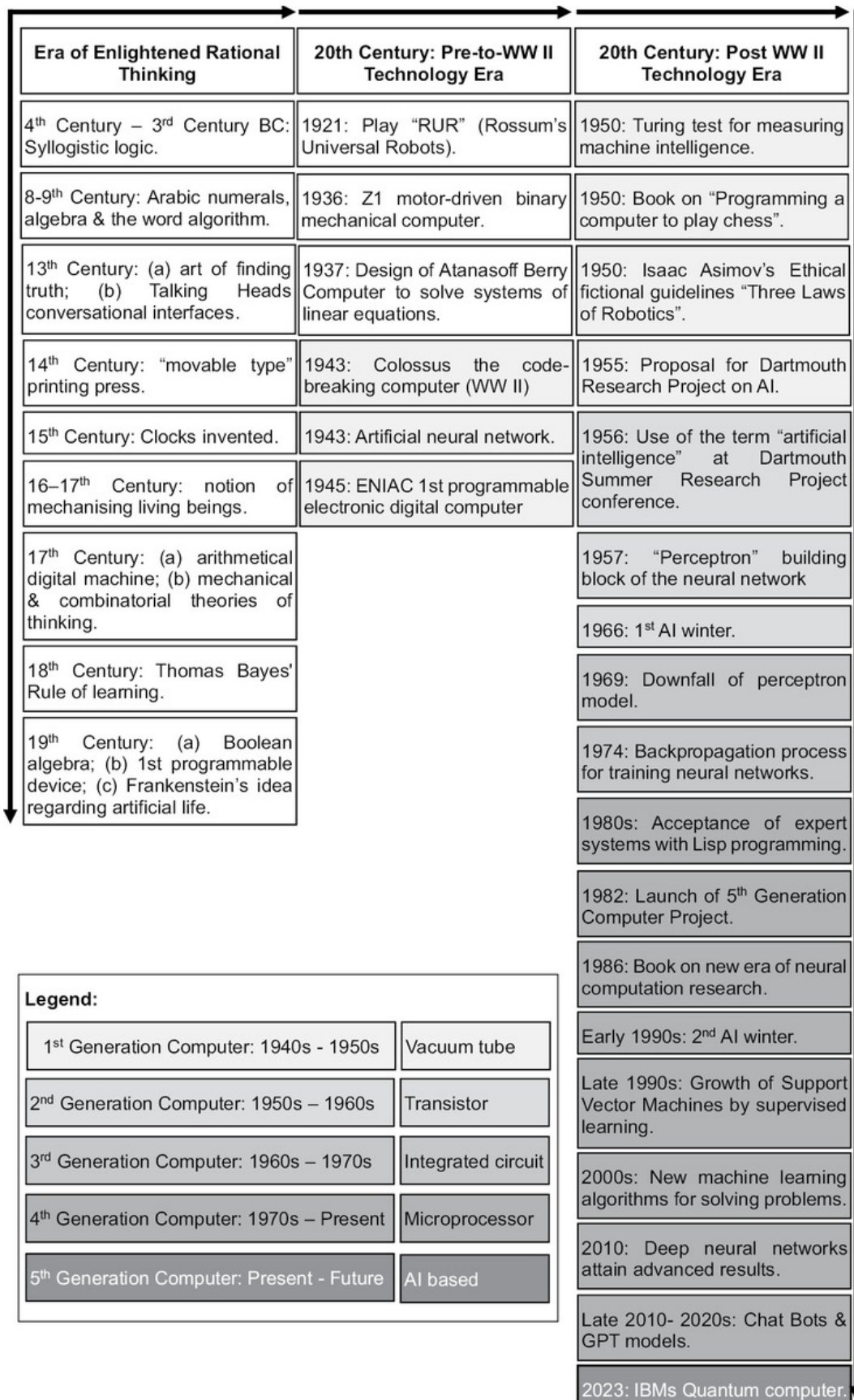


FIGURE 3.1 Summary of the Events Impacting the Evolution of Artificial Intelligence. 

There is no doubt that the ancient thinkers, be they philosophers, mathematicians, theologians, scientists or play writers have had a direct influence on the development of AI through the art of logical thinking and linguistic expression. The years leading to the WWII, and the war years, witnessed a giant leap in the technological development. This development was based on the mathematical discoveries of the ancient thinkers and expanded through the invention of other tools of a mechanical and electrical nature that utilised previously documented mathematical and scientific knowledge. This resulted in the first programmable computers and the pioneering work related to the artificial neural network. The launching of the first computers inspired writers, who were not necessarily scientifically orientated to kindle their creative imagination and conjure up artificial existence, such as the term robot. Admittedly, the idea of mechanised living beings was partly due to the vivid imagination of science fiction writers dating back to the 17th century and the frightening vision in 19th century with the creation of Frankenstein's artificial life.

However, the post-world war era witnessed an exponential leap in technological developments, particularly in electronics. Electronics formed the basis of all major inventions related to machines, aeroplanes, weaponry, and many others, all requiring design computations and production control systems, once they were ready for manufacturing. The improvement and the general demand for computers became a critical factor to enhancing development in every sphere. More powerful computers were continually being sought, driven by new concepts in robotics, neural networks, and general-purpose AI. Today, machine learning, deep neural networks, chatbots, amongst others are common knowledge to students, managers and other professions. They are not long viewed as mysterious. This is not to say that their underlying concepts are not complex, but that they appear to be accepted as the way to future economic development.

It is often said that history repeats itself. The timeline at [Figure 3.1](#) illustrates that the development of AI has at times been restrained or held back by the computing power available at a particular period. Moreover, whenever a new and powerful computing power has been developed, AI appears to make a forward leap. Hence, it is envisaged that the Quantum computer will likely provide an opportunity for AI to make an evolutionary leap into the realm of the Super AI. This will mean that it will be likely possible to develop and implement General AI or Strong AI, which is designed to perform any intellectual task that a human can do and is able to reason, learn, and understand complex concepts. Furthermore, it will also mean that Narrow AI or Weak AI that is designed to do a specific task or a limited range of tasks like facial, speech and image recognition; natural language processing; and recommendation systems will be improved and strengthened. However, the big leap will be the Super AI, which will likely exceed human intelligence in all areas and be able to solve complex problems that are beyond human abilities by learning and adapting at a rate that far exceeds human intelligence. This will also likely bring about new AI functionalities, such as Theory of Mind AI that will interact with the thoughts and emotions of humans, and Self-awareness AI, where AI will achieve enlightenment and becomes self-aware with independent intelligence that will not necessarily be under the control of the people that created it.

Whether this development is good or bad for humanity in general is debatable, and only time will tell. As previously explained, the critical concern with this type of development is the high well-known risk that criminals or states could use it to crack public key infrastructure that is used to secure a multitude of common web interactions. It could leave us open to intentional manipulation of critical facilities, such as credit card transactions; stock exchanges; air traffic systems and GPS signals; and the security of critical infrastructure, like nuclear power plants and their associated distribution grids; and weapons systems. Hence, many believe that AI Quantum computers will have the possibility of breaking available encryption systems, which are based on complex mathematical enigmas

that are difficult for current computers to solve. The general concern is that Quantum computers could be used by unauthorised users to access sensitive information such as financial records, medical data, and government secrets, including nuclear weapons systems. AI will no longer be a technical issue; it will basically become a philosophical debate. There will come a time and not into the distant future, that it will not be a question of what a computer system (supported by AI) can do, but what we human beings will trust these computer systems to do.

There is an urgent need today to consider and formulate a robust and safe legal mechanism on a national and international level to address the security risks that the future highly sophisticated AI systems supported by 5th, or even 6th, generation computers will pose to world security in general and the existence of humanity in particular. This legal mechanism needs to be explored and defined now before Super AI and Self-awareness AI are extensively developed and become readily available. As a further mitigation measure, it is also essential that research funding is forthcoming for the continuous development of security encryption systems or intrusion prevention system (IPS) to address the evolving threat landscape. These measures are doable but should be undertaken now before we are overtaken by future events.

REFERENCES

- Lefkowitz, Melanie. (2019). Professor's perceptron paved the way for AI – 60 years too soon. *Cornell Chronicle*, NY. <https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon>.↵
- Marr, Bernard. (2023). The evolution of AI: Transforming the World One Algorithm at a time. <https://bernardmarr.com/the-evolution-of-ai-transforming-the-world-one-algorithm-at-a-time/>
- McCarthy, John., Minsky, Marvin., Rochester, Nathaniel., and Shannon, Claude. (2006) “A Proposal for the Dartmouth Summer Research Project

on Artificial Intelligence, August 31, 1955”, *AI Magazine*, Vol. 27, No. 4, p. 12. doi: [10.1609/aimag.v27i4.1904](https://doi.org/10.1609/aimag.v27i4.1904).↵

Patowary, Upakul. (2023). Artificial intelligence and Mary Shelley’s *Frankenstein*: A comparative analysis of creation, morality and responsibility. *Integrated Journal for Research in Arts and Humanities*, Vol. 3, No. 4, pp. 121-127↵

Rockwell, Anyoha. (2017). The history of Artificial Intelligence. *Harvard University, Science in the News*.
<https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>

Roy, Tanya. (2023). The history and evolution of Artificial Intelligence, AI’s present and future. *Alltech Magazine*. <https://alltechmagazine.com/the-evolution-of-ai/#:~:text=This%20capability%20opened%20the%20door,think%20and%20learn%20like%20humans.>↵

Tucci, Linda. (2021). *A Guide to Artificial Intelligence in the Enterprise*. Large-Strategic Initiatives, SearchEnterpriseAI.com.↵

Wagner, Jacqueline. (2019). *Artificial Intelligence for Games: Programming a Computer to Play Chess in the 1950s*. University of Heidelberg.
https://hci.iwr.uni-heidelberg.de/system/files/private/downloads/1883894185/report_jacqueline_wagner.pdf↵

Chapter 4

Contrasting Human and Artificial Intelligence

DOI: [10.1201/9781003653950-5](https://doi.org/10.1201/9781003653950-5)

“Could one make a machine which would have feelings like you, and I do?”

Alan Mathison Turing OBE, FRS

English mathematician, computer scientist, logician, cryptanalyst, philosopher and theoretical biologist (1912-1954)

There is little consensus on the definition of AI but there is common agreement that AI has some of the abilities of human intelligence. Although AI is a human creation, the terms AI and human intelligence are sometimes confused and used interchangeably. It must be recognised that they are two distinct processes with their own distinctive features. While AI must be supported by powerful computing technologies that allow them to imitate cognitive processes such as learning and problem-solving, human intelligence is based on a living being with an assortment of mental attributes such as emotions, creativity, awareness, and memory, amongst others. With the technology that is currently available on the commercial marketplace (excluding possible covert state establishments), developing an

AI with human intelligence it is not yet doable. However, with the advent of more advance super computers, such as Quantum technology and sophisticated intelligent learning algorithms, there is the potential to create approaches significantly like those used by human beings.

Despite the differences between AI and human intelligence, most AI applications can perform effectively the specific tasks allocated to them with precise accuracy, however, with limited ability. Humans are restrained by their physique; for example, they must sleep, eat and rest. However, unlike humans an AI does not succumb to fatigue or stress and can operate on 24/7 basis with consistent performance. Human intelligence utilises brain power, memory, and cognitive abilities while AI is reliant on the data provided by humans. Having said this, it must also be recognised that once the seed (data from humans) is provided, an effective AI system is able to learn and expand its knowledge far beyond the data that was originally provided by humans. Stephen Hawking is reported to have said that the development of full artificial intelligence could spell the end of humans as it would take off on its own, and re-design itself at an ever-increasing rate. He argued that humans, who are limited by slow biological evolution, could not compete, and would be superseded ([Medeiros, 2017](#)).

AI makes extensive use of cognitive computing, which has made it possible to have expert systems, theoretically replacing the human expert. This type of technology has been found to be effective in such fields such as medicine and education. AI is viewed as an amalgamation of various disciplines, including computer science, electronic engineering, mathematics, biology, physiology, and philosophy. These disciplines have been founded by human knowledge built over many centuries. It must be emphasised that AI is the creation of humans. It is humans that have designed and programmed the algorithms which permit AI systems to execute very complex problems and continue to self-learn; thus, making AI more resourceful. However, currently there is no doubt that the human intellect is more developed than AI in terms of self-awareness, social interactions, and multitasking.

4.1 INTRODUCTION

It has been said that there is no comparison between human and AI for the simple reason that human intelligence is far more superior. However, one also needs to consider the fact that humans themselves differ in terms of their level of intelligence, which follows a normal mathematical distribution. The normal distribution illustrates that there is an average level of human intelligence, with a percentage of individuals being below average and others above average. One also must remember that it is humans that have created AI. However, I would suggest that the humans that have contributed to the creation of AI are likely to be on the above average intelligence level. Therefore, when comparing human and artificial intelligence, there is a need to establish that the comparison is between the average human and the average AI system. A point at issue is whether AI follows the normal distribution and whether the average level (and respective spread from the average) is higher or lower than that of human intelligence. Hence, whether it is possible to compare like with like is an issue.

Another important factor is the fact that there are different types of intelligence. For instance, if the level of memory (in terms of information on all types of topics) is a type of intelligence, then I would suggest that computers and their associated AI systems would be far more superior than that of an individual human being no matter what level of intelligence that individual may have. However, this is assuming that the comparison being made is between an individual human being and an AI system. However, this is not a fair comparison because the initial data that is inputted into an AI system is an assortment of human knowledge and not just from one specific individual. Therefore, if a collective approach is taken, where the collective knowledge (past and present) of humans is taken into consideration then it is doubtful whether an AI system is currently superior. The basic reason being that current AI systems are founded upon the past and present knowledge of humans, so at minimum AI would be equal but not superior to human intelligence.

However, when an AI system has extensive learning capabilities, this human advantage is likely to diminish gradually over time because an AI system will probably learn more swiftly. For example, in 1997, a chess playing computer program called Deep Blue was able to defeat the world chess champion, Garry Kasparov. It is likely that to reach world chess champion status, Garry Kasparov would have had many years of intensive practice playing chess. In fact, this was a key milestone for AI, because it demonstrated in a practical sense that computers could outperform humans at a task that required complex reasoning and strategic thinking.

Potentially an AI system will learn at a much faster rate than that of humans, particularly if humans take a back seat and allow the AI to do its thinking for them. For example, taking myself as a case in point. Before the simple calculator became cheaply available, I was able to do arithmetical calculations mentally very quickly without the aid of artificial tools, such as the calculator. Today, I would find it difficult to add a column of five or six numbers quickly, simply because of being reliant on the calculator or some other electronic tool, such as EXCEL. I also remember, that when I was completing my master's degree exam, I had to practice for hours in the task of plain handwriting because I became so dependent on using a word processor and keyboard that my handwriting abilities became very rusty indeed. The point that is being made is that humans need to be extremely cautious about being too dependent on a machine to do the things that they normally do, particularly thinking and attaining knowledge and eventually wisdom.

In the first chapter of this book reference was made to the lyrics of the song "In the Year 2525 (Exordium & Terminus)" by Zager and Evans. When I came to writing this chapter, I reread closely the lyrics of this song, which may seem an exaggeration to many, but the words above referring to being too dependent on computers and their associated AI systems reflect exactly the sentiment being expressed by this song. In fact, this song is conveying a message that if humans allow machines to do the normal things they do, both physically and mentally, even their biological qualities will diminish as the human being evolves in the future. Obviously, whether

humans will remain intellectually superior or not, only time will tell. Furthermore, before comparing human and AI it is appropriate that the meaning of intelligence and learning are explored and understood. When we refer to intelligence and learning, what exactly do we mean?

4.2 INTELLIGENCE AND LEARNING

[Lee \(2020\)](#) argues that while it may be unlikely that humans will have superintelligent computers or robots like the ones in films such as Terminator anytime soon, investigating the differences between human intelligence and AI can help us better understand the future of technology and our relationship with it. Lee contends that only when humans have created artificial life that can physically replicate by itself, that they have created truly AI. Until then, machines will always only be surrogates of human intelligence, which unfortunately still leaves open the possibility of abuse by people controlling the AI.

It is not the intention of this chapter to examine the many theories that are in existence regarding human intelligence. The objective of this chapter is to ascertain the general meaning of intelligence so that one can understand its underlying characteristics. According to [Sternberg \(2023\)](#), human intelligence is a mental quality that consists of the abilities to learn from experience, adapt to new situations, understand, and handle abstract concepts, and use knowledge to manipulate one's environment. He argues that the question of what exactly defines human intelligence is contested, particularly among researchers of AI, though there is broader agreement that intelligence consists of multiple processes, rather than being a single ability.

Sternberg contends that the main emphasis in a definition of intelligence, is that it is not a cognitive or mental process as such but somewhat a discerning combination of these processes that is purposively directed toward effective adaptation. He cites the example of a physician who learns about a new disease, adapts by perceiving material on the disease in medical literature, learning what the material contains, remembering the

crucial aspects that are needed to treat the patient, and then utilising reason to solve the problem of applying the information to the needs of the patient. Therefore, intelligence, in general is not a single ability but an effective combination of many abilities. Furthermore, measuring intelligence has attracted a lot of attention over many decades and standardised intelligence tests became popular during the 20th century. However, the limitations of intelligence tests have been extensively recognised.

The above definitions of intelligence refer to learning, yet intelligence and learning tend to have separate definitions and are often discussed individually. However, the difference between these two concepts is not well defined when it comes to the actual task of learning. Research has shown that there is a strong correlation between intelligence and learning, the likelihood being that the relationship is a causal one. It is important to understand that by applying their intelligence, humans are able to learn, articulate concepts, comprehend, and employ logic and reason. Moreover, research has shown that human intelligence incorporates the capability to recognise patterns, plan activities, be creative, solve complex problems, make rational decisions, retain and recall information, and use language to communicate. Intelligence can also be viewed as the human ability to solve complex problems or make rational decisions with the outcomes aiding the performer, and has evolved to adjust to different environments for their survival and reproduction ([Weder, 2020](#)).

Currently, intelligence is often viewed in terms of analytical and reasoning skills, which are vital and critical components to successful learning. However, there are other types of intelligences, such as language communication abilities or emotional awareness that influence an individual's learning process. Learning is the process of gaining new understanding, knowledge, skills, attitudes, values, behaviours, and preferences. It could be viewed as a mental exploratory expedition, where the individual gains from this experience, contributing to its intellectual and behavioural development. According to [Ambrose et al. \(2010\)](#), learning is a process that leads to change, which occurs because of experience and increases the potential for improved performance and future learning, with

the change in the learner to possibly occurring at the level of knowledge, attitude, or behaviour. As a result of learning, individuals come to see concepts, ideas, and/or the world differently. Learning is the direct outcome of how an individual understands, interprets, and reacts to their experiences. The learning process provides an individual with an opportunity to develop and apply intellectual skills and thinking processes, such as problem-solving; motor skills, values and attitudes; and develop interpersonal and social skills that are important for personal development. Intelligence and learning are two concepts linked with brainpower and cognitive functioning. Hence, the better the brain works, the greater the likelihood that an individual will learn and develop its intelligence.

[Mehta \(2021\)](#) argues that humans are not all naturally skilled at the same things. Some are more athletic and have better coordination, some are more linguistic, while others are good with numbers and visualizing patterns. Furthermore, he contends that most people do not fully understand their range of abilities, and as a result make the wrong career choice, or they might enjoy their jobs, but struggle to identify effective learning techniques that will help them excel further. Therefore, it may be concluded that individuals are not intellectually the same but have different talents and capabilities. Mehta argues that humans possess multiple intelligences and cites Harvard University Professor Howard Gardner's book *Frames of Mind* that was first published in 1983, which proposes that there are eight types of human intelligence, with each representing different ways of how a person best processes information (refer to [Figure 4.1](#)). Furthermore, Mehta maintains that a high level in one intelligence category does not necessarily influence the level one has in another. The eight types of intelligence identified by Gardner are as follows:

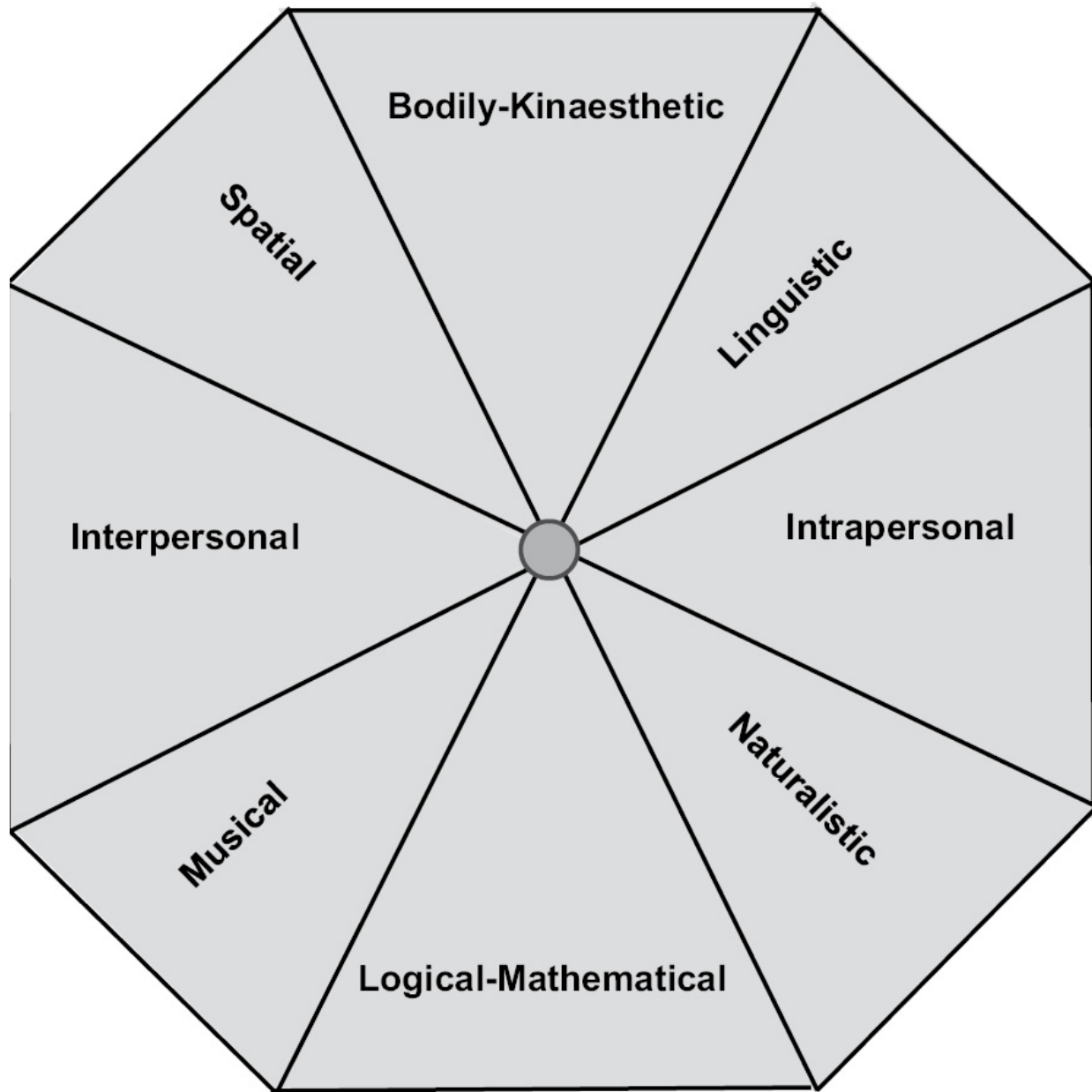


FIGURE 4.1 Howard Gardner's Eight Types of Human Intelligence.[📄](#)

1. Spatial intelligence: The ability to think abstractly and in multiple dimensions.
2. Bodily-kinaesthetic intelligence: The ability to use one's body in a way that demonstrates physical and athletic prowess.
3. Musical intelligence: The capacity of being sensitive to rhythm, pitch, meter, tone, melody and timbre, entailing the ability to sing and/or play

musical instruments.

4. Linguistic intelligence: The capability of being sensitive to the meaning of words, the order among words, and the sound, rhythms, inflections and meter of words, including memorising information and reading.
5. Logical-mathematical intelligence: The ability to analyse problems logically, carry out mathematical operations and investigate issues scientifically, including solving abstract problems.
6. Interpersonal intelligence: The ability to interact effectively with others; being sensitive to other people's moods, feelings, temperaments and motivations, including able to understand and relate to those around you.
7. Intrapersonal intelligence: The capacity of being sensitive to one's own feelings, goals and anxieties, including the capability to plan and act in the light of one's own traits.
8. Naturalistic intelligence: The ability to understand the distinctions in nature, including the differences between plants, animals, and other elements of nature and life.

[Figure 4.1](#) indicates that intelligence has many aspects which provide human beings with the various attributes that help the individual to undertake an assortment of tasks. Additionally, understanding these various aspects allows humans to have a comprehensive model of what intelligence really is and provides a baseline for comparing it with AI to ascertain the differences between the two. An important issue that needs to be addressed is the process of transforming data to information to knowledge and finally to wisdom. It is essential to evaluate this transformation process to see how intelligence through learning develops in humans and whether this differs in the case of AI.

4.3 THE JOURNEY OF TRANSFORMING DATA TO WISDOM

The general opinion is that humans tend to be superior to AI in varying environments, particularly at tasks that require empathy. Human intelligence incorporates the capability to understand and interact with the emotions and sensitivities of other human beings, which an AI is not capable of doing. This last statement needs a little more clarification. People can have empathy for others and this feeling can be genuine or made-up. For instance, an actor in a movie or play can mimic the emotions of the character that he/she is role playing. The more acting talent a human has, the better the individual's acting performance. For example, winning an Oscar award for best actor or actress is one of the highest measures of one's talent to act. Therefore, a human can have genuine feelings or can play act, but an AI can only possibly have the latter, play acting. It is argued that unlike humans, which is a living being, AI can only play-act the part of emotions. Additionally, a person can play act and still put feeling into the role. A robot can be programmed to make a sad or happy facial expressions but there is no true feeling and is likely to be amateurish even at role playing.

Before going into the specific differences between human intelligence and AI, it is appropriate at this stage to examine how data is transformed to wisdom. For example, consider an aubergine. It starts its life as a seed and evolves from the seed into a seedling and then a plant; and from a plant as the fruit itself, with each evolutionary phase adding to the development of the final product. The transformation process from raw data to wisdoms evolves in a similar fashion. Data is raw factual material, which is used as the basis for reasoning, discussion, or calculation. On the other hand, wisdom is the ability to think and act, applying the information, knowledge and experience, understanding, common sense, and insight that is accumulated over time. Therefore, data is at starting point of the data-wisdom transformation spectrum with wisdom being at the other end of the spectrum. It is argued that the process of transforming data to wisdom is different for humans and AI. Additionally, this transformation process brings into the open the human superiority over AI.

[Figure 4.2](#) demonstrates the evolutionary process that takes place as data is transformed to information; information is transformed to knowledge; and knowledge is transformed to wisdom. In fact, this transformation process may be viewed as the basis for both human and AI learning systems, with the transformation process reaching the summit in terms of wisdom, the accumulation of which generates intelligence.

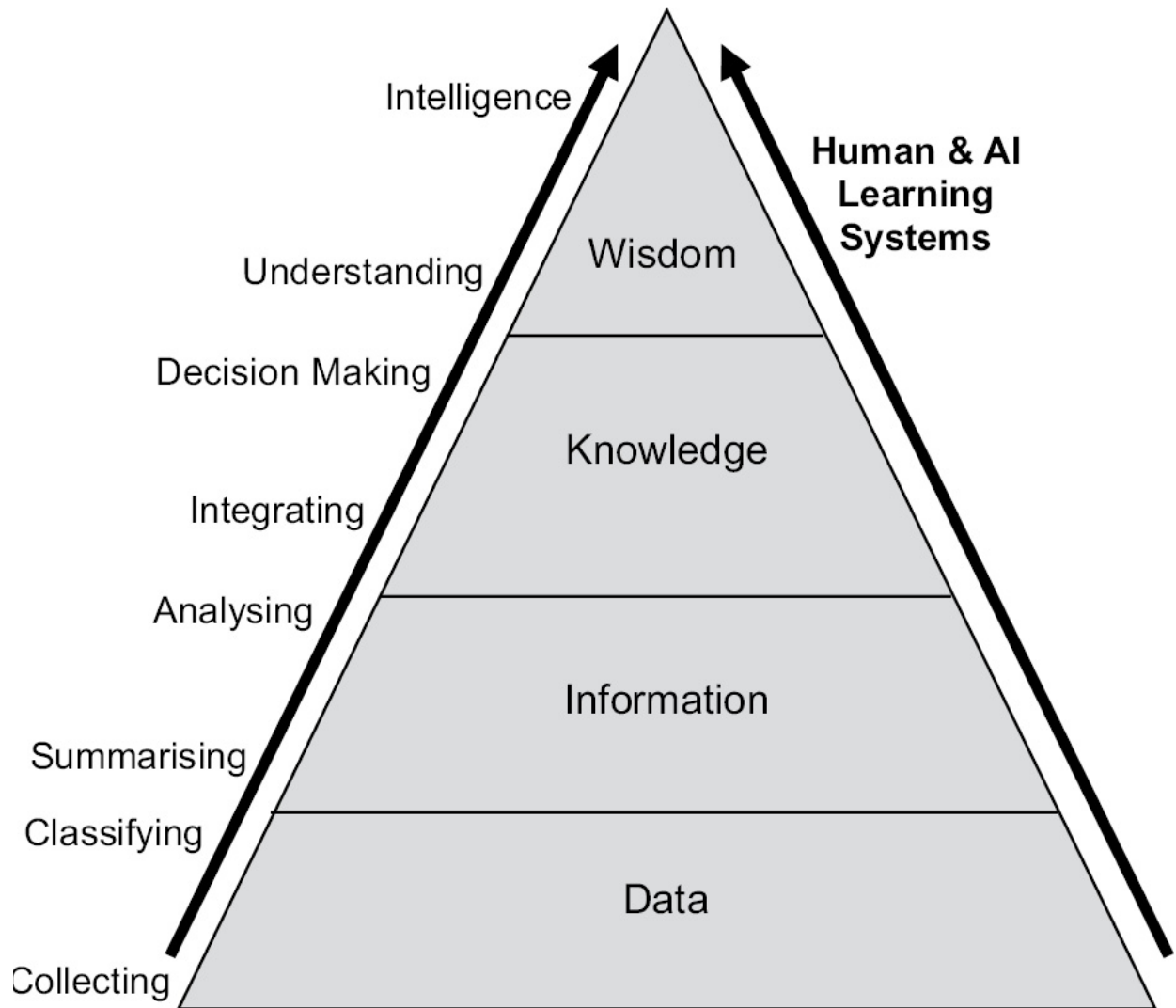


FIGURE 4.2 Data Transformation Process. [📄](#)

4.3.1 Data: Raw Material for the Data Transformation Process

Data are the basic raw facts that an individual and an organisation collect in their day-to-day activities. Both the individual and the organisation are continually being inundated with raw facts, which may or may not be relevant to their needs. At this level, the raw facts are classified into some kind of order to facilitate the determination of their relevance. Hence, the collected data is filtered for relevancy and is prearranged in a manner that allows it to be absorbed in memory in the case of an individual or inputted into a computer system in the case of an organisation to be accessed and utilised at any time by the individual and the organisation's AI systems.

Humans and AI systems alike are amassing an immense and escalating quantity of data in dissimilar formats and from diverse sources. Obviously, AI systems are able to handle much more data than an individual human being. What is absorbed into human memory is limited to what a particular individual deems important. What is stored in AI systems may include: (a) operational or transactional data, related to such things as revenue, expenditure, human resources, and accounting, amongst many others; (b) non-operational data, such as macro-economic data, industry sales and forecasts; and (c) meta-data that consists of descriptive data about the data itself, such as logical database design or data dictionary definitions. Moreover, at an AI level, data is typically prearranged into a repository that is referred to as a database and may take the form of numbers; words; observations or descriptions of things; pictures; and video and audio recordings. An individual's repository is its memory.

There are two primary sources of data for humans and AI systems, namely internal and external sources. The internal data source for humans is how they feel in terms of emotions and signals that their bodies may provide regarding their health. The AI systems obtain internal data about their organisation that reflects the internal environment of that organisation, such as its financial health. External data is generated from sources outside the human being and the organisation and reflects the external environment. However, this is where we start to see the major difference between humans and AI. Humans receive external data through the various senses, namely, sight, hearing, smell, taste, and touch. AI systems initially receive data that

humans feed it and then continues to learn from other data encounters. AI systems are smart enough to process different categories of data. For instance, they can analyse audio and video data that may include pattern recognition and image analysis and are applied for detecting abnormalities in terrorist exposure scenarios. However, AI systems do not have the extensive sensory facilities that humans possess, that carry nerve impulses from the sense organs toward or to the brain. AI system can be programmed to have some sensory facilities but nowhere as extensive as found in humans.

4.3.2 Information: The Transformation of Data into Information

Information is the transformation of raw data into something that is meaningful, significant and may have implications for the entity using it. For example, an individual may have several ingredients, but by themselves these ingredients are meaningless. However, by applying a recipe, the ingredients can be converted to a product that is cooked and eaten. On an organisational AI level, consider a digital file of taxpayers registered with the Inland Revenue Department (IRD) for a particular geographic region that contains their personal and financial details. This digital file may contain millions of entries. By itself this data has no value. However, data may be transformed into information by analysing or summarising it by a specific locality, income group, gender, marital status, employment position, age groups, education level, and so forth. This analysis (or transformation) may be needed to establish the average age of the taxpayers in a particular locality, and the minimum-maximum age per category (gender, position, marital status).

Hence, an entity may obtain a finer appreciation of the information that is generated in terms of better decision making. It should be noted that the actual decision-making function is entering into realm that is beyond the information level and is boarding on the knowledge-wisdom domains. However, at the transformation of data into information level, humans and an AI system are both capable, but humans have the added advantage of

their sensory abilities such intuition, whilst the AI systems have the advantage of speed.

4.3.3 Knowledge: The Transformation of Information into Knowledge

Data and information can be transformed into knowledge by examining historical patterns and future trends. Knowledge can be visualised as consisting of two corresponding and related components, namely the explicit details and the comprehension of these explicit details. Hence, the explicit details are acquired from the application of the data and information to address “how” questions to provide an insight about the magnitude and characteristics of some object. On the other hand, the comprehension of the explicit details is the objective of the analysis (i.e., why are we analysing these explicit details).

On a human level, a doctor may conduct research to examine whether there is a causal relationship between alcohol and various health problems. Through his research the doctor collects various pieces of evidence (data and information) and gains knowledge about specific situations. On an organisational level, through an AI system, a manager of a public health facility may want to know which services are to be encouraged and supported by this facility. Hence, data and information regarding the utilisation of the services offered at the health facility are analysed to provide the knowledge regarding the beneficiaries of the services being provided; in other words, the service recipient behaviour (how). This in turn will allow the manager to decide which health services should be supported and enhanced. This will aid the health facilities manager to select those services that are most responsive to any promotional effort that may be undertaken (objective or why).

The above example illustrates that knowledge is the extensive exploitation of data and information, in combination with the ability, competencies, ideas, intuitions, commitments and motivations of the manager involved. Hence, the transformation of data and information into

knowledge is the outcome of the synergy that combines human aptitudes and the application of the available data and information in an AI system. Admittedly, an AI could be capable of deciding itself, but once again the emotional feeling of a human is an essential feature for certain decisions, particularly where life and death is an issue.

4.3.4 Wisdom: The Transformation of Knowledge into Wisdom

Wisdom is the amalgamation of knowledge and experiences into insights that intensify an entity's (individual or organisation through its AI system) comprehension of relationships and the meaning of these relationships to enable one to make a correct judgement. Therefore, wisdom is the ability to make the right judgments and decisions, and hence transforms data, information, and knowledge into intelligence. In other words, knowledge is viewed as a tool, with wisdom being the skill needed to apply it. Wisdom is a subtle quality that is gained through extensive experience and when documented forms the basis for AI. Thus, as more experience is gained and documented, the higher level of intelligence that is generated. However, as one may appreciate, this is not easy to accomplish or sustain because to be wise implies a great deal of knowledge and experience that must continually be updated, maintained and sustained.

The objective of the data-information-knowledge transformation process is the attainment of wisdom. In other words, to have an evaluated understanding of what is going to take place. Hence, wisdom is concerned with the future and incorporates vision and design. With wisdom it is not a matter of grasping the past and present but to be able to forge the future. However, it must be recognised that wisdom is the pinnacle, thus it is not easy to achieve since one needs to move successively through the other transformation phases before it may be finally gained. In addition, once wisdom is gained it must be sustained with further knowledge; hence it takes an on-going demanding effort.

To place things into context, the previous sections have explained that an entity (individual or organisation) starts with raw data that must be gathered

and organised; processed and transformed to information through summarisation and analysis. This information is transformed to knowledge by integrating the data and information in a form that is suitable for decision making. Over time, the accumulation of knowledge becomes wisdom. Wisdom deals with the future because it incorporates foresight and creativity; therefore, it is not a matter of coming to terms with the past and the present. This process takes place at an individual level through brain power or at an organisational level through its AI system. However, in the absence of an AI system, the organisation attains wisdom through the brain power of its management team.

Those organisations, which do not have an AI system in place, are viewed as being at the information transformation stage, where decisions are made by those that possess individual knowledge. Hence, a key aim of an AI system is knowledge management with the objective of breaking the mind barrier, where decisions are made by individuals based on organisational knowledge. Organisations must strive to manage the knowledge gained over the years and transform it into wisdom, so that correct and effective decisions are made, preferably by the individual, supported by an AI system.

Hence, the central issue that must be addressed is related to what wisdom involves in a practical sense. [Lacewing_\(2009\)](#) in his examination of practical wisdom (from the Greek word ‘phronesis’; sometimes translated ‘prudence’) argues that Aristotle defines wisdom as a true and reasoned state of capacity to act regarding the things that are good or bad for man. He argues that while practical wisdom involves knowledge of what is good or bad, it is not merely theoretical knowledge, but a capacity to act on such knowledge as well. [Lacewing_\(2009\)](#) contends that this capacity requires:

1. Seeing things clearly. The ability to perceive, considering the general conception, and what is required in terms of feeling, choice, and action in a particular situation.

2. Acting in prudent and effective ways. Having a general conception of what is good or bad, which Aristotle relates to the conditions for human flourishing.
3. Acting with the well-being of the whole in mind (holistically). Practical wisdom involves understanding what is required in a particular situation in the light of a general understanding of what is good.
4. Deeply understanding the situation. Hence, having the ability to deliberate well.
5. Knowing when to act and when not to act. That is, the ability to act on that deliberation.

Furthermore, [Malone \(2003\)](#) contends that wisdom also involves being able to react to crisis calmly with an effective holistic response; and being able to anticipate potential problems and avoid them. Wise people generally share optimism that life's problems can be solved and experience a certain amount of calm in facing difficult decisions. It is posited that an organisation which has a management team that possess the above attributes and can repeatedly and consistently apply these attributes will have a tremendous competitive advantage.

One must recognise that an AI system has the advantage of speed and memory capacity when compared to an individual person but lacks emotions (feeling). An individual can come up with a balanced decision, considering factual tangible matters and intangible elements. The intangible elements might be the consequences for other beings. For instance, a decision regarding the downsizing of an organisation needs to consider the consequences of the decision on the motivation and commitment of the individuals working for the organisation, particularly those who will likely be retained. Hence, many organisational decisions in a practical sense do not just consider the numbers but also the consequences on those that will be directly affected by the decision. Furthermore, an organisation is normally managed by a management team that can express their feelings about a particular challenge being encountered. This is a weakness of the

current AI systems, the ability to feel (emotional perceptions). To overcome this weakness, it would be preferable if decisions are taken by humans with the aid of an AI system, where human emotions can play an important role in the final decisions.

4.4 THE GROWING CHALLENGE BETWEEN HUMAN INTELLIGENCE AND AI

The questions that are often raised are whether AI will ever outstrip human intelligence and whether there is something so exceptional about human intelligence that it may never be replicated. There is wide consensus that human intelligence experiences constraints that are not applicable to computers. However, computers also have their constraints as well, which are not applicable to humans.

Computers are exceptional when it comes to speed in number crunching. For example, AI has often exceeded human intelligence when confronted by an individual in circumstances such as playing chess and other games. Deep Blue the chess playing computer developed by IBM defeated in 1997 the chess world champion, Garry Kasparov. This triumph was viewed as a symbolic demonstration of the intellectual escalation of AI leading some researchers to believe that AI will surpasses human intelligence in every sphere. Such an event is what is known as “technological singularity.” The technological singularity is a hypothetical future point in time at which technological growth becomes uncontrollable and irreversible, resulting in unforeseeable consequences for the human civilization. The term was first used by John von Neumann in 1958, where the term “Singularity” described a hypothesis that developments in science, especially in the fields of informatics, nanotechnology, and biology will ultimately lead to the invention of an artificial general intelligence (AGI). [Kurzweil \(2005\)](#) predicts the singularity will coincide with the next epoch, “The Merger of Human Technology with Human Intelligence,” but after the singularity, the final epoch will occur, “The Universe Wakes Up.”

No one knows precisely when humans will create AGI (or superintelligence), but Kurzweil believes that the singularity will occur by approximately 2045. [Weder \(2020\)](#) argues that true intelligence requires life, which can be defined as a process of self-replication and sustains that superintelligence is either impossible or something in a very distant future. Weder maintains that true intelligence should promote but not interfere with the replication of the genes responsible for its creation, including the necessary hardware like the brain. She argues that without this constraint, there is no objective criteria for determining whether a particular solution is intelligent. Weder claims that machines will always only be surrogates of human intelligence, which unfortunately still leaves open the possibility of abuse by people controlling the AI.

Self-replication is not the only difficulty that will be encountered when creating an AI which will be more intelligent than humans. The intelligence of living beings, including animals is extraordinary because this type of intelligence can discern complex items and generate swift actions in different methods in various countless circumstances. Knowledge about how the brain works is not complete and will require perfect information about how the brain functions before AI systems can be developed to replicate human intelligence. [Weder \(2020\)](#) claims that the first signs of the emergence of superintelligence will occur when machines like Terminators come on the scene and surpass humans in acquiring energy and other resources that they need for their repairs and replications.

One needs to recognise that human intelligence is connected directly to the human brain, while an AI system uses algorithms to imitate and may even exceed human performance in some tasks, such as computations. Generally, human intelligence has natural cognitive abilities such as feeling, imagination, creativity, intuition, empathy, and reflexes, amongst others. These attributes are extremely problematic for computers to mimic fully or replicate accurately. However, today's AI systems do not need these natural abilities because the functions they perform are reliant on data processing power to store and retrieve information and execute rule-based software at a rapid and accurate rate that is much higher than humans can achieve.

Because of this processing power, AI systems learn new tasks swiftly by trial and error, and do not require instruction guides or advice from humans.

AI systems are better at recognising complex patterns or shapes through structural models like multi-dimensional entities. AI algorithms iteratively improve their comprehension of patterns by altering their internal constraints based on feedback. Hence, by comparing their extrapolations with known results, the algorithms improve their ability to identify subtle patterns. These techniques are very useful to improve overall quality of medical images in MRI and CT scans, improving diagnostic accuracy and contributing to better patient care. A shortcoming of an AI system is that it lacks the flexibility of human intelligence when it needs to think outside the box or improvising when caught unawares, which are very useful in emergency situations. While AI systems are improving at recognising emotions in photographs and videos, they still do not have any emotions themselves. This may raise ethical issues in the future, particularly given the fact that an AI system may at times be inaccurate or susceptible to simple deceptions. Moreover, since an AI system depends on human inputted data for training and feedback loops during its learning progression, it may be liable to unintended human influence, with the system not functioning as anticipated due to human biases and errors. However, this type of risk is also present in non-AI systems. There are many other differences between human intelligence and AI, which will be discussed in the next section.

4.5 SPECIFIC DIFFERENCES BETWEEN AI AND HUMAN INTELLIGENCE

Although there are some similarities between AI and human intelligence, [Table 4.1](#) illustrates, that there are also many specific differences between them. As has been shown previously, AI is an innovation created by human intelligence, while humans are biologically gifted with cognitive abilities, such as thinking and reasoning. Norbert Wiener who theorised on feedback mechanisms in the 1950s was a pioneer in the field of cybernetics, which is

recognised as the emergent scientific field of AI. Wiener is reported to have argued that modern “thinking” machines are much faster at working out certain problems than their human designers and are being constructed to think like humans. However, it was John McCarthy who used the term “artificial intelligence” for the first time at the Dartmouth Conference that specifically dealt with the theme of AI as the main topic.

TABLE 4.1 Specific Differences between AI and Human Intelligence [↗](#)

Features	AI	Human Intelligence
Emergence	Norbert Weiner who theorized on criticism mechanisms.	Humans are naturally gifted with cognitive abilities, such as thinking and reasoning.
State	Machines are digital that use binary code.	Brain is analogous, processing data in a continuous and non-linear fashion.
Essence	The aim of AI is to mimic humans and do the work that humans would normally do.	To combine a range of cognitive actions to adapt to new situations.
Physical Limitations	No physical limitations and can operate on a 24/7 basis.	Limited by physical capabilities and requires rest and maintenance
Energy use	Generally, only use 2 watts.	The human brain uses about 25 watts.
Processing	Based on algorithms and mathematical models.	Based on cognitive processes and biological structures.
Function	Rely on input of data and instructions.	Use of memory, processing capabilities, and cognitive talents.

Features	AI	Human Intelligence
Accuracy	Produces accurate results as it is programmed to optimise a given task or system based on rules or algorithms.	Room for “human error” but it has the capacity for creativity and innovation beyond the limitations of programming.
Adaptability	Can adapt to new data and situations but it takes time to adapt to new changes.	Can adapt to new states, learning from experience (decisions based on context).
Flexibility	Can perform fewer tasks at the same time and can learn duties one at a time.	Has multitasking ability and perform differing and concurrent roles.
Emotions	AI lacks emotions and empathy	Capable of feeling emotions & empathy.
Learning ability	Based on data and feedback loops and acquires knowledge by machine learning.	Human intelligence is based on experience, intuition, and creativity.
Speed	Able to process vast amounts of data and perform complex calculations rapidly.	Humans are much slower at processing large quantity of data.
Creativity	AI has limited ability to be creative or think outside of the box	Humans are capable of creativity, imagination, and innovation.
Decision making	Extremely objective because its analysis is based on factual accumulated data.	Decisions may be affected by subjective elements that are not figures alone.

Features	AI	Human Intelligence
Social Networking	AI has difficulty understanding the nuances of social and emotional cues, which are essential for effective human interaction.	Ability to process complex social and emotional information, interpret nonverbal communication, and respond appropriately to social situations.
Ethics	AI does not have a moral code or conscience	Human Intelligence has a moral code and conscience that guides decision-making

The human brain processes data in a continuous and non-linear fashion, whilst computers, which are the basis for AI are digital that use binary code. The primary aim of AI is to mimic humans, where they attempt to perform the work that is normally done by humans. In contrast humans combine a range of cognitive activities to adjust to new situations. This is a significant difference between AI and human intelligence. AI systems are designed to execute specific tasks and are optimised for them. While these systems can learn to perform new tasks, they still require a considerable amount of training and data to do this. In contrast, human intelligence is much more adaptable, with humans being able to learn quickly and adjust to new circumstances.

A strong advantage of AI is that it does not have physical limitations in terms of fatigue and can operate on a 24/7 basis. AI is no match for humans in terms of physical activity that uses bodily movement produced by skeletal muscles that requires energy expenditure. However, the human is limited in that it requires rest and maintenance. There is also a difference in energy consumption. While a computer uses about 2 watts of energy, the human brain uses about 25 watts. AI and human intelligence also differ in the way they process data. AI's processing is based on the use of human created algorithms and mathematical models. Whereas human intelligence is based on cognitive processes and biological structures. This difference

means that AI relies on input data and programming instructions, whereas human intelligence applies memory, processing capabilities, and cognitive talents.

This also means that AI and human intelligence are different in their creativity and intuitive capabilities. AI systems can furnish new ideas and solutions based on their specific training data and algorithms, but unlike humans their capacity to be creative and intuitive is lacking. Humans are creative and can discover innovative ideas and solutions that are not based on any existing processes or experiences. Humans can apply their sensitivity and experience to make associations and solve problems that AI systems would have difficulty with. This also has an impact on how AI and humans react. AI can adapt to new data and circumstances, but it takes time to adjust to new fluctuating situations. Whereas humans can adjust to new completely new states, learning from experience, and making decisions based on context.

An AI system can provide accurate results according to how it is programmed to optimise a given function based on rules. However, humans have the capacity for creativity and innovation beyond the limitations of programming but are prone to “human error.” This feature also means that humans tend to be more flexible and are able to perform in multitasking mode by executing differing and concurrent roles. On the other hand, AI can perform fewer tasks simultaneously and tend to learn tasks one at a time. One of the key differences between AI and human intelligence is their learning method. An AI’s learning approach is based on data and feedback loops and assimilates knowledge by machine learning. However, human intelligence is based on experience, intuition, and creativity. Therefore, AI systems apply pre-programmed algorithms and input data to learn from their circumstances and improve their performance over time. This is known as machine learning and involves training the system with large amounts of data to identify patterns and make extrapolations. On the other hand, human intelligence is a result of a complex interchange of genes, surroundings, and experience. Humans have an innate cognitive ability that may vary from one individual to another. Thus, humans develop and refine

their intelligence through exposure to different situations, experiences, education, and social interactions.

An important difference between AI and human intelligence is their capacity to comprehend language and context. Humans can understand the meaning behind words, double meanings, and phrases, as well as the social and cultural context in which they are applied. On the other hand, AI systems are not very proficient in natural language processing, since they still have difficulty comprehending the nuances of the human language and the context in which it is applied. However, once this difficulty is overcome, an AI system will have unlimited ability in this field in all languages concurrently. A sole individual may master several languages, but an AI system will have no such restriction. Furthermore, in terms of social networking, an AI system has difficulty comprehending the nuances of social and emotional cues, which are important for effective human interaction. On the other hand, humans have the capacity to process complex social and emotional information, interpret nonverbal communication, and respond fittingly to social situations. This also means that humans are able to make decisions that may be influenced by subjective factors and not just by facts and figures. However, an AI system is extremely objective because its analysis is founded on factual accumulated data.

It must be noted that currently, an AI system has very limited capacity to be creative or to think outside of the box, contrary to humans, who are very capable of imagination, creativity, and innovation. This is offset to some extent by the AI system's ability to process vast amounts of data and perform complex calculations rapidly, whereas humans are much slower at processing a large quantity of data. However, the greatest difference between AI and humans is the emotional factor. AI systems can identify and respond to emotions in humans, but they lack emotions and empathy, while humans are capable of both. This also means that an AI system does not have ethical behaviour through a moral code or conscience, whilst most humans do. Emotional intelligence is a vital feature of human intelligence since it permits humans to identify with others, recognise their emotions,

and encourage social relationships. Hence, humans are capable of a complex range of emotions that impact their decision-making and social interactions.

4.6 HARMONISATION BETWEEN HUMAN INTELLIGENCE AND AI

The human brain is the only source of intelligence that humans possess, which also controls their feelings, thoughts, memories, and actions throughout the human life span. On the other hand, an AI system depends on a computer (that can be replaced and upgraded) in the effort to think and learn to resemble a human. The human brain has the capacity for rational deliberation and decision-making, while an AI system can think and learn like humans according to a suite of executable programmed algorithms. However, an AI system can be viewed as a tool to be exploited by humans to help them increase their living standards. For example, AI systems are being used in healthcare to produce images of internal organs and diagnose diseases and/or suggest treatments; it is being applied in finance to forecast share prices on the stock market and execute econometric models for defining better economic policies, and a multitude of other important applications.

While the human brain is a very powerful organ, it is not capable of doing the computations at the speed that AI systems can do. However, one must recognise that it is the power of the human brain that invented these applications and many others. These are typical examples where humans and other living beings benefit from the collaboration between the human brain and AI systems, including their underlying equipment controlled by sophisticated computers, and operated by humans. An important issue regarding AI is how far should humans engage in its development. For instance, there is work currently being conducted regarding the integration of human AI. This refers to the integration of AI into the human brain and systems with the primary aim of making them smarter. This integration allows humans and an AI system to learn from each other much more

quickly to ensure that humans and the AI system constantly move forward as fast as possible. However, human AI will likely have its limitations, particularly in areas such as reasoning, social interaction, emotions, and common sense.

There are potential issues with human intelligence and AI but only if the AI application is abused by human operatives. Furthermore, ethical concerns are likely to arise regarding how an AI application will impact society in the future. For example, AI surveillance systems gather immense volumes of data, including personal information, without an individual's explicit consent. Additionally, data violations involving facial recognition data amplify the possibility for identity theft, stalking, and harassment because, unlike passwords and credit card information, faces cannot easily be changed. However, human centred AI, which focuses on the interactions between humans and AI systems attempts to interface with its users (humans) in a way that is natural for them, permitting individuals to use an AI that is more spontaneous, supportive, and satisfying. Nevertheless, several ethical concerns need to be considered when developing and using AI, including the possible risks presented by AI that negatively impact society, and have implications for human rights. These ethical considerations will be treated in detail in a separate chapter.

4.7 CONCLUSION

Currently there is little consensus on the precise definition of AI but there is common agreement that AI has some ability to think and mimic humans in certain aspects. It must be recognised that even though AI is a human creation, the terms AI and human intelligence are two distinct processes with their own distinctive features. Furthermore, while an AI system and human intelligence have some similarities, there are significant differences between the two that have been summarised in [Table 4.1](#). AI systems can perform many tasks quickly and efficiently that normally require human intelligence, but their biggest weakness is the lack of adaptability, imagination, creativity, intuition, and emotional intelligence that most

humans possess. Research has revealed that human intelligence is a complex interaction of genes, experience, and environment. This permits humans to comprehend, define and recognise concepts, and interact with the world around them in approaches that AI systems are not able to imitate. It is important to recognise that as AI continues to develop and grow in acceptance, one needs to appreciate the differences and comprehend the unique strengths and shortcomings of both AI and human intelligence.

As highlighted in the previous chapter, the word “robot,” which was first used in a 1920 in a play “R.U.R.” by Karel Čapek to describe an artificial automaton, derives from the Slavic word *robata*, which means “forced labour.” Often humans still think of (ro)bots and androids as monotonous, obedient slaves. It is important to recognise that as technology improves in processing capacity and AI continues to improve, humans must be able to maintain control over the AI systems and not simply think that an AI is mindless.

Often, humans feel themselves as being at a far superior evolutionary level in the biological creation order and view other species at a much lower level of realisation to be in complete control and to subject other living forms a lower status and few rights. Humans need to be conscious that all living things have feelings and sense pain and suffering. Furthermore, humans must not be complacent about artificial creations. This complacency stems from the fact that humans do not recognise the potency of AI. Humans often feel very frustrated whenever they have an experience of interacting with a dull online customer service bot or an automated phone service, reaching the general conclusion that AI is not particularly smart. Contact with a human is often more successful in getting a better customer service, where one can explain the situation without feeling that they are communicating with a non-existent entity.

In conclusion, AI and human intelligence are both extraordinary achievements of engineering and biological development, respectively. However, there are many critical differences between them that makes them very distinctive. Human intelligence is centred on the chemical and

electrical activity of the brain, while AI is dependent on computer systems. Furthermore, human intelligence is very well suited for tasks such as pattern recognition and natural language processing, while AI is more suited for tasks such as data mining and machine learning. It is unlikely that devices capable of thinking like humans will be created in the very short-term because technologists still do not fully understand how the human brain fully functions. Therefore, currently human intelligence has full control of how AI develops but this does not mean that this situation will not change. It is too early to know whether AI will eventually pose a serious risk to existence of humanity. Some experts believe this is a possibility due to the anticipate considerable growth of AI in the next decade or so. However, if legislative action is undertaken on a national and international level, and continuous and considerable investment in security systems are embarked on, then it is not likely that AI will pose a serious threat to humanity and any threatening risk is mitigated.

REFERENCES

- Ambrose, Susan A., Bridges, Michael W., DiPietro, Michele, Lovett, Marsha C., and Norman, Marie K. (2010). *How Learning Works: Seven Research-Based Principles for Smart Teaching*. John Wiley.[📄](#)
- Camilleri, Emanuel. (2019). *Information Systems Strategic Planning for Public Service Delivery in the Digital Era*. IGI Global.
- Kurzweil, Ray. (2005). *The Singularity Is Near: When Humans Transcend Biology*. Viking, Penguin Group.[📄](#)
- Lacewing, Michael. (2009). *Practical Wisdom*. Routledge; Taylor and Francis Group.[📄](#)
- Lee, Daeyeol. (2020). *Birth of Intelligence: From RNA to Artificial Intelligence*. Oxford University Press.[📄](#)
- Malone, Samuel, A. (2003). *Learning about Learning: An A-Z of Training and Development Tools and Techniques*. Chartered Institute of Personnel Development.[📄](#)

- Medeiros, João (2017). *Stephen Hawking: 'I Fear AI May Replace Humans Altogether'*. WIRED. www.wired.co.uk/article/stephen-hawking-interview-alien-life-climate-change-donald-trump?trk=article-ssr-frontend-pulse_little-text-block.↵
- Mehta, Kumar. (2021). *A Harvard Psychologist Says Humans Have 8 Types of Intelligence. Which Ones Do You Score the Highest In?* CNBC Make It. www.cnbc.com/2021/03/10/harvard-psychologist-types-of-intelligence-where-do-you-score-highest-in.html.↵
- Sternberg, Robert J. (2023). Human intelligence. *Encyclopedia Britannica*, 18 Nov. 2023. www.britannica.com/science/human-intelligence-psychology.↵
- Weder, Annika. (2020). *Q&A – What Is Intelligence? Fundamentals*. Johns Hopkins University. www.hopkinsmedicine.org/news/articles/2020/10/qa-what-is-intelligence#:~:text=Intelligence%20can%20be%20defined%20as,for%20their%20survival%20and%20reproduction.↵

II

The Technology Supporting Artificial Intelligence

Artificial intelligence (AI) like many other software-based applications requires robust technological support in terms of effective hardware and software tools that operate in a secure environment. [Part II](#) of this book consists of three chapters, with each chapter focusing on a distinct topic, namely AI associated hardware, software and security.

There is no doubt that most of the current AI applications execute on the traditional Von Neumann architecture. Unfortunately, this architecture executes instruction in a sequential manner that limits the capabilities of AI applications. It must be recognised that hardware is the fundamental pillar of all computer-driven functionalities and acts as a bottleneck to development. Hence, technological advancement is very much dependent on the capacity, capabilities, and sophistication of the hardware. [Chapter 5](#) is devoted to computer hardware focusing on fourth and fifth generation computers and tracing their development over the past six decades. The chapter discusses Quantum computing that aims to resolve the Von Neumann bottleneck by applying quantum states of subatomic particles to store information and execute multidimensional quantum algorithms.

Additionally, [Chapter 5](#) examines a completely different approach to computing, namely neuromorphic hardware that aims to emulate the human nervous system allowing parallel processing for specific tasks. Neuromorphic chips have extraordinary potential that are being exploited in

robotics; visual object recognition; gesture and smell recognition; medical monitoring; self-driving vehicles to identify and interpret their environments; and identification of diseases, amongst other applications. Neuromorphic chips have futuristic innovation features that when fully developed will revolutionise the way we view the capabilities of robotics. This chapter also examines current developments related to hardware chips that focus on specific dedicated tasks, which can be integrated as part of application systems that may include neuromorphic computing, such as Edge Computing Chips, Application Specific Integrated Circuits, Field Programmable Gate Arrays, and Complex Programmable Logic Devices.

It is an axiom that hardware on its own has no value unless it is supplemented by suitable software. Hence, [Chapter 6](#) is dedicated to software as an essential component of effective hardware and explores several AI programming languages that are crucial for developing AI applications. Moreover, this chapter examines IBM's Qiskit SDK, which is designed for the IBM Quantum Platform to execute specialised AI applications. However, the discussion on software would not be complete unless hardware description languages are also considered. Hence, Verilog and VHDL are examined since these are particularly useful for specific hardware related to AI, such as Field Programmable Gate Arrays. Furthermore, AI algorithms are discussed because of their importance in relation to executing AI tasks that would otherwise require human intelligence, such as speech and voice recognition; text recognition; image recognition; video analytics; generative AI; and natural language processing. Finally, the chapter focuses on neuromorphic computing and spiking neural networks. These are explicitly designed for parallel processing and have tremendous potential for AI applications since they are expected to be able to adjust to their environment, accessing external sources through the cloud if necessary to seek more information when required.

AI has great potential for humanity, but it has also a devastating ability to be destructive unless it is developed and executed in a safe environment. Hence, the closing chapter for [Part II](#) is dedicated to security. It is a fact that

the advancement of AI is attracting the attention of very clever threat actors that are continuously seeking ways to exploit weaknesses in cybersecurity. This chapter evaluates the evolving threat landscape and how the risk of cybersecurity penetration may be mitigated by implementing powerful intrusion prevention systems (IPS) that are located directly behind the firewall. An IPS offers an extra layer of protection by having powerful algorithms that continuously check network traffic to detect threats and take action to neutralise them.

This is not to say that a firewall is not essential. In fact, Next-generation firewalls (NGFW) act as the vanguard for cybersecurity by employing application-specific rules, protect against malware, block anomalous behaviour and prevent actors from entering the network through a nonstandard port. The aim is to keep ahead of threat actors. While IPS and NGFW protect the network, a public key infrastructure (PKI) defends data that is being transmitted by creating digital certificates that cryptographically connect ownership of a public key with the entity that owns it, with the specific aim of sharing these keys for data encryption and authentication. Finally, this chapter examines centres for AI safety that are nonprofit research organisations, which focus on the mitigation of AI risks and are continuously promoting innovative ways to continually come up with solutions because the war against threat actors will not cease.

Artificial Intelligence Hardware

DOI: [10.1201/9781003653950-7](https://doi.org/10.1201/9781003653950-7)

“There will always be rocks in the road ahead of us. They will be stumbling blocks or stepping stones; it all depends on how you use them.”

Friedrich Nietzsche

German philosopher (1844-1900)

AI systems are very much dependent on the processing power of the computers that are used to execute the various AI programmes. In the past, traditional computer hardware has been the greatest limitation to the development of AI. AI hardware refers to specialised computer hardware that is designed to facilitate and assist in the execution of AI related tasks effectively. Hence, AI hardware specifically supports the required infrastructure to execute algorithms and other essential models efficiently to optimise the hardware’s processing power. AI hardware includes a variety of items, such as specifically designed computer microprocessor chips. This type of hardware has the primary goal of executing AI programme suites faster and use less energy in the process. The AI hardware may support Lisp machines that are general-purpose computers designed to efficiently execute Lisp as their main software and programming language. They also support neuromorphic computing, physical neural networks, and event cameras that apply imaging sensors that responds to local changes in brightness.

Neuromorphic computing (or neuromorphic engineering) refers to the design of both hardware and software computing elements that have the objective of modelling bio-inspired computer systems that imitate the human brain and nervous system. An example of neuromorphic engineering is the self-driving car. A self-driving car needs to make prompt decisions to correctly navigate and avoid collisions. This type of technology requires considerable computing processing power. Therefore, by using neuromorphic hardware and software, self-driving cars can execute tasks quicker and consume lower energy in contrast to traditional computing.

On the other hand, a physical neural network entails the representation in a tangible form of artificial neural networks into physical systems, enabling real-time computation and interaction. In contrast to virtual neural networks, physical neural networks provide the possibility to link the gap between the digital and physical worlds. Whereas, traditional neural networks occur only in the digital domain, functioning via mathematical models and algorithms, physical neural networks integrate AI with tangible objects and environments, allowing real-time computation, interaction, and sensory feedback between the digital and physical domains. This chapter will discuss these hardware developments and others in detail to provide an insight into the potential that AI may provide through its applicability for the benefit of humanity.

5.1 INTRODUCTION

History has shown that AI applications have made considerable advances when more powerful computers have been developed and released in the commercial marketplace. As observed in a previous chapter, the term “Artificial Intelligence” was first formally applied at the Dartmouth Summer Research Project on Artificial

Intelligence in 1956. However, the conceptual models for AI were established in 1943 when American neurophysiologist and cybernetician Warren McCulloch and logician and cognitive psychologist Walter Pitts, published a paper “A Logical Calculus of the ideas Imminent in Nervous Activity.” This was the first mathematical model of a neural network. Their ideas were applied by John von Neumann (mathematical formulation of quantum mechanics and Game theory) and Norbert Wiener (the initiator of cybernetics, and the science of communication as it relates to living things and machines), which are viewed as being fundamental to some key concepts of AI.

McCulloch and Pitts are viewed as the pioneers of the artificial neural network and were a fundamental inspiration for the invention of the modern digital computer. It was also the era for the first-generation computers (1940-1956), at a time when in 1943 the Colossus computer became famous for its code-breaking capabilities, during the Second World War. John McCarthy and his colleagues instigated the Dartmouth Summer Research Project on AI in 1956. This came at the early beginning of the second generation of computers (1956-1963), which at that time were viewed as the most powerful computers, unparalleled with what had been discovered at that juncture. These so-called powerful computers provided the inspiration for McCarthy and his colleagues to push through their AI initiatives.

Nevertheless, despite the increase in the computing power that the second generation of computers provided, the computing processing unit and the external storage memory were very limited, and therefore, the processing power was restrained. This computing limitation and the substantial criticism of the Rosenblatt’s perceptron model are likely to have resulted to the termination of the machine translation project in 1966 due to poor AI outcomes, leading to the first “AI winter,” which denoted a period of reduced funding and interest in AI research. The second-generation computers moved away from the use of bulky vacuum tubes and applied transistor technology instead, and enhanced core storage. Transistors provided the computer with more computing power, including computational speed. This development generally resulted in the reduction of the size and cost of computers, including less heat generation when compared with vacuum tubes. Additionally, the second-generation computer brought with it the concept of the Central Processing Unit (CPU), enhanced internal and external data storage memory, programming languages, and input and output devices. Programming languages, such as FORTRAN, ALGOL, and COBOL simplified programming when compared with machine and assembler programme coding.

The third-generation computers (1964-1971) predicted (and realised) a shift from sizable transistors to integrated circuits (IC), made from silicon. An integrated circuit, also referred to as a chip, microchip, or microelectronic circuit, is a semiconductor wafer on which thousands or even millions of tiny resistors, capacitors, diodes and transistors are engineered and manufactured. In other words, an IC is an integrated system of multiple miniaturized and interconnected components embedded into a thin substrate of semiconductor material (typically a silicon crystal). It has proven to be a fundamental element of all contemporary electronic devices. Hence, its development provided the third-generation computer with enhanced speed and reliability. This era also saw the birth of the minicomputer as distinct from the mainframe computer.

The introduction of the minicomputer meant that small-to-medium firms could afford their own computer facilities due to lower hardware costs. The difficulty during this period was a scarcity of applications software that permitted organisations to use a computer to execute their administrative functions, such as accounting, payroll, stock control, amongst many others. These types of computers had an operating system that was designed to handle completely the inner operation of the computer itself. Additionally, it also had applications computer programmes (written in computer languages, such as COBOL and FORTRAN, amongst others) that were designed to perform a specific organisational function, such as executing a suite of programmes related to the accounting, stock control, and payroll. The operating system permitted the third-generation computer to execute multiple jobs simultaneously by allocating each application system fractions of a second of processing time sequentially. The high speed of these computers gave the impression that because of this time-allocation process, these multiple application systems were being executed in parallel. In fact, these computers were not designed with a parallel processing feature.

The release of the third-generation computer coincided to some extent with the first AI winter, with most research being directed towards traditional application software development. In the late 1960s, the development

of perceptrons and connectionism was contemporary and stimulating. Unfortunately, most of the research development funding for AI came from the United States Defence Advanced Research Projects Agency (DARPA). However, the Mansfield Amendment approved by the United State Congress obliged DARPA to fund research directed by a specific mission. This triggered DARPA to terminate funding of new AI research. Moreover, in 1973, the British Parliament commissioned James Lighthill to review the state of AI in the United Kingdom. The Lighthill report was very critical of the achievement attained by AI and declared that the breakthroughs did not fulfil the desired expectations and that the study of AI could be conducted in other sciences. AI was comparatively new in the 1960s and 1970s with high expectations that it would be epoch-making development. However, the termination of funding from DARPA, the damaging Lighthill report, together with unfulfilled high expectations, resulted in the unsustainable development of AI.

5.2 FOURTH-GENERATION COMPUTERS (1971-PRESENT): NEW PROCESSING CAPABILITIES

Fourth-generation computers were based on the development of the first microprocessors that exploited the IC. However, during this era, the large-scale of integration circuits (LSI) were also introduced. Genuine LSI circuits had about 10,000 transistors and came on the computing scene around 1974 and utilised for computer main memories and other microprocessors. The advantage of this technology enabled hardware manufacturers to group thousands of transistors and other related components on a single silicon chip about the size of a baby's fingernail to perform arithmetic, logic, and control functions. The computers using microchips were called microcomputers that provided ever smaller size computers, with larger capacities. In the late 1970s, the very large-scale integration (VLSI) was introduced that integrated hundreds of thousands of transistors on a single silicon semiconductor microchip. Technologies such as multiprocessing, multiprogramming, time-sharing, operating speed, and virtual memory made it a speedier and a more user-friendly device. The fourth-generation computers allowed the development of personal computers that gave birth to computer networks. This era ended the 1943 prediction of Thomas Watson, chairman of IBM that the world market for computers was around five. On the other hand, in 1975, Bill Gates and Paul Allen co-founders of Microsoft had a vision of a computer on every desk and in every home. A prediction that has come true over the past few decades.

The fourth-generation computers brought mixed fortunes for AI. The early years of its development overlapped with the AI winter of 1974-1980, which encountered a period that lacked AI research funding. However, even though AI research funding was virtually non-existent, private research through academic and industrial sources continued at a slower pace. Hence, significant results were still achieved due to these more powerful and faster computers that overcame some of the difficulties encountered by AI, which had up to that time gave disappointing outcomes. By 1974, these more powerful computers enabled Paul Werbos to initiate the backpropagation supervised learning algorithm for training neural networks. Additionally, by the 1980s the Lisp programming language was commonly used for AI research, paving the way for the acceptance of expert knowledge-based AI systems. However, in the 1990s AI was sidelined once again because these expert systems failed to meet (once again) the high expectations.

Nonetheless, despite this downturn, AI research witnessed the growth and acceptance of support vector machines (SVM) with supervised learning and the decline of neural networks. An SVM is mostly used in classification problems, such as text classification, image classification, spam detection, handwriting identification, gene expression analysis, face detection, and anomaly detection. By the 2000s these more powerful computers enabled new machine learning algorithms begin to solve a wide range of problems. However, one must recognise that even though fourth-generation computers are very fast and powerful, they are not specifically designed for AI systems.

Commonly, these computers can be viewed as general purpose computers that enable the execution of the current AI systems, such as deep neural networks that have since attained advanced results, including image recognition, natural language processing, and machine translation. More recently, Chat Bots and GPT models have made enormous headways in the industry. In other words, AI is an ancillary objective of fourth-generation

computers. As such, these computers are a major limiting factor to the development of AI. This means that General AI, which is designed to perform any intellectual task that a human can do, such as being able to reason, learn, and understand complex concepts; and Super AI that exceeds human intelligence in all areas and would be able to solve complex problems that are beyond human abilities and would be able to learn and adapt at a rate that far exceeds human intelligence, are not currently possible using this type of technology. Furthermore, far more advanced computers are required for Theory of Mind AI types that interact with the thoughts and emotions of humans and for having self-awareness AI that has an independent intelligence and will not be under the control of the people that created it.

5.3 FIFTH-GENERATION COMPUTERS (PRESENT AND BEYOND): NEW ERA FOR AI PROCESSING

The fourth-generation computer has relegated AI as an ancillary application. This has restricted AI's development to Narrow AI (or Weak AI), which is designed to do a specific task or a limited range of tasks, such as facial, speech, and image recognition; natural language processing; and recommendation systems. On the other hand, fifth-generation computers and other related technology are specifically designed with AI as their primary objective. The fifth-generation computers provide new opportunities for AI to edge towards General AI and Super AI that will enable these systems to possibly exceed human intelligence and interact with the thoughts and emotions of humans and thus achieve self-awareness.

In 1982, Japan's Ministry of International Trade and Industry launched the Fifth-Generation Computer Project with the objective to develop a new generation of computers to understand and use human language. This generation is based on parallel processing hardware and AI software, which exploit super large scale integrated (SLSI) chips that contains millions of components on an individual chip. Parallel processing is much faster than the multiprogramming serial processing operating systems used by fourth-generation computers. Apart from using VLSI and later SLSI architecture, the key technologies applied for the Fifth-Generation Computer System include, better data flow control, logic programming, knowledge basis based on relational database and applied artificial intelligence and pattern processing ([Moto-oka, 1983](#)). The Japanese initiative was meant for a new generation of computers that were envisaged for the 1990s.

According to [Moto-oka \(1983\)](#), the research and development period set for this project was ten years, divided into three phases as shown in [Figure 5.1](#). The first phase was related to the initial development of basic computer technology, which comprised of three basic modules, namely modules of individual functional mechanisms for the inference subsystem using parallel mechanism; modules of individual functional mechanisms for knowledge base subsystem based on knowledge-based technology; and basic software system with problem-solving, knowledge base, and intelligent interface capabilities utilising VLSI technology.

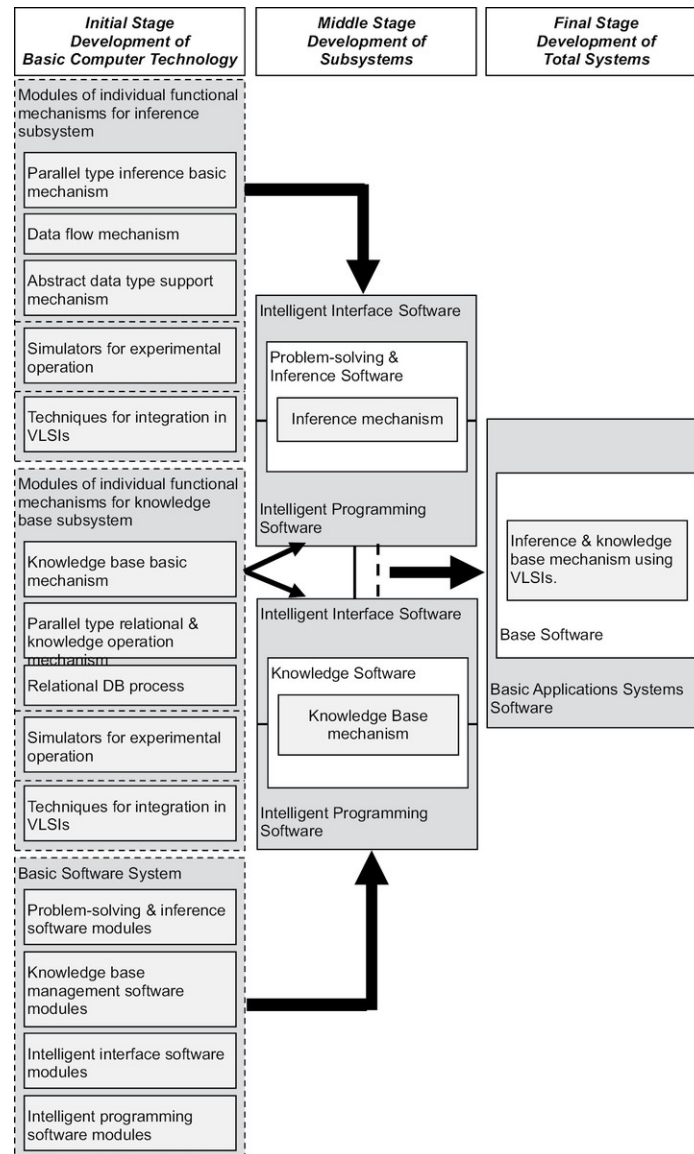


FIGURE 5.1 Fifth-Generation Computer Project Development Phases. [↗](#)

Moto-oka also explained the conceptual details of the project as shown in [Figure 5.2](#). He clarified that database machines and present mainframe computers will be networked to organise worldwide information systems. [Figure 5.2](#) illustrates that the fifth-generation computer was based on three aspects, namely the human applications subsystems, modelling software subsystems, and the machine hardware subsystem, with each aspect having a two-way communication interface. The access to the system is through the human applications subsystem utilising user language (speech, natural language, pictures, and images), which acts as the external intelligent interface of basic software system, communicating with the modelling software subsystems via the intelligent programming system and the knowledge-based system.

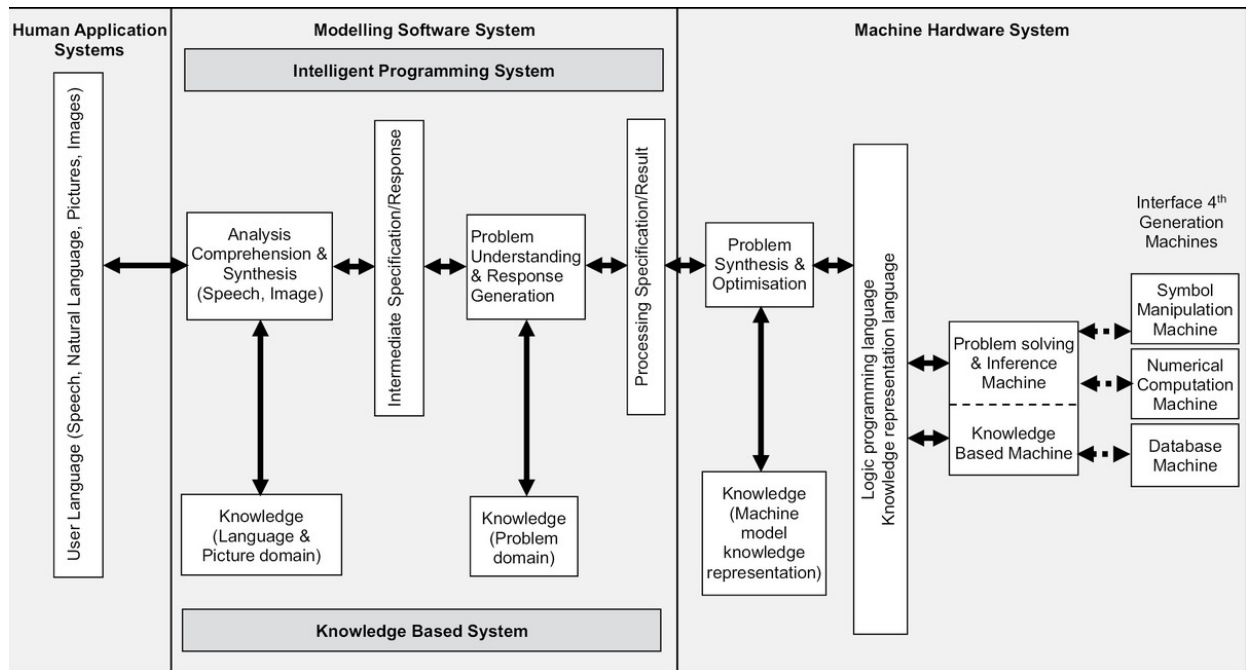


FIGURE 5.2 Fifth-Generation Computer Project Conceptual Details.

The modelling software system comprises of two key elements, namely, knowledge-based management system and the problem-solving and inference system, with the former communicating with the latter. The modelling software system communicates with the hardware system through three levels, namely the knowledge base machine, problem-solving, and inference machine, and the intelligent interface machine. According to Moto-oka the emphasis in computer systems is envisaged to shift towards the human system by significant enhancement of the logic level of the hardware system and by the positioning of the modelling system between the hardware and man. He explains that the interface between the software and hardware systems would be the kernel language. The entire software system would be realised in the kernel language, and the hardware system would directly execute the kernel language.

However, [Swinbanks and Anderson \(1992\)](#) reported in March 1992 that the Japanese project was in difficulties. They report that after ten years and an investment of more than \$400 million, all Japan had produced was a handful of parallel computers, the largest three of which have only a quarter of the target number of processors. By the end of that year, [Reid \(1992\)](#) reported that the Japanese government formally terminated the “Fifth Generation” computer project, after a decade-long research effort that was supposed to create a new world of computing power but turned out to have little impact on the global computer market. However, the Japanese Fifth-Generation Computer project provided an important and forceful impetus for computer engineers and scientists in Silicon Valley to enter a competitive race to produce their own version of the Fifth-Generation Computer. Therefore, despite its failure, the Japanese fifth-generation project had flung the gauntlet to the rest of the computer world to come up with a functional fifth-generation computer that has at its basis the AI objective. The key features of the fifth-generation computer include:

- Exploiting ULSI (Ultra Large-Scale Integration) technology in computer design.
- Expansion in superconductor technology and parallel processing.
- Multi-processor-based systems, with parallel computing.
- Having magnetic-enabled chips that stores information, encrypts transactions, and facilitates transaction processing.
- Use of optical fibre in circuits.
- User-friendly interfaces with multimedia features.

- Development of natural language processing (NLP).
- Massive development in storage capacity, with faster and larger storage capabilities.
- Interfacing powerful computers with AI.
- Use of AI, such as navigating, browsing, video analysing, and image processing, amongst others.

5.3.1 The Quantum Computer

According to [Nasrin \(2023\)](#), current fifth generation of computers are being created with an emphasis on quantum computing and artificial intelligence (AI). She contends that the move from computations to communication will be among the primary changes brought about by these fifth generation of computers. Nasrin believes that these computers will be able to link to one another and share information in ways that earlier generations were unable to do, making it possible to solve problems more effectively and collaboratively, resulting in the creation of fresh and original solutions.

Furthermore, [IDQ \(2024\)](#) assert that in October 2023, the Californian quantum start-up Atom Computing announced it had created the world's first quantum computer with over 1,000 qubits (1,180). They announced that this signified a substantial advancement on the previous record holder, IBM's Osprey with 433 qubits. However, this declaration conflicted with the IBM announcement of its own 1,121 qubit Condor computer model. IBM claims that Condor pushes the limits of scale and yield in chip design with a 50% increase in qubit density, advances in qubit fabrication and laminate size, and includes over a mile of high-density cryogenic flex IO wiring within a single dilution refrigerator. They maintain that Condor's performance is comparable to the 433-qubit Osprey but solves scale and paves the way for future hardware design.

A qubit is like a binary bit, which is the basic unit of information in traditional computing. A qubit (or quantum bit) is the basic unit of information in quantum computing. IBM currently uses a blend of aluminium and niobium, the two most used materials for a qubit. A superconducting qubit is typically a tiny loop or line of metal that behaves like an atom, which is an inherently quantum object.

In the meantime, in an issue of The Telegraph on 25 November 2023, Joe Pinkstone, Science Correspondent reports that the British government was to invest £2.5 billion in quantum technologies over the next ten years. UK Science Minister, Andrew Griffith, outlined a ten-year goal that would put quantum technology at the forefront of a plan to commercialise science. The UK Government's strategy aims to have quantum computers and networks running by 2035, and it is also anticipated that the technology will provide the NHS with new-era tests by 2030. The Minister claims that the Quantum can break current encryption and create incorruptible communication networks for defence purposes, while quantum clocks and sensors can revolutionise everyday applications like Sat Navs, mobile phones, and healthcare screening.

5.3.2 What is Quantum Computing?

Quantum computing is a multidisciplinary topic encompassing features of computer science, physics, and mathematics that applies quantum mechanics to solve complex problems faster than on classical computers. The discipline of quantum computing includes hardware research and application development. One must also make a distinction between AI and quantum computing. AI is a software routine or process, whereas a Quantum Computer is the hardware that executes the AI software. AI is like having an app (software application) on a mobile phone and the mobile phone (hardware) being the physical device that executes the app (process or application software).

Essentially, quantum computing is based on the principles of quantum theory and makes use of the quantum states of subatomic particles to store information. Quantum theory is the theoretical core of modern physics that explains the nature and behaviour of matter and energy at the atomic and subatomic level, which is commonly referred to as quantum physics and quantum mechanics. The notion at the root of quantum computing is to demolish the barriers that have limited the speed of computers by exploiting the counterintuitive physics of subatomic scales.

Researchers have found that certain features of quantum mechanics are counterintuitive and can seem paradoxical because they describe behaviour that is quite different from what one would expect. For example, hot

water can freeze faster than cold water, which may seem contrary to the normal intuition. This event is known as the Mpemba effect, which challenges the normal understanding of the physical properties of water. Scientists think that the increased evaporation rate of hot water and the formation of a thin insulating layer of ice may contribute to this counterintuitive behaviour. One of the most interesting examples of quantum mechanics predictions is that a pair of particles can be entangled. Hence, when two particles are entangled, it is impossible to describe the properties of one of them without referring to the other, since they behave as a single entity. It is not the intention of this text to prove the complex properties of quantum physics since this requires a much more detailed discussion on the topic. The author's intention is to highlight in general terms that quantum computing is revolutionary and is very different from traditional computing.

Traditional computers operate on binary bits however quantum computers operate on quantum bits called qubits. A bit (binary digit) is the smallest unit of data that traditional computers can process and store. A bit is always in one of two physical states, either on or off as in a light switch. The state of a bit is often represented by a single binary value, normally a 0 or 1. On the other hand, qubits are physically attained from minute quantum objects such as electrons and protons. Additionally, quantum computing is based on two quantum mechanical laws, namely superposition and entanglement. Furthermore, superposition does not restrict qubits to one of two possible values (1 or 0), but it can exist as a composite of both.

This feature facilitates quantum computers to function much faster than traditional computers. Entanglement is unique to quantum physics and unifies several separate particles, which ensures that what occurs to one particle will affect the other particles. When this principle is applied to quantum computers, this allows all particles to handle all their qubits together. This means that calculations are not executed in sequence but in parallel (i.e., simultaneously). These two important features (superposition and entanglement) combined with quantum computing provide the potential for an exponential speed enhancement to computing using quantum technology when compared with traditional processing.

5.3.3 How Quantum Computers Work

As stated previously, a traditional processor uses binary bits to perform its operations, whereas a quantum computer uses qubits (CUE-bits) to run multidimensional quantum algorithms. According to [IBM \(2024\)](#), Quantum algorithms take a novel and unique approach to solve complex problems, such as simulating molecular behaviour or chemical simulations, by creating multidimensional computational spaces. A Quantum computer is based on several innovative principles:

- *Superconductors*: While a desktop computer uses an air flowing device, such as a fan to keep it from overheating, the quantum processors necessitate a very cold temperature, about a hundredth of a degree above absolute zero, to retain their quantum states and avoid “decoherence” ([IBM, 2024](#)). Decoherence is the process by which the information of a quantum system is modified by the system's interaction with its environment that form a closed system, hence creating an entanglement between the system and heat bath (environment). On the other hand, quantum coherence is realised when an individual particle has wave functions that can be divided into two separate waves, thus the waves operate together in a coherent way to achieve quantum coherence. Decoherence is avoided by using super-cooled super-fluids to achieve ultra-low temperatures where certain materials exhibit an important quantum mechanical effect: electrons move through them without resistance, making them “superconductors.” According to [IBM \(2024\)](#), when electrons pass through superconductors they correspond, forming “Cooper pairs.” These pairs can conduct a charge across barriers, or insulators, through a process known as quantum tunnelling. Two superconductors placed on either side of an insulator form a Josephson junction.
- *Control*: IBM's quantum computers use the Josephson junctions as superconducting qubits ([IBM, 2024](#)). The behaviour of the qubits is controlled by firing microwave photons at them, which enable them to hold, change and read out individual units of quantum information.
- *Superposition*: On its own, a qubit is not very useful. However, by placing the quantum information a qubit holds into a state of superposition, which represents a combination of all possible configurations of the

qubit, a group of qubits in superposition can create complex, multidimensional computational spaces ([IBM, 2024](#)). This enables complex problems to be presented in new ways in these spaces.

- *Entanglement*: According to [IBM \(2024\)](#), quantum entanglement is an effect that correlates the behaviour of two separate things, hence when two qubits are entangled, changes to one qubit directly impact the other.
- *Interference*: An environment of entangled qubits placed into a state of superposition causes waves of probabilities that act as the outcomes of a measurement of the system. These waves can build on each other when many of them peak at a particular outcome or cancel each other out when peaks and troughs interact; these are both forms of interference ([IBM, 2024](#)).

[IBM \(2024\)](#) explains that a computation on a quantum computer uses the following process:

- Preparing a superposition of all possible computational states.
- A quantum circuit, prepared by the user, uses interference selectively on the components of the superposition according to an algorithm.
- Many possible outcomes are cancelled out through interference, while others are amplified.
- The amplified outcomes are the solutions to the computation.

A concern with quantum computers appears to be related to error correction or error mitigation, which has currently not been fully resolved. IBM have adopted Qiskit, an open-source, Python-based quantum software development kit (SDK). The quantum SDK is a complete quantum computing stack in simulation that includes an intuitive user interface; a compiler; and a quantum runtime environment optimised for running hybrid quantum-classical algorithms. Qiskit has modules that cover applications in finance, chemistry, optimisation, and machine learning. Python currently is at the vanguard of AI programming due to simplicity, versatility, and powerful libraries. Python is a high-level interpreted language, which allows rapid development and testing, which is a key feature in the iterative process of AI projects.

5.3.4 Neuromorphic Hardware

Neuromorphic hardware has the primary aim of emulating the human nervous system by applying dedicated computing architectures that imitate the structure (morphology) of neural networks. The idea is inspired by the way the human brain functions, where biological neurons and synapses work together. In the nervous system, a synapse permits a neuron to pass an electrical or chemical signal to another neuron or to the target effector cell. Hence, in a similar fashion, the network consisting of interconnected processors, known as neurons, allow it to have a parallel processing capability that permits the network to execute multiple tasks concurrently. The concept is based on a bottom-up approach where dedicated processing units imitate the behaviour of neurons directly in hardware, and a network of physical interconnections (bus-systems) allow a faster exchange of information.

These neuromorphic devices are not as flexible as the traditional central processing units (CPUs), however, due to their parallel processing ability they provide outstanding performance and improved energy efficiency, particularly when training and inferencing (predicting) of deep neural networks. The parallel processing ability is very important because the traditional computers that are based on the Von Neumann architecture, execute instructions sequentially, and process the data held in centralised memory. The outcome of such a process is that computing performance is restricted by the data transfer rate between the CPU and the external memory unit. This is known as the Von Neumann bottleneck.

Modern application systems are more demanding and require faster processing than non-parallel architectures cannot provide. Furthermore, the ability to parallelise computations is also limited by the access to shared memory resources. Additionally, the highly parallel structure of deep neural networks requires specific distributed memory access patterns, which are hard to map efficiently onto conventional computing technology. Hence, neuromorphic hardware addresses this problem by having AI integrated into these devices and systems. This allows neuromorphic hardware to be trained with wide-ranging datasets, making it suitable for far-reaching applications,

such as speech recognition, natural language processing, and image detection. Moreover, the neuromorphic hardware’s level of accuracy is extraordinary, since it rapidly learns from substantial volumes of data.

Neuromorphic architectures and associated neuromorphic hardware designs vary from networks with binarized weights in digital hardware that are similar to regular feedforward neural networks (NN) to analogue in-memory accelerators of convolutional neural networks and event-based spiking neural network accelerators. A convolutional neural network (CNN) is an artificial neural network that is utilised mainly for image recognition and processing because of its ability to recognize patterns in images. However, the optimal design choice depends on the specific application and its explicit requirements, such as energy efficiency, latency, flexibility, or peak performance.

The first attempt at designing neuromorphic hardware was in the 1950s with analogue circuits having been used to implement neural networks in hardware. With the analogue methodology, the real-valued quantities of the neural network model are epitomised by real-valued physical quantities like analogue voltages, currents, or charges. Operations such as multiplication and addition are realised directly by the application of the laws of physics that pertain to Ohm’s and Kirchhoff’s laws. Ohm’s law states that the current in a circuit is equal to the potential difference divided by the resistance of the circuit. While Kirchhoff’s voltage law states that the sum of the voltages around a circuit loop must equal zero, when following a consistent measurement direction of voltages for each element around the loop. Therefore, Kirchhoff’s Laws quantifies how current flows through a circuit and how voltage varies around a loop in a circuit. This means that the aggregate of the voltage steps as it goes around a circuit loop must end up back where it started. Moreover, Kirchhoff’s current law maintains that the sum of all currents into a circuit node (or a junction) must equal zero, meaning that the total current flowing into a node must equal the current flowing out of that node. This is a consequence of charge conservation.

Hence, circuit equations can be ascertained using Ohm’s Law, which presents the relationship between voltage and current in a resistor ($V=IR$), and Kirchhoff’s Current and Voltage Laws, which regulate the currents entering and exiting a circuit node and the sum of voltages around a circuit loop, respectively. This application of physical laws means that the numerous coefficients of the neural networks are either hardwired by suitable chosen resistive elements or can be programmed into novel memory cells distributed throughout the circuit, which radically diminishes the bottleneck of memory-transfer.

The development of neuromorphic chips is an important breakthrough, particularly the advances made in the 2010s in spiking neural networks (SNNs) and event-driven computing. The focus at that time was to improve efficiency and power-aware neuromorphic hardware designs. The Human Brain Project and the development of specialised neuromorphic chips have paved the way to demonstrate the potential of neuromorphic technology. In 2014, IBM developed the TrueNorth neuromorphic chip that is utilised in visual object recognition and has lower power consumption than traditional Von Neumann hardware, and in 2018, Intel developed the Loihi neuromorphic chip that has had applications in robotics as well as gesture and smell recognition. [Table 5.1](#) provides a chronological itemised details of various neuromorphic chips that have been developed in the past several years, showing the software and application areas. Neuromorphic hardware is currently rapidly evolving, with academic and commercial research projects focusing on improving the effectiveness and applicability of neuromorphic systems that are so vital for resolving complex tasks in AI. Additionally, [Table 5.2](#) provides a glossary of the terms applied for neuromorphic related technology.

TABLE 5.1 Chronological Itemised Details of Various Neuromorphic Chips[\[4\]](#)

Manufacturer	Features
Jetson Xavier NX	Designed specifically for autonomous machines, particularly for handling visual odometry, sensor fusion, localisation and mapping, obstacle detection, and path-planning algorithms used for specific robotics and edge AI applications.
AMD EPYC Embedded 3000 Series	Provide performance benefits of the “Zen” CPU architecture to bring special reliability, availability and serviceability features to networking, storage and industrial applications. Improved I/O integration, flexibility, and security capabilities for system designers to provide cost-effective embedded processing.

Manufacturer	Features
Jetson Nano	A small, powerful computer that allows multiple parallel neural networks processing for uses like image classification, speech processing, object detection, and segmentation.
ARM Cortex-M55	First processor to feature Arm Helium vector processing technology bring energy-efficient digital signal processing (DSP) and machine learning (ML) capabilities. The processor permits the integration of differentiating AI techniques, while staying within the system cost and energy constraints of battery powered IoT devices.
ARM Ethos-U55	A new class of machine learning (ML) processor, called a microNPU, designed to accelerate ML inference in area-constrained embedded and IoT devices.

TABLE 5.2 Glossary for Neuromorphic Related Technology [\[4\]](#)

Year	Chip	Manufacturer	Software	Application	Neurons	Synapses	Type	On Chip Learning
2024	Spiking Neural Processor T1	Innatera	Talamo	Smart-sensing, wearables, IoT	~500	~60K	Mixed-signal	
2023	Akidap	BrainChip	MetaTF	Smart sensing, one-shot learning	Configurable	Configurable	Digital	True
2022	BrainScaleS 2	Universität Heidelberg	hxtorch	Edge processing, robotics	512	130K	Mixed-signal	True
2022	ReckOn	Charlotte Frenkel	Verilog	Task-agnostic learning over second-long timescales at the extreme edge	256	256	Digital	True
2022	Speck	SynSense	Sinabs, Samna	Smart sensing	328	~278K	Digital, with analogue sensor	
2022	Xylo	SynSense	Rockpool	Smart sensing	1000	278K	Digital	
2021	Loihi 2	Intel	Lava	Research	1000k	120 million	Digital	True
2021	SpiNNaker 2	University of Dresden	PyNN, NEST	Real-time simulation of SNN; DNN; Symbolic; HPC	1000k	10K	Digital	True
2019	DynapCNN	SynSense	Sinabs, Samna	Smart vision processing	1000k	~2M	Digital	
2019	Odin	Charlotte Frenkel	Verilog	General-purpose experimentation platform for bio-inspired edge computing	256	256	Digital	True

Year	Chip	Manufacturer	Software	Application	Neurons	Synapses	Type	On Cl Learn
2019	DYNAP-SE2	Institute of Neuroinformatics	Rockpool, Research Samna		4096	65K	Mixed-signal	

Adapted from: <https://open-neuromorphic.org/>

Neuromorphic hardware based on analogue circuits are not flexible but are dedicated to a particular network architecture thus they are highly optimised for a specific purpose. This feature allows analogue accelerators to be exceptionally power-efficient and highspeed due to their asynchronous in-memory computing. This makes them ideal for specialised computing and signal processing that require ultra-low-power and real-time applications such as processing of low-dimensional sensor signals in audio, medical and condition monitoring applications. Other significant neuromorphic computing applications include self-driving vehicles that can apply neuromorphic computing hardware to augment their capability to identify and interpret their environments; in medical diagnosis, where neuromorphic hardware can help in image detection features for the identification of diseases; and may be utilised in Internet of Things (IoT) devices to compile and analyse data, supporting effective data processing and decision-making.

5.4 SPECIFIC HARDWARE ASSOCIATED WITH ARTIFICIAL INTELLIGENCE

Earlier in this chapter it was shown that AI hardware was viewed as specialised computer hardware designed to execute AI-related tasks efficiently. This specialised computer hardware also includes specific chips and integrated circuits. These provide much faster processing and consume less energy, providing the needed infrastructure to execute AI algorithms and models effectively. It is also observed that the function of AI hardware in machine learning is critical for supporting the execution of complex programmes for deep learning models.

In comparison with traditional computer hardware such as the central processing unit (CPU), AI hardware can execute numerous processes faster, thus significantly decreasing the time and cost needed for algorithm training and execution. Due to the growing demand and popularity of AI and machine learning models, the global AI hardware market size was valued at USD 53.71 billion in 2023, and it is projected to surpass around USD 473.53 billion by 2033 and is likely to grow at a compound annual growth rate (CAGR) of 24.3% from 2024 to 2033 (Source: Precedence Research, www.precedenceresearch.com/). Some of this specialised computer hardware is described in the following paragraphs.

5.4.1 Edge Computing Chips

Edge AI concerns the deployment of AI technologies directly on devices at the network edge, such as Internet of Things devices (IoT), smartphones, and edge servers. This allows real-time data analysis and decision-making to take place without relying on cloud infrastructure or other centralised systems. Therefore, Edge AI chips play an essential part in this concept by providing the necessary computational power to process AI algorithms efficiently and effectively. Additionally, these specialised processors are designed specifically to run AI models at the network's edge. They also function as dedicated hardware accelerators specifically designed to execute AI workloads at the edge, and are optimised for real-time processing, low-latency operations, and power efficiency. They greatly lessen latency because they process data near its source, improving the overall performance of AI ecosystems. By offloading AI computations from the central systems or the cloud to local devices, these chips enable faster response times, reduce network bandwidth requirements, and enhance data security. Data security is improved by minimizing the amount of data that needs to be transmitted to the central systems. Since, these chips mitigate data transmission to remote servers they reduce energy consumption and optimise power efficiency.

A key advantage of Edge AI chips is that chip architectures and designs, such as the neural processing unit (NPU) and specialised accelerators, are specifically customised to optimise AI workloads. Edge AI chips are used in various applications, such as in smart homes and IoT devices by supporting voice assistants, video analytics, and smart cameras that permit local processing and reducing dependence on central systems. Edge AI chips are

also used in healthcare and telemedicine by monitoring patients remotely helping to provide real-time diagnostics, and personalised healthcare recommendations. On the industrial front, Edge AI chips accelerate intelligent control systems, predictive maintenance, and quality control for automation and robotics on the shop floor. Edge AI chips also support autonomous vehicles and drones with real-time object detection, collision avoidance, and navigation.

However, Edge AI chips have limitations. For example, Edge AI chips have constrained computing power of edge devices, because unlike cloud servers, they have limited access to resources, including processing power, memory, and storage. This limits their processing capabilities when executing complex AI algorithms that require significant computational capabilities. Although Edge AI chips mitigate the transmission of private data to central systems, they often process sensitive data locally, such as personal health information or video surveillance material that still require protection from potential unauthorised infringements.

Compatibility issues may occur when integrating edge AI chips with existing systems and devices to ensure seamless integration with different hardware and software components. However, chip manufacturers and developers are aiming for standardisation and are providing appropriate frameworks and software development kits (SDKs) to assist seamless integration and ensure compatibility across different platforms. These limitations are being mitigated by the emergence of the NPUs. These are specifically designed for AI computations to handle more complex AI workloads and are optimised for executing deep learning models that deliver accelerated AI performance with improved power efficiency. [Table 5.3](#) provides some details regarding leading manufacturers of AI hardware in the field of edge computing chips.

TABLE 5.3 Leading Manufacturers of Edge Computing Chips AI hardware[↗](#)

Term	Meaning
Neuromorphic Hardware	Architecture designed to imitate the structure and function of the human brain, with low power consumption, parallel processing, and event-driven computation.
Spiking Neural Network (SNN)	An artificial neural network that emulates the spiking actions of biological neurons, using discrete spikes or pulses of activity for information processing.
Event-Driven Computation	A computing concept where processing occurs in response to specific events or stimuli, allowing for energy-efficient operation and asynchronous communication between components.
Synapse	The functional link between two neurons or between a neuron and another cell, where signals are transmitted through chemical or electrical means.
Plasticity	The ability of synapses to strengthen or weaken over time that enables learning and adaptation.
Memristor	A resistor with memory, that is able to process and store information, emulating synaptic plasticity found in biological systems.
Neuromorphic Chip	A specialised hardware component designed to implement neuromorphic computing norms, often featuring a large number of simple, interconnected processing units.
Neuromorphic Engineering	An interdisciplinary field combining principles from neuroscience, physics, computer science, and engineering to design and build brain-inspired computing systems.
Event-Based Sensor	A sensor that attains and transmits information in an event-driven manner, aligning with the principles of neuromorphic hardware for efficient and low-latency data processing.
SpiNNaker (Spiking Neural Network Architecture)	A neuromorphic computing platform designed for simulating large-scale spiking neural networks, with a focus on real-time processing and parallel communication.

5.4.2 Application Specific Integrated Circuits (ASIC)

Generally, application-specific integrated circuits (ASICs) are a diverse category of integrated circuits (ICs) that assist designers to optimise sophisticated electronic devices. There is no consensus amongst electronics

professionals regarding the definition of an ASIC. In fact, there is disagreement on what an ASIC is or whether a particular component should be classified as an ASIC. However, a working definition that portrays common acceptable features is that an ASIC is an integrated circuit designed for a specific customer, application, or market using cell-based techniques, where the necessary functional blocks are taken from a cell library, interconnected, and simulated to provide the desired system functions and level of performance ([Analog Dialogue, 1990](#), p.12). Furthermore, this definition excludes ICs designed with traditional “custom” design techniques.

According to [Keim \(2023\)](#), the full definition identifies an ASIC as an IC that offers customised functionality but does not require a fully customised design process; instead, the customised functionality is achieved through a process that is analogous to PCB design. Keim argues that for an ASIC, designers take functional blocks from a library, interconnect them, and verify functionality and performance via simulation. Keim also maintains that the term “application specific” in an ASIC is somewhat misleading because he argues that in electrical engineering, “application” normally refers to the practical purpose of electrical equipment (i.e., the type of useful work the device is intended to perform).

[Wasay \(2023\)](#), views ASICs as being designed for targeted tasks like image processing, speech recognition, and cryptocurrency mining. He argues that their objective is to accelerate AI procedures to meet the specific needs of an enterprise, thus delivering an effective infrastructure that boosts overall speed within the ecosystem. For example, distinct from traditional central processing units (CPUs) or graphics processing units (GPUs), ASIC miners are designed for a single goal, basically to execute the specific hashing algorithm that garbles data to make it unreadable, which are critically required for mining a particular cryptocurrency. These hashing algorithms are one-way programs, which means that the text cannot be unscrambled and decoded by anyone else. Hence, this singular focus permits these ASIC miners to deliver unparalleled performance and energy efficiency in comparison to general-purpose hardware. Apart from the fact that these integrated circuits simplify the deployment of complex machine-learning mechanisms, they are also capable of managing significant volumes of data, which makes them instrumental in training artificial intelligence models. These ASICs may be used in diverse areas that include natural language processing of texts and speech data.

[Keim \(2023\)](#) argues that ASICs are not application specific as they might seem and are not limited to a particular application. To support his argument, he cites that a highly integrated data converter ASIC, which may be designed primarily for medical imaging applications, can also be utilised for industrial video processing or a multichannel automotive sensor network. He also claims that a more generic ASIC, such as a system-on-chip (SoC) ASIC, which is initially designed for smartphones, incorporates enough functionality to be successful in a wide variety of applications. Therefore, Keim maintains that task-specific integrated circuit (TSIC) or functionality-specific integrated circuit (FSIC) is a more accurate term. Moreover, he argues that generally, an ASIC is designed so that one chip can efficiently perform a specific combination of tasks. However, he insists that even if this combination of tasks is initially needed for a specific application, there may be various other applications where this ASIC would be an effective and desirable alternative to off-the-shelf ICs.

ASIC development, particularly if it involves non-recurring engineering (NRE) can take a substantial amount of time and cost before a single chip can be manufactured. Essentially, NRE covers the research, design, development, testing, and other tasks required to create a new product or custom design. It also includes the up-front costs sustained during the new product development. However, the NRE is a once only cost that is normally encountered when designing, developing or manufacturing of a new design, product or customised devices. This essentially means that a customer requiring a new product or product modification would need to pay for the NRE costs or any other costs incurred for the design modification. [Figure 5.3](#) illustrates an example of a device that has gone through the ASIC design process applied by Intel and Google, which is a “programmable networking device” for “cloud and communication service providers.” According to [Keim \(2023\)](#), ASICs are a crucial addition to high-performance and high-volume electronic devices that cannot be optimally designed using off-the-shelf components.

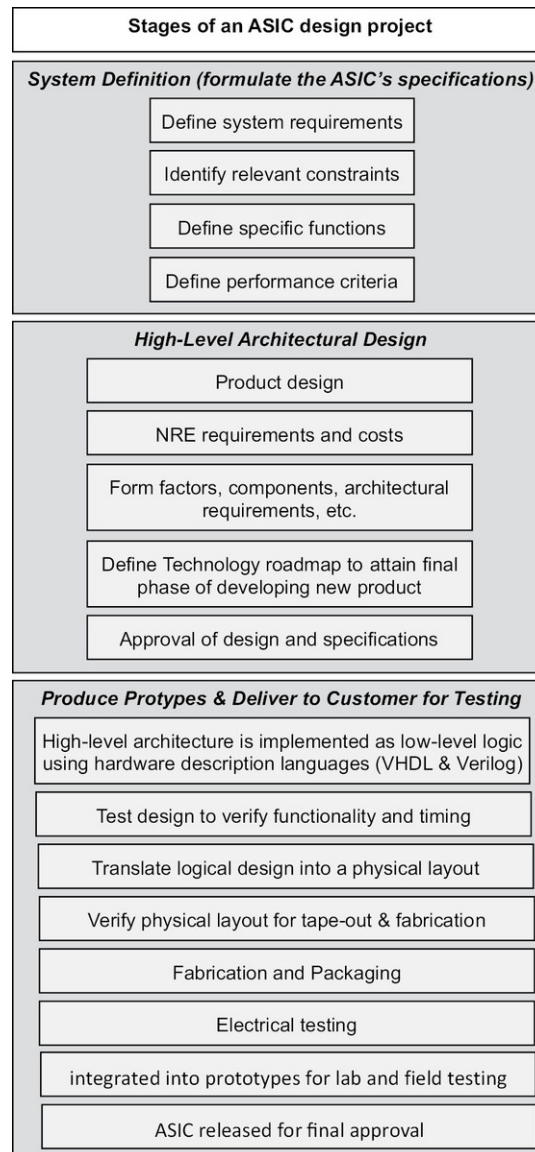


FIGURE 5.3 ASIC Design Process Applied by Intel and Google.

The decision to outsource the NRE depends on whether the component is considered to be a strategic product and whether the capacity, skills and experience required to design the product exist internally. Normally, companies do not tend to outsource strategic products. Keim argues that management can only justify NRE's high development costs for high-volume projects with demanding performance requirements. It is only in such circumstances that an ASIC can realistically become economically advantageous. An alternative option if a customised ASIC is not available is to seek a programmable logic device, such as a field-programmable gate array (FPGA) or complex programmable logic device (CPLD).

5.4.3 Field Programmable Gate Array (FPGA)

A Field Programmable Gate Array (FPGA) is an advanced integrated circuit that belong to a category of devices known as programmable logic. An FPGA is a very flexible device because it can be configured to be any digital circuit that is required by the user without any physical modifications, such as soldering or jumper wires. Therefore, the user can customise its function since all that needs to be done is to merely insert a configuration into

the FPGA and it will behave like the circuit the user requires. Hence, an FPGA is based on random access memory (RAM) through programmable logic and therefore can behave like any feasible circuit as needed and be reconfigured any number of times.

FPGA's flexibility makes it ideal for implementing AI software because it can be customised and programmed to meet the explicit requirements of the AI ecosystem thus being "field programmable." This is possible because an FPGA consists of configurable logic blocks (CLBs) that are interconnected and programmable thus providing it with the ability for running a wide range of AI applications. A CLB is the fundamental repeating logic resource on an FPGA, which when linked together by routing resources, the components in the CLBs perform complex logic functions, implement memory functions, and synchronise code on the FPGA. Moreover, a FPGA chip can be programmed to process operations of differing complexity levels, adjusting to the system's specific requirements. Although FPGAs operate like other read-only memory chips but a higher gate capacity, its key feature is its re-programmability capability, which means that it can be programmed various times. Another beneficial feature is that FPGAs provide better security because their all-inclusive architecture guarantees robust protection, making them trustworthy for secure AI applications.

A schematic diagram to create designs for FPGAs digital circuits is not advisable due to the magnitude and complexity the digital circuits within it. Typically, the user would define the function of the circuit that is required, and the device will automatically apply this to generate a circuit corresponding to the defined function. Therefore, FPGAs are like programming by inputting text, but the outcome is not software but hardware. An FPGA is ideal for rapid prototyping and final system design and are very suitable for applications, like AI, edge computing, system security, 5G, factory automation, and robotics.

A key advantage of FPGAs is their programmable characteristic that permits the designer to rapidly program and re-program the device to perform any function that is required. The re-programmability feature is doable at any time by using downloadable software updates during and after the FPGA has been integrated into an application. This feature allows a rapid time-to-market design that can be developed and modified concurrently with the overall system design. Furthermore, FPGAs have a parallel processing capability unlike non-sequential traditional processing through a "sea of gate" architecture. Parallel processing is highly suitable for high performance computing applications, such as AI that permit the FPGA to produce significantly higher performance exercising lower clock speeds that consume lower power.

However, FPGAs tend to be costly because of the design resources it consumes to cater for the diversity and complexity of its functional features. The cost of an FPGA depends on the amount of silicon needed to manufacture it; the extent of research and development required to design it, including the tooling; and the somewhat low volume demand when compared to other commonly used processors for everyday devices, apart from their high-power usage to operate it. On the other hand, to design a customised digital circuit out of discrete logic takes time and is costly, and with very little flexibility if a modification is required.

Although a circuit created directly in silicon has superior speed and efficiency in performance, its flexibility index is non-existent. Furthermore, custom silicon is very much dependent on economies of scale because it has extensive upfront costs for tooling and setup; and unless tens of thousands of these chips are made, it becomes a more expensive approach when compared with a FPGA chip. As stated previously, an FPGA has flexibility and may be modified without penalty. An FPGA is a tool and like every other tool it has its uses depending on whether the solution to use an FPGA makes sense, when other cheaper and more efficient chips are available. For instance, you do not use a sledgehammer to crack a nut.

FPGAs are very useful for AI associated applications because they can be configured (and re-configured) depending on the role that they are to play. FPGAs use dedicated software that is usually provided by the FPGA supplier. Developers would design the logic to be executed in the FPGA by using either graphical design capture or a hardware description language (HDL). Graphical design capture language (e.g., C and C++) is generally utilised for smaller FPGAs. On the other hand, HDL is used to describe the behaviour or structure of digital circuits (ICs) and might be applied to stimulate the circuit and check its response. Examples of HDL are VHDL (very high-speed integrated-circuit hardware description language) and Verilog. Other HDLs include JHDL (Java HDL) and Active-HDL (proprietary HDL by Cypress Semiconductor Corporation).

The HDL software compiles the design by integrating the logic that best fits the target FPGA, after which the software creates a bitstream that is downloaded to configure (program) the FPGA, making it ready to perform its defined role. Designers may opt for a rapid development strategy by using a suite of software design tools and IP (Intellectual Property), or pre-configured design blocks, such as those offered by Lattice Semiconductor (FPGA Design Software). Such a strategy would lessen the time, effort and complexity required in the design process to fast-track the creation of bitstreams that execute on Lattice FPGAs. Lattice FPGAs include AI and machine learning (ML) inferencing, data co-processing, embedded vision, and secure system control amongst many others.

5.4.4 Complex Programmable Logic Device (CPLD)

CPLD is another type of programmable logic device. It is similar to a SPLD (simple programmable logic device). However, a SPLD is the simplest, smallest, and least-expensive forms of programmable logic devices, with a complexity that is far below of a CPLD. A SPLD normally consists of a few hundred gates, whereas a CPLD contains of a few thousand logic gates. Hence, a CPLD is more likely to have implications for AI specific applications. However, CPLDs are less complex than FPGAs, which also makes them useful in AI applications.

CPLD comprises of multiple programmable array logic (PAL) structures known as function blocks (also known as macrocells). Hence, a CPLD is a logic device with completely programmable AND/OR arrays and macrocells. Macrocells can also be defined as functional blocks responsible for performing sequential or combinatorial logic. These function blocks are linked by a programmable interconnect, which is known as GIM (global interconnection matrix) that contains complex logic operations and logic for implementing disjunctive normal form expressions. Hence, by reconfiguring the GIM, different logic circuits can be achieved, which allow the CPLDs to interact with the outer world using digital inputs/outputs.

There are several similarities between CPLD and FPGA that include being a programmable logic device consisting of logic gate arrays, which may be programmed using HDLs such as Verilog HDL or VHDL. However, there are several critical differences. A CPLD normally has a few thousand logic gates, whereas an FPGA can have millions of logic gates. Therefore, an FPGA can accommodate complex circuits and systems, but this complexity comes at a higher cost, which make CPLDs more suitable for less complex applications. Additionally, CPLDs feature a built-in non-volatile EEPROM (electrically erasable programmable random-access memory), whereas FPGAs feature a volatile memory. Therefore, a CPLD can preserve its memory contents when powered off, while an FPGA does not retain its memory content. This means that because of this feature, a CPLD can resume operating straightaway after power-up, but most FPGAs require a bit-stream from an external non-volatile memory for start-up.

Furthermore, FPGAs have an irregular signal processing delay because of its highly complex architecture combined with the user's custom programming that affect its performance. However, the pin-to-pin delay of CPLDs is significantly lower due to the simpler architecture that makes it faster. Note that a signal processing delay is a crucial issue in the design of safety-critical and embedded real-time applications. FPGAs operate at higher frequencies and have more complex logic processes and therefore are likely to have higher power consumption than CPLDs. This means that a FPGA based system requires better thermal management. Note that the objective of thermal management is to maintain a system within its operating temperature tolerance. This is a critical issue with electronic devices because effective thermal management dissipates excess heat to prevent overheating. Therefore, thermal management is an extremely key concern in FPGA-based systems since they frequently use heat sinks and cooling fans that require bigger, more complicated power supplies and distribution networks. Additionally, CPLDs have a better information security feature because the memory is integrated into the chip itself. On the other hand, many FPGAs have external non-volatile memory that increases the potential data security risk even though they have data encryption algorithms included.

CPLD manufacturers include Lattice Semiconductor, Atmel, Xilinx, and Altera (owned by Microchip, AMD, and Intel, respectively). CPLDs are commonly used in many low-to-medium complexity digital control and signal processing circuits, such as bootloaders for FPGAs and other programmable systems. CPLDs are small in size and have low power consumption, which make them ideally suited for portable and handheld digital devices. They are also regularly used as address decoders and custom state machines in digital systems, and in safety-critical control

applications. [Figure 5.4](#) depicts the stages for CPLD programming and development. It illustrates that the programming and development stages of an CPLD consists of three phases, namely system definition and CPLD model selection; setting up the programming and testing environment; and finalisation for housing the CPLD and adjoining electronic components for prototype production. The CPLD goes through a meticulous testing process before it is sent for the production process.

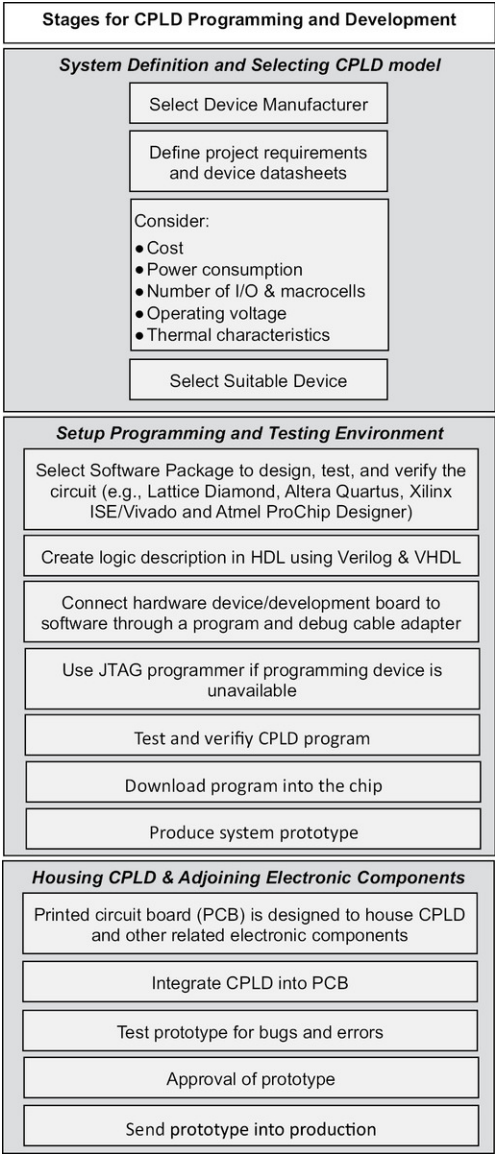


FIGURE 5.4 Stages for CPLD Programming and Development.

5.5 CONCLUSION

This chapter has traced the development of computer hardware in relation to the advancement of AI. The chapter examines the pioneering applications of AI using earlier generation of computers and then focuses on the fourth and fifth generation computers. Starting with largescale integrated circuits (LSI) in the early 1970s, moving to the very largescale integrated circuits (VLSI) of the late 1970s, and progressing to the super largescale integrated circuits (SLSI) in the early 1980s.

However, history has shown that the capability of AI applications is limited by the processing power of the computing architecture that is being utilised. Traditional computers that are based on the Von Neumann architecture, which execute instructions sequentially and process the data held in centralised memory have proven to be a serious limitation. This limitation is due to the restricted data transfer rate between the CPU and the external memory unit, known as the Von Neumann bottleneck. The Quantum computer is aimed at addressing this weakness and is making big strides in processing speed where the use of the quantum states of subatomic particles to store information is being applied to run multidimensional quantum algorithms. IBM has announced the development of the Condor quantum computer model with 1,121 qubits. Other manufacturers such as the Californian quantum start-up Atom Computing are also making significant progress in this area.

Another important development is in Neuromorphic hardware that has the primary aim of emulating the human nervous system by applying dedicated computing architectures that imitate the structure (morphology) of neural networks. The Neuromorphic hardware uses the laws of physics, namely Ohm's and Kirchhoff's laws to permit parallel processing for specific tasks. For example, the spiking neural networks (SNNs) and event-driven computing of the 2010s that focused to improve efficiency and power-aware neuromorphic hardware designs. Of particular importance is the development in 2014, when IBM developed the TrueNorth neuromorphic chip that is utilised in visual object recognition and in 2018, Intel developed the Loihi neuromorphic chip that has had applications in robotics as well as gesture and smell recognition.

Neuromorphic hardware is ideal for specialised applications such as medical and condition monitoring applications; self-driving vehicles to boost their capability for identifying and interpret their environments; in medical diagnosis for the identification of diseases amongst many other applications. The chapter also discusses other dedicated hardware, such as Edge Computing Chips, Application Specific Integrated Circuits (ASIC), Field Programmable Gate Array (FPGA), and Complex Programmable Logic Device (CPLD). This dedicated hardware allows the programmability of the chips for specific tasks and are integrated as part of applicable systems. This dedicated hardware may also be integrated with the neuromorphic computing. Developing computers that mimic the human brain is a futuristic objective that will transform the current limited AI applications to what is known as Super AI. However, developing chips that mimic the human brain is currently a reality but limited to specific tasks.

REFERENCES

- Analog Dialogue. (1990). *Mixed Signal ASICs Offer Designers an Alternative Solution*. www.analog.com/media/en/analog-dialogue/volume-24/number-3/articles/volume24-number3.pdf#page=12.[↗]
- IBM. (2024). *What is Quantum Computing?* www.ibm.com/topics/quantum-computing.[↗]
- IDQ. (2024). 2023 heralded the arrival of the 1000 qubit computer. *Quantum Computing Review Q4 2023*.[↗]
- Keim, R. (2023). What is an Application-Specific Integrated Circuit (ASIC)? *All About Circuits*. www.allaboutcircuits.com/technical-articles/what-is-an-asic-application-specific-ic/.[↗]
- Moto-oka, Tohru. (1983). *Overview to the Fifth Generation Computer System Project*. Department of Electrical Engineering, The University of Tokyo.[↗]
- Nasrin, S. (2023). *The Fifth Generation Computer: The Advancement of Technology*. Digital Marketing Manager at Fiverr.[↗]
- Reid, T. R. (1992). *Japanese Government Ends Development of Computer 'Fifth Generation' Falls Short of Goals*. The Washington Post.[↗]
- Swinbanks, David and Anderson, Christopher. (1992). Japan stubs its toes on fifth-generation computer. *Nature*, Vol. 356, 273–274.[↗]
- Wasay, Ali. (2023). *The Five Most Promising AI Hardware Technologies: The Future is Bright for AI*. www.makeuseof.com/most-promising-ai-hardware-technology/.[↗]

Chapter 6

Artificial Intelligence Software

DOI: [10.1201/9781003653950-8](https://doi.org/10.1201/9781003653950-8)

“Continuous improvement is better than delayed perfection.”

Mark Twain

Author (1835-1910)

Typically, hardware on its own is of no value but supported by the appropriate software it could become a very powerful tool. The previous chapter focused on different types of AI hardware. However, it also illustrated that for the AI hardware to be of any use it must have the software to make it functional. For instance, for its Quantum Computer, IBM have adopted Qiskit, which is an open-source, Python-based quantum software development kit (SDK). The quantum SDK is a comprehensive development tool that includes an intuitive user interface; a compiler; and a quantum runtime environment optimised for running hybrid quantum-classical algorithms. Additionally, Qiskit has application routines written in the Python software language regarding finance, chemistry, optimisation, and machine learning. Python is a high-level interpreted language, which allows rapid development and testing, which is a key feature in the iterative

process of AI projects. Furthermore, Python is currently at the forefront of AI programming due to its simplicity, versatility, and powerful libraries.

This is also applicable to the specific AI hardware related to Edge Computing Chips; Field Programmable Gate Array (FPGA); and Complex Programmable Logic Device (CPLD). These all require some type of applications software for them to provide the benefits that they have been designed for. For example, Edge Computing Chips that concern the deployment of AI technologies directly on devices at the network edge, such as Internet of Things devices (IoT), smartphones, and edge servers have standard software development kits (SDKs) to assist seamless integration and ensure compatibility across different platforms. Field Programmable Gate Arrays (FPGA) are programmed using either graphical design capture or a hardware description language (HDL). Graphical design capture language, such as C and C++ is normally used for the smaller FPGAs, while HDL is used for the larger FPGAs and utilise VHDL (very high-speed integrated-circuit hardware description language) and Verilog. Other HDLs include JHDL (Java HDL) and Active-HDL (proprietary HDL by Cypress Semiconductor Corporation). Complex Programmable Logic Devices (CPLD) apply similar programmable languages as a FPGA, namely Verilog HDL or VHDL. For further details regarding HDL refer to Keim (2020).

Neuromorphic computing to be of applicable use must also be supported by software. For example, the Spiking Neural Processor T1 chip produced by Innatera uses the Talamo programming language applicable to smart-sensing, wearables, and IoT. SynSense, which manufacture the Speck, Xylo, and DynapCNN chips use diverse programming languages, such as Sinabs, Samna, and Rockpool for smart sensing and smart vision processing. Charlotte Frenkel that manufacture the ReckOn and Odin chips use the standard Verilog programming language mainly for task-agnostic learning over second-long timescales at the extreme edge and general-purpose experimentation platform for bio-inspired edge computing applications.

This chapter will examine several significant aspects, namely AI programming languages; AI algorithms; and neuromorphic computing. When examining AI software, there may be several instances that overlap with the AI hardware aspects to ensure that a clear understanding is conveyed.

6.1 INTRODUCTION

Basically, software is a collection of instructions, data or programs that are utilised to operate computer hardware and execute specific functions. Software can be viewed as complementary to hardware. In fact, it may be envisaged as being the opposite of hardware or the other side of a coin, where hardware is the head, and the software is the tail side. Hardware describes the physical features of a computer, whereas software is a broad term used to indicate the applications, scripts, and programmes that execute on a hardware device. A software application can vary depending on the programme that is uploaded on the hardware, whereas the hardware has a fixed, non-flexible function.

There are two main types of software, namely system (operating) and application software. System software is designed to run the computer's hardware and provides a platform for applications to execute conjointly with the hardware. On the other hand, application software achieves a specific need and performs stipulated tasks. The other types of software consist of programming software, which provides the programming tools that software developers use to generate and create applications; middleware that lies between the system and applications software; and driver software that operates the computer devices and peripherals.

6.1.1 Software History

In the early days, software, whether operating or applications, was created for specific computers on a proprietary basis that was packaged as part of the hardware it executed on. In the 1960s and 1970s most hardware and

associated software was for mainframe computers and minicomputers. However, the mid to late 1970s witnessed a variety of microcomputers in the form of desktop or portable computers coming on the commercial market for business applications. During this period, software was sold and distributed on floppy disks, sometime later, around 1982, CDs were used and later still, around 1997, DVDs became a common software distribution medium. The main difference between a CD and DVD is basically the data storage capacity. DVDs use a different (shorter) wavelength laser that provide a much higher resolution of optical data allowing a DVD to store as much as 7 to 14 times as much data as a CD. Currently, most software is acquired and directly downloaded over the internet at either a once only cost or at a flexible usage licence cost, normally on a yearly basis. Moreover, software can be purchased directly from the software owner's website or application service provider websites.

It was not until the late 1950s that the term software was used. During this era, software was not commercially available even though varying types of programming software were being generated. Early software development was restricted to scientists and other technicians who working for the bigger organisations developed their own software. It should be noted that during this era, hardware was very costly and only large organisations like governmental entities or large corporations could afford to purchase the hardware and employ the scarce technical human resources.

[Figure 6.1](#) provides a general timeline regarding the history of software. Like hardware, computer software has developed through various generations. Early software, which is often referred to as first- and second-generation software languages were low-level machine languages. First-generation languages were machine language statements written in binary code (0 or 1), which basically reflected the computer hardware circuitry state of on (1) or off (0). These languages were difficult to write and understand, and very problematic for tracing programming errors. In June 1948, Tom Kilburn, who was a computer scientist wrote the world's first piece of software for the "Manchester Baby" computer at the University of

Manchester United Kingdom in binary code as part of his thesis “A storage system for use with binary digital computing machines.”

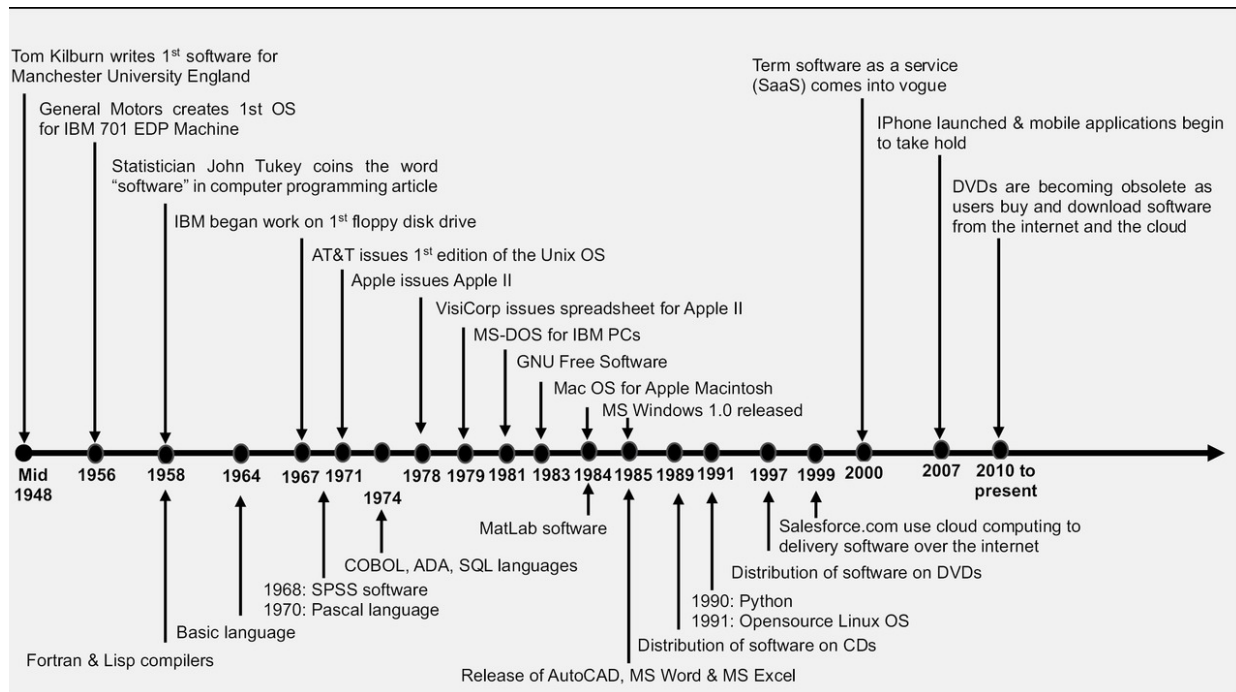


FIGURE 6.1 General Timeline Regarding the History of Software. [↗](#)

The second-generation languages were still machine based but were an improvement on the first-generation languages, called assembler languages. Assembly language contained human-readable statements that were converted to machine language using an assembler. Hence, an assembler converted written coded instructions to machine-level instructions. As early as 1947, Kathleen and Andrew Donald Booth working for the Institute for Advanced Study, at Princeton University developed the first assembly code in which a language is used to represent machine code instructions, namely coding for A.R.C (automatic relay computer). However, the word “assembler” is commonly accredited to Wilkes, Wheeler, and Gill in their 1951 book “The Preparation of Programs for an Electronic Digital Computer.” They used the word to represent “a program that assembles another program consisting of several sections into a single program.”

An important milestone was when in the early 1950s General Motors' Research division created the first operating system, GMOS for its IBM 701 and in 1956 for its IBM 704 computer. Third-generation computer languages came on the scene in the late 1950s. However, it was not until 1958 that John Tukey, a statistician used the word "software" when writing an article about computer programming. This era also coincides with the release of the FORTRAN and COMTRAN (precursor to COBOL) programme compilers in 1957, followed by the LISP compiler in 1958. The BASIC interpretive programming language came on the scene in 1964, followed by PASCAL in 1970, and later in 1974, by COBOL, ADA, and the first SQL language. During the 1960s and 1970s applications software were mainly developed in-house by organisations that could afford a structure to support its electronic data processing activities. However, several important commercial software packages such as PERT (program evaluation and review technique: ICL 1900 PERT) and SPSS (statistical package for social science) were released in 1966 and 1968, respectively.

In 1971, AT&T released the first edition of the Unix operating system and IBM introduced the first 8-inch (20 cm) floppy disk. However, a significant breakthrough for consumer software was the release by Apple of the Apple II personal computer (PC) in 1977, followed by the first spreadsheet software by VisiCorp that released VisiCalc for the Apple II. The 1980s witnessed an exponential growth in both hardware and software that saw Microsoft releasing its MS-DOS operating system, which became the standard for all IBM compatible PCs, paving the way for making commercial software being available to the average consumer. Furthermore, hard disc drives became the standard for PCs with manufacturers bundling software in their PCs. 1983 also saw the birth of the free software movement with Richard Stallman's GNU (GNU is not Unix) Linux project for creating a Unix-like operating system with the source code being made freely available. Apple released its Macintosh line in 1984, which was followed by Microsoft Windows 1.0 for the IBM compatible PCs. Additionally, the mid-1980s witnessed the release of important software applications, including AutoDesk's AutoCAD, and Microsoft Word and

Excel. Moreover, floppy disks that were introduced in the 1960s became a major software distribution media in the 1980s, with the CD-ROM being a major player in 1989.

The reader needs to recall that the 1980s witnessed the birth and acceptance of expert systems with the Lisp programming language and the launching of the 5th Generation Computer Project. Furthermore, David Rumelhart, Geoffrey Hinton, and Ronald Williams in 1986 introduced a new era of neural computation research through their book, “Parallel Distributed Processing: Explorations in the Microstructure of Cognition.” Additionally, the Salk Institute and the University of California launched an academic journal “Neural Computation” with the first issue being the Spring of 1989, with Terrence Sejnowski being the Editor-in-Chief. By 1990, the Python computer language was launched, which has become a popular programming language for AI applications. During the 1990s the internet became very popular and by 1999 Salesforce.com utilised cloud computing to pioneer software delivery over the internet, with the term software as a service (SaaS) coming into vogue by 2000. The iPhone was launched in 2007, and mobile applications begin to take hold. By 2010, DVDs became obsolete as users purchased and downloaded software from the internet and the cloud.

Concurrently with this development, fourth-generation languages were being developed as early as 1984, with MatLab from Matrix Laboratory; and Python in 1990, apart from Perl, Ruby, SQL. These fourth-generation languages consisted of commands that are similar to statements in the human language and are mainly used in database programming and scripting. There has also been a growth in the fifth-generation programming languages that are chiefly applied for AI. These programming languages have visual tools to develop a computer program. Examples of fifth-generation programming languages include OPS5, Mercury, ICAD, and KL-ONE. OPS5 is a rule-based computer programming language that has been used in successful expert systems, while Mercury is a functional logic programming language made for real-world applications. Additionally, ICAD (Intelligent Computer-Aided Design) has an open architecture that

can use all the power and flexibility of the underlying language. The original implementation of ICAD was on a Lisp machine (Symbolics). While KL-ONE is a language based on frame language that reflects a related idea. This computer language is what is referred to as a knowledge representation system applying semantic networks and frames. These fifth-generation languages will be discussed in more detail later.

The above software developments facilitated the creation and growth of supervised learning in the 1990s, and machine learning algorithms for solving problems in the 2000s, which are vital for AI applications. Furthermore, these programming languages allowed deep neural networks to achieve advanced results around 2010, and the current enhancement of Chat Bots and GPT models. Without the above software developments, today's AI software would not be possible. AI software is specifically aimed at processing voluminous amount of data to execute tasks that require human intelligence, including image recognition, video analytics, generative AI, voice recognition, text recognition, and Natural Language Processing (NLP) that enable a computer to understand, interpret, and generate human-like text.

6.2 AI COMPUTER PROGRAMMING LANGUAGES

This section will describe in some detail the key AI computer programming languages, including programming languages related to specific AI hardware that were discussed in the previous chapter. It should be noted that software programmes written in most of these computer languages execute on the traditional Von Neumann computers.

6.2.1 Lisp

Lisp (list processing) is a programming language that was specifically intended for manipulating data strings in a simple and effortless manner. It has been in use since 1958 and is considered as one of the oldest programming languages still in use. Lisp commenced as an interpretive

programming language, without the need of a compiler, similar to the original BASIC programming language. With interpretive programming languages, the source code is interpreted directly on the system. However, interpretive programming languages are slower to run than the compiled versions. Currently, most computer programming languages, including Lisp, have a compiler that generates an executable version that achieves faster programme execution velocity.

Lisp has influenced the development of other computer programming languages due to its different dialects. The current Lisp-based computer programming languages (Lisp dialects) include: Clojure; Emacs Lisp; Common Lisp; Julia; Racket; and Scheme. Lisp is very popular in higher education, where students may learn the relevant programming tactics in preparation for practical applications. Lisp is expressed as a function of at least one object. It processes symbolic information rather than numbers, thus making it very suitable for AI programmes. An object can be another function, data items either as constants or variables, or data structures. Lisp is easy to learn and use, making it very popular in the academic arena for teaching computer programming. Lisp can be applied for a variety of uses such as a basic programming tool, AI applications, machine learning, and quantum computing.

Currently, Lisp dialects are applied to generate code in a range of use-case scenarios. Use-case scenarios specify a meticulous description of how a certain use-case evolves. These scenarios are described by charting a sequence of interactions between the external actor and the system, showing the different steps and conditions that may arise during execution. Use-case scenarios provide a deeper or more detailed view to help developers and stakeholders visualise the specific user interactions. Use-case scenarios are defined by applying two basic elements, namely the flow of events; and use-case templates. The flow of events defines the step-by-step progression of actions within a use-case, with each step in the flow of events linking to a specific action or decision point. This provides an all-embracing view of the user-system interaction, identifying alternate flows and exceptions that are important for foreseeing different paths a use-case may take. On the

other hand, use-case templates present a standardised structure for documenting use-cases, by having areas for placing the use-case name, description, actors, preconditions, postconditions, and the leading and alternative flows. These, templates provide a systematic way to collect and manage information to ensure consistency across various use-cases within a project.

The application of a use-case scenario is best illustrated through an example. [Figure 6.2](#) shows an example for an online shopping system. The use-case scenario in this case is a situation where a customer is placing an order through an online shopping system. Hence, the use-case is the customer's order placement, and shows the customer instigating the process of purchasing the required objects online.

1. Use Case:

Title: Place Order

Description: Customer begins process of buying goods online.

2. Use Case Scenario:

Scenario: Customer adds items to shopping cart & goes to checkout.

Flow of Events:

- Customer logs in.
- Customer surfs goods & adds goods to shopping cart.
- Customer goes to checkout page.
- System urges customer to confirm order.
- Customer confirms and presents order.

3. Use Case Template:

Name: Place Order

Description: Customer begins process of buying goods online.

Actors: Customer

Preconditions: Customer is logged in.

Postconditions: Order is logged successfully.

Main Flow:

- Customer logs in.
- Customer adds goods to shopping cart.
- Customer goes to checkout page.
- System urges customer to confirm order.
- Customer confirms and presents order.

Alternate Flow: None in this case.

Use-case scenario Diagram: Online Shopping System

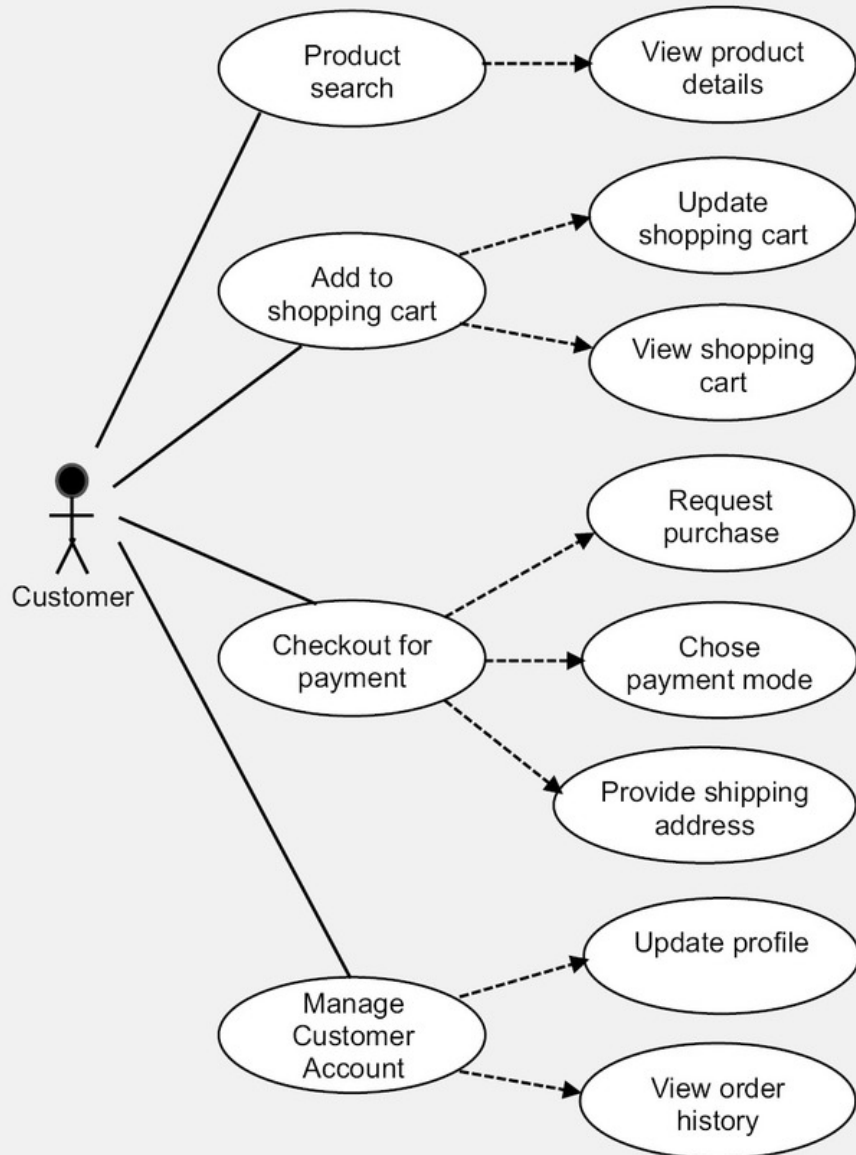


FIGURE 6.2 Use-Case Scenario: Example of the Common Online Shopping System. 

[Figure 6.2](#) illustrates that the use-case is the high-level description of the functionality, which is the placing an order. As can be seen, the use-case scenario breaks the task of placing an order into smaller, manageable steps, by detailing these steps from logging-in to confirmation. Additionally, the flow of events specifies a step-by-step sequence for the scenario, identifying the actions and decision points. The Online Shopping System use-case diagram is a clear visual representation of the main functions and interfaces within an online shopping platform. The diagram describes the various tasks and processes included in the system that allows users to understand the overall system structure and behaviour. The use-case diagram includes the important tasks of the system such as searching and viewing product details; adding goods to the shopping cart; and updating the shopping cart and viewing its contents. The diagram also illustrates the checkout and payment process such as placing an order; choosing the payment method; and providing the shipping address; apart from showing tasks related to managing user accounts, including updating profile information and displaying the order history. Applying the use-case diagram provides a homogenous structure for documenting the use-case, which ensures a consistent method across the different use cases in the system.

Common Lisp has several advantages such as the unmatched read–eval–print loop (REPL) also termed an interactive top-level or language shell that aids to interact with running a web app; the exception handling mechanism; performance; ability to build a self-contained executable process; stability; good threads story; strong typing, amongst others. REPL is a simple interactive computer programming environment that takes single user inputs, executes them, and returns the result to the user. Hence, a program written in an REPL environment is executed piecewise.

Lisp has been successfully used in several major applications for many years. For example, in 1984 Richard Stallman developed the text editor GNU Emacs that was written in a Lisp dialect, Emacs Lisp, which is also utilised to write Emacs configs and extensions. At around 2009, Common Lisp was used for Grammarly, which is an online service to process English texts. Grammarly applies AI to analyse the text and makes recommendations on how it may be improved. Apart from being a grammar and spell checker, it also checks the text for text brevity, vocabulary, and tone.

Another application written in Common Lisp is the Allegro Network File System (NFS) Server that is used by Boeing 747 and 777. An NFS is a means of storing files on a network, which is a distributed file system that allows users to access files and directories located on remote computers and handle these files and directories as if they were local. In the aviation field, both Boeing and Airbus use Piano, which is a software package written in Common Lisp for aircraft design development and analysis. Generally, Piano is a powerful tool for major airframe and aero-engine manufacturers, research and governmental institutions, and environmental organisations worldwide. ITA Software which is normally associated with the system for air tickets search and pricing is currently part of Google's travel industry department that develops most of its software using Common Lisp. Common Lisp is also the main programming language used by the Portuguese Siscog software provider, which develops decision support systems designed for resource planning, allocation, and management in the transportation sector. Their software is used for the London Tube.

CircleCI is a CI/CD tool that supports rapid software development and publishing, uses a Lisp dialect Clojure. They also use ClosureScript for frontend development. CircleCI permits automation across the user's pipeline, from code building, testing to deployment. Additionally, CircleCI integrates with GitHub, GitHub Enterprise, and Bitbucket to create builds when new code lines are committed. Furthermore, computer-aided design systems AutoCAD applies the AutoLISP dialect as the internal application development language. Riemann, which is a tool for distributed system

monitoring, is also mostly written in Clojure that offers greater flexibility and fault-tolerance. This system aggregates events from user servers and applications and combines them into a stream and then transmits them for further processing or storage. The above illustrates that the Lisp computer programming language is a very versatile software development tool for a variety of practical applications.

6.2.2 Python

Python is a general-purpose computer programming language, which can be used across technology and domains. It has been around for over three decades and has become a very popular programming language, particularly for AI based applications. Additionally, Python is a high-level programming language that can process data arrays, variables, objects, complex arithmetic, and Boolean expressions, as well as other abstract concepts. Hence, it is a very versatile computer programming language.

Python is an interpretive computer programming language, which means that it does not have a compiler to convert the programming code to machine language before it can be executed. Although, compiled languages tend to be faster to execute, Python is used as an emulator or a virtual machine, which is based on an existing machine's native code, therefore it can be directly understood by the hardware. Python is a powerful language because it has a dynamic-type system and automatic memory management to support various programming methods, such as object-oriented, functional, imperative, and procedural, amongst many others.

Python has been utilised for a range of functions, including machine learning; creating websites; software testing; and controlling self-driving cars. A big advantage of Python is that it is accessible for every operating system and is open source, therefore, it is free to use and distribute. Furthermore, Python can be exploited by developers and non-developers. Python can be applied to carry out complex statistical computations, manipulate and analyse data, and achieve other data-related tasks. It can also be used to generate data visualisations and create machine learning

algorithms. The data visualisation feature can facilitate the creation of graphs, histograms, pie charts, and 3D diagrams. Python is also supported by library routines that help developers to write their programs rapidly and efficiently involving data analysis and machine learning.

Python is very useful for developing the back end of a website or application that is unseen by the user and is normally responsible for storing and organising data, making certain that routines on the client-side work. Hence, its role is to communicate with the frontend, sending and receiving information to/from servers, processing data, and linking with databases, URL routing, and safeguarding security. Python has a replication feature for those tasks that are performed on a repetitive basis, using automated processes called scripting. This automation feature is exploited to execute simple math, convert files, test for errors across multiple files, and remove duplicates in data. Python is the friendliest language when compared with other object-oriented programming (OOP) languages.

OOP is based on the concept of objects that may contain data and code, where data is in the form of fields, and code in the form of procedures. [Figure 6.3](#) illustrates some of the key advantages of using Python for AI development that include access to an extensive library of routines, a high level of support from the developer's community, high flexibility, and is platform independent with an open-source classification. Python is considered as one of the best computer programming languages for AI because it is about 80% less complex as other OOP languages. Additionally, its key AI features include the availability of prebuilt libraries routines such as NumPy for performing scientific calculations, SciPy for doing high-end computations, and PyBrain for machine learning.

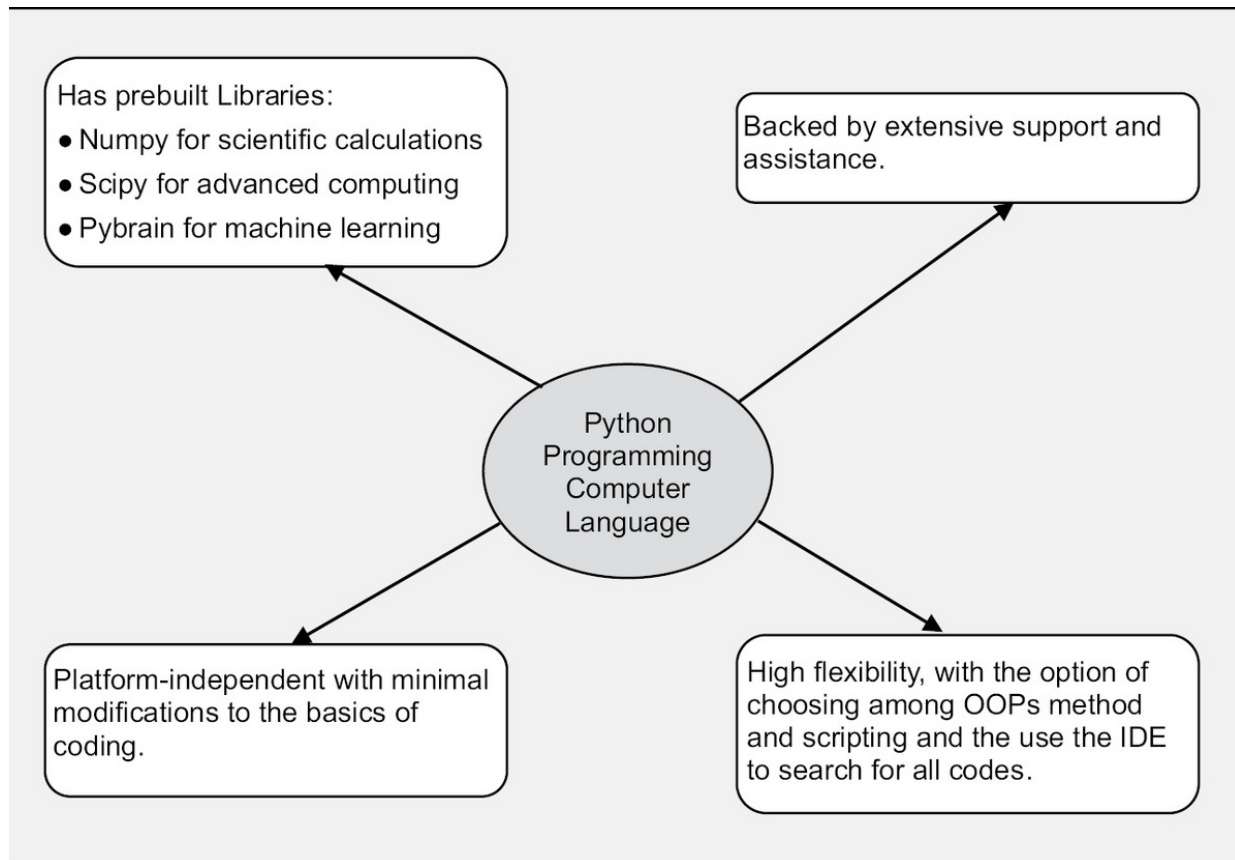


FIGURE 6.3 Advantages of Using Python to Create AI. [🔗](#)

The NumPy (Numerical Python) routine is used to store generalized data that consists of a one N-dimensional array and has tools for integrating C/C++ codes, Fourier transformation, random number capabilities, amongst many other functions. SciPy supplements Numpy allowing extremely fast and efficient processing since the high-level commands and classes permit for easy data manipulation and visualisation. The SciPy library is written in an assortment of languages such as C, C++, Fortran, and Python. Moreover, PyBrain is an easy to use open-source library of python for implementing Machine learning algorithms to solve problems quickly. Pandas is another open-source library built on top of Python for dealing with data structures and operations that are very powerful, flexible, and easy to use as analytical instruments, particularly for data analysis and manipulation. The Matplotlib library is used for producing high-quality 2D graphic representations and

can access up to six graphical user interface tools, web application servers as well as Python scripts.

[Figure 6.4](#) shows the main Python libraries for general AI. There are four main Python libraries for General AI, namely AIMA (AI: A Modern Approach), pyDatalog, SimpleAI, and EasyAI. AIMA is not a software library but a university textbook on AI regarding the Python algorithm implementation. This textbook was written by Stuart Russell and Peter Norvig in 1995, with the fourth edition published in 2020. pyDatalog is Python's logic programming engine that provides logic programmers with an extensive standard Python library that can express complex algorithms rapidly. This logic engine is a meticulous declarative language with strong academic foundations that has excellent features for managing complexity. The pyDatalog code is concise and used for simulating intelligent behaviour for games or expert systems; probing complex sets of related information such as data integration or natural language processing; and executing recursive algorithms, for instance on hierarchical data structures.

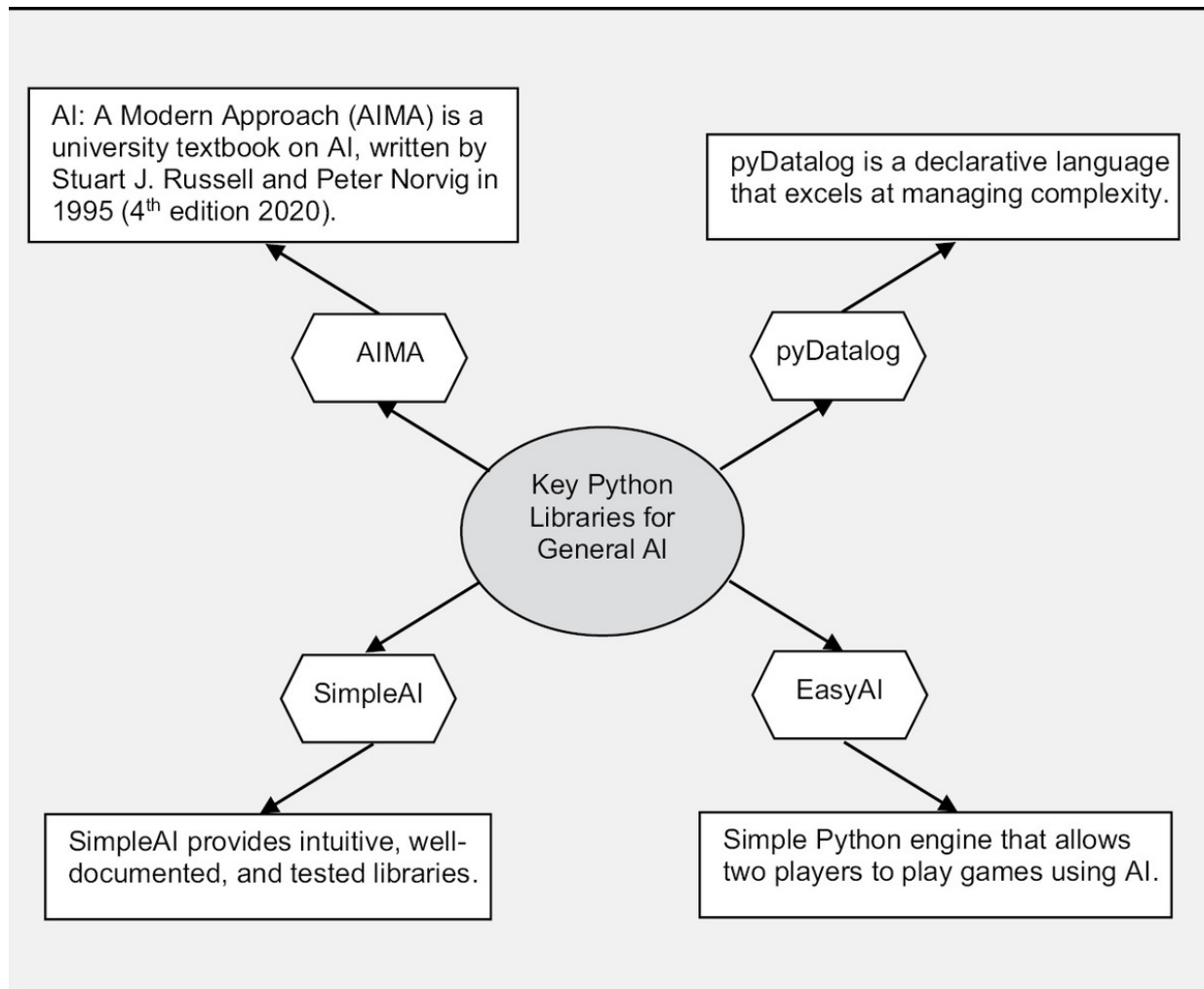


FIGURE 6.4 Python Libraries for General Artificial Intelligence. [📄](#)

The other two libraries that are regularly used are SimpleAI and EasyAI. SimpleAI is Python's application of various AI algorithms that are explained in AIMA and provide intuitive, well-documented, and tested libraries. Whilst EasyAI is a minimal Python engine that permits two players to play games using AI, namely Negamax and transposition tables for Game-solving. Negamax is the search variant form of minimax search that depends on the zero-sum property of a two-player game. Basically, Negamax is minimax but with less code.

Python is an important component of AI programming languages because it has well established frameworks like its scikit-learn machine learning library. This library meets most of all the requirements of current

AI development due to its user-friendly visualisation features. The scikit-learn library consists of highly efficient tools for machine learning and statistical modelling including classification, regression, clustering, and dimensionality reduction. It integrates a broad range of machine learning algorithms for medium scale supervised and unsupervised problems, particularly applicable to non-specialist users by having a general-purpose high-level language. Scikit-learn has the following important features:

- Supervised learning algorithms. The scikit-learn toolbox has generalized linear models, Support Vector Machines (SVM), Decision Trees, and Bayesian methods, amongst many others.
- Cross-validation. This feature implements several approaches for checking the accuracy of supervised models on unseen data.
- Unsupervised learning algorithms. This is supported by several algorithms such as clustering, factor analysis, principal component analysis, unsupervised neural networks, and many others.
- Assorted toy datasets. These are fictional dataset for exploratory data analysis (EDA) and to test simple prediction models. These datasets can be used to train fine-tuned models for toy detection by exploring datasets and export the dataset into a required valid format.
- Feature extraction. The module can be used to extract features in a format supported by machine learning algorithms from datasets consisting of formats such as text and image (e.g. Bag of words). A bag-of-words model (BoW) is a method of extracting features from text in documents for use in modelling in machine learning algorithms.

In Python, it is an extremely popular language because nearly every concept is easily validated with 20 to 30 lines of code making it a very powerful and rich language to facilitate the programmers' effort in the development of AI applications.

6.2.3 Intelligent Computer-Aided Design (ICAD)

ICAD is a knowledge-based system (KBE) that could capture the intention underlying the product design. Basically, KBE is the integration of object-oriented programming, AI, and computer-aided design (CAD). Hence, KBE systems have the objective of gathering product and process information to enable engineers and other professionals to reproduce or model engineering processes. Therefore, KBE systems allow engineers to generate a geometric design and other essential design information from a model of a product or process. This means that when a particular product or process specifications are modified or new design versions are required, the rules in their current model evaluate the new specifications and a new geometric design is created automatically.

The outcome of a KBE is to automate parts of the process or its entirety from a specified product model. As a result, the product model functions as a computer representation of the design process, which describes the how, why, and what of a design. This also means that a KBE product or process model can contain legacy programs, physics-based analysis, spreadsheets, databases, cost models, and other information from its external environment. Generally, KBE becomes a powerful way of integrating the knowledge from many entities, divisions, departments, and people that together contribute towards the creation of an organisation's products. [Figure 6.5](#) illustrates the ICAD process in which the engineers would specify the input requirements, and the information is extracted from the organisation's internal and external environment and existing CAD. All this information is exploited by the ICAD system as its primary input data for processing, which generates the CAD models, process design, material requirements planning data, bill of materials, and the associated cost analysis, and ancillary reports

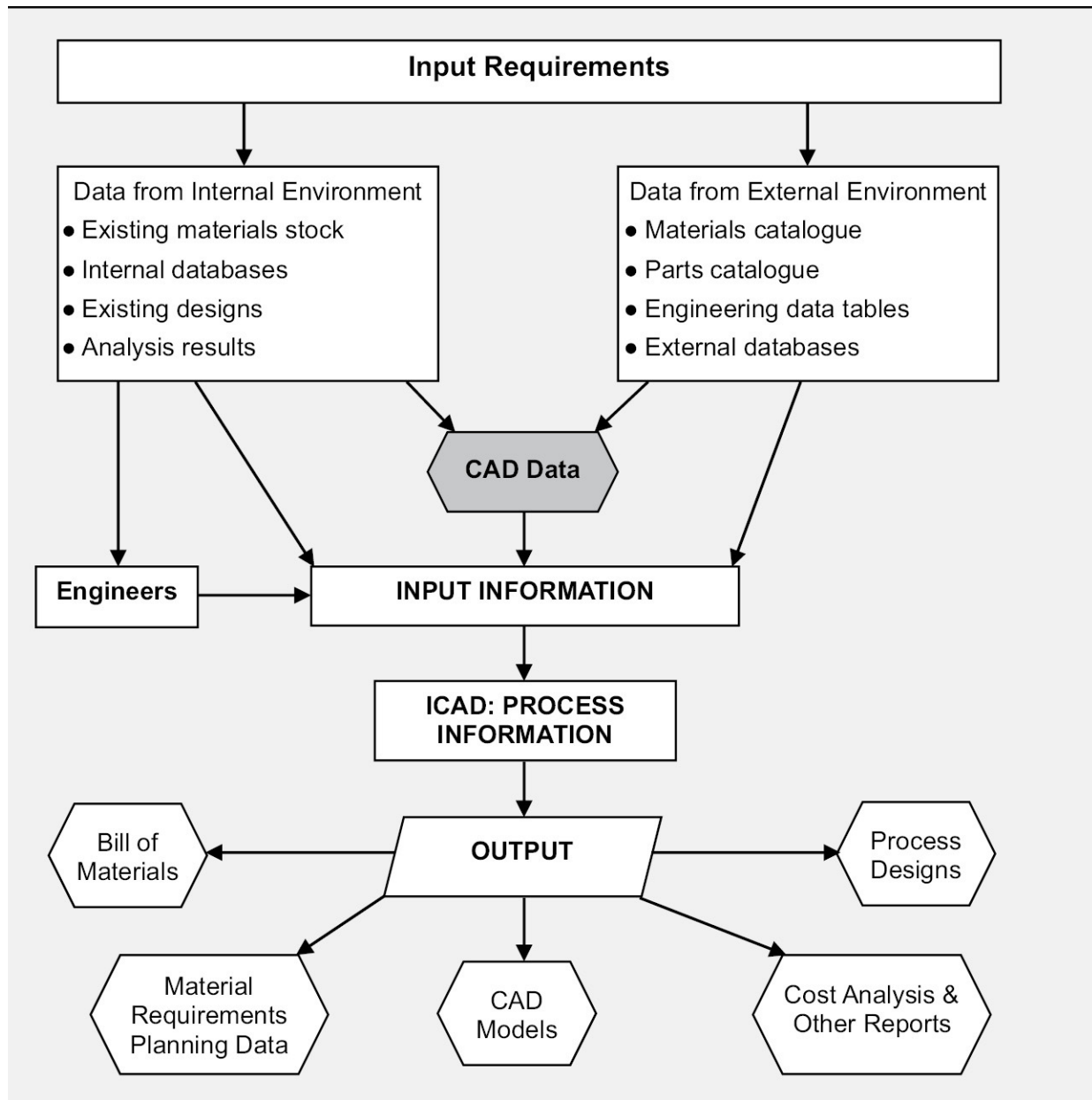


FIGURE 6.5 ICAD Knowledge-Based System. [🔗](#)

An example of an ICAD system is the Fujitsu Manufacturing Industry Solution COLMINA Design and Manufacturing Support iCAD SX(1) for machinery design, which is claimed to be powered by the world's fastest 3D CAD engine. The system reduces design work for large-scale industrial machinery. Fujitsu's ICAD package is an integrated system that processes everything from conceptual design studies to shop drawings using 3D

models of machinery, steel parts, and conduits. Their latest model released in 2020 implements significant enhancements to memory access processing and processing speed of ultra-largescale data expanding the number of parts that can be handled to as many as 3 million parts in the design examination and verification of machine equipment.

Fujitsu has also developed and implemented new design functions that allow mechanical, electrical, and control designers to retain and use operation and manufacturing information, as well as functions that enable the use of 3D data in manufacturing and maintenance services. This feature facilitates the efficient data sharing between departments in a digital environment by linking common 3D data from design upstream to manufacturing and maintenance services, thus aiding in the decrease of lead time. From the initial design stage, the system can directly add to the 3D data, the motion information including the amount of movement, sequence of movements, and timing of a 3D device model. This enables the examination of movements, processes, and cycle times while confirming the positional relationship with peripheral parts. Additionally, the 3D equipment models, and machining information make it possible for the system to calculate the cost of parts and check the consistency of surface roughness and fitting tolerance between parts in the early phase of design.

Fujitsu's ICAD system helps to shorten the development period and reducing the number of manufacturing processes in the manufacturing industry. It enables design departments to hasten product development and enhance quality, concurrently with improving the efficiency of processing and assembly operations. Intelligent Computer-Aided Design (ICAD) techniques have been used as a basis to facilitate the development of Integrated Product Design (IPD) systems by providing a more "complete" design tool to assist the designer in all phases of the product design. One should note that the main goal for implementing an IPD system is to achieve an integrated approach, where all the primary constituents, such as the CAD tools, numerical analysis tools, data bases, knowledge bases, and expert systems, work cooperatively to perform various tasks associated with that design task.

6.2.4 KL-One

KL-One is a knowledge representation system used in AI, which was developed in the early 1980s by John McCarthy and Patrick Hayes. KL-One is a frame language based on the formalism of description logics, in the tradition of semantic networks and frames. KL-One has an automatic reasoning engine that can authenticate a concept (referred to as a frame) ontology and generate new information through logical reasoning about the ontology based on the initial information provided by an expert in a particular field of knowledge. Ontology is a means of how to determine if things exist or not, as well as the classification of existence. It attempts to take things that are abstract and establish that they are, in fact, real.

The principal unit of information in KL-One is called a Concept, which denotes a set of objects often referred to as Frames. These form hierarchies using subsume-relations. The terminology of KL-One views a super class as subsuming its subclasses, with multiple inheritance being permitted. A concept (frame) is said to be well-formed if it inherits from more than one other concept; and all the concepts, except the top concept (usually `THING`), must have at least one super class. In KL-One, descriptions are separated into two fundamental types of concepts, namely primitive and defined. The primitive type belongs to the domain concepts, which are not fully defined. In other words, the given and known properties of a concept are not adequate to classify it. Primitive type concepts are also viewed as incomplete definitions. On the other hand, defined concepts have complete definitions where their properties are fully known. Hence, these properties (attributes) provide all the required (or adequate) knowledge and conditions to classify the concept. Moreover, a slot-concept is called roles, and the values of the roles are role-fillers. There are several types of roles that may be applied in different situations. However, the most common and important role type is the generic RoleSet, which describes the fact that the role may be filled with more than one filler. For example, one can construct a Concept representing the set of all people where each person has a birthdate, hence `PERSON` is a Concept and has a Roleset `BIRTHDATE`.

The best way to illustrate the above principles regarding KL-One is through an example. Assume that one wants to denote that all horses are essentially mammals. This is achieved by stipulating that the Concept denoting mammals subsumes the Concept denoting horses. Hence, a subsumption is represented by a link between Concepts called SuperConcept. KL-One maintains a catalogue (taxonomy) of Concepts with the top of the taxonomy being THING, which is defined to subsume every other Concept. On the other hand, there is no corresponding bottom Concept. Additionally, Concepts inherit the elements from their subsumers, which means that each member of the set represented by a Concept inherits the same properties that are specified by the elements of the Concepts subsumers, including the elements of the Concept itself. Therefore, the HORSE Concept inherits all the elements of MAMMAL. In other words, each horse has all the properties or attributes that are stipulated for mammals and that mammals essentially include horses.

An important feature of this Concept aspect is what is known as Classification. Classification is a process that reveals through logical reasoning that a set signified by some Concept automatically includes the set represented by a second Concept, but where no subsumption relation between the Concepts has been explicitly recorded. Therefore, classification determines the subsumption relations between Concepts and authenticates SuperConcept links between them. Hence, through the classification process, the system automatically categorises a Concept at its proper location in a KL-One taxonomy since it determines the appropriate SuperConcept links between it and other Concepts. To summarise, KL-One consists of five core factors as follows:

- Concepts: These are abstract objects that represent universal classifications or groups of objects within a domain that can be further specified or subclassed using several taxonomic relationships (e.g., subconcept, superconcept) and inheritance properties (e.g., attributes, roles).

- **Relations:** These are binary predicates that describe an activity and depict meaningful links between the different concepts or objects within the domain. These can be used to define various constraints or restrictions on concept attributes and relationships (e.g., role hierarchy properties).
- **Individuals:** These are explicit instances of concepts or objects within the domain, which can be categorised or ordered according to their taxonomic relationships and inheritance properties and be designated with various attributes or roles based on their conceptual definitions and constraints.
- **Axioms:** These are logical statements or rules that describe the essential and adequate conditions for concepts or individuals within the domain to be categorised according to certain taxonomic relationships or inheritance properties.
- **Inference engine:** This has the function for processing and evaluating the different logical statements or rules related to concept definitions, constraints, and relationships within the domain. These are able to be applied for deriving new facts or conclusions based on the given knowledge base.

[Figure 6.6](#) provides an example regarding the KL-One frame language concepts. The THINK concept encompasses the taxonomy of the natural world into five kingdoms, namely Animal, Plant, Fungi, Protists, and Monera. The diagram illustrates that within the animal kingdom only four aspects are mentioned, each of which are depictions that may be applied to describe many individual things in the world. As stated previously, several of these Generic Concepts may be formed out of the others (e.g., the Generic Concept HUMAN would have MAMMAL). There are also some structure-forming operations presented for building Concepts, which blend one or more general Concepts, and a set of restrictions on those Concepts.

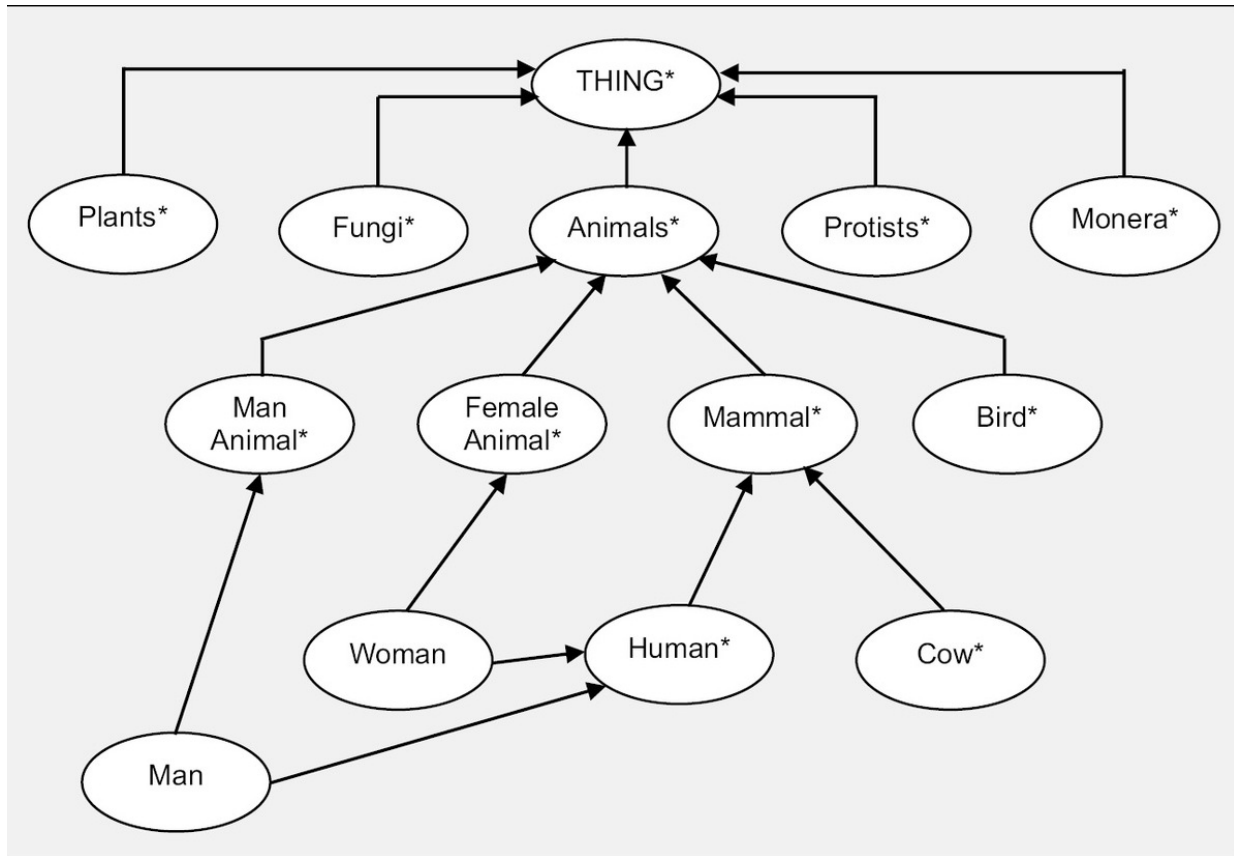


FIGURE 6.6 Example Illustrating the KL-One Frame Language Concepts. [📄](#)

As explained previously, the components of a Concept consist of its subsuming Concepts (i.e., superConcepts) and its local internal structure expressed in Roles. Roles describe possible relationships between occurrences of the Concept and those of other closely associated Concepts (i.e., its properties, elements, etc.), and structural descriptions that articulate the interrelations among the Roles. Additionally, a well-formed KL-One Concept should have more than one superConcept (if there are no local restrictions), must differ from its superConcept in at least one restriction, or be a primitive type. A Concept with no local restrictions is defined as the conjunction of its superConcepts. Hence, once a Concept like HUMAN is specified in terms of one like MAMMAL, one is indirectly identifying that the “more general” subsumes the “more specific.” This means that if Concept X subsumes Concept Y, then every individual that can be described by Y can also be described by X. Hence, when representing the Concept

HUMAN so that the Concept MAMMAL subsumes it, the attributes of being a human includes the characteristics of being a mammal and therefore denoting that any human must be a mammal.

It should be noted that KL-One taxonomies will always have one root Concept, typically identified as THING. A root Concept subsumes all other Concepts and has no superConcepts. Primitive Concepts (denoted with an * in [Figure 6.6](#)) are not fully defined, however Concepts not denoted by an “*” are fully defined by their components. Furthermore, [Figure 6.6](#) illustrates that a Generic Concept can have various superConcepts and subConcepts. For instance, the Concept MAN has both MAN-ANIMAL and HUMAN as its superConcepts. Hence, both MAN-ANIMAL and HUMAN subsume MAN. In this example, the Concept MAN is defined to be simply the combination of the two. Moreover, [Figure 6.6](#) shows that a Concept such as MAMMAL does not receive any of its properties from the Concept HUMAN. In other words, a Concept’s representation is precisely determined by its subsuming Concepts together with the information related explicitly with the Concept. These principles suggest that KL-One provides several benefits regarding AI, including:

- **Articulateness.** KL-One is a highly expressive language that can denote a wide diversity of knowledge, including factual and procedural intelligence. This feature enables it to portray complex ideas and relationships, making it a very powerful instrument for AI applications.
- **Legibility:** Representation of knowledge in KL-One is both human-readable and machine-readable, thus facilitating the understanding and manipulation of knowledge by humans and AI systems.
- **Declarative Language:** KL-One is a declarative language, which means that knowledge is denoted independently of any specific reasoning or inference algorithm. This facilitates the development of new reasoning algorithms, and to port existing algorithms to new domains.

- **Rational Classifier:** KL-One applies a deductive classifier. Classifiers are commonly used in supervised learning systems where the correct class for each input example is known during training. The purpose of a classifier is to learn from the training data to enable it to make accurate predictions on unseen data. Hence, a deductive classifier acts as an automated reasoning engine that validates a frame ontology and deduce new information to allow the system to infer new knowledge based on the existing knowledge base, augmenting its problem-solving capabilities.
- **Effective Toolkit:** KL-One comes with a well-developed collection of tools that includes editors, compilers, and inference engines. This toolkit facilitates the development process of applications, mitigating the time and effort needed to develop applications.
- **Extensibility:** KL-One is designed to be extended effortlessly. Hence, it has the inherent ability to extend a system and reduce the level of effort required to implement the extension. This principle allows improvements without impairing existing system functions. Hence, new representations and reasoning tools can be added to the language as required, which enables KL-One to adapt to a wide range of AI applications, boosting its versatility and utility.
- **Broad Applicability:** KL-One has been applied in several AI applications, including natural language processing, expert systems, and machine learning, illustrating its effectiveness and adaptability in different AI environments.

There are several main features of KL-One in AI that makes it a rich and powerful knowledge representation language for use in AI. These include the following:

- KL-One acts as a knowledge representation system in AI and the benefits outlined previously enable it to represent a broad scope of knowledge, depicting complex ideas and relationships.

- KL-One has a unique naming structure called “unification” for easy identification of objects and relations. This feature helps in amalgamating new knowledge into the representation.
- KL-One is based on well-defined semantics that permits for vibrant and consistent reasoning, using inference rules to draw new conclusions from the knowledge being represented.
- KL-One’s deductive classifier with its automated reasoning engine validates a frame ontology and determines new information, enhancing its problem-solving abilities.
- KL-One’s comprehensive toolkit facilitates, including its extensibility and flexibility features facilitate the AI development process.

These key features allow KL-One to be applied in many industrial fields. For instance, it can be applied to depict a medical knowledge base, where the patients are shown as frames, and their symptoms and treatments are denoted as predicates (i.e., describing the activity not the patient). This permits for useful probing and reasoning about patient data and facilitates the development of intelligent systems for diagnosis and treatment planning. KL-One’s knowledge representation features make it ideal for applications regarding medical information processing; expert systems; natural language processing; computer vision for identifying and understand objects and people in images and videos; and predictive modelling.

KL-One is ideal for AI applications development because it allows researchers and developers to definite complex domain-specific knowledge within formalised ontologies or conceptual models, especially for representing taxonomic relationships and concept hierarchies. Additionally, it provides several key concepts and mechanisms for describing, categorising, and deduction about diverse concepts or objects within an application domain, including taxonomy and inheritance attributes; controls and restrictions on concept features; and logical rules and inference abilities. One should note that KL-One is based on the notion of frame-based systems that partition knowledge into small, self-contained units

called frames, with each frame containing information about a specific aspect of the world, and the frames are interconnected to form a network of knowledge.

6.2.5 Official Production System (OPS5)

OPS5 is a fifth-generation programming language that has the objective of solving problems using constraints provided to a program, rather than using an algorithm written by a programmer. It belongs to a family of rule-based computer programming languages. This feature distinguishes it from fourth-generation programming languages. The primary difference between fourth generation and fifth generation programming languages is that the former are designed to create specific computer programs while fifth-generation computer languages are designed to make the computer solve the problem for the user.

This means that the programmer, instead of writing and implementing an algorithm to solve a problem, must only concern itself about what problem needs to be solved and what conditions need to be met to solve the problem. OPS5 and other fifth-generation computer languages are used mainly in AI research. OPS5 being a rule-based programming language is suitable for designing for implementing efficient expert systems in disciplines such as AI, medical diagnosis, and decision support systems.

Scientists Herbert Simon and Allen Newell from Carnegie-Mellon University who for many decades were studying human problem-solving processes, came up with the notion of the production system. In their view, a production system denoted knowledge as a set of condition-action rules (also referred to as “if-then rules”). Hence, if the conditions for a rule are fulfilled, the specified action stipulated by the rule is carried out. Simon and Newell’s idea was taken up by Charles Forgy in the late 1970s who developed the OPS5 programming language specifically to cater for the rule-based problem-solving idea. [Forgy_\(1981\)](#) claims that OPS5 is an interpretive computer programming language implemented in LISP and

BLISS that is mainly used for applications in areas of AI, cognitive psychology, and expert systems.

[Figure 6.7](#) illustrates that basically OPS5 stores data in working memory, and the “if-then rules” (Rules) are stored in production memory. The Rules in OPS5 are completely independent of one another and can be placed in production memory in any sequential order. OPS5 works by matching the data in working memory with the rules (and their associated conditions) in the production memory. If there is a match between the data and a specific rule (i.e., its conditions) in production memory, the actions associated with the specific matching rule are executed. Possible actions may include altering the contents of working memory that might then match the conditions of another rule; reading data from a file; displaying data on the monitor; and calling external algorithms (programs). There are versions of OPS5 that permit an action of a Rule to create a new rule, which permits systems to learn.

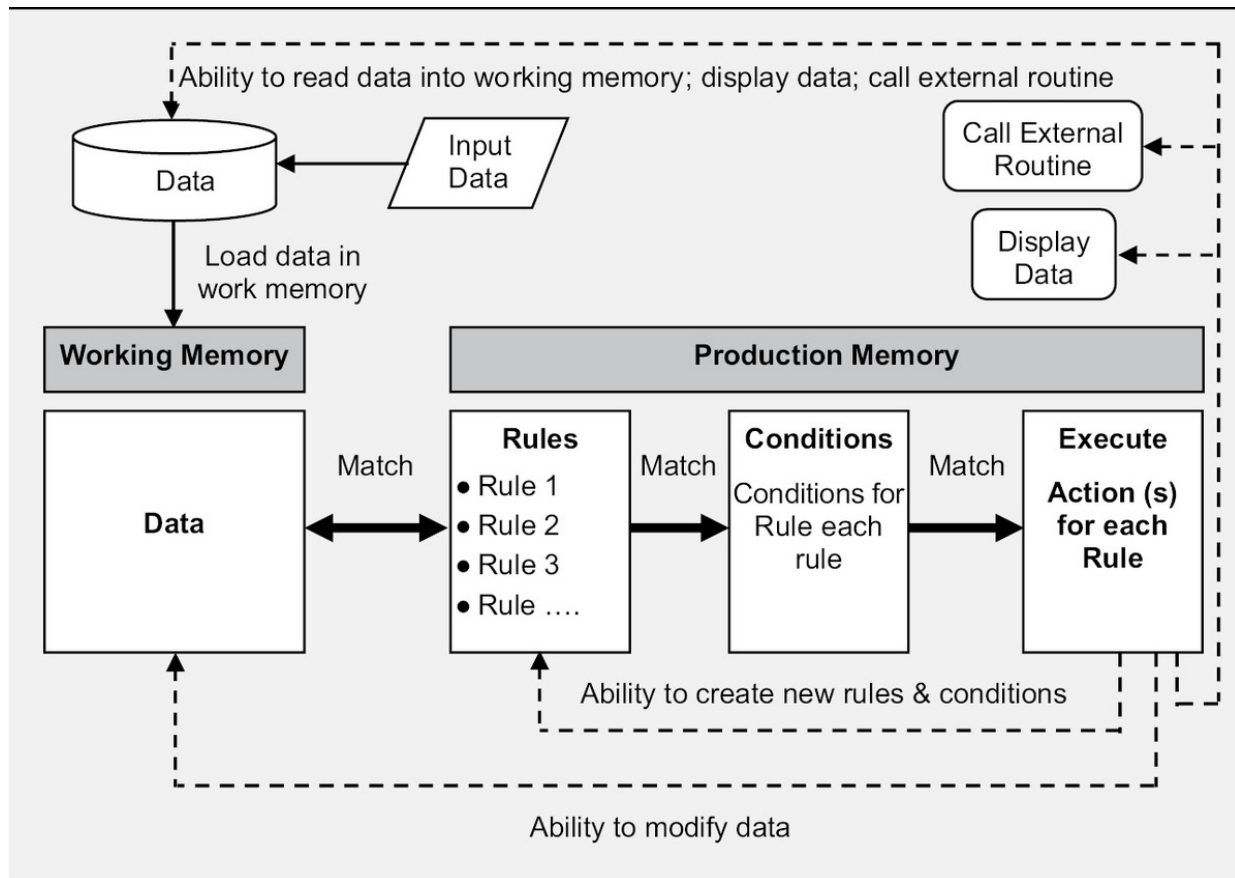


FIGURE 6.7 OPS5 Language Processing Concept with Self-Learning Ability. [↗](#)

[Figure 6.7](#) implies that OPS5 has an integrated deduction procedure and a forward chaining process, which is a form of reasoning for an AI expert system that starts with simple facts and applies inference rules to extract more data until an objective is achieved (i.e., ability to change rules, and modify and add data). OPS5 begins from a set of data in working memory and initiates a sequence of rules to be applied until a goal (conditions) is satisfied. Although backward chaining is not directly available in OPS5, this feature can be implemented with greater effort. Backward chaining is a situation in which a particular goal to be accomplished triggers subgoals to be fulfilled until a specific problem solution is attained.

Allen Newell's research group in artificial intelligence had been working on production systems for some time, but *Forgy's* implementation, based on his Rete algorithm, was especially efficient so that it was possible to scale up to larger problems involving hundreds or thousands of rules. OPS5

adopts the blend of basic principles of the execution engine for a Petri Net extended with inhibitor arcs and the Rete algorithm. Petri Nets were originated by Carl Adam Petri in 1962 with the underlying concepts being broadened and enhanced for application in a variety of fields, including programming languages.

[Figure 6.8](#) provides a simple example of a Petri Net illustrating the basic concepts. A Petri Net should be visualised as a collection of directed Arcs (symbolised by an arrow) that link Places (symbolised by a circle) and Transitions (symbolised by rectangle). However, Places may hold Tokens (symbolised by a black spot), and the status of a net is determined by its assignment of Tokens to Places. Furthermore, Arcs have capacity 1 by default; if the capacity is not a 1, the capacity is denoted on the arc. In the example at [Figure 6.8](#) the capacity is 1 by default. Additionally, Places have an infinite capacity by default, and transitions have no capacity at all, and cannot store tokens. Moreover, Arcs can only link Places to Transitions and vice versa.

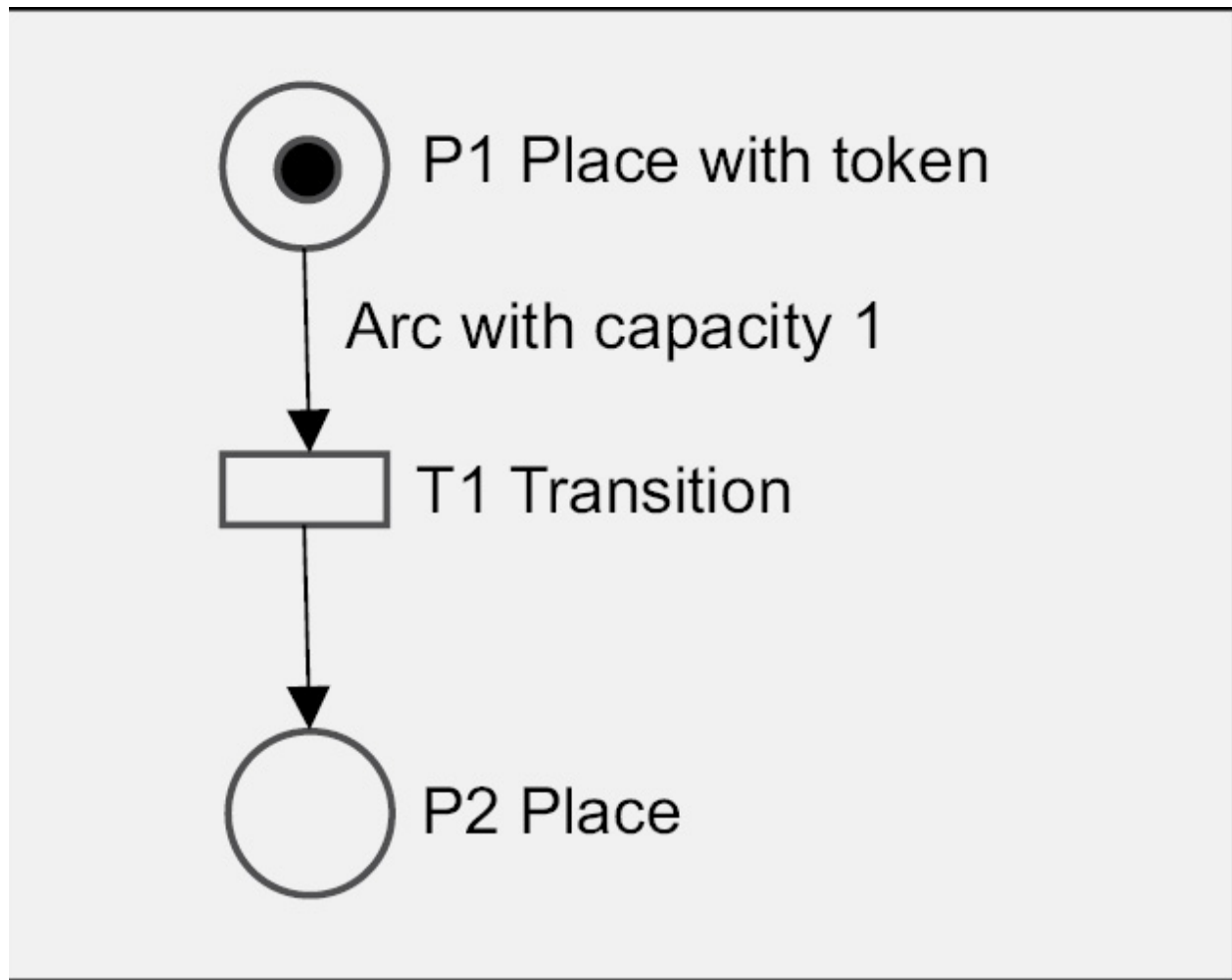


FIGURE 6.8 Example of Petri Net Illustrating Basic Concepts. [↗](#)

[Figure 6.9](#) depicts a series of Petri Nets that show how the status of a Petri Net changes depending on the assignment of Tokens to Places through the Transition condition. A Transition is enabled when the number of tokens in each of its input places is at least equal to the Arc capacity (weight) going from the place to the Transition. [Figure 6.9](#) (A) illustrates that in this case, the number of Tokens in each of its input Places is 1 and the Arc weight has a default capacity of 1, which satisfies the condition for the Transition to be enabled. Moreover, an enabled Transition may fire at any time. [Figure 6.9](#) (B) shows the changed status of the Petri Net when the Transition is fired, with the Tokens in the input Places being transferred to the output Places, corresponding to the Arc weights and Place capacities.

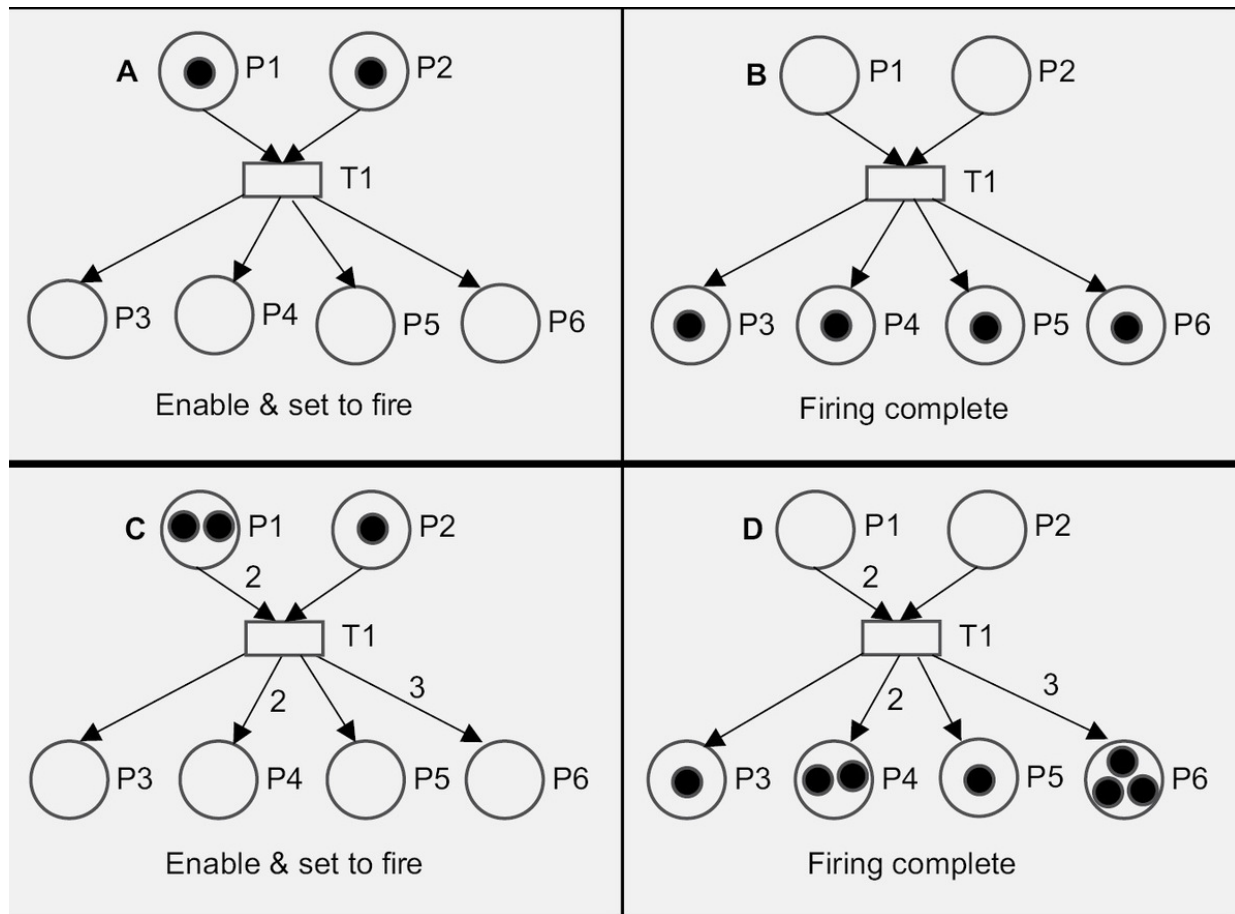


FIGURE 6.9 Example of Petri Net Illustrating Various Additional Basic Concepts. [📄](#)

[Figure 6.9](#) (C) illustrates what happens when Places and Arcs have different token capacity and its status is ready to fire, while [Figure 6.9](#) (D) shows the outcome when fired. Notice that in [Figure 6.9](#) (C), Place P1 (with 2 Tokens) has an Arc capacity of 2, while P2 (with 1 Token) has the default Arc capacity of 1. However, the Arc capacity after Transition is 1 for P3 (with 1 Token), 2 for P4 (with 2 Tokens), 1 for P5 (with 1 Token), and 3 for P6 (with 3 Tokens). Hence, the outcome shown at [Figure 6.9](#) (D) corresponds to these Arc capacities when fired.

In other words, when a Transition fires, it receives the Tokens that enabled it from the input Places and disseminates Tokens to output Places matching the Arc capacity weighting. If the Arc weights are all identical, it gives the impression that the Tokens are transferred through the transition.

However, if Arc weights differ, it gives the impression that Tokens may vanish or be generated. In fact, in [Figure 6.9](#) (C) the number of Tokens is 3 (P1, 2 and P2, 1) but after firing at [Figure 6.9](#) (D), the number of Tokens is 7 (P3, 1; P4, 2; P5, 1; and P6, 3). Hence, four additional Tokens have been generated (producing output Tokens according to Arc weight). There is also an exceptional type of Arc known as an inhibitor Arc, which is applied to reverse the logic of an input Place. [Figure 6.10](#) shows the application of an inhibitor Arc, which is symbolised by having a small box at the end of the arrowhead. [Figure 6.10](#) illustrates that the Transition cannot fire since the Token in P2, which has an inhibitor Arc, prevents it from doing so. The above provided a concise description of the Petri Net concepts.

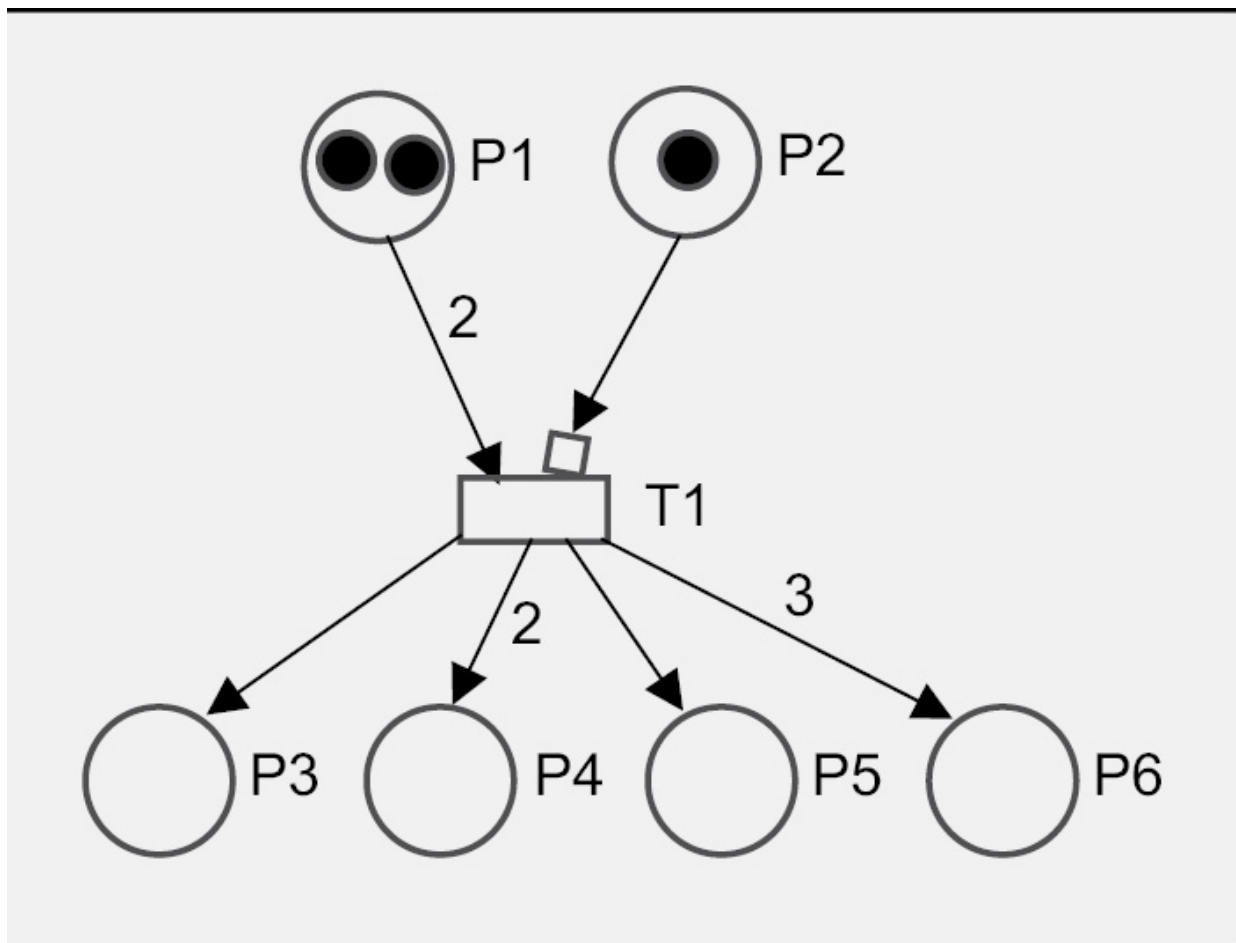


FIGURE 6.10 Example of Petri Net Illustrating Inhibitor Arc Concept. [📄](#)

As stated previously, OPS5 adopts a blend of Petri Net principles and the Rete algorithm. The Rete algorithm is named after the Latin word “net.” It was designed by Charles Forgy in a working paper submitted in 1974. This algorithm is typically a pattern matching algorithm for implementing rule-based systems to apply many rules or patterns to many objects, or facts, in a knowledge base efficiently. The Rete algorithm has the objective of ascertaining which of the system’s rules should fire based on the stored factual data.

The Rete algorithm structures its data as a tree-like formation (similar to the Petri Net). Each node in the tree represents a test that is executed when matched with the data being added or removed from memory. Every node will have one or two inputs but possibly many outputs. As an additional item of data enters the tree, it is filtered through the tree until it reaches a terminal node. If the data item reaches a terminal node, this signifies that a new activation record is created with all the data points that triggered the rule to be true. Rules in Rete systems must be as explicit as possible to achieve better performance. This is achieved by composing rules that search for very detailed items in a given piece of data. This produces a path in the Rete tree to be navigated in a rather determinant approach, which reveals the efficiencies that are fundamental to the Rete algorithm.

Like other methods, the Rete approach has a set of conventions or terminology for explaining its application. A ruleset is a knowledge base consisting of one or more business rules. Each rule in the ruleset denotes some kind of knowledge. Rules are normally defined as an if-then or condition-action rule. The if-then rules are called Rete Rules. An example of a Rete Rule is: *If age > 49 then assign screening-status = 1.*

Note that the words “if,” “then,” and “assign” are keywords. The “If” part signifies a condition and the “then” part represents action. It is possible to have more than one condition, which need to be logical operators. The “then” part denotes one or more actions. Therefore, in the above example, Rule means that if a person’s age is more than 49 then that person is to be screened (0 = not to be screened; 1 = to be screened). Therefore, if the person’s status as to be screened (or not) needs to be verified, then data

(called fact) is required about the person's age. Hence, as the number of distinctive variables or definitions increases, the number of facts also increase. For example: *If age > 49 and gender = "female" then assign screening-status = 1*

To execute (perform) the above rule, two data items or facts are required, namely the "age" and the "gender." If the status of two or more persons is being verified then there must be a data set for each person. This particular example is a simple event but in real life circumstances there may be hundreds and even thousands of facts where the rule should be applied. Furthermore, there may be a voluminous number (thousands) of different rules as well.

A Rete algorithm requires that a complete ruleset must be given to the rule engine to broaden processing. Typically, the rule engine matches each rule (i.e., condition) in the ruleset with specified facts to determine whether to fire (i.e., execute) the rule (or not). This procedure is referred to as the pattern matching process, which is conducted repeatedly. Additionally, the list of facts can be altered for each cycle. In other words, the list of facts is continually being updated by the inclusion of new facts or/and deletion of old facts from the list of facts.

Therefore, these modifications may trigger (cause) previously unsatisfied (unfulfilled) patterns to be fulfilled (satisfied). Additionally, during each cycle, the set of rules fulfilled must be updated and retained (stored). Often, the actions of the rules change just a few facts in the list of updated facts. This process is known as temporal redundancy. The algorithm is highly efficient because the rule engine does not verify each rule by directing the search for all the facts since most of them are not modified and doing so would delay the process. This is achieved because the algorithm is designed to remember (recall) what had already been matched from cycle to cycle and only process the necessary changes for the newly added or removed facts.

The Rete algorithm structure consists of a network of nodes. As described previously, it is designed to achieve high processing speed by storing in memory the state of each matching process from cycle to cycle

and only computing those with modified facts. Therefore, the status of the matching process is only updated when facts are added or/and removed. Hence, a lower number of modifications result in a faster matching process.

[Figure 6.11](#) illustrates that each rule has a corresponding inference (deduction) cycle that consists of three stages, namely the match, select, and execute phases. Each stage has a distinct function. For instance, in the match phase, the conditions of the rules are compared to the facts to ascertain which rules are to be executed (fired). Therefore, the rules whose conditions have been satisfied (fulfilled) are stored in a list called agenda-ready for executing (firing). Finally, one of the rules from the list of rules is selected to execute (fire). The selection(s) strategy is contingent on several factors, namely the priority, recentness of usage, rule specifications, or on other specified conditions. The rule selected from the list is executed (fired) by conducting (performing) the actions associated with the rule. The action may consist of several events, such as verification (confirmation); performing a user defined or built-in function; or implementing a decision table through a decision table engine, or otherwise.

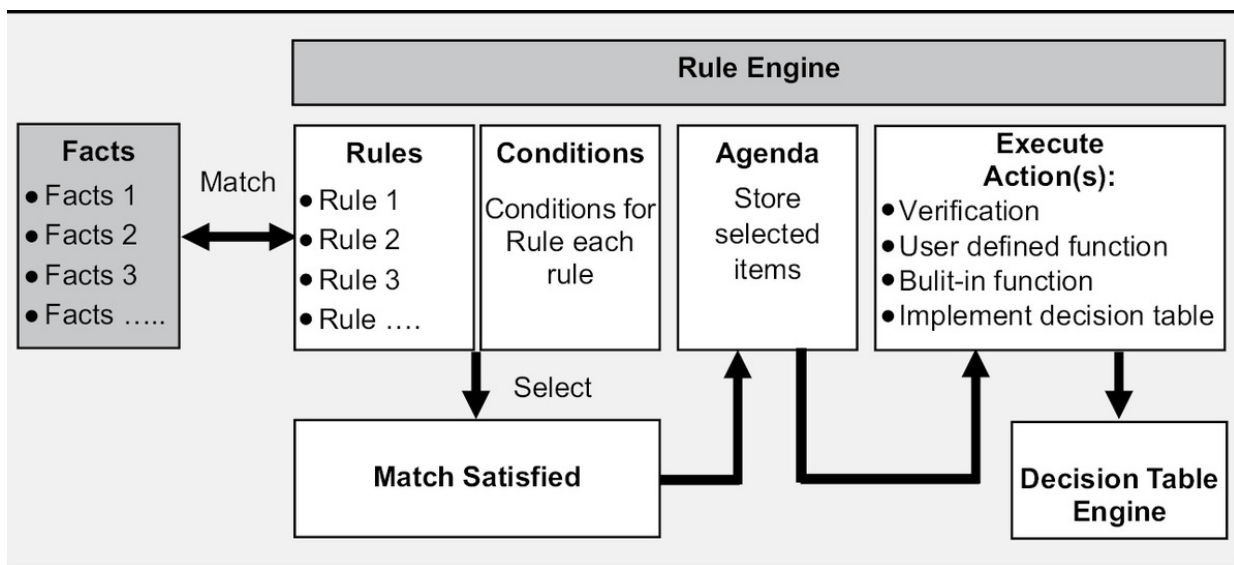


FIGURE 6.11 Illustration of a Rule Engine Inference Cycle. [↗](#)

The construction of a Rete network requires the adherence to certain conventions. A Rete network is a direct acyclic graph that consists of nodes

representing patterns in the conditions of the rules and is bipartite with no graph cycles. Moreover, its nodes are directionally related to each other and do not form a directional closed loop. The nodes act as filters to test (check) the incoming tokens, and firing (sending) only those tokens that qualify (i.e., pass the test). The Rete network consists of two segments, namely the alpha network and beta network. The alpha network comprises of nodes known as alpha nodes, with each node having one input that defines intra-elements. The beta network comprises of beta nodes with each node having two inputs to define inter-element conditions.

A Rete network begins with the Root node known as the Rete Node. Following the Root node are Kind nodes, with a Kind node for each fact type. Alpha nodes are connected to the corresponding Kind node and are created for each pattern. If a condition has two patterns, then two alpha nodes are created, with each alpha node being associated with a memory known as alpha memory. The alpha memory is used to recall the matching facts. The alpha nodes are linked to beta nodes. Beta node can only receive two inputs. Therefore, if there are three alpha nodes then the first two nodes will be linked to one beta node. Subsequently, the output of this beta node and the third alpha node will be linked by another beta node. Hence, the beta nodes may support partial matching, with each beta node having a memory to store joined patterns.

In the beginning, the tokens move into the root node. Later, the affirmation of each fact generates a token. The network then divides a branch for each token type, with each Kind node receiving a copy of the token and executing the SELECT operation to select tokens of its kind. The kind node sends a copy of the token node it receives to the alpha nodes. On receiving the token, the alpha nodes perform a PROJECT operation and take out elements from the token that match with the variables of the pattern. Basically, the alpha node assesses the conditions, and the beta node then ascertains the feasible cross product for a rule. When this is achieved, the actions in the rule will be performed.

The construction of a Rete Network is best illustrated through an example. Assume that the local health authority has a breast cancer

screening scheme for females between the 50 and 70 years of age. The age facts are java type, while the gender facts are xml. The rule for this would be as follows:

If age>49 or age<71 or gender = “female” then screen-status = 1

This rule would be depicted by the Rete Network shown at [Figure 6.12](#). The example at [Figure 6.12](#) has two kinds of nodes with two types of facts, namely: Kind 1 (java) and Kind 2 (xml). There are three patterns, namely: age>49, age<71 and gender=female and therefore three alpha nodes will be generated. The example shows that Alpha 1 and Alpha 2 denote the first two patterns and are linked to Kind 1, while Alpha 3 denotes the third pattern and is linked to Kind 2. Alpha 1 and Alpha 2 are linked to Beta 1, while Alpha 3 is linked to Beta 2.

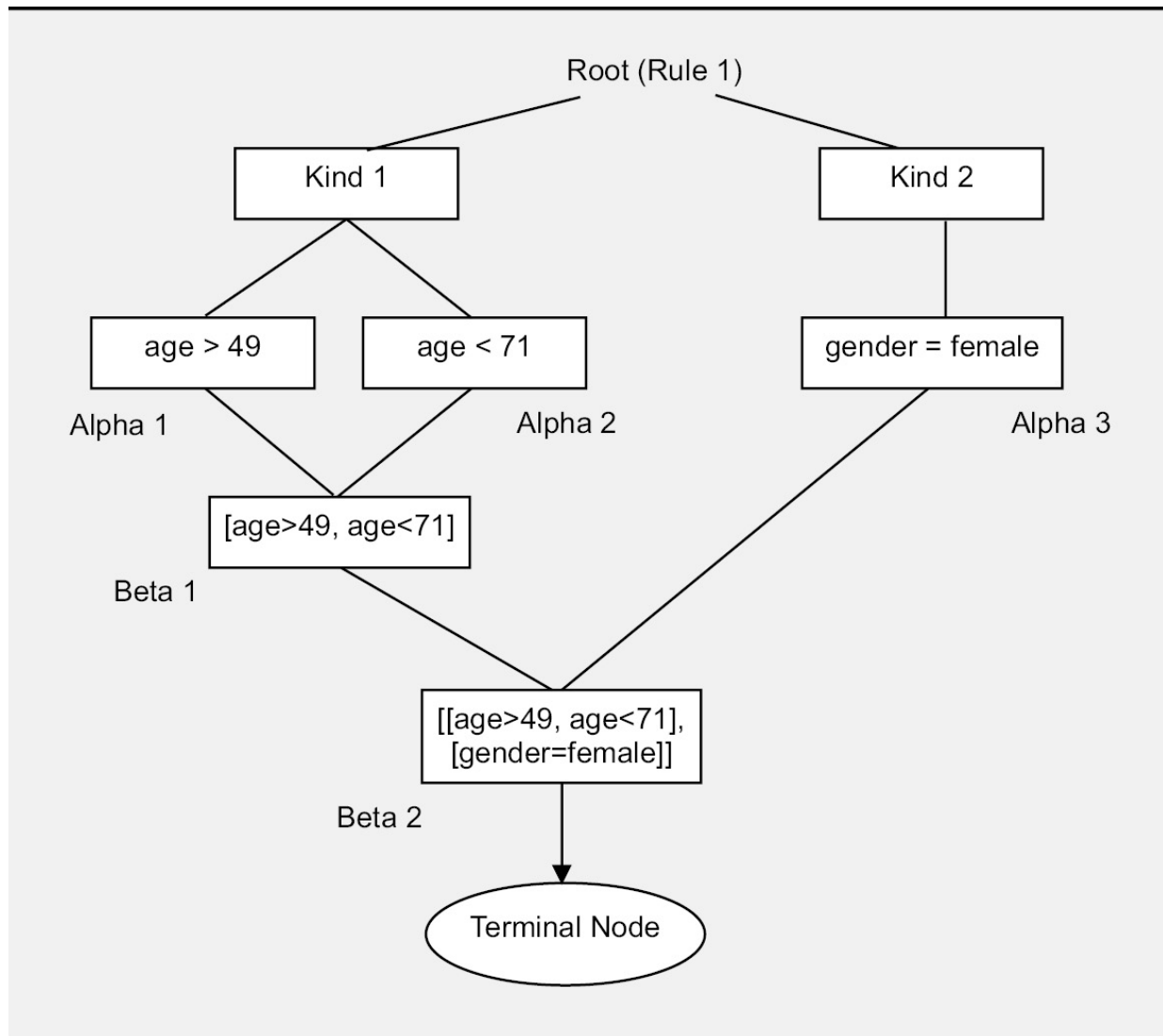


FIGURE 6.12 Example of Constructing a Rete Network.

A token is generated as a value for age enters the root. Additionally, this token will be copied and disseminated to kind nodes. Hence, Kind 1 will accept the token because the fact type is java and forwarded to Alpha 1 and Alpha 2. If the value fulfils (satisfies) the constraint, then the token will be forwarded to Beta 1 and then to Beta 2. In this intervening time the value of gender enters the root, which is received by Kind 2 (since it is xml type). Alpha 3 obtains it and verifies if the value fulfils (satisfies) the constraint, gender=female. If this is the case (i.e., yes) then it permits the token to be forwarded to Beta 2. Moreover, if the values (facts) match with the

condition in the Beta 2 then the rule will be added to a list for firing (execution). Therefore, if a particular case has age between 50 and 70, and the screen-status is 1, a letter will be sent to invite the individual for a breast screening test.

The Rete algorithm is very efficient in terms of speed. However, it needs a lot of memory because of its feature of saving the state of the system using pattern matches and partial matches. Hence, the main aspects of the Rete algorithm are its ability to retain partial matches between consecutive updates because of the new factual data, and it endeavours to reuse overlapping rules in the matching process. The demand for memory computed by the number of rules (R) multiplied by the number of declared facts (F) multiplied by the average number of patterns per rule (P). Therefore, if $R = 100$, $F = 10,000$, and $P = 10$, the storage needed is 10mb. Research has shown that the Rete algorithm can be successfully applied with Petri nets and that Petri nets can be applied as an underlying mechanism in rule-based systems ([Burdescu and Brezovan, 2001a](#), [2001b](#); [Hura, 1993](#); [Murata and Yim, 1995](#)). Additionally, [Dohlborn \(2011\)](#) illustrated that Petri nets can be applied as efficiently as a rule-based technique created by the Rete algorithm with temporal extensions.

6.2.6 Mercury and Prolog

Both Mercury and Prolog are declarative computer programming languages (or logic programming). Prolog has been in use since 1972 and was specifically designed for developing logic-based AI applications. With Prolog, developers can set rules and facts around a problem, and then Prolog's interpreter will use that information to automatically infer solutions. Mercury was released around 1995 and is a souped-up version of Prolog that has enhanced features to provide better performance.

Declarative programming is a high-level programming concept that basically means that one would state the outcome of a task, and the compiler determines the steps that need to be executed to achieve the outcome. Hence, a programmer would write code that describes what is to

be done and what outcomes (results) is to be achieved, rather than how they will achieve the results. This is opposite to imperative programming where functions are completely coded for each step required to achieve the desired outcome. In declarative programming it is the compiler that works out the steps required to achieve a task. In other words, declarative programmes depend on preconfigured facilities in the language without specific case-by-case instructions on what steps need to be executed to complete a particular task.

Declarative programmes are context (circumstance) independent because a programmer only needs to declare the eventual goal (the what) and not the logical steps involved to attain the goal (the how). This implies that the same programme expressions have the same meaning in different circumstances and consequently can be applied in different contexts. Therefore, declarative programs such as Mercury and Prolog are much simpler to work with because they formulate computation logic without complicating the programming code with control flows. These types of programmes are easier to understand because they have less ambiguous processes, implied dependencies and changeable states that simplify programming code and readability.

Mercury being a more recent development has more advanced static analysis and error detection features. It is also supported by a highly optimised execution algorithm that is far more efficient than other logic programming systems and achieves similar performance to conventional programming systems without complex control flows. Mercury is considered as a pure logic programming language that is designed to generate sizeable, speedy, reliable programs. The coding syntax of Mercury is based on Prolog, but the way it relates to the different meanings of words or signs or symbols (semantically) make the two languages somewhat different because of Mercury's clarity, style, approach, conditions (determinism) and module systems. Although Mercury and Prolog are both logic programming languages there are several differences between them:

- Syntax. Mercury's syntax is similar to modern imperative languages. Prolog uses symbolic syntax, which is a more conventional approach.
- Type system. Mercury has a robust type of system that permits for more stable type verification making programmes more efficient than Prolog.
- Mode and determinism. Mercury permits additional control over the computation process since it supports both declarative and imperative semantics, thus permitting the use of non-deterministic predicates. Prolog only supports declarative semantics to express logical relationships, but it does not have the ability to specify how those relationships are to be computed.
- Performance. Prolog is not as efficient as Mercury. Mercury focuses on static type checking to verify the type of safety of a program based on analysis of a program's text (source code), mode-directed compilation, and automatically creating specified code.
- Concurrency: Mercury supports concurrent and parallel execution of predicates that permit the efficient use of multi-core processors, whereas Prolog does not have this feature.
- Applications: Prolog was designed for AI applications and is used for knowledge representation and natural language processing. Whereas Mercury is used mainly for coding large, complex, and performance-critical systems.

Since, the syntax code for Mercury is based on Prolog, it is appropriate to use Prolog syntax to illustrate certain basic aspects of these programming languages. It should be noted that Prolog is a robust and flexible computer programming language, which is very suitable for creating logic-based AI applications. In fact, it was specifically designed with AI development as a primary objective. Like Mercury, Prolog permits the programmer to specify the rules and facts about a problem domain. Once this is done, the Prolog interpreter will use the given rules and facts to automatically suggest solutions to problems. An important aspect of Prolog is its facility to handle uncertain or incomplete information. Hence, a programmer can define a set

of rules and facts that are accepted as being true, but they can also define rules and facts that may possibly be true or false. The Prolog interpreter will use the defined rules and facts and automatically deduce (reason) about the problem domain and discover solutions that are probably correct, given the available information.

Prolog programs consist of two key components, namely facts and rules. Facts are declarations that are accepted as being true, such as “Peter is a man” or “the capital of Germany is Berlin.” Rules are logical declarations that designate the relationships between different facts, such as “If Peter is a man and Rebecca is a woman, then Peter is not Mary.” The syntax for Prolog programs is like natural language. For instance, the above case might be expressed in four lines of code as shown in [Table 6.1](#). In this case, the first three rows are facts, while the fourth row is a rule. Observe that the rule uses the *not/2* predicate to declare that if *X* is a man and *Y* is a woman, then *X is not Y*.

TABLE 6.1 Examples of Prolog Syntax Statements [↗](#)

Row	Syntax	Type
1	man(peter).	Fact
2	woman(rebecca).	Fact
3	capital_of(germany, berlin).	Fact
4	not(X1,Y1):- man(X1), woman(Y1).	Rule

Prolog permits a wide-ranging variety of different programming styles and approaches. This means that there is no specific single “syntax” for Prolog. However, there are several common conventions that are frequently used as described below:

- Facts are declarations that are presumed to be true; hence, they are written using a predicate name succeeded by a list of arguments enclosed in parentheses (e.g., woman(rebecca) also see [Table 6.1](#), rows 1 to 3).

- Rules are logical statements that specify relationships between facts. Rules are expressed using the predicate name succeeded by a list of arguments enclosed in parentheses, followed by a colon and a hyphen (:-) and the rule (e.g., not(X1,Y1):- man(X1), woman(Y1) also see [Table 6.1](#), row 4).
- Variables denote values that may change or be determined. Variables are normally written using a name that begins with an uppercase letter (e.g., X1).
- Queries are utilised to find solutions to problems based on the defined rules and facts in the program (e.g., not(peter, rebecca)?).

As mentioned previously, facts are statements that are accepted as being true and provide Prolog with information about the problem domain, which will be used to automatically deduce answers to problems. Facts are expressed by using a predicate name followed by a list of arguments enclosed in parentheses. For example, at [Table 6.1](#), row 1 declares that peter is a man; row 2 declares that rebecca is a woman, and row 3 declares that berlin is the capital of Germany. [Table 6.2](#) shows a simple computer program that illustrates some basic Prolog principles. The program at [Table 6.2](#) identifies several facts and rules about a problem domain. It applies these facts and rules to provide answers to two queries.

TABLE 6.2 Example of Simple Prolog Program

Row	Syntax	Type
1	% Facts	
2	man(peter).	Fact
3	woman(rebecca).	Fact
4	boy(jake).	Fact
5	girl(sue).	Fact
6	river_of(germany, danube).	Fact
7	capital_of(germany, berlin).	Fact

Row	Syntax	Type
8	% Rule	
9	not(x1,y1):- man(x1), woman(y1).	Rule
10	not(x2,y2):- boy(x2), girl(y2).	Rule
11	% Query 1	
12	not(peter, rebecca)?	Query
13	% Query 2	
14	river_of(germany, X)?	Query
15	% Query 3	
16	capital_of(germany, Y)?	Query

The program defines six facts at rows 2 to 7, namely, man(peter), woman(rebecca), boy(jake), girl(sue), river_of(Germany, Danube), and capital_of(germany, berlin). These facts basically declare that Peter is a man; Rebecca is a woman; Jake is a boy; Sue is a girl; Danube is a river in Germany; and Berlin is the capital of Germany. Note that row 1 (% Facts) declares that the rows following it are facts. At row 8 (% Rule), the program signifies that the rows following it are rules. The rules are defined at row 9 and row 10. The program declares that the rules are using the not/2 predicate, declaring that if X1 is a man and Y1 is a woman, then X1 is not Y1; and that if X2 is a boy and Y2 is a girl, then X2 is not Y2. The final segment of the program describes the queries. These queries are declared at rows 11 to 16. Note that each query is numbered and headed by the statement %Query 1 at row 11, %Query 2 at row 13, and %Query 3 at row 15 before the actual query statements made are at row 12, 14, and 16.

The query at row 12, not(peter, rebecca)?, requests the interpreter to ascertain whether Peter is not Rebecca, based on the not/2 rule and the man and woman facts. The program will apply these rules and facts to determine that Peter is not Rebecca, and it will respond with true as the solution to the query. The next query at row 14, river_of(germany, X)?, solicits the interpreter to establish the main river of Germany. The program will apply

the `river_of/2` fact to ascertain that the main river of Germany is Danube, and it will respond with the answer Danube for the variable X as the result. The final query, `capital_of(germany, Y)?`, seeks the interpreter to disclose the capital of Germany. The program will apply the `capital_of/2` fact to reveal that the capital of Germany is Berlin, and it will respond with the answer of Berlin for the variable Y as the result.

The Prolog program at [Table 6.2](#) shows how to define and express rules and facts about a problem domain, and how the rules and facts are applied to automatically deduce solutions to problems. The next example of a Prolog program is shown at [Table 6.3](#), which illustrates a simple AI application. This computer program demonstrates a basic diagnostic mechanism to check for three medical conditions, namely Covid-19, Flu, and Cold. The program consists of three main segments. The first segment defines the facts for each specific medical condition (rows 1 to 23). The facts describe the particular symptoms for the three medical conditions, some of which may overlap. The second segment defines the rules, which in this case has only one rule. It declares the rule by applying the `illness_condition/2` predicate, which declares that if a patient has a particular symptom, then they might have the medical condition that is linked with that symptom.

TABLE 6.3 Example of Simple Prolog Program AI Application [🔗](#)

Row	Syntax	Type
1	% Facts	
2	<code>illness_indicator(covid, fever).</code>	Fact
3	<code>illness_indicator(covid, chills).</code>	Fact
4	<code>illness_indicator(covid, cough).</code>	Fact
5	<code>illness_indicator(covid, shortness_of_breath).</code>	Fact
6	<code>illness_indicator(covid, difficulty_breathing).</code>	Fact
7	<code>illness_indicator(covid, fatigue).</code>	Fact
8	<code>illness_indicator(covid, muscle_aches).</code>	Fact

Row	Syntax	Type
9	illness_indicator(covid, body_aches).	Fact
10	illness_indicator(covid, headache).	Fact
11	illness_indicator(cold, sneezing).	Fact
12	illness_indicator(cold, cough).	Fact
13	illness_indicator(cold, stuffy_nose).	Fact
14	illness_indicator(cold, sore_throat).	Fact
15	illness_indicator(flu, fever).	Fact
16	illness_indicator(flu, muscle_aches).	Fact
17	illness_indicator(flu, body_aches).	Fact
18	illness_indicator(flu, chills).	Fact
19	illness_indicator(flu, fatigue_weakness).	Fact
20	illness_indicator(flu, fatigue_weakness).	Fact
21	illness_indicator(flu, cough).	Fact
22	illness_indicator(flu, chest_discomfort).	Fact
23	illness_indicator(flu, headache).	Fact
24	% Rule	
25	illness_condition(Y, C):- illness_indicator(C, Y).	Rule
26	% Query 1	
27	illness_condition(sneezing, Y)?	Query
28	% Query 2	
29	illness_condition(stuffy_nose, Y)?	Query
30	% Query 3	
31	illness_condition(headache, Y)?	Query
32	% Query 4	
33	illness_condition(chills, Y)?	Query
34	% Query 5	
35	illness_condition(difficulty_breathing, Y)?	Query

The third and final segment(s) of the Prolog program declare the queries to determine the underlying medical condition, whether it is covid-19, flu, or the common cold. A specific query requests the interpreter to ascertain the possible medical condition that a patient might have based on their symptoms. In this example there are five specific queries since the medical conditions may have some common symptoms. The first two queries are meant to determine whether the symptom being experienced by the patient indicate a common cold. The next two queries have the purpose of indicating whether the symptoms indicate flu or Covid-19 since these have common symptoms. The final query will determine whether the patient is experiencing flu or Covid-19, since there is a high probability that the difficulty breathing symptom is applicable to Covid-19. Each individual query is requesting the interpreter to determine which medical condition the patient is likely to be experiencing. As stated previously, the program will apply the `illness_condition/2` rule and the `illness_indicator/2` facts to establish that the patient has either the covid-19, flu, or the common cold, and it will respond with the appropriate solution to each query.

The above example illustrates how a Prolog program can be applied for an AI application. As one may appreciate, in a practical scenario the program is likely to be more sophisticated by accessing a database of symptoms and have more complex rules for diagnosing a medical condition. It should be noted that Prolog has been used instead of Mercury because Prolog is a sub-set of the Mercury computer language. This section has demonstrated that both Mercury and Prolog are powerful and flexible languages for developing AI applications, with Mercury being more recently developed having many added features.

6.2.7 IBM Qiskit Software Development Kit (SDK)

IBM have adopted Qiskit, which is an open-source, python-based quantum SDK. The Qiskit SDK is designed specifically to work on the IBM Quantum Platform or on simulators on a local computer at the level of circuits and algorithms. Qiskit has been developed by IBM Research with

contributions from academic organisations to facilitate software development for IBM's cloud quantum computing service. Apart from the Python version, Qiskit versions were considered for Swift and JavaScript, but these initiatives were not pursued. Qiskit also supports Quantum Artificial Intelligence (QAI), which is a contemporary interdisciplinary research field that combines quantum computing with AI that aims to use the unique properties of quantum computers to enhance the capabilities of AI systems.

Qiskit is composed of several integrated components that enable quantum computing with the primary objective of creating a software package that facilitates users to exploit quantum computers, irrespective of their level of skill or sphere of interest. Hence, Qiskit provides a computing platform for users to create test prototypes and applications to execute on quantum computers and/or conventional simulators. This is achieved by generating quantum software both at the machine code level of OpenQASM, and at abstract levels that are appropriate for end-users without quantum computing expertise.

The functionality for Qiskit is supported by several components which provide some very powerful features. For instance, the Qiskit Terra component is the basis for the other Qiskit modules and provides the necessary tools to generate quantum circuits at or close to the level of quantum machine code. This permits the generated processes that execute on quantum hardware, which are specifically created in terms of quantum gates. These provide the means for the quantum circuits to be optimised for a specific device, apart from managing batches of jobs and executing them on remote-access quantum devices and simulators.

An important component is Qiskit Machine Learning. This component provides the basic computational modules, such as Quantum Kernels and Quantum Neural Networks that are applied in several applications including classification and regression. Qiskit Machine Learning is user friendly, allowing initial prototypes to be created rapidly without having comprehensive quantum computing knowledge. Additionally, it is very flexible and can be broadened to support innovative quantum machine

learning research. Qiskit Machine Learning modules, such as FidelityQuantumKernel that applies the BaseStateFidelity algorithm can be utilised to directly compute kernel matrices for given datasets or can be forwarded to a Quantum Support Vector Classifier (QSVC) or Quantum Support Vector Regressor (QSVR) to solve classification or regression problems.

Qiskit Machine Learning has a generic interface for neural networks that is executed by various quantum neural networks that permits users to link parametrised quantum circuits with quantum mechanical observables that may apply Qiskit's circuit library. Furthermore, Qiskit Machine Learning has a range of learning algorithms such as the NeuralNetworkClassifier and NeuralNetworkRegressor for addressing classification or regression problems. Additionally, it supports the TorchConnector module for integrating all of IBM's quantum neural networks directly into the PyTorch open-source machine learning library.

Machine learning has become an exceptional tool in contemporary decision making and supported by further developments in quantum computing it is expected to drive this capability to new heights. The integration of quantum principles in neural networks provides the potential for considerable computational advantages, specifically in the sphere of AI and deep learning. In fact, a detailed recognition and understanding of this integration is likely to provide added capability in how complex computations are undertaken.

By combining the Qiskit environment with PyTorch, a framework is created that facilitates efficient prototyping and testing of variational quantum algorithms, which leverage both classical and quantum computing resources to find approximate solutions to problems. Furthermore, performance is significantly enhanced by providing a quicker and more scalable approach to quantum computations. These benefits are achieved through the availability of advanced tools that facilitate simple integration approaches for existing projects. This allows researchers to explore quantum neural networks without significant overhead that spans the gap between quantum computing and traditional neural network methodologies.

Hence, quantum neural networks are viewed as having great potential for augmenting the computational abilities of AI systems, fostering further innovations in the field.

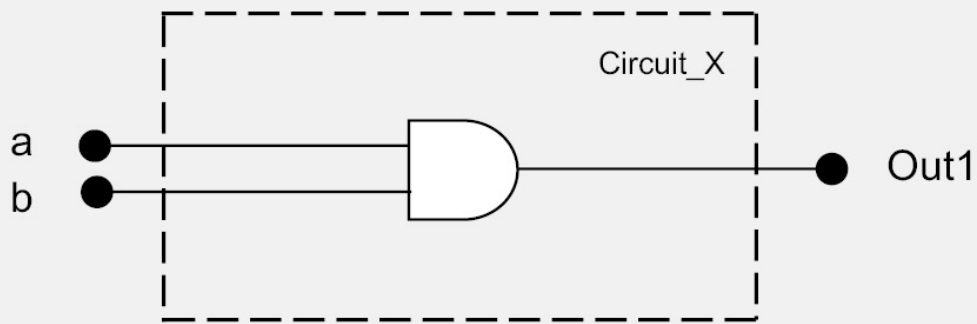
6.2.8 Other Programming Languages for Specific Hardware Associated with AI

This section provides a brief description of key programming languages for specific hardware associated with AI that have the objective of controlling digital circuits to perform a particular function. The hardware itself was covered in a previous chapter. This chapter provides additional information regarding the programming languages that drive this specific hardware. Generally, hardware refers to the external and internal devices and equipment that enable users to perform major functions such as input, output, storage, communication, processing, and amongst other functions. Hence, hardware refers to the physical parts. However, for these physical parts to function they must receive and execute instructions.

Inherently, hardware has digital circuits as a primary component, that chiefly consist of interconnected transistors. In fact, a central processing unit (CPU) consists of a massive conglomeration of transistors. More specifically, a transistor can function as a switch or gate for electronic signals, opening and closing an electronic gate numerous times per second. Hence, a transistor ensures that the circuit is switched on if the current is flowing and switched off if it is not flowing.

As such, a digital circuit may be depicted by having interconnected diagrams that are referred to as a schematic diagram, which could be quite complex to visualise. In fact, as the complexity increases it becomes impractical to describe the interconnected digital circuit diagrams in a visual style. However, an easier and more practical approach to describing digital circuits is to use a textual language that is specifically aimed to clearly and concisely describe the defining features of digital design. Such, textual languages are termed hardware description languages (HDLs).

In the chapter addressing hardware, it was mentioned that the common hardware description languages are Verilog and VHDL (Very-High-Speed Integrated Circuit Hardware Description Language). These are commonly applied in combination with Field Programmable Gate Arrays (FPGAs) that are digital devices that are specifically designed to aid and therefore expediate the creation of customised digital circuits. FPGAs are integrated circuits that are often sold off-the-shelf and are known as “field programmable” because they give developers the capability to reconfigure (modify) the hardware to meet their specific use case requirements after the manufacturing process. In other words, HDLs allow developers to define a circuit using words and symbols to create software that can convert the word-based-symbolic description into configuration (design) data that is loaded into the FPGA to execute the required functionality. [Figure 6.13](#) provides a simple example of singular circuit schematic diagram and the corresponding HDL code.



HDL Code:

```

1 library ieee;
2 use ieee.std_logic_1164.all
3 entity Circuit_X is
4     Port (a: in STD_LOGIC;
5           b : in STD_LOGIC;
6           OutX : out Std_LOGIC);
7 end Circuit_X;
-----
8 architecture Behavioural of Circuit_X is
9 begin
10 out1 <= ( a and b );
11 end Behavioral;
```

FIGURE 6.13 Example of Singular Circuit Schematic Diagram and Its HDL code. [↩](#)

One may ask whether there is a difference between computer programming languages and HDLs (hardware description languages)? HDLs are very similar to high-level programming languages such as Python or C, however there is an important difference. HDL statements represented by HDL code entail having parallel operations, whereas programming languages embody sequential operations. In other words, with a computer program or firmware module the processor executes the program statements (code) one at a time, with a top-to-bottom chain of command like an

individual reading text on a page. HDL programming code describes digital hardware, and therefore discrete segments of this hardware can operate concurrently even though the matching lines of code are written using a top-to-bottom approach.

As stated previously VHDL and Verilog are commonly used hardware description languages. The example at [Figure 6.13](#) represents a simple HDL for a basic digital circuit using VHDL. VHDL was instigated by the US Department of Defence around 1981 in conjunction with private sector companies like IBM and Texas Instruments that piloted the launching of its first version in 1985. This development was actively supported by the first FPGA that was released in 1984; consequentially contributing to its advancement as an established language in digital circuit design, simulation, and synthesis. The VHDL lines of code are relatively easy to understand. For instance, taking the example at [Figure 6.13](#), the following explanation will clarify what each line of code is achieving:

- Line 1: Adds the library “ieee.” This library contains key definitions regarding data types and operators and may consist of several packages that are applicable to the circuit design.
- Line 2: Specifies that package “std_logic_1164” from the “ieee” library is required. This signifies the data type being used. A “std_logic” data type can denote a one-bit signal, whilst the “bit” data type can accept a logic one or zero. However, the “std_logic” option is more powerful since it is not limited to just logic high and low but can have a tri-state inverter condition. This means the circuit outputs a voltage signifying the opposite logic-level to its input, thus inverting the input signal applied, such that when an applied input is low then the output becomes high and vice versa. Hence, the “std_logic” data type supports the “don’t care” state, which enables better hardware optimisation for look-up tables.
- Line 3: This specifies an arbitrary name for the circuit being described. The word “circuit_X” is placed between the keywords “entity” and “is,” thus naming this module.

- Line 4 to 6: These define the input and output ports of the circuit. Matching these lines to the circuit diagram, it is noted that the ports of the circuit together with their features are listed after the keyword “port.” For example, line 4 denotes an input port called “a” as indicated by the keyword “in” after the colon with a `std_logic` data type to describe a one-bit digital signal.
- Line 7: This line establishes the end of the “entity” statement.

Therefore, the first seven lines of code describe the library to be applied, and the programme entity, consisting of the circuit name being described; and the circuit ports and their attributes, such as input/output and the data type to be passed on by these ports. This basically, describes the interface of a module with its surrounding environment. The next lines of code (lines 8 to 11) describe the operation of the circuit. The operation of the circuit is described by the “architecture” section that specifically relates to `circuit_X`, which was defined by the entity statement in Line 3.

- Line 8: This line establishes the “Behavioural” segment for the architecture to be described by the next lines and is placed between the keywords, “architecture” and “of”. It relates and describes “`circuit_X`.”
- Line 9: This specifies the start of the architecture description.
- Line 10: This line describes the circuit’s operation. The AND of the two inputs a and b is placed within the parentheses, and the outcome is designated to the output port using the assignment operator “`<=`.”
- Line 11: This line specifies the end of the architecture description.

The above lines of code describe the circuit’s internal operation that depict a simple AND gate as shown in [Figure 6.13](#) using VHDL. For further information regarding VHDL refer to Arar (2017). The other hardware descriptive language that has already been mentioned is Verilog. Verilog is an alternative to VHDL, and both are viewed as powerful tools for designing a complex digital circuit. Verilog stems from the combination of

“verification” and “logic” because the language was initially envisaged as a simulation and verification tool.

Verilog has taken longer to emerge as a popular HDL option. It was initially designed in early 1984 by Gateway Design Automation a few years after VHDL was released by the US Department of Defence. However, with the support from Department of Defence, VHDL was adopted in 1987 as an IEEE standard with Verilog lagging. Verilog HDL was conceived by Phil Moorby and Prabhu Goel around 1984 and was initially a proprietary hardware modelling language owned by Gateway Design Automation. Verilog was continuously upgraded with revisions being released between 1984 to 1990. The breakthrough for Verilog materialised in 1990 when it was acquired by Cadence Design System, which was an important supplier of electronic design technologies and engineering services in the electronic design automation (EDA) industry.

Cadence realised that Verilog could not progress as a proprietary product since it would lack standardisation status. Hence, in 1991 Cadence released Open Verilog International (OVI) and transferred it to the public domain. This was followed by a submission to IEEE to attain standard status, which was achieved in 1995 and became IEEE standard 1364-1995 (Verilog-95). Verilog has been continually revamped with important added features, attaining IEEE standard 1364-2001 (Verilog-2001) and IEEE Standard 1364-2005 (Verilog-2005). In 2005 System Verilog was issued, which was a superset of Verilog-2005, with many new features and functionalities to aid design verification. Additionally, in 2009, SystemVerilog and Verilog language standards were amalgamated into SystemVerilog 2009 (IEEE Standard 1800-2009), which is currently one of the most popular languages for IC design and verification.

The syntax for Verilog can be illustrated by using the circuit example shown previously when explaining VHDL. [Figure 6.14](#) shows the circuit design (same as shown in [Figure 6.13](#)) and the corresponding Verilog lines of code. Recall that the circuit design has two inputs and one output with the ports being all one bit wide.

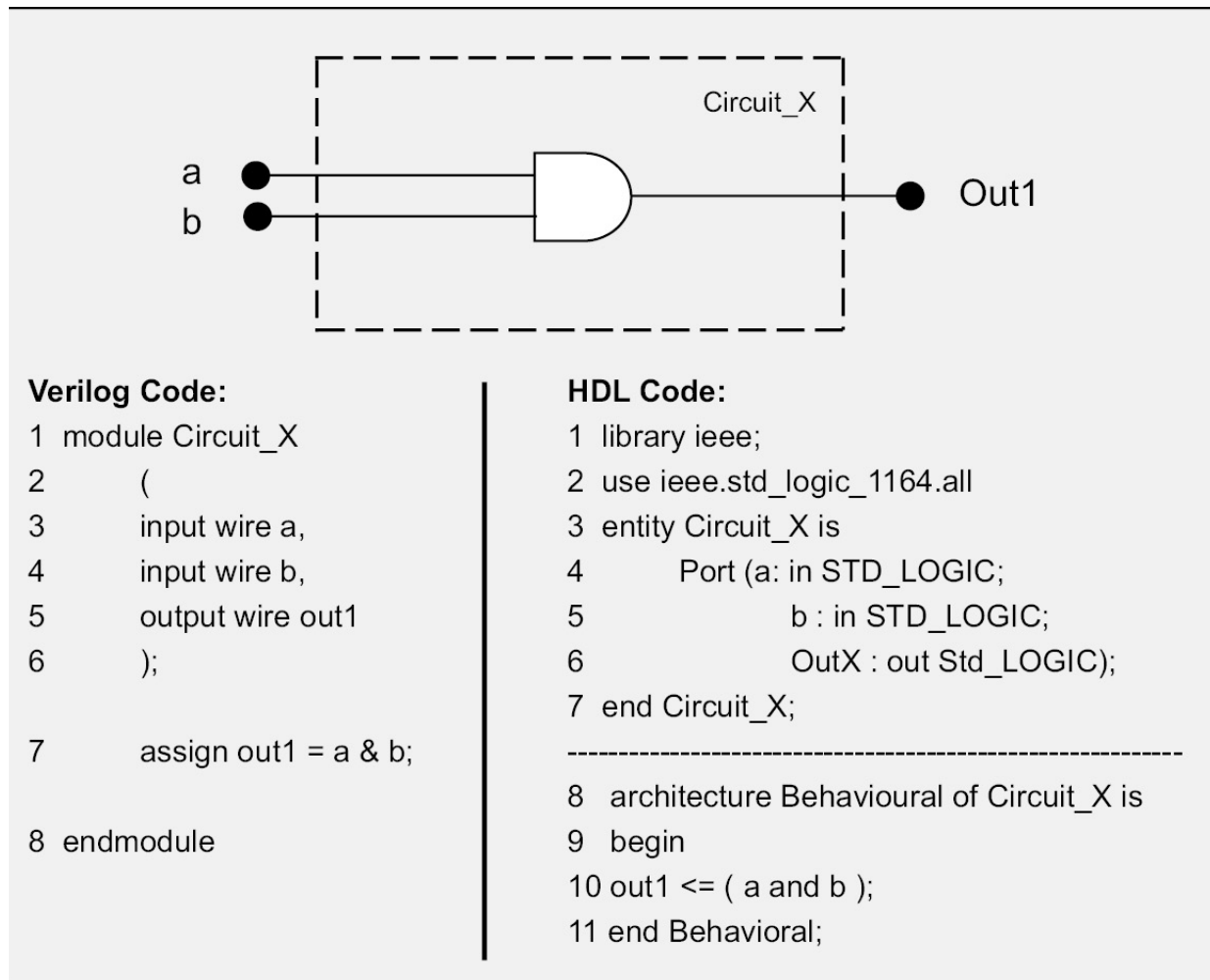


FIGURE 6.14 Example of Circuit Design and Corresponding Verilog Code. [🔗](#)

The circuit has an AND gate with two inputs and passing the result to the out1 port. The Verilog lines of code are as follows:

- Line 1: This line uses the keyword “module” to specify that the line will start to describe a circuit named “Circuit_X”. This is an arbitrary name for referencing the circuit.
- Line 2: This specifies the start of the interface of “Circuit_X” with its surrounding environment.
- Line 3: This specifies the input port described with using the keyword “input”. This keyword is followed by another keyword “wire” and

then the name of the port in this case “a.” “Wire” is a Verilog data type that represents a physical wire in the design.

- Line 4: This is like Line 3, but it specifies the input port “b”.
- Line 5: This specifies the output port using the keyword “output”. Once again, this keyword is followed by the keyword “wire,” and then the name of the port (out1).
- Line 6: This specifies the end of the module interface of “Circuit_X”.
- Line 7: This line applies the bitwise AND operator using the “&” symbol to describe the functionality of the circuit. The keyword “assign” puts a&b on the output port out1. This type of assignment is known as a continuous assignment because it is always active whenever an operand on the right-hand side changes, a&b is evaluated and assigned to out1. A continuous assignment is viewed as a combinational circuit whose output is continuously updated depending on its inputs, which in this case is a simple AND gate.
- Line 8: This line applies the keyword “endmodule” to specify that end of the description of the circuit named “Circuit_1”.

[Figure 6.14](#) shows the programme coding for both Verilog and VHDL for the same circuit description. Verilog is more concise and does not make use of predefined library routines as VHDL. As you may recall the libraries in VHDL provide the definition for the data types and operators, which for std_logic data type can have one of nine different values. However, the most used options for std_logic data type are 0, 1, Z, and ‘-’ (‘-’ denotes a don’t care). On the other hand, Verilog has four basic data types that are often applied. Hence, the wires defined in the example may have one of the following: 0 for a logic low or a false condition; 1 for a logic high or true condition; z for the high-impedance state, which can be used to describe the output of a tri-state buffer; and x for a don’t-care or unknown value, when an input is not initialized or two or more conflicting outputs are connected together.

It should be noted that currently, VHDL and Verilog are the prevalent HDLs that are viewed as industry standard with FPGA design tools

ordinarily supporting both languages.

6.3 AI SOFTWARE ALGORITHMS

This section will address AI software algorithms, which is a category of computer software that enables the implementation of AI to process substantial volumes of data for performing tasks that would otherwise require human intelligence. These tasks include speech and voice recognition, text recognition, image recognition, video analytics, generative AI, and natural language processing (NLP).

In recent years, the strategic importance of AI technology has been acknowledged, and its application has increased exponentially across all types of industries. Many private and public sector organisations are studying and investing in AI solutions to strengthen their competitive position in a continually turbulent global market and to improve their organisational processes. It is a well-known experience in technological development that late movers will likely be at risk of being outdated in terms of innovation and will critically lag their competitors in terms of servicing their customers.

6.3.1 Voice and Speech Recognition (Conversational AI)

The thrust for Conversational AI is currently from the smart home market. Google, Amazon, Microsoft, and others are sustaining their investment in this technology to develop smart assistants as an integral part of people's lifestyle through smartphones and personal computers. This investment is not aimed at incremental change but to revolutionise the way tasks are executed at home and in the longer term in organisations. Although more needs to be done before this technology is widely accepted, substantial improvement is being made in terms of enhancing the complex mathematical models that drive and control voice and speech recognition technology. However, for these types of mathematical models to be fully complete they must be able to harness vast volumes of data that

organisations have accumulated over the years to be able to learn and recognise diverse speech patterns.

It must be recognised that these models are very complex because they must have the capacity to decipher not only vocabulary, regional pronunciations, and colloquialisms but also the context of the conversations. They emphatically need to speedily analyse all recorded conversations of call centre agents with customers and the voluminous daily interactions with a digital assistant. This is highly dependent on having a permanent and fast internet link to provide the necessary processing integration and reliable high-velocity connectivity.

Currently, applicable voice technologies are available from well-known providers, such as Google Assistant, Amazon Alexa, Microsoft Cortana, amongst others. However, the current technology is still limited in what it can do. For instance, most chatbots and virtual agents are typically inclined towards having action links (i.e., keys/buttons) that are applied for moving the conversation forward or having external links to divert the customer from a conversation entirely. This is likely to create a problem for voice-only exchanges. Another issue with AI voice recognition technology is that it is still not as robust as the chat-based equivalent and has difficulty in interpreting elusive linguistic distinctions that are commonly used in speech.

It must also be realised that voice and speech recognition are two different functions. Voice recognition is based on technology that recognises the voice of the speaker. On the other hand, speech recognition is the technology that recognises the actual words. This difference in definition illustrates that both functions have a different task. For example, voice recognition includes security features such as voice biometrics, whereas speech recognition is a device that generates automatic transcripts and accurate instructions. Basically, AI voice recognition recognises and decodes human speech patterns and permits a smartphone, computer, or virtual assistant to know what one is saying and respond to this. In contrast, speech recognition technology transforms speech into text by identifying words, and in advanced AI applications may interpret differences based on

context. For example, automatic speech recognition (ASR) applies this technology to transcribe voicemails and has an accessibility tool that permits persons unable to type to talk to their computer to compose emails, perform searches, and many other functions.

Furthermore, voice recognition and ASR have distinct roles and process and respond to audio in a different way. Voice recognition is a device like the digital assistants, Amazon Alexa or Google Dot, which execute certain tasks by listening to a person's voice in real-time and responding to it. For example, Alexa resides in the cloud and is, therefore, accessible from anywhere at any time, responding to requests or answering questions like reading the latest news headlines, playing a specific song, and dimming lights in a particular room, amongst many others. Basically, voice recognition identifies who said what, while ASR captures what people say. Hence, ASR does not identify voices but recognises speech by applying natural language processing (NLP) to produce a transcript of the audio and generating real-time captioning. ASR is not perfect in its interpretation, but it is quick and is a cheap method for converting speech to text.

The benefits of exploiting voice recognition at an individual or organisational level include its effectiveness as a user verification tool, convenience, and better performance. For instance, banks have implemented voice biometrics as a security measure on certain accounts that have resulted in substantial savings from their fraud prevention endeavours. Additionally, voice-based technology mitigates the requirement for error scans by having more accurate communication facilities permitting faster, and better task management. Furthermore, speech recognition AI is being applied to provide business solutions ranging from voice-activated audio content assistants to ATMs by facilitating the interaction of people with technology providing better data transcription accuracy.

Speech recognition is being used at call centres by allowing organisations to listen to what customers have to say and then applying the gained information via cloud models to provide a quick and proper response. It is also being applied for voice or audio biometrics (voice patterns) to identity and authorise users to access systems without the need

for passwords, fingerprints, or eye scans. Fintech companies are using speech recognition to help customers resolve basic business queries. Additionally, clients can interact with organisations in real-time 24/7 from anywhere using voice transcription solutions or text messaging applications to resolve their concerns.

6.3.2 Text Recognition

AI is also being used for text recognition. Text recognition software does not only decipher typed text through optical character recognition (OCR) but also interprets handwritten documents. OCR basically recognises text within a digital image such as scanned documents and images and convert them into an accessible softcopy version with text. Traditionally, OCR was restricted to printed scanned documents. However, AI handwriting recognition software can be trained to decrypt handwritten documents as well as printed texts, which requires a more complex model. It is being applied for deciphering historical documents that may contain linguistic difference from their modern counterparts.

For instance, Transkribus a company supplying an AI based handwriting recognition software claims that its software can be trained to decipher both handwritten historical documents and printed texts. They maintain that its pre-trained AI model is able to transcribe large number of documents very quickly and make them accessible to be read, searched, and prepared for further applications such as research, data analyses, or publishing. Amazon Textract is another text recognition software for extracting text and structured data (tables and forms) from documents using AI, with no configuration or templates being necessary; and having robust data privacy, encryption, and security controls. This software is being applied by Fintech companies to root out critical information from mortgage applications regarding asset and property valuation, and credit rating or extracting information from insurance forms, claims, and receipts that have varying document layouts and formats.

Text recognition software has been around for some time, but with the support of AI, its transformation and extraction features are improving rapidly, particularly in recognising historical and modern handwritten documents (in various languages). Algorithms to decipher historical and modern handwritten documents are very complex and require extensive machine learning features that can only be resolved by applying AI technology.

6.3.3 Image Recognition

AI is being applied to image recognition, sometimes referred to as picture or photo recognition. Image recognition has the specific purpose of identifying and classifying objects of interest within a given image. The human brain is capable of automatically classifying objects that it visually observes, and to link them with individual definitions. While as humans we take this function for granted, it is extremely difficult for machines to achieve this task. However, with the support of AI such a function is being performed but it requires a complex algorithm sustained by considerable processing power.

A primary objective of image recognition is the classification of the discovered (or observed) objects into different categories to assess which group the image is associated with, hence earning the name deep learning object recognition. With the ever-increasing processing power of the computer, machine learning, in particular deep learning technology, has made important and successful advances in computerised vision and image processing. Currently, AI models for image recognition supported by deep learning algorithms are being successfully applied in many private and public industrial sectors. When discussing AI image recognition, it is important to clarify the terminology applicable. For instance, it is necessary to distinguish between image recognition and computer vision; object localisation; and image detection. Although, there are some overlaps in their meaning, there are also some important differences.

Image recognition is the application of computer vision. Image recognition is viewed as a subset of computer vision that focuses on detecting, analysing, and interpreting images for decision-making. It includes tasks like image tagging (identifying and labelling the contents of an image), object detection, and guiding autonomous vehicles using neural networks trained on annotated datasets. Computer vision is a field of AI that applies machine learning and neural networks to train computer systems to obtain meaningful information from digital images, videos, and other visual inputs and to also make suggestions or act when they detect defects or issues. Hence, image recognition performs several functions, such as classification, tagging, detection and segmentation using machine learning.

Object localisation is also a subset of computer vision. It refers to identifying the location of one or more objects in an image and depicting each within a drawn box (demarcation) to show clearly its location, but it does not include the classification of detected objects. On the other hand, image detection is a function that uses an image as input data and searches and discovers various objects within it. For instance, face detection algorithms aim to seek face patterns in images irrespective of whether the detected objects are significant. The reason for this is that the aim of image detection is merely to distinguish one object from another, thus assessing the number of distinct entities that are present within the image. Therefore, boundaries are drawn around each separate detected object. With image recognition, the aim is to pinpoint the objects of interest within an image and identifying with which category they are associated.

The growth and importance of computer vision algorithms has been significant over the years due to its accuracy and speedy processing. The two well developed algorithms are the Mask RCNN (Mask Region-based Convolutional Neural Network) released around 2017, and YOLO (You Only Look Once) released around 2021. The latest version of the YOLO algorithm is YOLOv9, which was released in early 2024. Mask R-CNN is a deep learning instance segmentation method that executes pixel-level segmentation on detected objects and can accommodate multiple classes and overlapping objects by precisely delineating their boundaries at the

pixel level. Mask R-CNN applies Feature Pyramid Network (FPN) and the Region of Interest Align (ROIAlign) to attain high accuracy and resilient performance. FPN is a feature extractor that receives a single-scale image of an arbitrary size as input, and produces proportionally sized feature maps at multiple levels, in a fully convolutional fashion as outputs. Hence, it functions as a generic solution for creating feature pyramids inside deep convolutional networks that are applied in executing tasks like object detection.

FPN works in conjunction with ROIAlign, which extracts a small feature map from each RoI (Region of Interest) used in detection and segmentation tasks by removing the harsh quantisation of the RoI Pool and accurately aligning the extracted features with the input. However, Mask R-CNN requires substantial processing power due to its computational complexity and high memory utilisation during training and inference. This may hinder accuracy in segmenting very small objects or handling heavily obscure scenes and may be more resource reliant for training and fine-tuning the model. The YOLO models are a real-time object detection mechanism that identify and classify objects in a single pass of the image. The YOLO model (like Mask R-CNN) uses the convolutional neural networks (CNNs) to classify the objects and regressor networks to accurately predict the bordering (bounding) box coordinates for each detected object. However, YOLO and Mask R-CNN differ in object detection approach, the former uses a single-shot detection (SSD) approach, while the latter applies a two-shot (or multi-shot) method.

An SSD object detection algorithm predicts the bounding box and the class of the object in one single shot through a single forward pass of the network that simultaneously predicts the presence of an object and its respective bounding box. YOLO segments the input image into a grid and predicts a certain number of bounding boxes and class probabilities for each grid cell. On the other hand, Mask R-CNN predicts bounding boxes and class probabilities at multiple scales in different feature maps. Hence, it uses a two-step approach, with the first step suggesting a series of bounding boxes that may possibly contain an object (using a method called region

proposal); and the second step entails processing these suggested regions through a CNN to classify the object classes within the box. Hence, an SSD algorithm is very fast and efficient making it very suitable for tasks requiring real-time detection. However, the two-shot detectors are slower but are more accurate because they are more computationally intensive since speed is not their primary objective. Therefore, they can consume added computational resources on each likely object.

YOLOv9 is an improvement on previous versions since it addresses the issue regarding information loss inherent in deep neural networks and ensuring the retention of vital information throughout the detection process, to attain outstanding accuracy and performance. This is achieved by integrating Programmable Gradient Information (PGI) and the versatile Generalized Efficient Layer Aggregation Network (GELAN) architecture. PGI is an innovative concept to mitigate the issue of information bottleneck by safeguarding crucial data across deep network layers and to permit the creation of reliable gradients, expediting accurate model updates, and enhancing the overall detection performance. This concept is combined with GELAN to attain greater parameter application and computational efficiency. The integration of PGI and GELAN permits YOLOv9 to have an adaptable mixture of various computational blocks, allowing it to be used in a wide range of applications without forfeiting speed or accuracy.

The early image processing methods of two decades ago required vast quantities of training data, while the more recent deep learning approaches need much smaller learning samples. Despite this development, deep learning involves manual labelling of data to mark good and bad samples, a process called image annotation. This labelling process to train AI models is known as supervised learning, which entails time-consuming human work. Imagine the amount of human effort that is required to label images and annotate typical traffic conditions for autonomous vehicles.

The early image recognition methods depended on a sequence of image filtering, image segmentation, feature extraction, and rule-based classification (computer vision pipeline). The difficulty at that time was that engineering such pipelines, involved deep expertise in image processing

and computer vision, which took a great deal of development time and testing, in addition to manual parameter fine-tuning. This meant that the early computer vision methods supported by pixel-based image recognition systems were not very flexible, hence their reusability in varying scenarios or, and locations was very restrictive.

However, machine learning (with unsupervised learning) resulted in a high level of scalability, since it made use of algorithms to learn hidden knowledge from a dataset of good and bad samples. Additionally, deep learning applies multiple hidden layers of a neural network in its model. Deep learning would not be possible without the support of very powerful AI hardware and graphics processing units (GPUs). These tools provided the much need breakthroughs for the development of robust image recognition technology. AI image recognition technology has many practical uses. In healthcare, image recognition models are applied for identifying diseases or abnormalities in medical images. It has practical uses in security surveillance systems for real-time object detection, tracking, and classification, with face detection being adopted for biometrics, security, and facial recognition systems. In the agricultural industry it is applied to detect and classify crops, pests, and diseases, particularly facilitating precision agriculture techniques and automating farming processes. Although still at the experimental stage, it is used in autonomous vehicles for detecting objects such as other vehicles, pedestrians, traffic signs, and signals in real-time.

The constant challenge with this type of technology is to find an equilibrium between performance and computing efficiency, since an essential prerequisite for deep learning models is the impeccable affiliation of hardware and software to mitigate cost concerns related to computer vision. This impeccable affiliation of deep learning, image classification and deep neural network face recognition algorithms have succeeded in providing extraordinary performance and real-time object detection that exceeds human capabilities.

6.3.4 Video Analysis

Video analysis provides a user with the capability of extracting useful information from a video recording, such as counting the number of people or/and identifying specific objects or individuals. Like image recognition, video analytics makes use of computer vision technology that utilises AI for handling the analysis of digital images and videos. Video analysis employs deep learning techniques that use neural networks to learn patterns in data. Deep learning, which is a subset of machine learning, exploits neural networks that consist of multiple layers of interconnected processing nodes. This enables AI video analysis using deep learning algorithms to detect, track objects and recognise specific actions in video footage.

As already explained in the previous section when discussing image recognition, deep learning technology is the basis for this type of analysis using the latest Mask R-CNN and YOLO algorithms to detect objects in video. These algorithms can be custom trained with collected video data or pre-trained on large image datasets such as MS COCO. The MS COCO (Common Objects in Context) dataset is a large-scale object detection, segmentation, and captioning dataset that is designed to encourage research on a wide variety of object categories and is commonly used for benchmarking computer vision models. MS COCO image dataset contains over 320,000 images of everyday objects and humans with annotations that can be utilised to train machine learning models to recognise, label, and describe objects. These deep learning models facilitate the analysis of video by detecting and tracking objects, including object counting and rule-based analysis. For instance, they can be fine-tuned to count people in large crowds.

Another important feature of deep learning models is action recognition, which entails identifying specific actions in a video sequence or real-time video streams. These types of models can be taught to classify actions being executed in different contexts or environments. There are two categories of software systems related to video analysis, namely video management systems (VMS) and computer vision systems. A VMS typically manages

camera streams from security cameras by providing a single interface for observing live and recorded video from several cameras, with a search, playback, and export features. Additionally, VMS applications are normally intended to function with a specific type of camera, such as IP cameras or CCTV cameras. These systems precede advances in AI and are considered old-fashioned since many organisations are currently demanding video analytics functionality by including face and person detection and automated tagging or alerts.

On the other hand, computer vision systems focus on video recognition and the usage of complex real-time video analysis. These systems typically apply image processing algorithms using multi-step computer vision pipelines to evaluate images for extracting information from video footage, such as person or object detection, facial recognition, activity recognition, quality inspection, amongst many others. Many organisations opt for a customised computer vision system to satisfy their specific needs regarding system integration, performance, cost effectiveness, scalability, and data privacy.

AI video analytics is being used in private and public industrial sectors, particularly concerning security such as traffic monitoring, incident detection, intrusion management, people counting, augmented reality (AR), and facial recognition. It is also utilised in automatic number plate recognition (ANPR) that uses optical character recognition on images to read vehicle registration plates to create vehicle location data and ego motion that estimates a camera's motion relative to a rigid scene. Video analytics software can be taught to recognise the subtle differences in motion direction between the regular and irregular behaviour, such as fence claiming and other forms of intrusion by evaluating the difference in direction between horizontal and vertical (claiming) motion. Video analytics software in conjunction with X-ray security screening can be trained to perform object classification on real-time footage of baggage at security check-ins to recognise specific objects of interest, such as weapons and sharp instruments.

AI video analytics can also be applied in healthcare environments such as monitoring elderly patients at their private residence or in care homes for detecting falls, and unusual positions or periods of time in which a person is on the floor or incapacitated. They can also be trained to determine whether an elderly person has taken a given medication when scheduled. Furthermore, facial emotion analysis with AI using video analysis can be taught to identify differences in normal and abnormal facial or physical behaviour and alert designated responsible persons. The above examples illustrate that video analysis is a valuable technology that has a wide range of applications and is somewhat mature in its practical development.

6.3.5 Generative AI

Generative AI is a category of AI that can generate new content and ideas, including conversations, stories, images, videos, and music. Generative AI was initiated in the 1960s when the first chatbots were introduced. However, the breakthrough came in June 2014 when Ian Goodfellow and his colleagues at Google Brain developed the concept regarding generative adversarial networks (GANs). GAN is a category of machine learning structure, which has become an important concept in managing generative AI. Generative AI applies a machine learning algorithm that enables it to generate convincing authentic images, videos, and audio of factual people. While Generative AI has provided a powerful tool in the education and movie industry, it has also raised concerns regarding deepfakes. Deepfakes involve digitally fictitious images or videos that may be used to fraudulently launch damaging cybersecurity attacks on an individual and businesses, with the possibility of criminal extortion.

Innovative recent developments such as transformers and large language models (LLMs) have made generative AI an acceptable standard technology. Transformers being a machine learning tool allows technologists to train models that are constantly increasing in size and intensity without the need to initially labelling all the data. Hence, newly initiated models can be trained using an unimaginable volume of pages

containing text to provide responses with more depth. This is possible because transformers through the concept of “attention” do not just track the connection of words in individual sentences but allow models to track the linkages between words across an immense volume of text and could track these linkages to examine code, proteins, chemicals, and DNA.

Furthermore, LLMs have billions or even trillions of parameters that strengthen generative AI models to write appealing text, generate impromptu entertaining situation comedy, and may even portray photorealistic images. Therefore, applying the features of transformers and LLMs, generative AI has developed into a very powerful and robust tool that can create content through various types of media, including text, graphics, and video. For example, there are tools like Dall-E (named after the surrealist artist Salvador Dalí) that automatically generate images from a typescript or create text descriptions from images. However, generative AI is still at its early development phase. It requires much enhancement before it may be perfected for generating readable text and photorealistic stylized graphics that is considered highly accurate and unbiased, and free from hallucinations.

[Figure 6.15](#) illustrates that generative AI begins with an input prompt, which could be text, a design, an image, musical notes, a video, or any other input that the AI system can process. The input could be from a scanned copy and/or from a repository of stored data. This input is received by the various AI algorithms, which then create new content as a response to the input prompt. The generated content, depending on the input prompt, may take the form of compositions, specific answers to questions or problems, or realistic fakes generated from photographs or images or audio of an individual. Recent versions of generative AI allow users to define their request in plain language rather than using application languages such as Python. This makes them more user friendly, where the user can also modify the initial outcome regarding form, tone, and other specific aspects.

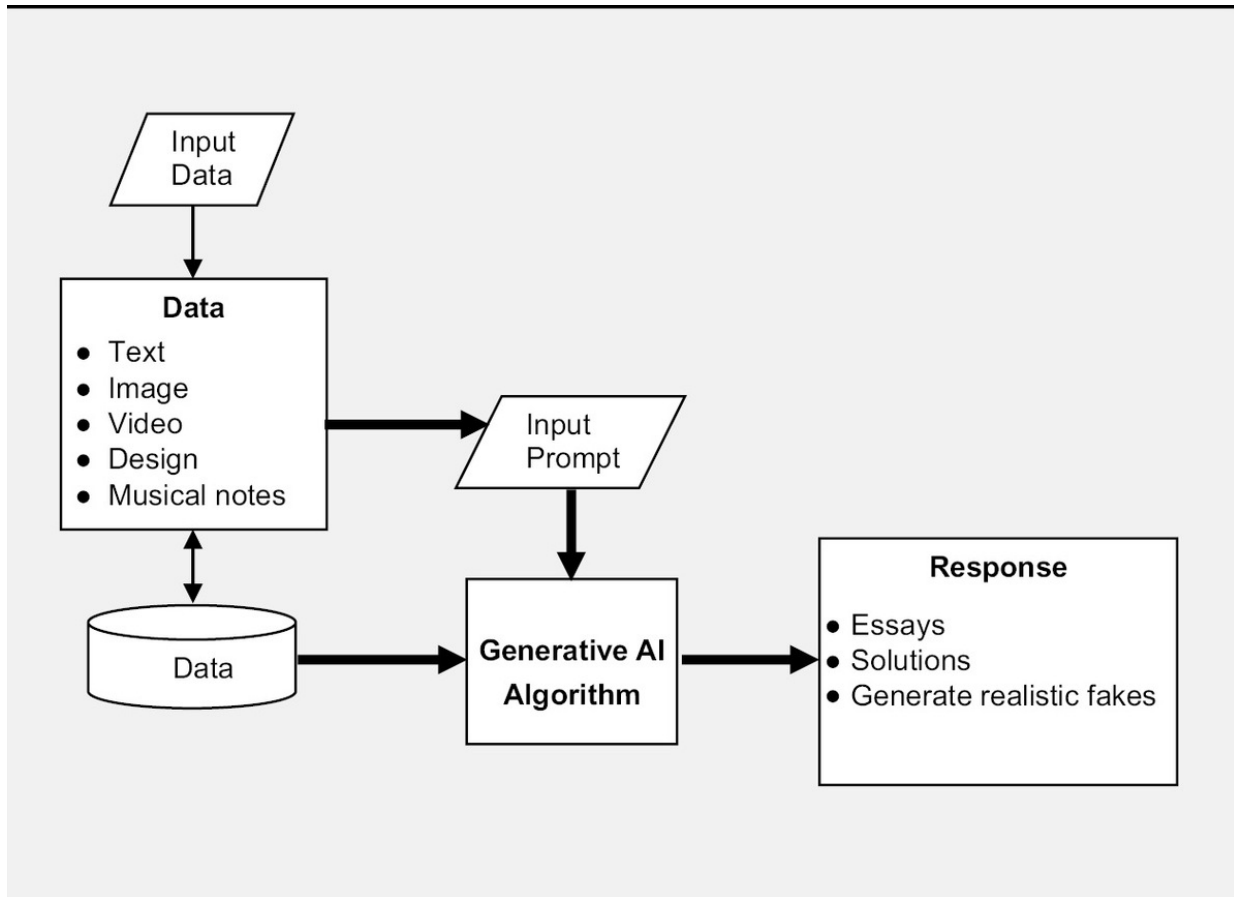


FIGURE 6.15 Example of Generative AI Process. [↗](#)

Typically, generative AI models apply a specific neural network to produce new content in reply to a query or prompt using techniques such as generative adversarial networks (GANs) and variational autoencoders (VAEs). These neural networks with a decoder and encoder can create realistic human faces, fake data for AI training or even replicas of individuals.

GANs consist of two neural networks, the generator and the discriminator that are united in a perpetual duel. The generator network has the function of creating data samples by commencing with random noise and endeavours to create data that is as realistic as possible. On the other hand, the discriminator network examines the data with the purpose of distinguishing between real and fake samples. Therefore, the training process is an iterative back-and-forth method with the intention of

generating more and more convincing data, while the discriminator refines its abilities in distinguishing between real from fake. Over a span of time, the generator develops into a master counterfeiter, creating data that is almost impossible for the discriminator to distinguish from real examples.

VAEs work differently from GANs. These are probabilistic models that are created on the framework of autoencoders, which are normally applied for dimensionality reduction and feature learning. The autoencoders consist of an encoder and decoder. An encoder network takes the input data and encodes it into a lower-dimensional space by mapping the data into a probabilistic latent space. On the other hand, the decoder considers the points from this latent space and tries to reconstruct the original data. This is achieved by sampling from the latent space's distribution to statistically generating data, leading to a deeper examination of the data's generative potential. GANs are inclined to create sharper, higher quality outcomes, but can be somewhat erratic during training, whereas VAEs are more consistent to train but may experience lower quality.

Several tools that apply transformers are now available such as BERT (Google's Bidirectional Encoder Representations from Transformers), OpenAI's GPT and Google AlphaFold. These neural networks have the capability to encode language, images and generate new content, with ChatGPT, Dall-E and Bard being common generative AI interfaces.

ChatGPT was first released towards the end of 2022 as an interface to OpenAI that provided the ability to interact and fine-tune text responses through a chat interface with interactive feedback. The ChatGPT interface has been improved with refined new versions and has been integrated in Microsoft's Bing search engine. ChatGPT provides human-like responses creating conversations that are engaging and appear more natural. OpenAI also introduced Dall-E for creating images from textual descriptions. It was trained on a considerable data set of images and their linked text descriptions and is classified as a multimodal AI application that recognises connections across multiple media, such as text, vision, and audio and creates images in multiple forms driven by user prompts.

Google Bard is a multimodal model that exploits the enhanced abilities of the Language Model for Dialogue Applications (LaMDA) to improve conversational AI. Bard is designed to recognise multiple data types, such as text, audio, images, and video that enhances its integrational reasoning abilities. This integration capability permits Bard to execute functions that include summarising text, translation, and code analysis. Google adopted a hurried strategy when initially releasing Bard to catch up with Microsoft. However, this rushed launching resulted in a negative initial acceptance of Bard when it incorrectly stated that the Webb telescope was the first to discover a planet in a foreign solar system. But Microsoft experienced similar outcomes with the ChatGPT implementations when it provided erroneous responses and behaved in an inconsistent manner. In the meantime, both Google and Microsoft have made significant progress in addressing similar unfortunate experiences, but more still needs to be done. The latest version of Bard has some exceptional features that include a multimodal function that is trained on a various data sets enabling Bard to comprehend and generate content over various formats; advanced neural architecture that applies a transformer model with effective attention structures that enable Bard to manage long contexts essential for in depth and complex conversations; and broad safety testing feature to mitigate risks of bias and harmfulness. Bard applications have well developed user interaction capabilities to facilitate the developers' efforts for generating innovative software.

Generative AI has a lot of potential, but it does have several significant limitations. For instance, generative AI is not always able to identify the content's source, which could be problematic for evaluating the bias of original sources, particularly for realistic-sounding content that renders it difficult to recognise incorrect information. Therefore, it may present incorrect and misleading information with users being concerned about the reliability of the information provided. Furthermore, when generating content, particularly for academic assignments, it may overlook copyrights and thus inadvertently support plagiarism. Additionally, generative AI facilitates the creation of fake news and may also create an environment

where it becomes easier to discredit authentic photographic evidence or imitate individuals as an effective means to achieve social engineering cyber-attacks. [Table 6.4](#) presents several generative AI tools for various applications regarding text, image, voice synthesis, music, and code generation; whilst [Table 6.5](#) shows the various AI chip design tool companies that support the various generative AI tools.

TABLE 6.4 Example of Generative AI Tools for Various Applications[📄](#)

GenerationApplication	Examples of Generative AI tools
Text	GPT, Jasper, AI-Writer, Lex, GrowthBar, ContentShake
Image	Dall-E 2, Midjourney, DreamStudio, Stable Diffusion, Craiyon
Voice Synthesis	Descript, Listnr, Lovo.ai, Murf, Podcast.ai, Synthesys, Fliki
Music	Amper, Dadabots, Loudly, Mubert, MuseNet, Soundraw,
Code	Codex, Codeium, Gemini, GitHub Copilot, CodeT5, Tabnine

TABLE 6.5 Example of Generative AI Chip Producers[📄](#)

AIChip Producers	Application
Nvidia	Autonomous driving, AI training & inference in data centres.
AMD	AI accelerator products, AI training, machine learning.
Intel	AI accelerator products, AI training & inference in data centres.
Google Alphabet	Machine learning accelerator chip that powers Google products like Translate, Photos, Search, Assistant, Gmail.

AIChip Producers	Application
AWS	Model training (Tranium chip) and inference (Inferentia chips).
IBM	Deep learning chip (highlighted use for fraud detection).
Alibaba	Hanguang 800 for inference in data centres.
SambaNova Systems	Systems for high volume generative AI workloads.
Cerebras Systems	For pharmaceutical companies to accelerate genetic and genomic research and shortens the time for drug discovery.
Groq	Focused on inference related to solutions for financial services.

Generative AI is often viewed as a general-purpose type of technology because it is applicable to a variety of industries for different purposes. [Table 6.6](#) provides some examples of how generative AI is used within some industry segments. However, it must be emphasised that generative AI is continually being enhanced, and as stated previously, has many challenges to overcome related to precision and reliability, trustworthiness, bias, hallucination, and plagiarism. A specific chapter will address ethical and legal issues in greater detail later. Such a chapter is extremely important because generative AI portrays the generated content so realistically that it is exceedingly difficult to identify incorrect information, particularly if the user is not familiar with the subject matter content being presented. Additionally, generative AI content is not transparent, therefore a user needs to be aware of the original source the results have been attained from and be extremely cautious regarding copyrights infringement risk.

TABLE 6.6 Example of Using Generative AI for Various Industrial Sectors [📄](#)

Industry	Application

Industry	Application
Finance	Monitor an individual's transactions history to develop more robust fraud detection systems.
Manufacturing	Amalgamate data from cameras, X-ray, and other metrics to detect defective parts and conduct root cause analysis.
Medical	Identify prospective drugs with higher usefulness value.
Legal firms	Draught and interpret contracts, evaluate evidence, and propose arguments.
Architectural firms	Create, model, and modify prototypes faster.
Gaming enterprises	Produce, design and model game content and levels.
Film and media companies	Produce cost-effective content and converting it into other languages with the actors' own voices.

Generative AI is here to stay, and its future development will mostly focus on making it more acceptable to users by mitigating the concerns mentioned previously as far as possible. Future enhancement of generative AI will likely place a greater emphasis on integrating the capabilities of the various available tools to provide better accuracy, less bias, and more transparency of the content being generated and enable organisations and individuals to exploit this technology to automate and strengthen the performance of tasks normally undertaken by humans.

6.3.6 Natural Language Processing (NLP)

NLP is a segment of AI where a computer program is developed to understand human language (natural language) in its spoken and written form. Although there is some overlap with what has been discussed previously, it is appropriate to address it separately due to its importance. The origin of NLP evolved from the field of linguistics, computer science,

and computational linguistics. Work with NLP commenced in the 1950s when Alan Turing created the Turing Test to assess whether (or not) a computer is truly intelligent. The Turing test entailed the computerised interpretation and creation of natural language as a critical benchmark of intelligence.

Up to the 1990s, NLP was mostly rules-based, applying human generated rules created by linguists who study every aspect of language, including vocabulary, grammar, and the sound of language to determine how computers would process language. An important milestone occurred on 7 January 1954 with the joint Georgetown University–IBM experiment. This experiment was a significant illustration of machine translation concerning the full automated translation of more than sixty Russian sentences into English. The subsequent several decades witnessed the development of rule-based parsing to identify patterns of text that represent specific concepts; morphology to analyse internal structure of words and the rules governing the formulation of words; and semantics to study meanings of words amongst other forms of natural language understanding.

In the 1990s, computers had higher processing speed that enabled developers to apply a rule-based approach for natural language processing through a statistical approach (i.e., data-driven) without a linguist generating all the rules. Hence, NLP became an engineer-based approach rather than linguist-based that applied a broader application of scientific disciplines. NLP has experienced phenomenal growth for the past few decades. Starting in the early 2000s, developers exploited further enhancements in computer processing power that enabled them to introduce unsupervised and semi-supervised machine learning algorithms. Thus, combining classical linguistics and statistical approaches. The outcome of this development is the initiation of practical applications and improved human interaction through chatbots and virtual assistants. The development cycle of NLP is approaching maturity, but it still has to overcome several concerns related to accuracy, accessibility, and widespread applicability in diverse industrial sectors, such as business intelligence, search engines, linguistics applications, and medical research.

Essentially, an NLP management system consists of two functions, namely the data preprocessing and developing the algorithm to process the data. Data preprocessing is the process of preparing and cleaning the dataset. It involves detecting, correcting, and/or removing corrupt or inaccurate details from a dataset. Hence, it identifies erroneous, incomplete, irrelevant segments of the dataset, and then amends, replaces, or deletes the corrupt or abrasive data content. The objective of data preprocessing is to convert the given dataset into a content and format that enables the computer programme (through an algorithm) to analyse and use it. This is accomplished by several methods that include the following:

- Tokenisation: This is the process of creating a digital representation by substituting sensitive information with nonsensitive information to protect sensitive data or to efficiently process large amounts of data. For example, it is used in cybersecurity related to payments to disguise the identity of the payment to prevent fraud.
- Stop word removal: This is the process of removing the words that commonly occur across all the documents permitting only unique words that offer the most information about the text to be retained.
- Stemming and lemmatisation: Stemming is the process of stripping attachments from words to obtain reduced word strings, and lemmatisation is the process of reducing morphological variants to one dictionary base form. Hence, lemmatisation clusters mutually unique modified versions of the same word. For instance, the word “sleeping” would be reduced to its basic form “sleep.”
- Part-of-speech (POS) tagging: This involves assigning a grammatical type such as noun, verb, adjective to each word in a sentence. The purpose is to comprehend the syntactical structure of a sentence and identify the grammatical roles of individual words.

When the dataset has been through the preprocessing function, a computer programme (algorithm) is used to process it. The commonly applied natural language processing algorithms fall under two key types,

namely rule-based and machine learning based systems. A rule-based system applies meticulously conceived linguistic rules, whilst a machine learning-based system applies statistical methods. Both methods learn to achieve tasks based on training data that are made available and fine-tune their processes as more data is dealt with. Therefore, an NLP algorithm refines its own rules through an iterative process by adopting a combination of approaches such as machine learning, deep learning, and neural networks. NLP as a system is a significant development because enterprises, irrespective of size, require an effective means of processing a considerable volume of unstructured and text-heavy data. There are two chief processing techniques used in NLP, namely syntax and semantic analysis.

Syntax refers to the sentence structure and how words are organised in a sentence according to set rules to make grammatical sense. Hence, NLP applies syntax (sentence structure language rules) to evaluate connotation from a specific language. In NLP, syntax techniques include the following:

- Parsing: This refers to analysing a sentence by dividing it into grammatical parts and identify the parts and their relations to each other. For instance, when an NLP algorithm is given the sentence, “The cat meowed.” Parsing breaks down the sentence into segments of speech, such that cat = noun, and meowed = verb.
- Word segmentation: A word is viewed as a linguistic unit made up of one or more morphemes. Hence, word segmentation determines the word boundaries in a sentence or a document deriving word forms. For instance, an algorithm can analyse a handwritten scanned document and recognise that the words are divided by white spaces.
- Sentence breaking: This is also known as sentence boundary disambiguation (SBD), sentence boundary detection, and sentence segmentation, where NLP decides where sentences begin and end. Hence, the algorithm can analyse text and by using sentence breaking can recognise the full stop (period) that separates the sentences.

- **Morphological segmentation:** This is a method for identifying and analysing the structure of words, which is valuable in morphologically-rich languages such as Arabic and Hebrew, where patterns of prefixes, suffixes, and roots form words. It is also helpful in machine translation and speech recognition. For instance, the word *unaccountable* would be split into [un[[account]able]], where the algorithm identifies “un,” “account,” and “able” as morphemes.
- **Stemming:** Stemming is the process of reducing inflected form of a word to one so-called “stem,” or root form, also known as a “lemma” in linguistics. For instance, in the sentence, “The cat meowed” the algorithm recognises that the root of the word “meowed” is “meow.” This is useful when a user is analysing text for all occurrences of the word *meow*, as well as all the variation of the form of a verb (i.e., its conjugations). The algorithm is able to recognise that they are basically the same word, despite the fact that the letters are different.

The other processing technique is semantic analysis, which essentially entails the use of and meaning behind words. Hence, an NLP algorithm is used to figure out the meaning and structure of sentences. There are basically three main semantic techniques that are applied:

- **Word sense disambiguation:** This is the process of determining the intended meaning, or sense, of a word within a specific context. The process is important for machine translation, information retrieval, question answering, and sentiment analysis. For instance, in the sentence, “The horse is in the stable.” The word “stable” has many meanings. Hence, the NLP algorithm will apply this method to understand that the use of the word here refers to a horse barn, and not any other word denoting stability as in steady or constant.
- **Named entity recognition (NER):** This technique focuses on identifying and classifying specific data points from textual content. For instance, an algorithm can analyse a document and identify all references to a specific establishment or product. By applying the

semantics of the text, it can distinguish between objects that appear to be the same. For instance, in the sentence, “John McPherson the farmer went to purchase a McPherson tractor,” the algorithm can identify that the two occurrences of “McPherson” as two distinct things, one being an individual and the other being a tractor manufacturer.

- Natural language generation (NLG): This is the process of translating data into text or speech by applying a database to establish the semantics underlying words and create new content. Hence, NLP is the process of accurately translating what is said into machine-readable data, which would enable NLG to apply that data to create a reply.

NLP employs deep learning as a fundamental approach, which is a type of AI that analyses and exploits the patterns in data to improve an algorithm’s interpretation. Deep learning models make use of a massive volume of labelled data, which could be rather challenging in terms of logical and processing power. The pioneering rule-based methods were cumbersome since the algorithms were informed what words, idioms, and phrases they are to search for in the text and provided with specific responses when they found a match. Deep learning models can automatically identify patterns and features in data without being explicitly programmed to do so. These models apply an intuitive (untaught) approach in which the algorithms learn to recognise the writer’s or speakers’ intention from numerous examples that mimics how a child would learn human language. This makes deep learning models immensely powerful and flexible approach. Several opensource tools that are commonly applied for NLP, and which are based on Python, include Natural Language Toolkit (NLTK) that has data sets and tutorials; Gensim containing a library for topic modelling and document indexing; and NLP Architect by Intel that uses a library for deep learning topologies and techniques.

The practical applications using NLP include the analysis of social media assessments to be used for customer feedback analysis; speech

recognition to recognise what the customer is saying over the phone for directing customer calls to the appropriate location; translating text, audio, and documents into another language; and analysing vast volumes of academic research material based on the metadata for academic research, amongst many others. These and other application areas illustrate that NLP's key advantage is its ability to enhance the means that humans and computers communicate with each other. However, the main challenge of NLP is to keep pace with the evolution of natural language and its inherent obscurity of meaning in the way words are used and the emotions expressed by the content. This translates to ensuring that an NLP algorithm is accurate and precise in interpreting human speech and text content that may in itself be imprecise and ambiguous; correctly interpreting tone and inflection of content (speech and text); and eliminates bias. NLP systems are not perfect and requires more effort to addressing the above issues.

6.3.7 Expert Systems

An expert system is a knowledge-based computer system that has the primary objective of addressing complex problems by extracting knowledge from its knowledge base applying reasoning and inference rules according to user queries, thus providing a decision-making capability like a human expert. However, it is critical to understand that an expert system is not intended to replace human experts but is used as an aid to humans in making a complex decision. Hence, expert systems do not mimic human thinking abilities but merely operate because of the knowledge base of the particular domain for which they are designed.

Expert systems are not a new phenomenon and were first developed over half a century ago by Edward Feigenbaum, a computer science professor at Stanford University and founder of Stanford's Knowledge Systems Laboratory. These systems are viewed as the first successful use of AI that initially originated in the medical domain. Over the years the knowledge base of a particular domain naturally increases with a resultant significant improvement in performance. Today, expert systems apply machine

learning and AI to replicate the behaviour or wisdom of domain experts, and their performance continuously intensifies as they attain more knowledge and experience. Hence, expert systems continuously accumulate knowledge and experience to convert these into wisdom by integrating them with an inference engine to respond to queries.

An inference engine basically interprets and assesses the facts in the knowledge base to provide an answer to a user designed query (refer to [Figure 6.16](#)). The basic function of an inference engine is to enable the expert system to derive deductions from the rules in the knowledge base. The inference engine typically applies two approaches for extracting information from the knowledge base, namely forward and backward chaining. The function of forward chaining is to read and process a set of facts to reach a logical extrapolation about a future occurrence, such as predicting the price of commodities (e.g., grain, steel, fuel). On the other hand, backward chaining has the function of reading and processing a set of facts to arrive at a logical conclusion about the reason something had occurred, such as assessing apparent symptoms to reach a medical opinion.

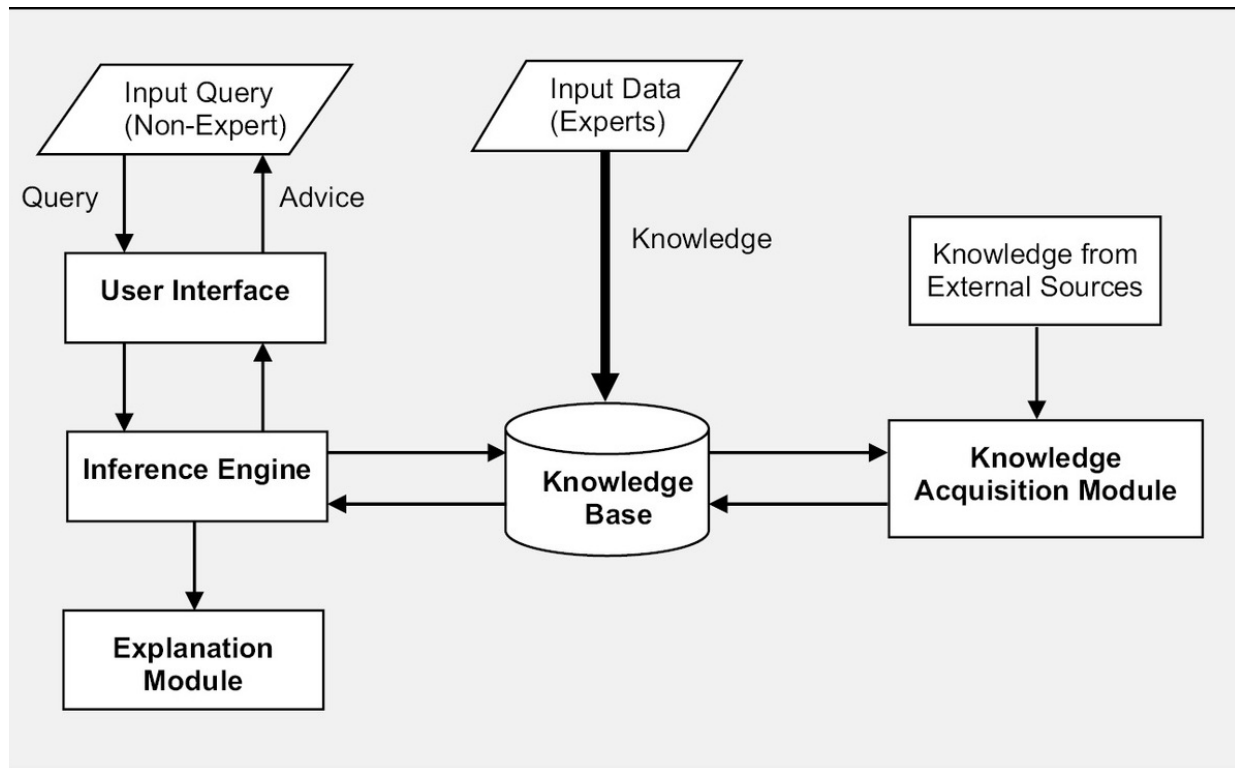


FIGURE 6.16 Expert System: Illustration of Inference Engine. [📄](#)

Both approaches are dependent on having a significant and accurate knowledge base. In a practical sense, human experts continually add information to the knowledge base with the aim of having perfect information, where nonexperts can utilise the system to solve complex problems that would usually require a human expert. Obviously, attaining perfect information is difficult to achieve in many domains because knowledge continually grows. The holistic expert system process is generally referred to as knowledge engineering. Knowledge engineering refers to the integrated process of creating and maintaining an expert system with essential information to solve a problem using knowledge representation methodologies, such as symbolic patterns where rules and axioms are applied to make inferences and deductions. Hence, an expert system's competences can be enriched by intensifying the knowledge base or/and generating additional sets of rules.

6.4 NEUROMORPHIC COMPUTING

In a previous section of this chapter the topic of neuromorphic hardware was discussed in detail. This section discusses the topic of neuromorphic computing, which goes beyond neuromorphic hardware and includes the software segment within the neuromorphic approach. However, some overlap with neuromorphic hardware is likely to cover the topic completely. Before, specifically discussing neuromorphic computing it is firstly appropriate to discuss neural networks, since they are an important aspect of neuromorphic computing.

6.4.1 Neural Networks

Artificial neural networks are algorithms (or software programs) that refer to a network of artificial neurons or nodes, which are inspired by the biological neural networks that form the human brain. Neural networks operate differently from conventional computers in that they could do many things concurrently known as parallel processing. The Von Neumann architecture, which forms the basis of the traditional computer, process instructions in a sequential manner. The concept used in a multiprogramming operating system (allocating a minute time sequence to each individual software application module) makes it appear that many things are happening simultaneously.

The artificial neural network is an inherently multiprocessor-friendly architecture that is specifically designed for parallel processing, which allows the execution of instructions in less time. However, the processors must be specifically designed for the neural network, which means that they perform a specific predetermined task. Additionally, artificial neural networks function differently from traditional computers because artificial neural networks can function through images, pictures, and concepts. On the other hand, traditional computers function logically according to a programmed set of rules and calculations. Hence, traditional computers learn by rules, while artificial neural networks learn through examples, by

performing a function and subsequently learning from it. Therefore, the applications adaptable to traditional and artificial neural networks are fundamentally and exceedingly different.

Furthermore, traditional computers function through programmes written in a higher-level computer language such as C and then transforming these programmes into the machine's assembly language. However, artificial neural networks are constantly adaptable by modifying their own programming, thus are able to "program themselves." As mentioned previously, neural networks require the use of multiple chips that are customised for the application, which result is extremely high-speed processing when compared to the traditional Von Neumann processors.

6.4.2 Neuromorphic Computing Concepts

Neuromorphic computing is the ability to model a computer processor to represent the human brain and nervous system that includes both hardware and software computing elements. Moreover, it makes use of several fields that include computer science, mathematics, electronics, biology, and physics to generate computer systems and hardware. Neuromorphic architectures are frequently modelled after neurons and synapses that are viewed as the fundamental units of the brain.

The human brain consists of many billions of neurons that communicate to each other using a combination of electrical and chemical signals to send information between different regions of the brain and the rest of the nervous system. The linkage-communication points between neurons are known as synapses. Research has shown that neurons and synapses are extremely flexible and much more energy-efficient information processors than the traditional Von Neumann computer systems. Neuromorphic computing is an emerging field that is currently the domain of research programmes at academic institutions and technology companies in a strategic alliance with national defence establishments.

Neuromorphic processors are viewed by many experts as overcoming the confines of Moore's Law established in 1965, which state that the

number of components on a single chip doubles every two years at minimal cost. The main thrust for neuromorphic research is the creation of artificial general intelligence (AGI), which has practical military and commercial implications. The conceptual basis underlying AGI is the notion of having an AI computer that can mimic the human's ability to understand and learn, by imitating the biological human brain and nervous system with equivalent cognitive ability. Such a significant development will unlock tremendous potential for next-generation semiconductors; transistors and accelerators; deep learning applications; and autonomous systems regarding AI, robotics, self-driving cars, and drones.

As stated previously, neuromorphic computing applies hardware that is based on the configuration, faculties and processes of neurons and synapses in biological brains by applying a spiking neural network (SNN). This approach means that spiking neurons (nodes) process and store/retain data like biological neurons with artificial synaptic devices being used as the linkage-communication points between the spiking neurons. Hence, these spiking neurons use analogue circuitry to transmit electrical signals that mimic brain signals. Whilst traditional Von Neumann computers convert data through a binary system, neuromorphic computing use spiking neurons to measure and encode the discrete analogue signal changes themselves.

6.4.3 Characteristics of Neuromorphic Computers

Neuromorphic computers have several unique characteristics. For instance, neuromorphic chips instead of having separate areas for the process and storage tasks, group these tasks on each individual neuron, known as collocate processing. The outcome of this approach is to resolve the Von Neumann bottleneck (using the same channel to retrieve data and instructions, so that only one task is executed at a time), thus increasing performance and decreasing energy consumption. Some neuromorphic chips (e.g., Intel Loihi 2) may have up to one million neurons that makes them extremely parallel by executing as many functions at one time as there are neurons, thus each neuron performs different tasks concurrently.

An advantage of neuromorphic computers is their scalability because one only needs to increase the number of active neurons to have larger networks. Hence, there is literally no limit to their scalability. In other words, neuromorphic computers have the unique ability to achieve significant performance under an expanding workload. Additionally, neuromorphic computers thrive on event-driven computation that focuses on processing events as they occur since the individual neurons and associated synapses process in response to spikes from other neurons. In other words, only those neurons that are processing spikes consume energy, while the remainder are inactive, which results in exceptionally efficient use of energy.

Neuromorphic computers are highly flexible because they are designed to respond to external everchanging stimuli through their spiking neural network (SNN) where each synapse is allotted a voltage output, and adapts this output based on its specific task. Hence, SNNs provide neuromorphic computers with the potential to rapidly adjust to new and changing environments by constantly learning and resolving innovative problems. Furthermore, due to the numerous neurons that process independently and concurrently, neuromorphic computers become extremely fault tolerant, where the failure of one component does not prevent the computer from functioning, since information is stored at various points. It should be noted that neuromorphic computers are mainly utilised in spiking neural network-based machine learning applications that are Turing-complete. This means that theoretically they can perform all general-purpose computation. Whilst neuromorphic computing has the potential to transform the applicable capacity of AI, it needs to overcome many technical difficulties. Neuromorphic computing is currently at its early stage of development and requires significant investment for it to be truly a viable computing platform.

Neuromorphic computers are chiefly available to experts. This is not surprising, and one could say that history is repeating itself. Up to and most of the 1970s traditional computers were accessible to technical experts only, behind closed doors. The computer room was like a holy sanctum for the

computer initiated elite. In the 1980s with the growth of personal computers and later the introduction of the internet, this exclusive barrier was being demolished. Neuromorphic computing is going through a similar phase. Once it becomes accessible to nonexperts and software developers create application programming interfaces, programming models and languages, it will make neuromorphic computers more widely available and acceptable. However, before this can happen, neuromorphic computers need to be more accurate, less technically complex, and cost effective. Once, again this is reminiscent of the past (the 1960s) when the cost of computers and their associated operating systems and technical complexities were only practically affordable for large organisations. Once the cost of computing came down and computers were more technically and user friendly in the 1980s and 1990s, the level of acceptance increased exponentially. This is likely to happen with neuromorphic computers in the not-too-distant future, but much further development is necessary.

Before neuromorphic computers can become widely used and accepted more applications software and algorithms must be released. Once again, this is not new. In the 1960s and 1970s, organisations had no option but to develop their own in-house software applications. Very few commercial packages were available at that time, even for rudimentary applications such as the payroll, stock control, and accounting. This is similar to what is happening with neuromorphic computing because most neuromorphic research is carried out with standard deep learning software and algorithms developed for Von Neumann hardware. There is a need to break away from the Von Neumann approach and come up with novel methods that go beyond the Von Neumann techniques. However, to do this there is a need for neuromorphic computing to have clearly defined performance standards. This would permit current and potential users to assess whether progress is being achieved in this field.

Neuromorphic computing has a lot of potential and has attracted substantial investment and research funding. Because of the smaller size and low power consumption, it is envisaged that neuromorphic computers will be utilised to execute AI algorithms at the edge instead of in the cloud.

In other words, the AI algorithms would be implemented directly on local devices at the “edge” of the network, rather than depending on centralised cloud servers. This means that processing of data would be conducted at or near the point where data is collected, eliminating the transmission of data over the network, resulting in faster response times and enhanced data security. Therefore, AI applications executing on neuromorphic computers would be able to adjust to their environment and would memorise what is essential, accessing external sources through the cloud if necessary to seek more information when required. This makes them ideal for Internet of Things (IoT) Devices such as security cameras and other linked devices that process data locally; for wearables, like smartwatches; and localised servers that enable on-site processing capacity for manufacturing plants or retail chains systems.

However, more important potential future applications include robots; drones; driverless cars; smart home appliances; natural language, speech, and image processing; data analytics amongst many others, where edge AI is essential. Neuromorphic computing, which may eventually feature emotional awareness in machines, is generating considerable debate about the ethics and legal issues surrounding the treatment of sentient machine behaviour. There are those who argue that sentient machines should be legally considered as a nonhuman animal. However, others contend that a sentient machine should be legally and morally viewed as a person to be protected by the same laws as human beings. These ethical and legal issues are considered extremely important for the future of AI development and will be addressed in separate chapters in greater detail.

6.5 CONCLUSION

Whereas the previous chapter was dedicated to computer hardware in relation to the advancement of AI, this chapter has focused on the other side of the coin, namely the development of computer software and its impact on the development of AI. Software is what drives hardware, therefore, hardware on its own is of no value unless it is supported by the appropriate

software and the reverse is also applicable. Hence, it is essential to recognise that hardware and software complement each other. Whereas, hardware defines the physical features of a computer, software is a broad term that is applied to indicate the applications, scripts and programmes that execute on a hardware device. This means that software is a logical collection of instructions, data or programs that are applied to operate computer hardware and execute specific functions. Software applications vary depending on the computer programme that is uploaded on the hardware, whereas the hardware has a predetermined, non-flexible function. This distinction between hardware and software is important.

This chapter has explored various AI programming languages that execute on the traditional Von Neumann computers and are essential for developing current AI applications. The AI programming languages examined included LISP, Python, Intelligent Computer-Aided Design (ICAD), KL-One, Official Production System (OPS5), Mercury and Prolog. Each language has its distinguishing features that take advantage of the current hardware architecture. Additionally, the chapter considers hardware description languages, namely, Verilog and VHDL that are applicable for specific hardware related to AI, such as Field Programmable Gate Arrays. These computer languages are used to create software algorithms on specialised hardware associated with AI that have the objective of controlling digital circuits to perform a particular function.

Although the traditional Von Neumann computer architecture is still currently the dominate hardware architecture platform for AI, it is slowly but surely giving way to other more sophisticated computer architecture, namely the quantum and neuromorphic computers. These are predicted to revolutionise AI development in the not-too-distant future. In this regard the chapter briefly discussed IBM's Qiskit software development kit (SDK) that is specifically designed for the IBM Quantum Platform to execute specific AI applications. IBM have adopted Qiskit, which is an open-source, python-based quantum SDK and has been designed for the IBM Quantum Platform or on simulators on a local computer at the level of circuits and algorithms. Qiskit has been created to facilitate software development for

IBM's cloud quantum computing service and supports Quantum Artificial Intelligence (QAI), which combines quantum computing with AI that aims to use the unique properties of quantum computers to enhance the capabilities of AI systems.

The chapter also discussed in some detail AI algorithms that are viewed as an essential category of software that enables the implementation of AI for performing tasks that would otherwise require human intelligence. Currently, these AI algorithms have mostly been designed to execute on the traditional von Neumann computer architecture. AI algorithms include several tasks such as speech and voice recognition; text recognition; image recognition; video analytics; generative AI; and natural language processing. However, it is predicted that these AI algorithms will also execute on quantum and neuromorphic computers.

Finally, the chapter discusses neuromorphic computing and spiking neural networks that are explicitly designed for parallel processing. The topic of neuromorphic computing goes beyond neuromorphic hardware and includes the software segment within the neuromorphic approach. However, some overlap with neuromorphic hardware was necessary to cover the topic more fully. This section examines the underlying computing concepts and characteristics of neuromorphic computing. AI applications executing on neuromorphic computers can adjust to their environment by memorising what is important and seek information from external sources through the cloud if necessary. This makes them ideal for AI applications that require an on-site processing capacity such as driverless cars; robots; drones; and smart home appliances where edge AI is essential. Neuromorphic computing, which may eventually feature emotional awareness raise ethical and legal concerns regarding sentient machine behaviour. However, ethical, and legal issues will be addressed in separate chapters in greater detail later.

REFERENCES

Arar, Steve. (2017). What is VHDL? Getting started with hardware description language for digital circuit design.

www.allaboutcircuits.com/technical-articles/hardware-description-language-getting-started-vhdl-digital-circuit-design/.

Burdescu, Dumitru, Dan, and Brezovan, Marius. (2001a). Algorithms for high level petri nets simulation and rule-based systems. *Annals of the University of Craiova*, Vol. 25, pp. 95–105. [↗](#)

Burdescu, D. D., and Brezovan, M. (2001b). High level petri nets and rule-based systems for discrete event system modelling. *International Journal of Smart Engineering System Design*, Vol. 3, pp. 81–97. [↗](#)

Dohlhom, Anders. (2011). *Petri Nets for Situation Recognition*. University of Skovde. [↗](#)

Forgy, Charles, L. (1981). *OPS5 User's Manual*. Department of Computer Science, Carnegie-Mellon University. [↗](#)

Hura, Gurdeep, Singh. (1993). Representation and processing of rule-based expert system using petri nets: A viable framework. *In Proceedings of the 36th Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 934–937, Detroit, MI, USA, 16–18 Aug 1993. [↗](#)

Keim, Robert. (2020). What is a Hardware Description Language (HDL)? www.allaboutcircuits.com/technical-articles/what-is-a-hardware-description-language-hdl/.

Murata, Tadao., and Yim, Jaeyeol. (1995). Petri net methods for reasoning in real-time control systems. *In 1995 IEEE International Symposium on Circuits and Systems (ISCAS'95)*, pp. 517–520, Seattle, WA, USA, 30 Apr – 3 May 1995. [↗](#)

Chapter 7

Artificial Intelligence Security

DOI: [10.1201/9781003653950-9](https://doi.org/10.1201/9781003653950-9)

“You have your way. I have my way. As for the right way, the correct way, and the only way, it does not exist.”

Friedrich Nietzsche

German Philosopher (1844-1900)

Security of computer systems has always been a challenge and currently this challenge has grown to be vastly bigger and tougher. Cutting-edge cyber attackers have a broad strike field, with an ever-growing assault domain, complex, and expanding infrastructure, and an upsurge of information. AI systems because of their devastating power have an exceedingly higher need to be protected. Furthermore, AI itself can be used as a powerful tool by cyber attackers to penetrate the cyber fortress. This hazardous environment is making it extremely difficult for organisations to safeguard data, manage system access, and promptly detect and respond to security threats. This chapter will examine this turbulent security environment and will specifically address the evolving threat landscape, intrusion prevention systems, crack public key infrastructure (PKI), and the centre for AI Safety.

7.1 INTRODUCTION

[Allan \(2023\)](#) from *Cyber* magazine noted that in 2023 the cybersecurity landscape witnessed an extraordinary transformation. She observed that the growing demand for effective, worldwide threat intelligence continues to increase as geopolitical and economic developments generate an ever more complicated and uncertain world for both businesses and consumers.

It is a well-known fact that technology can be used for the good or as a destructive instrument. This has always been the case. Unfortunately, the development of new technology and their associated tools are providing malevolent perpetrators with the means to assault the cyber fortress. Hence, organisations need to keep ahead of these malicious perpetrators and continually updating their security strategies and related implementation tactics. [Allan \(2023\)](#) argues that cybersecurity professionals must recognise that no one is absolutely safe from cyber-attacks and the need to monitor these rapidly evolving threats is becoming increasingly critical.

According to [IBM \(2024\)](#), an intrusion prevention system (IPS) monitors network traffic for potential threats and automatically blocks them by alerting the security team, terminating dangerous connections, removing malicious content or triggering other security devices. IPS is an advancement from intrusion detection systems (IDSs) that they detect and report threats to the security team. However, an IPS apart from having the threat detection and reporting functions also introduces an automated threat prevention capabilities that are at times known as intrusion detection and prevention systems (IDPS).

Basically, an IPS can promptly impede malicious traffic and therefore facilitate the work of the security teams so that they may divert greater effort on the more complicated threats. An IPS plays a crucial function of the enterprise security system, particularly in a multi-cloud network that has numerous access points and manages high volume of traffic, which renders manual monitoring and response an impractical alternative. Furthermore, enterprise security systems are under constant threats from malicious but expert cyber-attackers that also have access to state-of-the-art tools.

Therefore, the automated facilities of an IPS permits the organisation to address threats swiftly without placing undue pressure on the IT administration.

However, IPS is just one segment of a robust security system and needs to coherently integrate with other technology aspects for maximum effectiveness. One should note that an IPS is normally located behind the firewall and works in conjunction with it to give added protection by capturing threats that have escaped the firewall detection mechanism. Therefore, an IPS provides an additional level of security by recognising and filtering out threats that other segments of the security infrastructure have not detected. Public key infrastructure (PKI) complements the security systems and refers to the tools that are applied to generate and manage public keys for encryption. This is viewed as a conventional way of securing data transfers on the internet. In fact, PKI is commonly integrated with all web browsers and provide secure public internet traffic.

Organisations apply PKI to secure the communications they send back and forth internally and to ensure that connected devices can connect securely. PKI's fundamental concept is related to cryptographic keys that are part and parcel of the encryption process that authenticate the different individuals or devices trying to communicate with the network. PKI is important because the encryption and authentication it scrutinises makes it possible to ensure trustworthy secure communication online. Therefore, an PKI can prevent an intruder acquiring access to the network through a connected device by keeping away a potentially malicious and hazardous threat from the organisation.

The other important topic to be addressed by this chapter is the Center for AI Safety (CAIS), which is a non-profit organisation based in San Francisco. CAIS views AI as having the potential to overwhelmingly benefit the world, contingent on the premise that AI can be developed and applied safely. CAIS recognise that extraordinary progress has been made in AI but there are many basic problems in AI safety that have yet to be addressed. Hence, CAIS's mission is to mitigate societal-scale risks

associated with AI by conducting safety research, building the field of AI safety researchers, and advocating for safety standards.

7.2 EVOLVING THREAT LANDSCAPE

[Cyber Security Magazine \(2024\)](#) claim that 2023 witnessed an increasing adoption of AI in both offensive and defensive cybersecurity strategies. It maintains that AI-driven threat actors will become more sophisticated, and organisations will dedicate more resources and AI-driven tactics to stay ahead of the evolving threats. Additionally, it viewed AI as becoming a profound battleground for cybersecurity. This analysis supports Deloitte's argument that the interplay between attacker and victim has become a cat-and-mouse game in which each side perpetually learns and adapts, leveraging creativity and knowledge of the other's motives to develop new offensive tactics and defensive postures ([Deloitte, 2014](#)).

According to [Deloitte \(2014\)](#), the relatively static compliance or policy-centric approaches to security found in many financial services companies may be long outdated. They claim that the critical issue is whether the industry can generate a forceful, intelligence-driven approach to cyber risk management not only as a preventative measure, but also to detect, respond to, and recover from the potential damage that results from these attacks. Yakir Kadkoda, Lead Security Researcher from Aqua Security, predicts that in 2024 the cybersecurity landscape will experience a substantial change because of the strategic inclusion of AI by cyber attackers. For example, 'Package Illusion' strikes will likely be a notorious feature where AI will be applied to manipulate software dependency chains. This will unintentionally expose applications to provide unexpected or factually incorrect responses that do not align with its machine learning training data. This departs and conflicts with the normal and planned behaviour for an AI model, which should impart a reply based on the data it was trained upon.

Hence, Kadkoda argues that this tactic is expected to be part of a broader trend where AI is not just a tool for defence but also a weapon in the attacker's arsenal. Thus, he asserts that by abusing the trust that software

developers place in automated dependency management and suggestion systems, attackers can initiate a new class of supply chain vulnerabilities that are difficult to detect and mitigate. Such AI-driven incidents are anticipated to become a regular occurrence that will constrain the cybersecurity industry to actively transform itself to counter such threatening development. However, the cybersecurity industry cannot afford to be reactive but must take proactive and robust action against AI-powered threats by focusing on the creation of new standards and best practices that are supported by an environment of international alliance across industries.

The above concerns are also supported by [Allan \(2023\)](#). She observes that cyber threats are emerging across the globe because advances in technology have enabled malicious players to exploit innovative methods that such state-of-the-art technology can provide. [Allan \(2023\)](#) claims that ransomware due to its growing scope and complexity, alongside alliances with other threat actors through underground forums, remains a continuing hindrance for organisations worldwide. The growing sophistication and complexity level of these devious tools make it very difficult for current security tools to detect and identify them. [Allan \(2023\)](#) endorses Kadkoda's view that all industries need to embrace a cyber strategy that is constantly vigilant, actionably comprehensible, and adaptable to new threats.

The most serious threat that is likely to be posed by AI is through the creation of malicious large language models (LLMs). LLMs are a leading-edge natural language processing (NLP) that are designed to understand and generate human language. These advanced AI models are trained in massive amounts of text data, enabling them to recognise linguistic patterns, comprehend context, and produce coherent and contextually-relevant responses. LLMs embody a focused approach that has significantly enhanced the machine's capability to mimic human-like language understanding and generation. Hence, LLMs can be applied for many beneficial purposes. However, they also have the potential to be exploited for malicious purposes, such as spreading misinformation, creating fake news, and conducting cyberattacks. Furthermore, organisations need to be extremely aware that cybercriminals are continually modifying their

methods to exploit new weaknesses and capitalize on their gains. Therefore, organisations need to be responsive and keep abreast of these illicit developments to continuously safeguard their systems and data repository.

It must be recognised that the upsurge of AI has expedited matters for cybercriminals to instigate tactical strikes with less expertise and resources. [BlackBerry \(2024\)](#) analysing cybersecurity events from March to May 2023 focusing on critical infrastructure and large organisations found that threat actors deployed approximately 11.5 attacks per minute across all sectors, with the healthcare and financial sectors being the most targeted. It is noted that the healthcare and financial sectors are prime targets because of the critical and sensitive information held by these industries, such as bank account information, social security numbers, and personally identifiable information. Such information is viewed as having money-spinning potential for cybercriminals aimed at blackmail and identity theft. Additionally, threat actors will target healthcare facilities because they are fully aware that hospitals are unable to function offline for lengthy periods and will likely yield to ransom demands hastily to get back online. [BlackBerry \(2024\)](#) also found that attacks during the period under consideration were mainly through ransomware attacks that targeted local and state government systems.

[Kellerman \(2022\)](#) itemised five trends regarding cybersecurity in the evolving cyber threat landscape, namely: ransomware is getting worse; AI is enhancing cybersecurity; mobile is the main target; data breaches and network intrusions; and increased social engineering attacks. Furthermore, he argues that cyber threat actors present a financial threat, mainly by stealing intellectual property and proprietary information, including personal information that enables further criminal deeds such as identity theft and financial fraud. [Kellerman \(2022\)](#) maintains that ransomware is getting worse. The process with ransomware is to firstly appropriate the data, encrypt it and then demand a ransom. Moreover, he claims that AI is being implemented in all industries and is being merged with machine learning to bring about transformational changes to the cybersecurity landscape. AI is being used to develop applications to automatically detect

threats. However, it is also being utilised to create smart malware and launch strikes to circumvent security protocols that manage data.

Kellerman's research shows that cybersecurity trends suggest a significant increase in mobile banking malware and attacks. Hence, handheld devices are attracting cyber threat actors, since these devices apart from posing a financial threat through mobile banking applications, also contain personal data such as photos, messages, and transaction histories. This is likely to prompt a demand for anti-virus and malware software for smartphones. Additionally, Kellerman notes that data breaches and network intrusions remain a primary source of apprehension for organisations and individuals given the sensitivity of the data being held. Moreover, there are some cyber-attack practices that disguise intrusions such as multi-routing, covert scripts, protocol impersonation, and traffic flooding that increase the vulnerability of users.

Increased social engineering attacks are also on the increase with more sophisticated cyber-attack techniques being applied providing cybercriminals with further access to sensitive and confidential information. Some of these methods include phishing, spear phishing, fraudulent communications, and distributed denial-of-service (DDoS) attacks. A DDoS attack is particularly devastating because the attacker literally floods a server with internet traffic to prevent users from accessing connected online services and sites. As for phishing emails, these pretend to be renowned services or organisations such as Amazon or Apple so that threat actors can collect login credentials. [Kellerman \(2022\)](#) suggests that organisations assess whether they are adequately safeguarded against these threats and threat actors and recommends that they reflect on the following questions: Do you know who might be targeting you and why you could become a target? Is there anything specific to your organisation or operations that they may be after? Answering these questions in a formal and organised focus group discussion will likely help organisations to initiate a way forward to tackle this growing cybersecurity concerns.

[Cyber Security Magazine \(2024\)](#) notes that Aqua Nautilus an organisation focusing on cybersecurity research in the cloud native stack,

captures more than 80,000 cloud native attacks every month, specifically those unique to containers and microservices that other platforms cannot see. Cloud-native stack describes the layers of cloud-native technologies that developers use to build, manage, and run cloud-native applications. The infrastructure layer that is managed by third-party cloud providers forms the basis of the cloud-native stack, which consists of the supporting components of cloud-native application development, including operating systems, storage, network and other computing resources. A growing concern is that as the cloud computing landscape continues to expand, the attacks within its perimeters are also being refined.

[Cyber Security Magazine \(2024\)](#) reports that there are early indications of state sponsored threat actors who have devised methods to execute code directly in user space memory without triggering a system call (i.e., `execve` and `syscalls`) that are usually scrutinised by security systems, consequently evading traditional detection mechanisms. A “syscall” is a programming method in which a computer program requests a service from the operating system on which it is executed, while a “`execve`” system call is used to execute a new program and transforms the calling process into a new process. Hence, these sophisticated approaches are striking at the heart of the system. [Cyber Security Magazine \(2024\)](#) predicts a significant shift in tactics from cloud attackers by exploiting userland execution methods that implement code that generally replicate what the `execve` system call does.

Hence, these new security evasion methods need to be countered by implementing nuanced behavioural security measures that include using AI and machine learning algorithms, which are able to recognise normal user behaviour and detect anomalies, along with improving memory scanning and process monitoring technologies. Such a proactive approach is critical for revealing and mitigating threats that circumvent conventional detection procedures to safeguard the ever-evolving cloud ecosystem. However, AI should also be viewed as a double-edged sword since (as stated previously) AI can be applied for the creation of malicious large language models (LLMs) that poses a complex and serious threat.

Another potential threat target area is the result of the proliferation of eBPF technology, which can execute programs in a privileged context such as the operating system kernel. eBPF is being used in areas related to networking, security, or observability, where conventional tools are unable to execute with adequate efficiency or detect events with sufficient granularity. Hence, this feature permits users to apply eBPF filters to specific sockets beyond just packet filtering. For instance, custom scripts can be called with minimal overhead at various points in the kernel. eBPF programs can be attached to various kernel hooks to gather data about what is taking place. This makes eBPF a very powerful tool that enables a whole new class of observability tools within the operational frameworks. However, eBPF is likely to attract the attention of threat actors as it gains broader acceptance across an assortment of industries and is implemented in a variety of tools. As an outcome of this development, [Cyber Security Magazine \(2024\)](#) foresee a heightened prevalence of eBPF deployment assessments, evasion and disabling tactics. This will compel cyber security developers to take additional rigorous security measures and proactive threat mitigation strategies.

Combating cyber threat actors will continue to be a running battle. In fact, it is a lasting state of war with no clear conditions that would lead to its conclusion. However, the technology industry has a responsibility to their users to ensure that the cyber network and the associated systems and data are safeguarded from all types of potential threats. The industry must keep a step (or more) ahead of the threat actors. This can only be achieved with the collaboration of all stakeholders on an international level, roping in governments, technology providers and a broad range of industry.

7.3 INTRUSION PREVENTION SYSTEMS (IPS)

An IPS is a security system that continuously monitors network traffic to identify threats and incorporates an intrusion detection prevention system (IDPS). Hence, an IPS system is more inclusive in the way it addresses potential security threats. An IPS is placed inline, in the flow of network

traffic between the source and destination and usually sits just behind the firewall. [Figure 7.1](#) shows the board design of a robust cyber security system regarding the location of an IPS in relation to the Firewall. [Figure 7.1](#) illustrates that the IPS and the Firewall have different functions. The firewall is designed to prevent unauthorized access into or out of a computer network and to make sure internet users without access are not able to interface with private networks, or intranets, connected to the internet. On the other hand, the IPS examines network traffic and specifically spots threats and takes appropriate action to address the incoming threat.

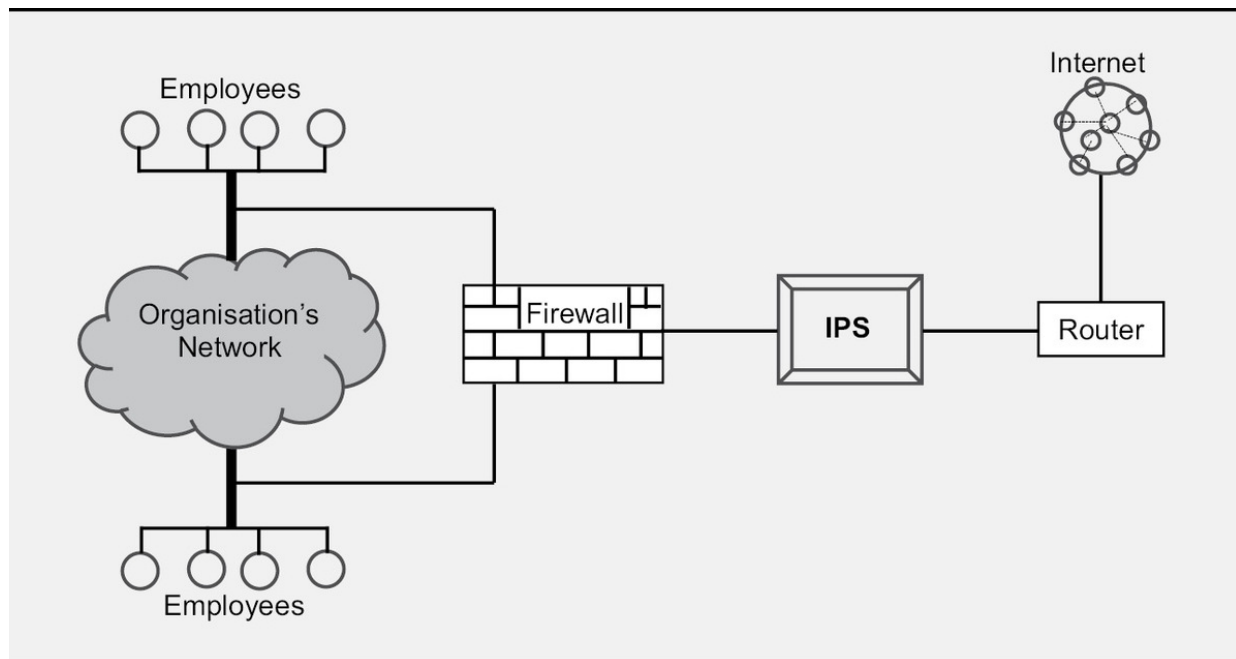


FIGURE 7.1 Board Design of a Robust Cyber Security System. [↗](#)

Hence, an IDPS monitors a network for threats and then takes the necessary action to stop any threats that are detected. Moreover, an IDPS is similar to an intrusion detection system (IDS) but while both systems detect threats and forward alerts about them, an IDPS also endeavours to provide a solution for those threats. An IDPS is synonymous with an IPS and the terms are frequently used interchangeably. However, when referring to IPS it is normally assumed that one is referring to the threat hunting function of

an IDPS. Therefore, an IPS is viewed as a security system that continuously monitors network traffic to identify threats and incorporates an intrusion detection prevention system (IDPS). The functionality of an IDPS depends on the provider, the desired deployment method, and the explicit requirements of the organisation using it. Basically, there are two types of IDPS, namely network-based and host-based.

Network-based IDPS (NIPS) is a type of IDPS that is installed at specific points within a network to monitor all of that network's traffic and scan for threats by analysing the activity taking place and matching each activity with a database of known attacks that is created manually by a security expert. Therefore, if an activity finds a match with a known threat in the database it is not permitted to carry on through the network. A NIPS is often implemented at the boundaries of networks, such as in routers or modems, behind firewalls, and at network remote access points. Furthermore, there are two categories of NIPS, namely wireless intrusion prevention systems (WIPS) and Network behaviour analysis (NBA).

An WIPS monitors wireless networks for the presence of rogue access points and unrecognized devices by examining the network's radio frequencies and are implemented in wireless networks and in locations that are vulnerable to unauthorised wireless access. On the other hand, network behaviour analysis (NBA) systems assess network traffic for unusual patterns of activity such as a distributed denial of service attack (DDOS), where thousands of requests are sent to the network to overwhelm it. The individual requests themselves might look valid, but collectively they suggest a problem. NBA systems often support a NIPS of an organisation's internal network.

A host-based IDPS (HIPS) are implemented on a single host such as a key server with critical data or public servers that are gateways to the organisation's internal network. They can also be used on mobile devices. These types of systems monitor and analyse traffic flow on its host system to detect host operating system activity and internet protocol suite (TCP/IP) activity. HIPS have the specific objective of detecting and blocking malicious activities.

IDPS applies several approaches to detect threats, namely signature-based, anomaly-based, and protocol-based. Signature-based detection examines network traffic, compares it to known signatures on a database of signatures, and generates an alert when a match is made. This approach is effective for detecting well-known threats, but new or unusual threats are likely to avoid detection. Anomaly-based detection is used to identify both computer and network threats by training the system with a normalised baseline that depicts how the system normally works and should work. This approach is continually being improved by advancements in AI and machine learning algorithms. Basically, the approach compares a random selection of network activity with a baseline standard of network activity. If the random selection varies sufficiently from the baseline, then the threat triggers action. However, while this detection method reveals and apprehends novel threats, it also generates more false positives than signature-based threat detection. The protocol-based (or policy-based) threat detection approach is very similar to the signature-based method in that it uses a database of specific protocols defined manually by a security expert and blocks any activity that violates those protocols. When a perceived threat is detected, the system may take several alternative actions contingent on its set up and the nature of threat detected. The preventative actions include:

- Alerting administrators when an automatic action might not be suitable, or when the system is uncertain if there is a false positive.
- Employ banishment vigilance by barring incidents before they occur. This is achieved by blocking traffic or flagging users from a threatening IP address such as blocking an IP address that has failed a password check too many times.
- Change the security environment. This approach modifies the security setup of the network to prevent the threat from gaining access such as reconfiguring a firewall.
- Modify the attack content. This method entails automatically altering the content of the attack. For example, if a suspicious email is

identified, the system would eliminate any facet that may have malicious content and pose a threat to the network, such as the email attachments.

An IDPS has the primary objective of revealing threats that might slip through by assessing system activity for potential threats and taking action without human intervention for the simpler threats. Complex threats are highlighted for human intercession to enable security teams to rapidly respond to these threats before they can cause harm. Additionally, an IDPS ensures that user and security procedures are continuously enforced and the rule-based feature guarantees that threat detection is applied in a consistent way. Furthermore, the outcome of having far less human intervention is that fewer individuals have an opportunity to interact with private data, which may be a regulatory requirement in various industries.

As stated in the beginning of this section, the IDPS features are a subset of an IPS. It should be noted that an IPS works as part of other security solutions (i.e., the firewall), and is able to identify threats that those other solutions cannot detect. This means that it provides a layer of additional security and offers superior protection due to a high level of application awareness. An IPS is placed directly in the flow of the network traffic between the source and destination (refer to [Figure 7.1](#)). This is what differentiates an IPS from the traditional intrusion detection system (IDS). It should be noted that an IPS is proactive, while an IDS is a passive system that scans traffic and reports back on threats. Hence, an IPS usually sits directly behind the firewall and analyses all traffic flows that enter the network, acting automatically, when necessary, which saves a lot of time effort for the IT team. Additionally, an IPS sifts for malicious traffic before it can approach other security devices and controls, hence it mitigates the workload for those controls and permits them to function more efficiently. It is also noted that an IPS can be configured to reflect the organisation's specific security policies making it a very robust and flexible security tool.

An IPS is critical for having a secure system because a multi-cloud network is a highly linked environment with numerous access points and

high traffic volume and has an expanded attack surface making it highly vulnerable and exposed to greater threats. This means that manual monitoring with a quick response in this type of environment becomes impractical and unmanageable. [Table 7.1](#) provides details of typical attacks that may be detected and prevented by an IPS. According to [Palo Alto Networks \(2023b\)](#), an IPS must have several vital features that include the following:

- IPS offers vulnerability protection by analysing and monitoring the attack lifecycle for breaches, infections, and ransomware.
- Antimalware protection by having a stream-based scanning engine that identifies and detects known malware and its unknown varieties and barring them swiftly.
- Broad command-and-control protection by blocking and preventing unknown C2 inline. A C2 channel serves as the communication link between the compromised device (bot) and the C2 server. This channel is essential for issuing commands, receiving instructions, and exfiltrating data. Hence, this IPS feature is extremely important because after an initial infection, cyber actors link with the host machine through a covert C2 channel, which is used to pull down additional malware, issue further instructions, and steal data. It is noted that these channels cannot be stopped with traditional signature-based approaches, therefore an IPS having this capability is critical.
- IPS includes an automatic multi-factor authentication feature that requires additional credentials beyond a simple password before the end user can gain access to an application or system. Security operations teams are able to act swiftly, quarantine, and validate policy to control potential infections.
- IPS allows incident response teams to immediately determine which systems are under attack and which users are potentially infected so that it is more efficient than guessing based on IP addresses. This greatly simplifies network security policy creation and management.

- IPS can automatically assess threat intelligence to keep abreast of potential attacks.

TABLE 7.1 Details of Attacks Detected and Prevented by IPS [📄](#)

Attacks Detected and Prevented	Attack Details
Address Resolution Protocol (ARP) Spoofing	Re-directs traffic from a legitimate system to the attacker. Fake ARP messages sent by an attacker create a link between the attacker's MAC address and IP address of an attacked system.
Buffer Overflow	Uses vulnerabilities in the buffer overflow to overwrite memory and corrupt the execution of an application.
Distributed Denial of Service (DDoS)	Initiates a massive flood of traffic from distributed computers meant to overwhelm a system making it unavailable for legitimate requests.
IP Fragmentation	Exploits datagram fragmentation mechanisms confusing the targeted system about how to reassemble TCP/UDP datagrams.
Operating System (OS) Fingerprinting	Exploits vulnerabilities in the OS.
Ping of Death	Using a ping command, an attacker sends oversized or malformed packets that crash a system.
Port Scanning	Scans for an open, unprotected port to exploit.
Server Message Block (SMB) Probes	Captures SMB protocol authentication requests to relay them to the attacker's host.
Smurf (a DDoS attack)	Using Internet Control Message Protocol (ICMP) packets to overwhelm a system.

Attacks Detected and Prevented

Secure Sockets Layer (SSL) Evasion	Exploits SSL and Transport Layer Security (TLS) encryption that hides malicious content to avoid detection and get past network security.
SYN Flood	Considerable volume of SYN (synchronize) packets are sent as connection requests to overwhelm a server or firewall.

Source: [Fortinet \(2024\)](https://www.fortinet.com/resources/cyberglossary/what-is-an-ips). What Is Intrusion Prevention System (IPS)? Definition and Types: Understand how IPS works, attacks prevented by IPS, and differences to IDS. Fortinet, Inc. www.fortinet.com/resources/cyberglossary/what-is-an-ips

7.4 NEXT-GENERATION FIREWALL

As stated previously, an IPS sits directly behind the firewall. Hence, it is appropriate at this stage to briefly discuss the next-generation firewall (NGFW) to have a complete picture of the security systems process. According to [VMware \(2024\)](#) the NGFW is within the 3rd generation of firewall technology that are designed to address advanced security threats at the application level through intelligent and context-aware security features. An NGFW is designed to include the traditional firewall capabilities but is enhanced to facilitate better decision making about what traffic is to be allowed. VMware maintains that the NGFW can filter packets based on applications and to inspect the data contained in packets, rather than just their Internet Protocol (IP) headers. VMware predict that NGFW will likely provide the following features depending on the NGFW provider:

- Ability to filter traffic and apply complex rules based on application and not just based on port.
- Block traffic from specific applications, as well as maintain greater control over individual applications.

- Allow deep-packet inspection that directly inspects the data contained within the packets. It should be noted that traditional firewall technology only inspects a packet's IP header to determine its source and destination but not its data.
- Have an inbuilt IPS to monitor the network for malicious activity and blocks it where it occurs.
- Improve performance by allowing the firewall to monitor large volume of network traffic without slowing down the organisation's operations and thus avoid business functional disruption.
- Have an external threat intelligence feature or the ability to communicate with a threat intelligence network to ensure the latest threat information and to identify threat actors.
- Have an integrated antivirus and malware protection feature.
- Have access to a Firewall as a Service (FWaaS) offered by a service provider, which is cloud-based and provides scalability and easier maintenance. This allows resources to scale automatically to meet processing demand and release enterprise IT teams from dealing with the burden of handling patches, upgrades, and sizing.

The above illustrate that NGFW will deliver enhanced robust security when compared to a traditional firewall that have limited capabilities. Traditional firewalls are not able to apply application-specific rules, protect against malware, or detect and block anomalous behaviour. Hence, experienced threat actors can easily evade detection by entering through a nonstandard port, which an NGFW is designed to prevent.

7.5 CRACK PUBLIC KEY INFRASTRUCTURE

According to [Antony_\(2023\)](#), the main objective of a Crack Public Key Infrastructure (PKI) is to manage digital certificates and provide a powerful security tool that supports numerous operations. He argues that the application of a PKI by an organisation unequivocally exhibits a commitment to cybersecurity since it authorises passwordless (not having a

password) authentication and encrypted communication. Moreover, PKI is viewed as a foundational piece of a modern, zero-trust security framework by the National Security Agency (NSA), Cybersecurity and Infrastructure Security Agency (CISA), and National Institute of Standards and Technology (NIST). These institutions are of paramount importance in the United States and are responsible for maintaining the nation's security and resilience against cyber threats.

Antony maintains that at a very high-level, PKI permits organisations to employ encryption for their various security requirements by allowing phishing-resistant authentication to applications or Wi-Fi and permitting the client's web servers to communicate in encrypted tunnels. PKI is specifically used to secure credit cards and car keys because it supports an all-inclusive infrastructure that encompasses many extremely complex components by generating digital certificates that replaces passwords, which take up to 300 trillion years to crack with a supercomputer. [Antony \(2023\)](#) explains that these widely used PKI certificates utilise public-private key cryptography that enable everyone to use a public key to encrypt a message and transmit it to someone else, without fear that the message will be made known to a cyber attacker even though they may know the public key. Additionally, Antony argues that frequently these private keys are secured on the hardware of the devices, which are practically impossible to steal, unlike the use of passwords that are extremely vulnerable.

PKI use Asymmetric Encryption where the sender encrypts the data, and the recipient decrypts it, however with the sender and the recipient using two different keys. In other words, this method uses a public key-private key pairing, where data is encrypted with the public key can only be decrypted with the private key.

7.5.1 What is a Digital Certificate?

A digital certificate is a public key certificate that cryptographically connects ownership of a public key with the entity that owns it, with the specific aim for sharing these keys for encryption and authentication. These

digital certificates apart from having the public key being certified, also includes information identifying of entity that owns it, metadata (information that describes the data) relating to the digital certificate, and a digital signature of the public key that the certificate issuer created.

PKI has three key functions, namely distribution, authentication and revocation of digital certificates. As stated previously, public key cryptography is based on a system of key pairs, namely: (a) one private key to be retained by the owner and used for signing and decrypting; and (b) one public key that can be applied for encrypting data sent to the public key owner or authenticating the certificate holder's signed data. Thus, the digital certificate allows entities to share their public key so it can be authenticated. Hence, a digital certificate is a phishing resistant system of verifying one's identity with a trusted third party and is one of the most reliable and safe method of ensuring that the organisation's authentication mechanisms are phishing resistant.

Digital certificates are commonly used for priming the Secure Sockets Layer (SSL) connections between web browsers and web servers, to share keys for encrypting and decrypting web content, and provide assurance that unauthorised actors have not modified data content. They are supported by various operating environments including mobile devices, laptops, tablets, internet of things (IoT) devices, and networking and software applications that assist to protect websites, wireless networks and virtual private networks. Digital certificates are also used in online and offline environments for giving cryptographic assurance and data privacy.

There are several specific applications for exploiting digital certificates, these include the following:

- Domain validation regarding websites to disclose that these are trusted and authentic.
- Ensuring secure email by identifying one user to another and for electronic document signing by having the sender digitally signing the email, and the recipient verifying the signature.

- Ensuring that credit and debit card transactions are secure and authentic by having chip-embedded digital certificates in a credit and debit card that connects with users and banks.
- To authenticate transactions for digital payment companies related to automated teller machines (ATMs), kiosks and point-of-sale devices connected with a central server at their data centre.
- Prevent the theft of broadband service through device cloning by having computer hardware manufacturers embed digital certificates into cable modems.

Many companies have recognised the growth of cyberthreats and are finding ways of attaching digital certificates to IoT devices that operate at the edge and within their enterprises with the aim of preventing cyberthreats and protecting intellectual property.

7.5.2 Key Components of a PKI

[Antony \(2023\)](#) argues that the components of a PKI comprise of everything an organisation requires to feel comfortable managing the entire certificate lifecycle. Hence, he identifies the following contemporary components of a PKI:

1. Certificate Authority (CA)

- A CA issues digital certificates to the authenticate owner of the public key.
- PKI requires the user (client) to generate a public-private key pair.
- User sends the public key and information to be imprinted on certificate to CA.
- CA creates a digital certificate with the user's public key and certificate attributes.
- CA sign certificate with the private key.

- Root and Intermediate CAs are issued.
- Root CA is securely stored away from public internet.
- Intermediate CA is sent to user as the working copy.

[Figure 7.2](#) illustrates the CA process. It should be noted that there are two trusted types of certificate authority, namely root and intermediate. A certificate authority passes through two phases, the intermediate phase and the root phase. The root CA is authorised to verify the identity of a person and signs the intermediate certificate that is issued to a user. Hence, the root certificate is not issued directly to the user, but instead an intermediate CA is issued. This means that the immediate CA is the working copying, while the root CA is stored securely away from the public internet. In other words, the root CA is completely safe and cannot be compromised under any circumstances.

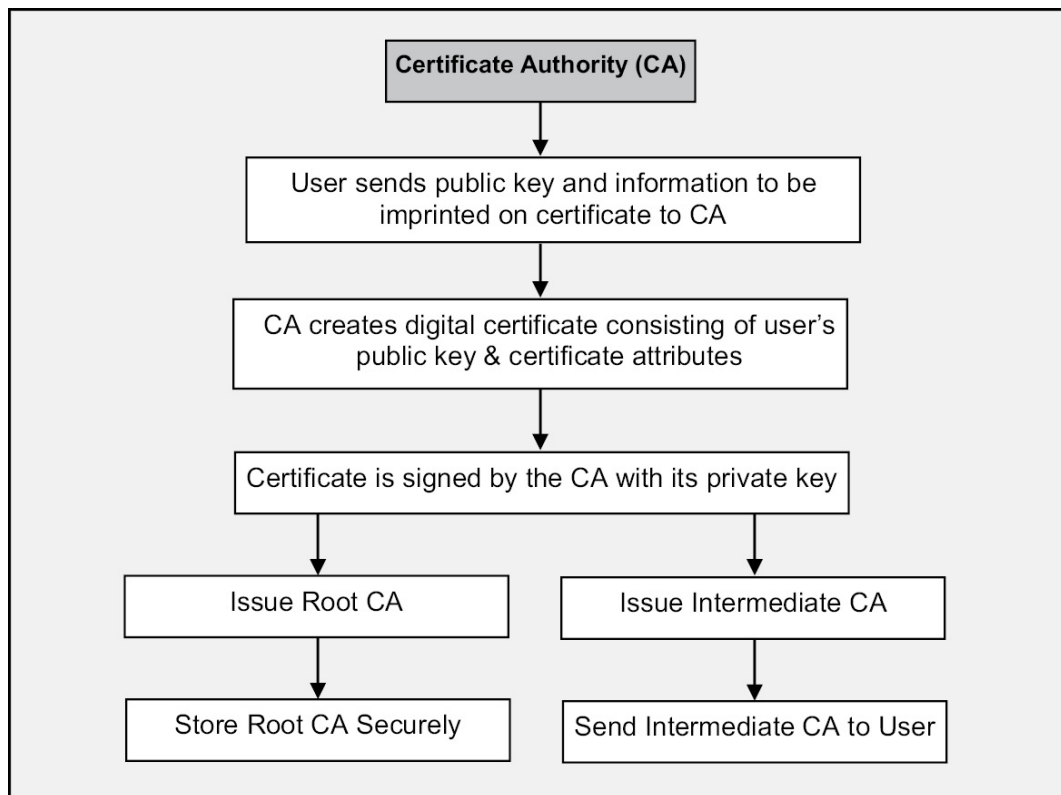


FIGURE 7.2 Certificate Authority Process. [↗](#)

1. **Certificate Templates.** These define the format of the certificate, what it will be used for, and the permissions required for receiving a certificate. For example, the certificate template will contain information regarding the validity period and the overriding validity period of the certificate; key size; signature algorithm; and use of the certificate such as server authentication, client authentication, email protection, amongst others. Having a misconfigured certificate template may lead to serious compromises.
1. **Certificate Lifecycle.** [Figure 7.3](#) illustrates that the certificate lifecycle process consists of several stages described below.

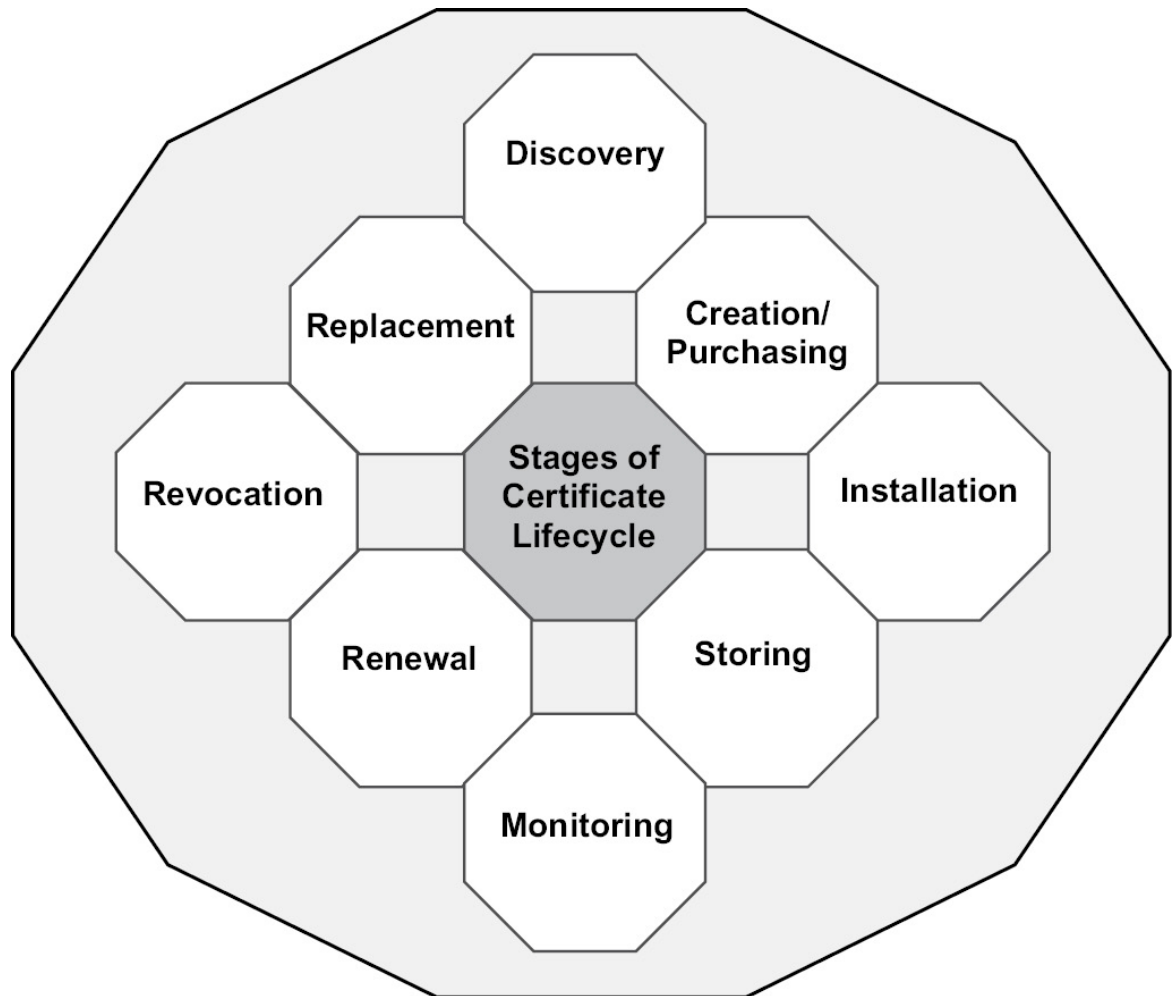


FIGURE 7.3 Certificate Lifecycle Process. [↗](#)

- **Discovery.** This stage searches the network for missing, expired, compromised, or unused certificates that must be revoked, renewed, or replaced. The aim is to find gaps in the security of certificates to relay these gaps to the monitoring phase to seal these breaches. Certificates are inventoried for reference in future Discovery phases and for auditing purposes.
- **Creation/Purchasing.** This stage creates a certificate as requested by an online user, organisation, or device from the CA, which contains the public key and other enrolment information needed to enrol the user. This permits the CA to confirm the given information and if it is authentic creates the certificate. The CA used to create the certificate can be owned by the organisation that desires the certificate, or by a third-party. If the certificate is obtained from a third-party, then it must be purchased from that third-party.
- **Installation.** This stage installs the certificate in a secure but accessible location to permit users trying to verify the authenticity of the certificate to have access to it. Once the certificate is installed, the CA consigns the policies in place to ensure the security and proper handling of the certificate.
- **Storage.** Once the certificate is installed it is kept in a secure location that prevents it from being compromised in any way but still be accessible to users for appropriate utilisation. [Figure 7.2](#) illustrates the use of intermediate and root certificates to always ensure having an uncompromised certificate copy.
- **Monitoring.** This is a continuous activity where the certificate management systems (either automatic or manual) scrutinise for breaches, expirations, or compromises of digital certificates by utilising the inventory of certificates created during the Discovery stage to keep track of when certificates should be revoked, renewed, or replaced. The system is then able to move

these certificates to the next stage, which can be renewal, revocation, or replacement.

- **Renewal.** A certificate is renewed when its expiry date is attained but the cumulative life including renewals should not exceed a period of more than 5 years. Certificates can be automatically renewed, or renewals can be initiated by an administrator.
 - **Revocation.** A certificate is revoked if it is compromised, stolen, or otherwise negatively affected. When revoked, a certificate is placed on a Certificate Revocation List (CRL) to ensure that other CAs know that this is no longer a valid certificate.
 - **Replacement.** The certificate is replaced when users transfer from paying for certificates to creating their own PKIs and CAs. However, this is seldom done because renewing a certificate from the original provider is much easier.
2. **Hardware designed for cryptographic functions.** A Hardware Security Module (HSM), which is not mandatory, generally enhances the security of the PKI when implemented since it safeguards and manages digital keys. It serves as the groundwork for building a secure enterprise PKI infrastructure that includes creation, rotation, deletion, auditing, and support for application programming interfaces (API's) to integrate with various applications.

TABLE 7.2 Commonly Used PKI Algorithms[↗](#)

PKIAlgorithm	Details
AES 256 Certificate	This algorithm is the current encryption standard that keeps track of vulnerabilities. A new standard would be implemented when the encryption is breached, which is virtually impossible.

PKI Algorithm	Details
Diffie Hellman	A mathematically complex encryption method that creates a private encryption key, which is shared between two parties over a secret communication on a public network.
RSA Key Exchange	Similar to the Diffie Hellman algorithm but factors large integers that are the product of two large prime numbers. This method uses public and private keys. The public key can be shared with everyone, but the private key must be kept secret. In RSA cryptography either of the public or private key can be used to encrypt a message while the other is used to decrypt.
Digital signature algorithm (DSA)	This is used to create digital signatures. DSA was created in 1991 by the National Institute of Standards and Technology and is the standard for government agencies.

[Table 7.2](#) provides several PKI algorithms that are commonly used. It is noted that both the RSA and Diffie Hellman algorithms exceed the recommended key length for encryption systems in that they sit at 1024-bit keys while the current standard is 256. However, the Deffie Hellman algorithm is vulnerable to the notorious Man-in-the-middle (MITM) attack as it does not authenticate either party in the exchange. Hence, the Deffie Hellman method is best applied in combination with another authentication method such as digital signatures. An infamous case of a MITM attack occurred in 2015, when Europol defeated a group of 49 “cyber fraudsters” who operated by intercepting communications between certain businesses and their clients across Europe, causing victims to unwarily transfer money to illegitimate bank accounts. On the other hand, although RSA can be used for signing digital signatures, even their 2048-bit key lengths have been vulnerable to Man-in-the-browser (MITB) attacks, which makes use of a Trojan Horse by injecting malicious software (malware) into a victim’s web

browser. This malware typically exploits vulnerabilities in the browser or its plugins to intercept and manipulate data exchanged between the browser and the websites the user visits.

The DSA algorithm also has its pitfalls in that it can only execute digital signatures and not public key encryption. However, its advantage is the algorithm's speed of producing a digital signature. The configuration of a PKI on a network provides robust security and also applied tools that simplify and expand the network management scope. Its importance rests on the fact that users furnished with certificates and browsing on the secure network are protected from all types of over-the-air attacks. PKI saves time and effort for the network administrators who do not have to manually configure each device for certificates. Hence, they can simply monitor network activity, inspect connection concerns, and revoke certificates when they are no longer required. As one may appreciate, generally an IPS and Firewall protect the network, while a PKI protects the data that is being transmitted between networks.

7.6 CENTRE FOR AI SAFETY

The Centre for AI Safety (CAIS) is a nonprofit research and field-building organisation based in San Francisco. CAIS view AI as having the potential to intensely benefit the world, provided that it can be developed and used safely. However, CAIS's management recognise that many basic problems in AI safety have yet to be resolved and is not keeping pace with the extraordinary progress being made in AI developments. They see their mission as one that aims to reduce societal-scale risks associated with AI by conducting safety research, building the field of AI safety researchers, and advocating for safety standards.

At the beginning of this book, it was specifically stated that AI could be beneficial, but it could also be destructive unless there is a secure and safe environment within which AI can be developed. It was also emphasised that international cooperation was essential at governmental, institutional and corporate levels. CAIS is seen as an important institution for promoting and

creating a safe AI environment, which should be supported by the creation of similar institutions across the globe. According to [Hendrycks, Mazeika and Woodside \(2023\)](#), AI has seen rapid advancements in recent years, raising concerns among AI experts, policymakers, and world leaders about the potentially catastrophic risks posed by advanced AIs. They classify these risks under four categories, namely, malicious use, AI race, organisational risks, and rogue AIs.

Malicious use results when threat actors intentionally exploit powerful AIs to instigate widespread harm. The type of risks caused by malicious use include bioterrorism using AI to assist humans generate deadly pathogens; the deliberate spreading of uncontrolled AI agents; and the application of AI capacities for propaganda, censorship, and surveillance. These risks may be mitigated by enhancing biosecurity, limiting access to the most dangerous AI models, and ensuring that AI developers are legally accountable for damages caused by their AI systems.

[Hendrycks, Mazeika and Woodside \(2023\)](#) argue that rivalry could force nations and corporations into an AI race, where they adopt a hurried AI development strategy that yield power to AI systems. They picture a scenario where armed forces may come under pressure to develop autonomous weapons and apply AIs for cyberwarfare, enabling a new kind of automated warfare where accidents can escalate unrestrained before humans can intervene. Industry is likely to encounter similar pressure and may automate human labour to focus on profits rather than safety, which may possibly lead to mass unemployment and dependence on AI systems. They also envisage the possibility of natural selection among AIs that may lead to egocentric behaviours, where the gains of AIs have over humans could ultimately lead to the displacement of humanity. Although this depiction of possible future AI developments is disheartening, Hendrycks, Mazeika and Woodside believe that such risks can be mitigated by the implementation of safety regulations, international coordination, and public control of general-purpose AIs.

Another source of risk they identify are organisational risks. Hendrycks, Mazeika and Woodside contend that catastrophic organisational risks have

happened in the past such as Chernobyl, Three Mile Island, and the Challenger Space Shuttle disaster. In a similar manner, they maintain that organisations creating and implementing advanced AIs could experience similar accidents, especially if they do not have a strong safety culture. For example, AIs could be inadvertently leaked to the public or stolen by malicious actors, particularly if such organisations do not adequately invest in safety research; lack understanding of how to reliably improve AI safety faster than general AI capabilities; or suppress internal concerns about AI risks. Hendrycks, Mazeika and Woodside are confident such risks can be mitigated by better organisational cultures and structures being established, including internal and external audits, multiple layers of defence against risks, and the latest information security.

The final category of AI related risks is regarding rogue AIs, where the human race might lose control over AIs as they become more intelligent than human beings. Hendrycks, Mazeika and Woodside maintain that AIs could optimise flawed objectives to an extreme degree in a process called proxy gaming. Thus, they argue that AIs may experience goal drift as they adapt to a changing environment, like how people acquire and lose goals throughout their lives, and in some instances, it might be instrumentally rational for AIs to become power-seeking. Additionally, Hendrycks, Mazeika and Woodside examine how and why AIs might engage in deception, appearing to be under control when they are not. Hendrycks, Mazeika and Woodside provide suggestions how these technical risks may be mitigated to ensure AIs are controllable.

For instance, they suggest avoiding developing and implementing AIs that carry excessive risk until safety has been irrefutably verified; AIs should be trained never to make threats to lessen the likelihood of individuals being influenced by them; and AIs should not be implemented in sites that would make shutting them down extremely costly or infeasible such as critical infrastructure. Another suggestion is having countries around the world collaborating to establish a symmetric international off-switch for AI systems to rapidly deactivate AI systems globally if deemed necessary, particularly if there is an urgent risk of extinction. Hendrycks,

Mazeika and Woodside argue that a successful off-switch would require increased transparency and monitoring in AI development and operations, such as know-your-customer systems, so creating an off-switch also creates important infrastructure for mitigating other risks.

Hendrycks, Mazeika and Woodside also suggest legal liability for cloud compute providers to guarantee that such providers take steps to safeguard that their platforms are not helping rogue AIs survive and spread. Legal liability would motivate these providers to ensure that agents running on their hardware are safe. Furthermore, they recommend that to strengthen the liability framework, a similar accord to international agreements for cyberattacks can be devised, which would basically create a decentralised off-switch that would allow for swift interventions if rogue AIs started spreading. The CAIS's research exclusively focuses on mitigating societal-scale risks posed by AI by creating introductory benchmarks and methods that establish the groundwork for the scientific community to address technical challenges.

CAIS has actively supported research on the robustness of safety guardrails in large language models (LLMs) that emphasises the need to restrict third-party developers from bypassing safety controls. This research has focused on developing technical methods to detect and measure the inclination of LLMs to hallucinate, thus facilitating the AI systems to be more truthful and reliable. This work focused on ascertaining the degree that AI systems act based on reward systems as opposed to ethics to assist AI researchers know how AI systems perform according to ethical considerations.

Having this, those AI safety researchers who want to experiment on the latest LLMs encounter a predicament in that performing such research requires access to the latest graphics processing unit (GPU) infrastructure to run experiments resembling real-world scenarios. It is noted that a GPU is used in a wide range of applications, including graphics and video rendering. The difficulty is that the cost, complexity, space, and infrastructure skills required to create such a system establish high entrance barriers for most AI-safety researchers. However, the CAIS Compute

Cluster, which is a dedicated GPU-accelerated cluster, offers AI safety researchers with sponsored, on-demand access to the latest infrastructure for LLM training, and other AI safety projects. Furthermore, this cluster is purposely intended for researchers working on the safety of machine learning systems and supports a wide range of research activities and collaborators.

7.6.1 Stanford Centre for AI Safety

The mission of the Stanford Centre for AI Safety (SAFE) is to develop rigorous techniques for building safe and trustworthy AI systems and establishing confidence in their behaviour and robustness, thereby facilitating their successful adoption in society. SAFE has similar objectives to CAIS and focus on three key aspects, namely:

- Formal methods. Promoting the use of precise mathematical modelling to certify the safety, security, and robustness of conventional software and hardware systems.
- Learning and control. Designing systems that intelligently balance learning under uncertainty and acting safely.
- Transparency. Understanding safety in the context of fairness, accountability, and explainability for autonomous and intelligent systems.

SAFE have several leading research projects regarding the development of a responsible AI (RAI) assessment framework; error pocket retrieval in computer vision from unlabelled data; verifying learning-based controllers using Neural Lyapunov barrier certificates; and safe navigation with learned environment maps.

Research regarding the RAI assessment framework supports regulatory compliance, risk mitigation, and continuous improvement in AI governance and ethical responsibility. It consists of evaluating metrics and mitigation strategies, which aims to combine industry and academic perspectives to create a comprehensive guide for building RAI. This research reflects the

growing need for regulation by governments regarding transparency, accountability, and data protection in reaction to advances in AI technology. It is self-evident that compliance with such regulations is critical to avoid substantial fines and penalties. Hence, relevant metrics and key performance indicators (KPIs) within a robust framework are essential for organisations. Therefore, regular assessments provide useful insights into the current states of AI systems that reveal areas of excellence and required enhancements so that targeted actions can be identified to improve AI governance. Additionally, knowing RAI weaknesses would assist in obtaining funding for such initiatives. Furthermore, RAI assessments mitigate risks associated with bias, discrimination, and privacy violations that allow proactive risk management, which lessen possible harm to individuals and society.

Error pocket retrieval in computer vision from unlabelled data concerns finding the most suitable models for integration into larger pipelines, or to cross-check fine-tuned models against their predecessors for desired improvements. Computer vision is becoming increasingly mission-critical across many sectors, from facial recognition to self-driving cars. However, creating effective computer vision models is difficult, particularly when handling large amounts of data and ensuring that the data is of high quality. SAFE's objective is to improve the current model comparison pipeline in computer vision by removing the need for in-depth image annotation to gather validation sets, and by making available more informative and actionable insights than conventional accuracy metrics. SAFE are creating a tool that bypasses the need for human annotations. Their tool identifies and produces "Error Pockets" that are thematic groups of images where two models disagree, which can be coherently described by humans. Hence, these error pockets permit practitioners to assess the systematic strengths and weaknesses of any model in the context of their specific application, providing a more detailed audit that goes further than the simplistic analysis provided by numerical metrics and offers prescriptive data for model improvement with human-understandable explanations. This approach is aimed at improving performance on the identified error pockets.

An important research project is related to verifying learning-based controllers using Neural Lyapunov barrier certificates, applying deep reinforcement learning (DRL) for generating agents that control autonomous systems. A concern with DRL is its implementation in real-world safety-critical applications. SAFE is experimenting with an approach using Neural Lyapunov Barrier (NLB) certificates that provides strong guarantees on an agent's behaviour. NLB certificates are learned functions over the system whose properties indirectly imply that an agent behaves as desired. SAFE are exploring the possibility of having joint training of neural controllers and NLB certificates, using a formal neural network verification tool (Marabou) to check the certificate and provide counterexample-guided feedback to the training. This approach has important implications for autonomous vehicle applications.

Another research project undertaken by SAFE is regarding safe navigation with learned environment maps. According to SAFE, 3D neural maps learned from sensor data have exhibited astounding results in visual realism and detail, as well as more compactness and efficiency compared to traditional map representations. However, further improvements are needed because neural maps are still limited by incompletely observable and highly dynamic environments, making their use in safety-critical systems challenging. Hence, SAFE are enhancing current technology to develop safe methods for downstream navigation applications (localisation, error forecasting, and path-planning) that are robust to diverse sources of map errors. Their aim is to quantify and mitigate uncertainty in map regions, provide bounds on localisation error, and ensure safety guarantees for trajectory planning.

While CAIS tend to sponsor AI safety related projects and recommend ways to mitigate underlying risks, SAFE appear to undertake the research projects in the safe application of AI directly themselves. However, both organisations have as their eventual aim to ensure the safe use of AI, and the minimisation of the risks involved in its application.

7.6.2 European Network for AI Safety

The European Network for AI Safety (EN AIS) connects researchers and policymakers for the safe development and application of AI in Europe. The network appears to be an informal one, in the sense that each organisation has its own AI interests but is willing to voluntarily cooperate and share knowledge with the network members. Additionally, in some instances there are also joint research collaboration between selected country partners. [Table 7.3](#) shows the location of network membership in terms of countries and cities. It illustrates that there are 23 member countries, two of which are non-European Union (EU) members. The network membership is open to educational institutions, non-profit organisations, and individuals.

TABLE 7.3 Countries and Cities within the European Network for AI Safety [📄](#)

Country	Location
Austria	Vienna
Belgium	Brussels
Bulgaria	Sofia, Potsdam, Groningen
Czech Republic	Prague
Denmark	Copenhagen
Finland	Helsinki
France	Paris, Lyon, Nantes
Germany	Berlin, Aachen, Cologne, Dresden, Freiburg, Greifswald, Hamburg, Karlsruhe, Munich, Tübingen, Uppsala, Würzburg
Greece	Athens
Hungary	Budapest
Ireland	Cork, Dublin
Italy	Milan, Torino, L'Aquila

Country	Location
Netherlands	Amsterdam, Delft, Eindhoven, Groningen, Maastricht, Utrecht
Norway	Bergen, Trondheim, Tübingen
Poland	Prague, Warsaw, Cracow, Kraków, Poznań
Romania	Bucharest
Serbia	Belgrade
Slovakia	Bratislava
Spain	Madrid, Barcelona
Sweden	Stockholm, Gothenburg, Stuttgart
Switzerland	Zurich, Lausanne
Turkey	Istanbul, Ankara
United Kingdom	London, Cambridge, Oxford

ENAIIS (www.enais.co) claim that it represents a dynamic community of researchers and policymakers from across Europe, united in their efforts to advance AI safety. The network members have the common aim of working towards a common goal that of ensuring that AI systems are safe, ethical, and positively impactful for humanity. ENAIIS differentiates itself from related organisations by its focus on existential risks from AI. Like CAIS and SAFE, ENAIIS are committed to researching and promoting best practices in AI safety, ensuring that advanced AI systems are developed and deployed responsibly. However, the ENAIIS website does not provide any details of the research projects being directly undertaken or sponsored by it. The ENAIIS website refers to technical AI safety research as investigating possible solutions to ensure AI is safe, fair, and transparent. However, in this regard, the ENAIIS website only refers to a research paper submitted by [Hendrycks et al. \(2022\)](#). ENAIIS recognise that as AI technology continues to advance, the potential risks and challenges that come with this development needs to be seriously considered.

[Figure 7.4](#) shows an analysis of the predominate areas of interest of the ENAIS members. It is interesting to note that 26.1% of the network members are mainly interested in AI interpretability at 13.4% and AI policy (EU) at 12.7%. These two aspects of interest scored the highest. The analysis also reveals that of all the ENAIS members only 3.3% registered an interest in AI safety empirical research; 2.3% theoretical AI safety research; 2% S-risk reduction (risks involving an astronomical amount of suffering such as terrorism involving advanced technology); and 0.5% AI and Biosecurity. This is somewhat disappointing when one considers that the objective of ENAIS is to connect researchers and policymakers for the safe development and application of AI in Europe. The analysis suggests that ENAIS is attracting members who are interested in AI developments but not necessarily for AI safety reasons. From the AI safety aspects ENAIS needs to take a more forceful approach to attract members that are specifically researching safety in the deployment of AI technology and to promote direct research like CASIS and SAFE.

Areas of Interest of Network Members

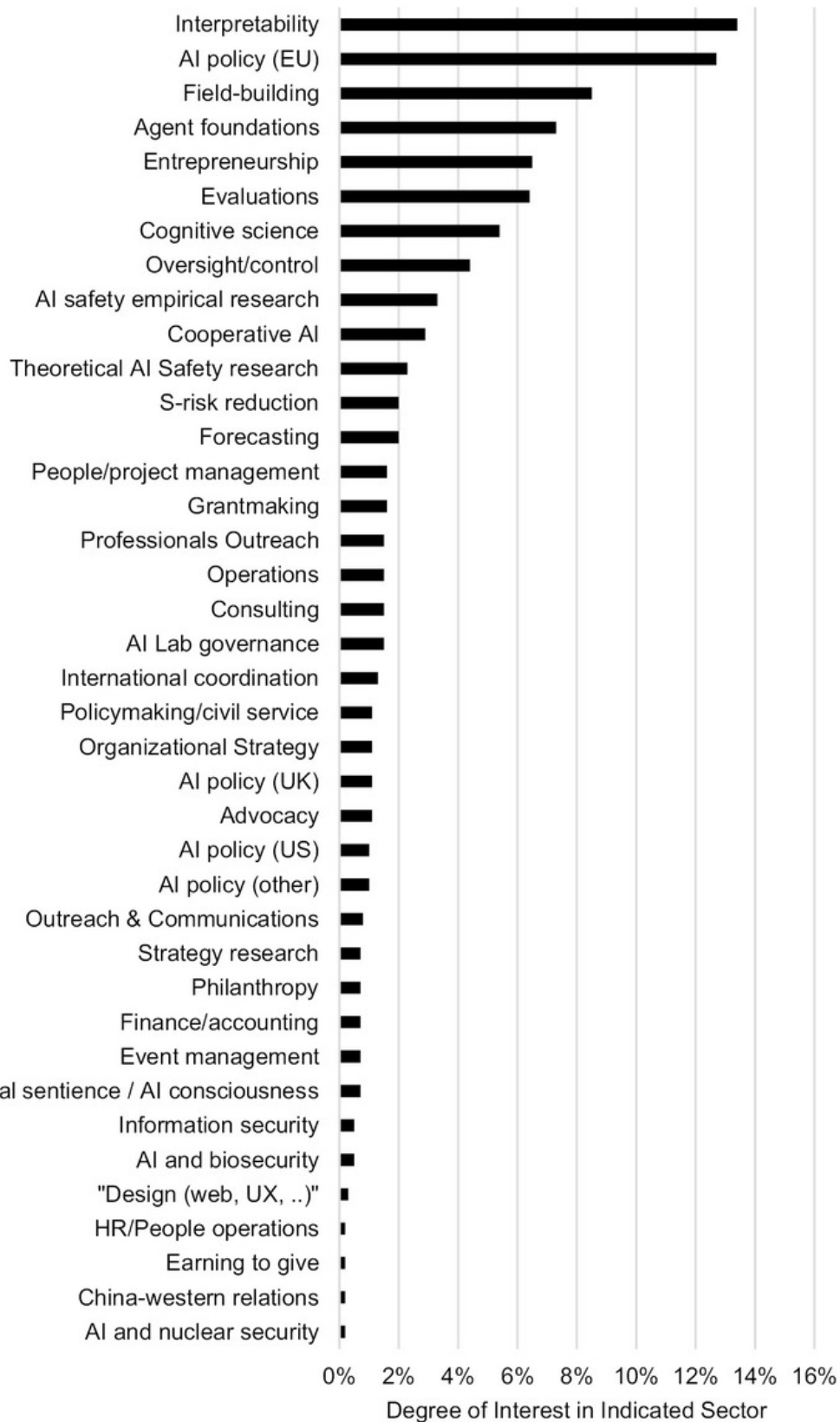


FIGURE 7.4 Predominate Areas of Interest of ENAIS Members. 

7.6.3 First International AI Safety Report

AI safety is attracting international attention with the first international scientific report on the safety of advanced AI being released in January 2025. This report was commissioned immediately after the inaugural AI Safety Summit hosted by the UK government at Bletchley Park in November 2023. The report has the primary objective of facilitating constructive and evidence-based discussion about the decisions to be taken by societies and governments. [Skelton \(2025\)](#) contends that this report will be used to inform upcoming diplomatic discussions around how to mitigate a variety of dangers associated with AI. He argues that the report highlights an environment where there is a high degree of uncertainty around the exact nature of many threats and how to best deal with them. The main findings of the report ([AI Action Summit, 2025](#)) are summarised below:

- The capabilities of general-purpose AI are rapidly increasing and have developed to a stage where it can write computer programs, generate custom photorealistic images, and engage in extended open-ended conversations.
- Investing in the development of general-purpose AI agents has become a common trait for many organisations due to the potential for significant advancement, such as using computers to complete longer projects. This provides added benefits but also unfastens further risks.
- It is difficult to predict the pace of advancement since it can range from slow to extremely rapid depending on the capacity to rapidly deploy yet more data and computational power to train new models, and whether research breakthroughs to surmount current limitations become available. Research indicates that rapidly scaling up models may not be physically feasible for some time.

- Hazards from general-purpose AI have been identified and researchers have developed mitigation measures to address them. However, up to now no combination of techniques applied is capable of fully resolving these concerns, which continue to sprout.
- Further risks are progressively emerging with the increased capabilities of general-purpose AI, such as large-scale labour market impacts, AI-enabled hacking or biological attacks, and loss of control over general-purpose AI. There is little consensus among experts regarding the interpretation of the timeline for these events. There are those who view such risks as being decades away, while others predict societal-scale harm within the next few years. Recent tests related to scientific reasoning and programming suggest that the potential risks for AI-enabled biological attacks have been reassessed by a leading AI company using results for its best model from “low” to “medium”.
- Risk management practices regarding general-purpose AI are emerging. However, current methods to assess and mitigate these risks all have limitations. For instance, current interpretability techniques for explaining why a general-purpose AI model yields any given output continue to be severely limited. On the other hand, researchers and policymakers are striving to standardise risk management approaches, and to coordinate internationally.
- Due to the pace and unpredictability of advancements in general-purpose AI, policymakers are faced with an “evidence dilemma” and must weigh potential benefits and risks of imminent AI advancements without having a large body of scientific evidence available. As such, risk mitigation measures based on limited evidence might end up being ineffective or unnecessary. However, waiting for stronger evidence may leave society unprepared or even make mitigation impossible. Developing early warning systems and risk management frameworks may lessen this dilemma by initiating specific mitigation

measures when there is new evidence of risks or obliging developers to provide evidence of safety before releasing a new model.

- There is broad consensus among researchers that knowledge is required regarding: (a) the prediction of future progress of general-purpose AI capabilities and to find a means of reliably measuring such progress; (b) having reasonable risk thresholds to trigger mitigations; (c) the best means for policymakers to gain access to information about general-purpose AI that is relevant to public safety; (d) a reliable approach for researchers, technology companies, and governments to assess the risks of general-purpose AI development and deployment; (e) the manner general-purpose AI models work internally; and (f) how general-purpose AI may be designed to behave reliably.

Finally, the report emphasizes that it is the choices made by people that determine the future of general-purpose AI. Moreover, the report argues that decisions made by societies and governments on how to navigate the uncertainty emanating from general-purpose AI determines which path is taken that may include both very positive and very negative outcomes. [Skelton \(2025\)](#) cites Clara Chappaz, the French minister delegate for AI and digital technologies, who stated that AI is a central topic of our time, and its safety is a crucial foundation for building trust and fostering adoption.

7.7 CONCLUSION

It must be recognised that the deployment of AI applications is becoming a dangerous minefield for cybersecurity. Generative AI (GenAI) in particular is a magnet to threat actors who are continuously seeking ways to cause havoc by producing disinformation and generating false narratives to influence individuals in their election choices, sharing ideological propaganda, and conducting fraudulent ploys. This chapter has examined various aspects of cyber security. Firstly, it has explored the evolving threat landscape due to the tremendous advancement of AI applications. The

evolving threat landscape is attracting AI-driven threat actors that have developed clever and avantgarde methods to exploit weaknesses in current cybersecurity systems. This is forcing organisations to employ more resources and AI-driven tactics to stay ahead of the evolving threats.

This evolving threat landscape is prompting the cybersecurity industry to take a proactive and vigorous action against AI-powered threats by creating new standards and best practices that need to be supported by an environment of international alliance across industries. Preventing cyber threat actors will continue to be an everlasting battle without a conclusion. One can only hope that the technology industry keeps ahead of threat developments to ensure that the cyber network and the associated systems and data are safeguarded from all types of potential threats. This can be facilitated with the collaboration of all stakeholders on an international level, roping in governments, technology providers, and a broad range of industry.

Once the evolving threat landscape was assessed the chapter examined intrusion prevention systems (IPS). These continuously scrutinise network traffic to identify threats and prevent them from entering the network. An IPS is placed directly behind the firewall to offer an additional measure of protection by monitoring the network traffic for threats and taking the necessary action to stop any threats that are detected. The preventative actions undertaken by an IPS include several measures such as alerting administrators; employing banishment vigilance by barring incidents before they occur; changing the security environment to prevent the threat from gaining access such as reconfiguring a firewall; and modifying the attack content by automatically altering the content of the attack, like removing attachments from suspicious emails.

The cyber system is also protected by a firewall, which is designed to prevent unauthorized access into or out of a computer network and to make sure internet users without access are not able to interface with private networks, or intranets, connected to the internet. The next-generation firewall has added protection features that facilitate better decision making about what traffic is to be allowed. These apply application-specific rules,

protect against malware, or detect, block anomalous behaviour, and preventing threat actors from entering through a nonstandard port.

While the network is being protected by a firewall and an IPS, the data that is being transmitted between the networks is safeguarded by having a Crack Public Key Infrastructure (PKI). A PKI manages digital certificates and provides a powerful security tool that supports numerous operations. PKI is viewed as a zero-trust security framework by NSA, CISA and NIST. The PKI certificates use public-private key cryptography that enable everyone to use a public key to encrypt a message and transmit it to someone else, without fear that the message will be made known to a cyber attacker even though they may know the public key. PKIs use Asymmetric Encryption where the sender encrypts the data, and the recipient decrypts it, however with the sender and the recipient using two different keys.

In combination, IPS, Firewall, and a PKI provide a comprehension method of providing a robust cybersecurity system. However, as stated previously, organisations must keep ahead of threat actors. This is the *raison d'être* of the concept underlying the creation of the centres for AI safety. This chapter has examined three organisations regarding a centre for AI safety, namely CAIS, SAFE and ENIAS. These organisations are nonprofit entities that encourage research for minimising AI security risks. While CAIS tend to sponsor AI safety related projects and recommend ways to mitigate underlying risks, SAFE appear to undertake the research projects in the safe application of AI directly themselves. However, both these organisations have as their final aim the safe use of AI and the minimisation of the risks involved in its application. On the other hand, ENAIS connects researchers and policymakers in an informal way to promote the safe development and application of AI in Europe.

REFERENCES

AI Action Summit. (2025). *The International Scientific Report on the Safety of Advanced AI*. UK Government, Department for Science, Innovation and Technology and AI Safety Institute.

www.gov.uk/government/publications/international-ai-safety-report-2025.^[4]

Allan, Katy. (2023). The rapidly evolving threat landscape of 2024. *Cyber Security, Cyber Magazine*.^[4]

Antony, James. (2023). *What is a PKI (Public Key Infrastructure)?* 2023 Update. Secure W2. www.securew2.com/blog/public-key-infrastructure-explained.^[4]

BlackBerry. (2024). Global threat intelligence report. Reporting period: March 1 – May 31, 2023. www.blackberry.com/us/en/solutions/threat-intelligence/2023/threat-intelligence-report-august.^[4]

Cyber Security Magazine. (2024). The evolving threat landscape in 2024: AI takes centre stage. *River Publishers*. <https://cybersecurity-magazine.com/the-evolving-threat-landscape-in-2024-ai-takes-centre-stage/>.^[4]

Deloitte. (2014). *Transforming Cybersecurity: New Approaches for an Evolving Threat Landscape*. Centre for Financial Services.^[4]

Fortinet. (2024). *What Is Intrusion Prevention System (IPS)? Definition and Types: Understand How IPS Works, Attacks Prevented by IPS, and Differences to IDS*. Fortinet. www.fortinet.com/resources/cyberglossary/what-is-an-ips^[4]

Hendrycks, Dan., Carlini, Nicholas., Schulman, John., and Steinhardt, Jacob. (2022). *Unsolved Problems in ML Safety*. Cornell University. <https://arxiv.org/pdf/2109.13916>.^[4]

Hendrycks, Dan., Mazeika, Mantas., and Woodside, Thomas. (2023). *An Overview of Catastrophic AI Risks*. CAIS. <https://arxiv.org/pdf/2306.12001>.^[4]

IBM. (2024). *What is an Intrusion Prevention System (IPS)?* www.ibm.com/topics/intrusion-prevention-system.^[4]

Kellerman, Robert. (2022). *Cybersecurity in the Evolving Cyber Threat Landscape: Five of the Latest Trends and Predictions*. STAGE2DATA. <https://stage2data.com/cybersecurity-in-the-evolving-cyber-threat-landscape-five-of-the-latest-trends-and-predictions/>.^[4]

Palo Alto Networks. (2023a). *Navigating the Evolving Threat Landscape: Resilient Cybersecurity Tactics for CISOs*. Palo Alto Networks, Tannery Way.

Palo Alto Networks. (2023b). *What is an Intrusion Prevention System?* Palo Alto Networks, Tannery Way.
www.paloaltonetworks.com/cyberpedia/what-is-an-intrusion-prevention-system-ips.↵

Skelton, Sabastian, Klovig. (2025). First international AI safety report published. *ComputerWeekly.com*.
www.computerweekly.com/news/366618170/First-international-AI-safety-report-published.↵

VMware. (2024). *What is a Next-Generation Firewall (NGFW)?* Broadcom Inc. www.vmware.com/topics/glossary/content/next-generation-firewall.html.↵

III

The Practical Implications of Artificial Intelligence

There is no doubt that AI has the potential to provide great benefits for society in general and to humans in particular. To illustrate the impact of AI, [Part III](#) commences with a detailed discussion of the knowledge-based economy. The concept of a knowledge-based economy is not new, but the latest developments in AI have elevated this concept to newer and greater dimensions. The foundation of a knowledge-based economy is information that is utilised for the innovative creation and delivery of services, which are focused on knowledge-intensive activities. These activities consequently have the impact of speeding up the evolution of technological and scientific development. This in turn means that there is a much greater dependence on intellectual capabilities than physical efforts or geographical resources. AI augments the intellectual capabilities of humans by its capacity to monitor and analyse the immense volume of data and information, and to learn from these for the development and benefit of nations, organisations, and humans. Therefore, by supplementing human intelligence, AI can propel the knowledge-based economy to greater and unprecedented levels.

Aligned with the concept of the knowledge-based economy is Enterprise AI that is revolutionising the way private businesses and many public sector entities operate. Enterprise AI involves the amalgamation of AI tools to create a complete tailored solution for the particular requirements of all

types of organisations. It gives these organisations specific benefits related to customer relationship management, minimising operational expenditure, maximising revenue, and facilitating the formulation of operational strategies resulting in greater competitive advantage. However, like every other high-technology methodology, Enterprise AI needs to overcome several challenges associated with algorithm bias, better data security, and ensuring data privacy. Additionally, Enterprise AI is resource hungry and is likely to be only affordable and suitable for large organisations.

Whilst Enterprise AI is applicable to organisations that operate on a commercial basis, Public Sector AI is relevant for public administrations, whose priority is to provide their services to citizens, civil society, and the commercial sector. [Part III](#) provides a dedicated chapter to explain and demonstrate Public Sector AI use cases that are applicable to diverse governmental backdrops. This chapter applies a particular framework for the implementation of AI systems in public administrations at every government level. Public Sector AI encounters major challenges in its deployment, namely, trustworthiness, transparency, public sector employee skills upscaling, and interoperability. Individual governments in deciding to implement Public Sector AI must assess how AI will impact economic growth, employment displacement, technology transfer, healthcare, education, and ethical behaviour.

The final chapter in [Part III](#) is dedicated to ethical, moral, and legal considerations. There is consensus by governments and commercial organisations that AI must be trustworthy and developed and deployed in a safe manner. This implies that there is also a consensus that AI might be harmful. No one has specifically defined or identified the extent and impact of this harm to individuals and for national security. While most governments have endorsed voluntary ethical and moral frameworks, there is a reluctance by most governments to introduce binding legislative measures that specifically address AI. This reluctance appears to be based on the premise that specific laws might hinder AI innovation. Most governments have revised certain provisions in existing legislation to address specific AI legal issues. The EU has emerged as the global leader in

AI regulation, while most other countries are adopting a wait-and-see attitude.

World events in Ukraine, Gaza, and particularly Lebanon (regarding the detonation of communication devices) raise concerns as to whether AI can be used as a weapon of mass destruction. Mass destruction not in the sense that AI can be used to physically destroy buildings or infrastructure, but whether it can be applied to obliterate the systems that most critical infrastructure depends on, including the destruction of life support systems that sustain the lives for hundreds of thousands (if not millions) of people. It should be kept in mind that world leaders (national government or corporate) will normally act in the interest of the wellbeing of their own nation or organisation, with ethical behaviour playing a secondary role. Therefore, there is an urgent need for each country to have in place strong and forceful legislation to govern the development and deployment of AI systems and ensure that it is safe and trustworthy. There is an even a greater need for an international binding treaty committing nations, as a unified body, to regulate and govern all types of AI systems on an international level, and to classify them by the magnitude of risk they pose and have a continuous monitoring mechanism for such AI systems.

Chapter 8

Knowledge-Based Economy and Artificial Intelligence

DOI: [10.1201/9781003653950-11](https://doi.org/10.1201/9781003653950-11)

“To know what you know and what you do not know, that is true knowledge.”

Confucius (Kong Qiu)

Chinese philosopher of the Spring and Autumn period (551-479 BCE)

According to [Nguyen \(2024\)](#), AI has the potential to revolutionise the way people work and live in the coming years. He argues that while AI may disrupt some industries, it also presents opportunities for growth and innovation. Hence, he maintains that as the changes brought about by AI are adopted, it becomes critical for all to continue to upskill and reskill to remain relevant in the knowledge economy. Additionally, those with an open mind for adapting and embracing AI will position themselves better for future success. Nguyen maintains that the question is not whether AI will disrupt the knowledge economy, but rather, whether entities are ready to harness its potential?

In support of this, [CNBC \(2022\)](#) reports that the integration of AI helps firms with competitive data analysis helping them to stay ahead of their

competitors. Their view is that from production to customer relationship management, AI plays a vital role in transforming an entity to an effective organisation. However, the question is to what type of transformation are they referring? There are those who maintain that society and industry are witnessing the end of the knowledge economy and are experiencing a shifting towards an automation-based economy.

There is no doubt that current global economies are governed by knowledge-driven roles. Economic sectors, such as healthcare, finance, gaming, law, technology, and many others owe their existence to the knowledge industry and are intensely dependent on proficient information management. It is in the sphere of information management that AI dominates and outrivals any other related technologies. However, there are fears that AI may pose a serious threat to human workers. On the other hand, current signs suggest that AI is a useful partner assisting human workers in performing their tasks more effectively and enhances their productivity and decision-making abilities. Whether these signs will materialise for the benefit of society and industry, only time will tell.

8.1 INTRODUCTION

The Covid-19 pandemic has illustrated that automation enabled much knowledge-based industries to survive and prosper. In the post Covid-19 era, recovery of economies in various countries was partly due to the knowledge economy. Many view a knowledge-based economy as being sturdier and better prepared to act in response to specific problems and economic challenges. There are indications to suggest that the Covid-19 pandemic hastened the use of automation in the day-to-day lives of both individuals and businesses. This is particularly the case for industries that were able to continue their operations and permitted their employees to work from home in a remote information technology (IT) environment.

Technology is viewed as having grown to the extent that it is difficult to differentiate it from businesses. The two have become synonymous, particularly for those industries that provide or are dependent on state-of-

the-art, leading-edge solutions that are intended to fully automate tasks that encounter high cost due to their labour-intensive nature. Where does AI feature in this environment? Currently, AI is not viewed as replacing human reasoning. It is viewed as a tool to support business functions that may not be inclined to human intervention, such as automated cybersecurity where businesses can operate in a secure environment. However, this does not mean that this situation will not change in the future.

It must be recognised that information management has enabled the accumulation of data to be transformed to information, and the accumulation of information to be transformed to knowledge. Hence, the foundation for the emergence of knowledge-based economy. However, there is another transformation phase, namely transforming the accumulation of knowledge to wisdom. In the past and the immediate present this transformation of the accumulation of knowledge to wisdom has been mostly a human characteristic. In fact, in a knowledge-based economy the accumulation of knowledge is viewed as a primary asset for organisations to be used by its employees.

[Figure 8.1](#) illustrates the phases in the data-to-wisdom transformation process. [Figure 8.1](#) shows that the knowledge-based economy encompasses the data-to-knowledge transformation phase. The quicker that new knowledge is generated in combination with enhanced access to that knowledge, the better the organisation is prepared to combat and keep ahead to its competitors. Hence, knowledge becomes the primary resource leading to greater efficiency and competitiveness. This is the fundamental basis of a knowledge-based economy. [Figure 8.1](#) demonstrates that a knowledge-based economy consists of several steps that include data collection; data classification; analysis and data integration to create information; generating a knowledge repository; and decision making.

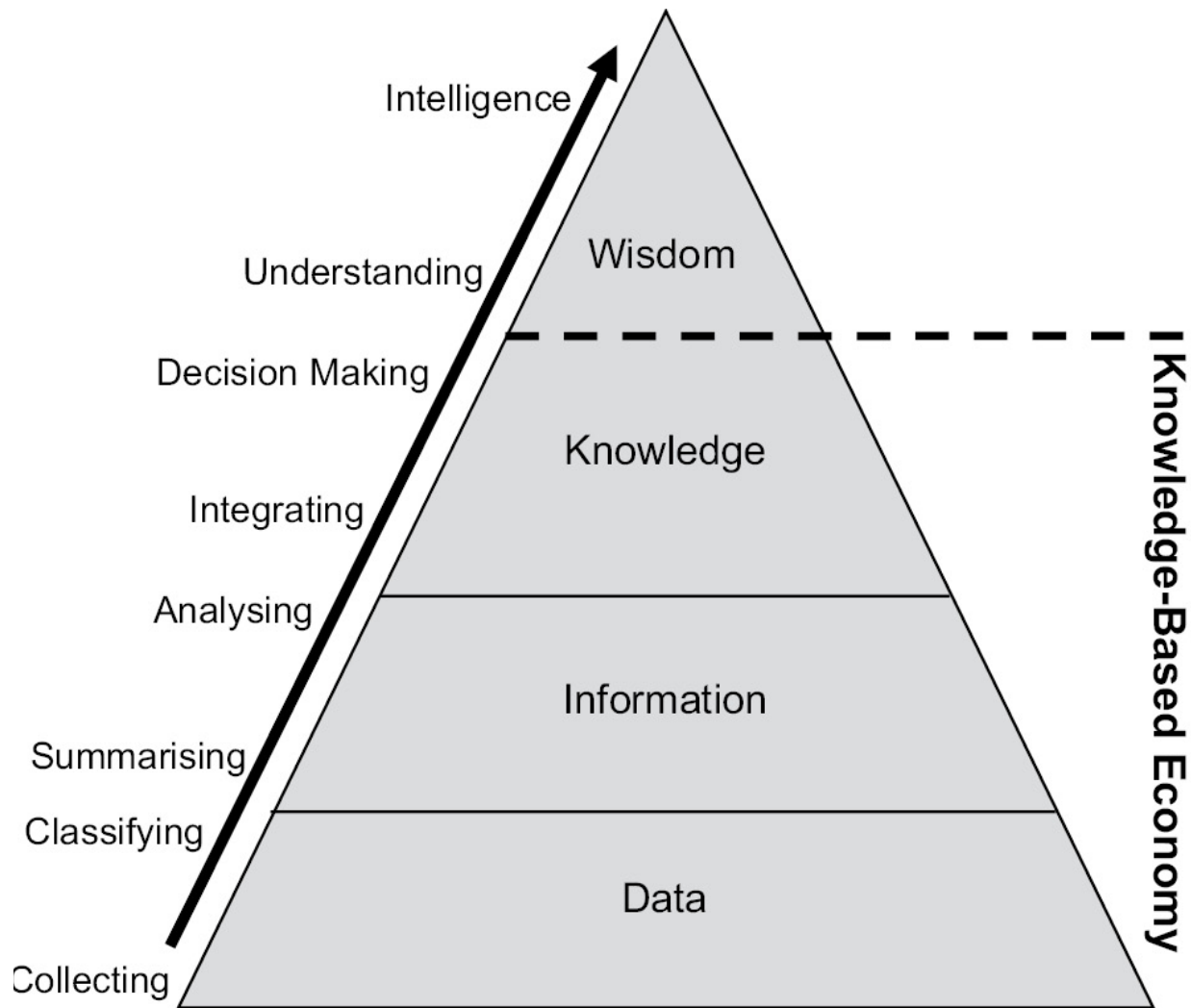


FIGURE 8.1 Phases in the Data-to-Wisdom Transformation Process. [↗](#)

It should be noted, that under the traditional model the decision-making aspect is a human task. Additionally, [Figure 8.1](#) illustrates that there is one further phase, which is wisdom that consists of two key functions, namely understanding and intelligence. However, humans are being overwhelmed with data and information through offline and online sources, and the rapid transformation of the information to knowledge is making it even more difficult for humans to make significant sense of all this input to make the appropriate decisions. Wisdom is normally associated with elder men and women who have accumulated much knowledge over many years to become wise and make better decisions. However, the modern world is

generating so much knowledge from the accumulated information that the human brain cannot possibly cope and requires technological tools to do this. As AI's sophistication develops and flourishes the transformation of the accumulation of knowledge to wisdom will likely be automated using AI supporting tools. Hence, it is foreseen that the outcome of this will be a shift from a knowledge-based economy to an automation-based economy soon, with AI having a critical role in this transformation.

An automation-based economy will utilise AI as a sophisticated tool to assist humans in the understanding of the vast amount of generated knowledge and transform it to wisdom, which is beyond human capacity. This integration of the knowledge base with AI capabilities will likely provide the means for organisations to keep ahead of their competitors. However, it should also be recognised that a technological advantage is short-lived because the competitive arena is not static, and technology is easily attained and implemented. The competitive race is a continuous one and a leader today can be a follower tomorrow (and vice-versa).

8.2 KNOWLEDGE-BASED ECONOMY

The knowledge-based economy, which is often referred to as knowledge economy is not a new concept. In fact, in 1967, Peter Drucker made first reference to the knowledge economy in his book "The Effective Executive" and later in 1969 had a chapter dedicated to knowledge economy in the book *The Age of Discontinuity* ([Drucker, 1967](#); [Drucker, 1969 Chapter 12](#), p247). Drucker in his publications emphasised knowledge, data analysis and measurable performance, and strategic management by objectives (MBO), which appear to have been too advanced to be truly appreciated during the time he proposed them.

[Powell and Kaisa \(2004\)](#) define knowledge economy as the production and the provision of services based on knowledge-intensive activities that contribute to an accelerated pace of technological and scientific advance, as well as its equally rapid obsolescence. They contend that the key components of a knowledge economy include a greater reliance on

intellectual capabilities rather than on physical inputs or natural resources, combined with efforts to integrate improvements in every stage of the production process, from the R&D laboratory to the factory floor, to the interface with customers.

[Figure 8.2](#) shows the main attributes of a knowledge economy. It illustrates that generally there is a global diffusion of knowledge but there must be the appropriate institutional structures to provide the incentives for entrepreneurship and the application of knowledge, which encourage the development of industrial clusters in specific geographic regions. A core attribute is having the latest information and communications technology (ICT) infrastructure that encourages the exchange of knowledge between industries and innovation-driven users and producers. This would also promote an active innovation landscape that incorporates a wide range of stakeholders such as industrialists, academic institutions, and civil society. However, this holistic development needs to be supported by a good education system that generates the required skilled labour, which in turn increases the demand for STEM (science, technology, engineering, and mathematics) workers and a sharp growth in the quantity of patents.

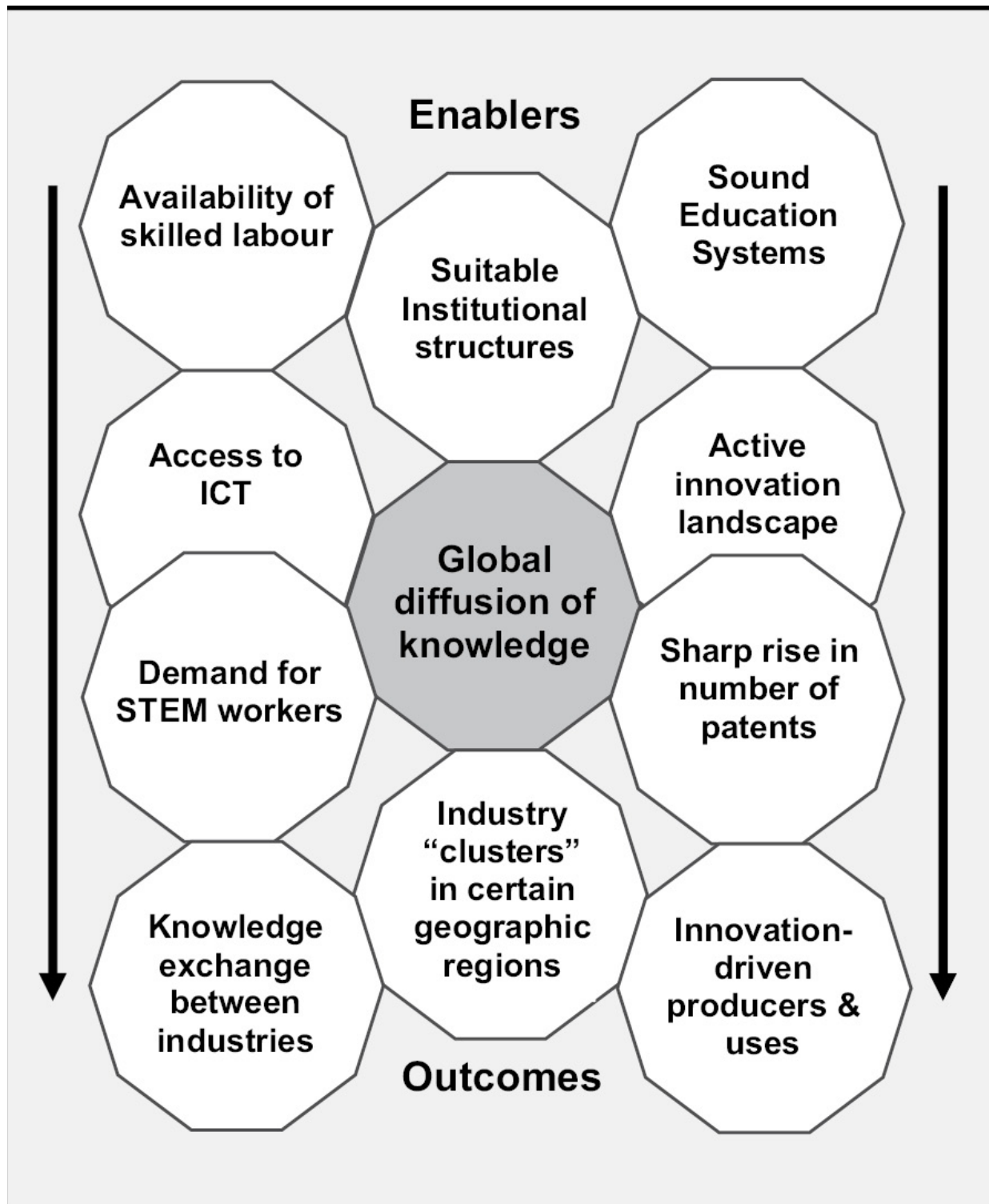


FIGURE 8.2 Main Attributes of a Knowledge-Based Economy. [📄](#)

The knowledge economy places high importance on human capital as being not only critical but essential in the 21st-century economy. The tremendous growth of knowledge, the increasing dependence on information management and big data analytics are transforming the economy of the developed world to one that is more reliant on intellectual capital and skills, and less dependent on the production process. This changing environment views the knowledge economy as an economic approach that depends more profoundly on intellectual capabilities and not so much on physical inputs or natural resources. Hence, this economic method views big data analytics and automation as being a fundamental input to the production process.

The emergence of the knowledge economy has resulted in the growth in the generation of intangible assets such as patents, copyrights, or proprietary software or processes. This is in contrast with previous economic eras. For instance, the agrarian economy had land as the primary asset, and in the industrial economy the key assets were manufacturing plants and equipment. At the core of the current economic growth is the technology industry, particularly in the AI and robotics sectors.

Economic indicators suggest that when compared with other investment types, ICT investment is a major factor in productivity growth but only if it is combined with organisational changes. Reference to these organisational changes should be seen in the light that the knowledge economy is characterised by having a greater percentage of highly skilled employees whose jobs require special knowledge or skills that entail the analysis of data, creating and working with financial models, the ability to innovate and apply rational judgement. This implies that an important organisational change may include increased employee involvement in decision-making.

The main thrust for the knowledge economy is technological progress driven by persistent innovation and research. Hence, those that work within the knowledge economy are well qualified with high computer related competencies and capable at developing business and financial models. All these skills depend on having a reliable data repository through a comprehensive data collection and analysis process supported by the

development of algorithms and AI applications. This implies that for organisations to survive with the possibility of keeping ahead of their competitors, they must renew themselves by the continual search for doing things differently through innovation and the development of improved business methods. This also implies having an endless collaborative cycle between academic institutions and industry that promotes research to ensure having a continual flow of discoveries from the innovation pipeline.

This knowledge-based environment, particularly the innovation pipeline has generated a huge demand for STEM jobs. STEM is an approach to learning and development that integrates the areas of science, technology, engineering, and mathematics. These types of occupations, which are in high demand, are associated with high compensation packages and a rapid escalation of the organisational hierarchical career ladder. The above illustrates that knowledge is the dominant basis that is driving the high demand for technical expertise, data collection and analysis skills, and innovative management practices that is assisting organisations in the private and public sectors to compete in the contemporary global economy. On a macroeconomic level, those countries that have seriously embraced the knowledge economy have excelled in specific areas, such as computer technology in Silicon Valley in the United States, the electronics industry in South Korea, and the automotive engineering industries in Germany, amongst others.

ICT is focused on integrating communication services and information technology that has the primary aim of increasing the efficient application of information by connecting computer servers and data repositories, with the resultant transmission of information to fixed and mobile devices. This has resulted in the knowledge economy becoming heavily dependent on ICT developments. The knowledge economy has revolutionised several industrial sectors such as the use of robotics and just-in-time inventory management systems in the manufacturing of automobiles. This has made automobiles more affordable to a wider range of consumers and encouraged the current developments in driverless cars. The healthcare industry has benefited from the knowledge economy through more research and the

creation of new medicines, a greater use of 3D and robotic surgical aids, and the expansion of telemedicine services.

[Figure 8.3](#) shows the 2023 global knowledge economy index by country rankings for 133 countries (UNDP, 2023) compiled by United Nations Development Programme (UNDP). The global knowledge economy index assesses whether the environment in a country is conducive to the creation of knowledge. The index ranking is assessed on seven aspects, namely pre-university education, technical and vocational education and training, higher education, research development and innovation, ICT, economy, and enabling environment. It is interesting to note that Malta, which is the smallest European Union member state ranks 18th, with its strength being pre-university education, ICT, and an enabling environment. However, its weaknesses are technical and vocational education and training, higher education, and research development and innovation. These weaknesses are also apparent for Singapore and for the Republic of Korea, who rank 12th and 17th respectively. The two small island states (Singapore and Malta) rank in the first top twenty countries out of 133. It is also important to recognise that with no exception, all countries have scored weakly in research development and innovation.

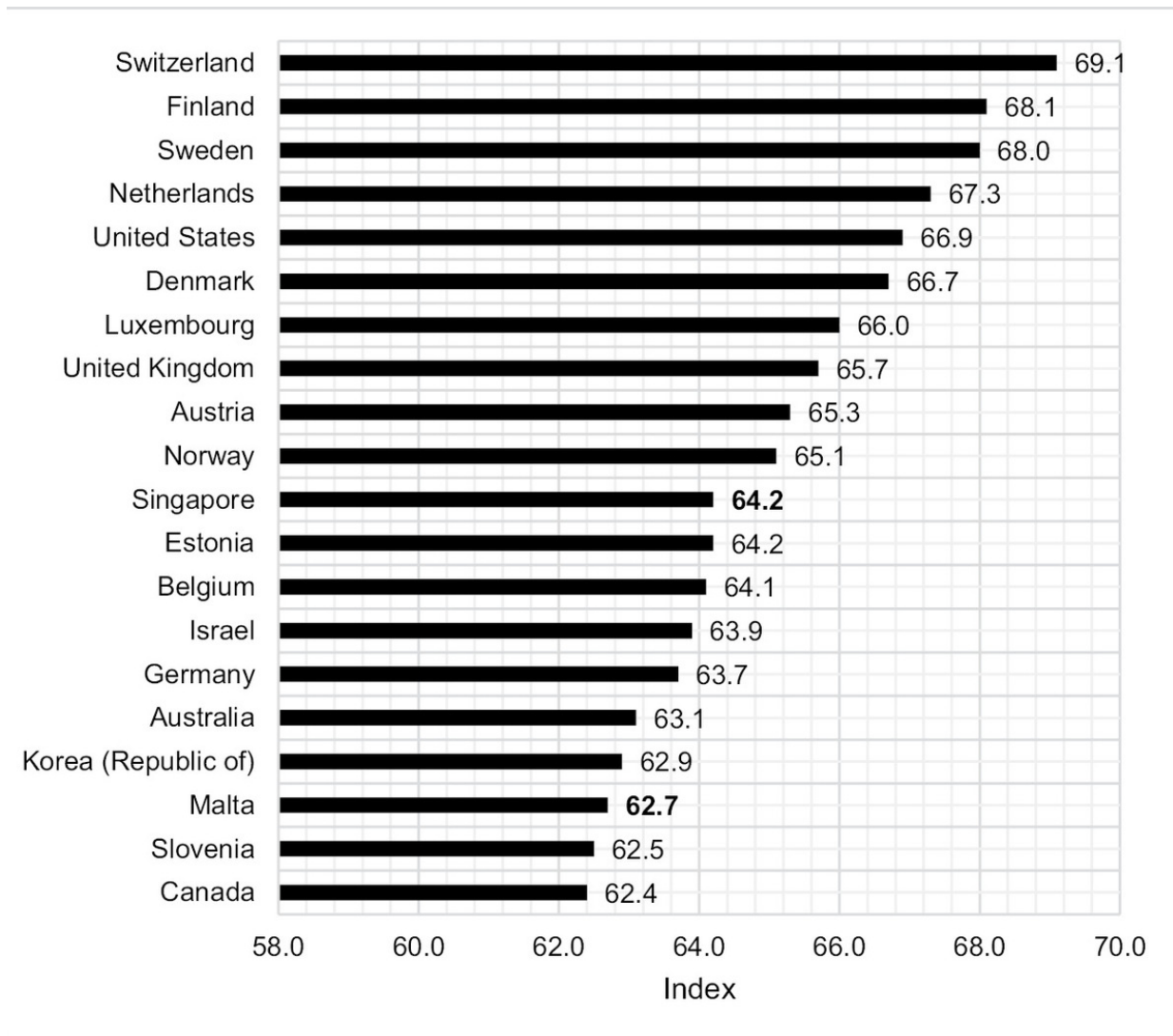


FIGURE 8.3 Global Knowledge Economy Index by Country Rankings. [🔗](#)

The greatest challenge encountered by the knowledge economy is to continually strive to ensure that those who lack the requisite competencies to function and be optimally productive in a knowledge economy are trained to the appropriate level. This concern requires the collaboration of education institutions and the private sector for the development of suitable courses that address the knowledge economy needs, particularly those related to STEM jobs. Furthermore, the private sector needs to design and implement on-the-job training programmes and support employees in upgrading their skills and education by sponsoring them to attend education

institutional courses. National governments can also facilitate matters by subsidizing industry through higher tax concessions to sponsor employees to attend specific formal training courses that address STEM careers.

8.3 DATA-TO-WISDOM TRANSFORMATION PROCESS

[Marchand, Kettinger and Rollins \(2000\)](#) contend that many organisations are having difficulty understanding how to apply information to improve their performance. They argue that these organisations have placed greater emphasis on developing systems and processes to classify and store information rather than improving the way people apply this information. This situation is applicable to both the private and public sectors, where huge investments on ICT have been made. However, management is finding it extremely difficult to exploit the vast amount of collected and generated information and the associated analytical methods to increase organisational performance and keep ahead of their competitors.

[Charan \(2005\)](#) argues that organisations are experiencing a profound transformation in that never has so much power been computerised and made available so broadly. He maintains that sophisticated software and complex algorithms, combined with new tools capable of collecting and storing vast amounts of data, can predict changing patterns in everything from consumer behaviour to the maintenance needs of machinery, thus becoming a critical source for improving performance and attaining a competitive advantage. [Parker \(2016\)](#) cites an example where a hospital conducted traditional rostering of staff in terms of shift allocations instead of using optimisation analytical modelling techniques to test a wide range of alternatives. Thus, he maintains that this hospital missed an opportunity to make substantial savings through more analytical approaches. Information must be viewed as a corporate asset that needs to be comprehensively used and exploited in the decision-making process by applying various analytical tools that are available. It should be noted that the analytical power as applied to data is only limited by one's imagination.

In above paragraphs the terms data and information have been loosely used. Hence, the first issue that needs to be addressed is whether there is a difference between data, information, knowledge, and wisdom. [Figure 8.1](#) shows the process that takes place as data is transformed to information; information is transformed to knowledge; and knowledge is transformed to wisdom. This transformation process may be viewed as the foundation for AI applications, with the transformation process reaching the pinnacle in terms of wisdom. In other words, it is the accumulation of wisdom that generates intelligence.

8.3.1 Raw Data

Data is viewed as the basic raw facts that an organisation gathers in its day-to-day operations. However, an organisation is constantly being besieged with raw facts, which may or may not be relevant to the organisation's needs. Hence, typically the collected data is sifted for relevancy to be accessed and utilised when needed. Data is the factual material utilised as a basis for discussion about what the facts reveal; deciding about a particular issue; and calculating or measuring a specific thing, such as the cost per litre of fuel.

Data may consist of operational data related to such things as revenue, expenditure, inventory, payroll, and accounting, amongst many others; non-operational data, such as macro-economic data, industry sales, and forecasts; and meta-data that consists of descriptive data about the data itself, such as logical database design or data dictionary definitions. Additionally, data is typically organised in a database and may take the form of numbers; words; observations or descriptions of things; pictures; and video and audio recordings.

Generally, there are two sources of data, namely internal and external. The internal data is generated by the organisation itself through its information systems and reflects its internal environment, such as the accounting transactions. External data is generated from sources outside the organisation and reflects the external environment, such as economic

trends; commodity prices; and data provided by the National Statistics Office. External data is generated by the information systems of other organisations.

There are also different categories of data, with the most common type being quantitative, such as values and quantities. This type of data is typically utilised for computational purposes. Another type is classification data, such as passport number or component part number. Classification data is normally used for identification purposes, for instance to check whether a particular item is in stock. An important data type is reference data, such as an order number or cheque number. Reference data are used for control purposes, such as tracking an order or payment. The fourth data category is qualitative data, such as remarks and observations that is normally unformatted textual data used for documenting general remarks. The final two data categories are audio and video data. These are analysed by dedicated statistical and mathematical algorithms, applying AI tools. These AI tools use supervised and unsupervised learning for pattern recognition and image analysis for detecting abnormalities, such as terrorist exposure scenarios.

There are several ways that data may be collected that in some cases can be effortless, while in others it can be demanding. For example, data may be collected automatically by using meters or counters, such as palm readers for recording employee attendance or electricity meters for determining energy consumption or production counters for measuring output of manufactured components. Another method may be through direct data entry on a display screen when serving a client, such as entering the client details at a hospital registration desk. Another common method is direct observation by viewing certain general attributes of persons entering a particular shop, such as gender, and general age group. A popular approach of collecting data is conducting market surveys from a number of relevant respondents and asking pertinent questions about particular issues.

8.3.2 Information

As explained earlier, data are facts, numbers, text, pictures, video, and audio recordings that are stored in a digital format to allow them to be processed by a computing device, traditionally known as data processing. Information systems process the raw data to generate information. Basically, raw data is transformed into information, which is meaningful, significant and may have implications for the organisation. For example, consider a digital file of customers that may contain their purchasing habits and some personal details. This digital file may contain thousands of entries. However, by itself this data has no value. On the other hand, data may be transformed into information by analysing it by a specific geographic region or locality, typical types of purchases, gender, marital status, age groups, and so forth. This analysis (or transformation) may be needed to establish the typical profile of the customers in a particular locality and their spending habits. Hence, an entity may obtain a better appreciation of the information that is generated in terms of better decision making, such as product or service promotions.

Furthermore, for data and information to be of value to an organisation it must meet three basic criteria, namely: accurate, timely, and relevant. Therefore, a key objective of the data-information transformation process is to get the right information to the right person at the right time. Typically, information is data that is processed to be useful and provides answers to who, what, where and when questions. Additionally, information is generated by summarising and analysing the raw data through the examination of patterns and the relationships among data items. For example, an analysis of retail point of sale transactions can yield information on the products that are being sold, when they are being sold, and who is buying them (particularly if a loyalty card system is in place).

8.3.3 Knowledge

The next level of the data-information transformational process is knowledge. Knowledge is the conversion of information into historical patterns and future trends. To convert information to knowledge there is a

need to synthesise (integrate) the data and information gathered from the various sources and decide how this knowledge is to be applied. For example, a national tax authority may compare the financial ratios of several companies within a specific industry over number of years to examine financial and income trends.

Knowledge may be viewed as comprising of two aspects, namely the explicit details and the comprehension of these explicit details. The explicit details are obtained from the application of the data and information to address “how” questions to provide an insight about the magnitude and attributes of some object. On the other hand, the comprehension of the explicit details is obtained through the analysis (i.e., why are we analysing these explicit details). For example, a manager of a public health facility may want to know which services are to be encouraged and supported by his/her facility. Hence, data and information regarding the use of the services offered are analysed to obtain the knowledge regarding the beneficiaries of the services being provided. In other words, the service recipient behaviour (how). This in turn will allow the manager to decide which health services should be supported and enhanced. Furthermore, this will assist the health facilities manager to select those services that are most responsive to any promotional effort that may be undertaken (objective or why). The above example illustrates that knowledge is the extensive exploitation of data and information, in combination with the ability, competencies, ideas, intuitions, commitments, and motivations of the manager involved. Hence, the transformation of data and information into knowledge is the outcome of the synergy that combines human aptitudes and the application of the available data and information.

There are various types of knowledge but the most common are tacit and explicit knowledge. Tacit or implicit knowledge is not easy to document since it is more idiosyncratic, it cannot be codified, it is applicable to a particular circumstance, and it is difficult to describe or explain to others. For example, learning a musical instrument takes talent and a lot of practice. Explicit knowledge can be learnt from the application of precise rules and can be codified. For example, take an activity like baking a cake.

If a person precisely follows the recipe and baking instructions, the outcome should always be consistent. Therefore, gaining and documenting explicit knowledge is not a concern. The concern is the gaining and transferring of tacit knowledge (as distinct from explicit knowledge), which is a complex matter and can be gained through extensive experience, socialisation, mentorships, and face-to-face communication. It is in this sphere that AI tools can be made useful. However, there is a long road ahead for all issues to be resolved.

8.3.4 Wisdom

The accumulation of a great amount of knowledge is wisdom. Through knowledge, an entity learns how to act in its own best interests. Wisdom takes a step further, it reinforces and reveals the principles and standards that an entity applies to its knowledge. Wisdom is the ability to make the right judgments and decisions, and hence transforms data, information, and knowledge into intelligence. Wisdom is gained through extensive experience and when documented forms the basis for supervised learning in AI applications. Therefore, as more experience is attained and documented, the higher level of intelligence that is created. This is not easy to achieve or sustain because to be wise implies a great deal of knowledge and experience that must continually be updated, maintained, and sustained.

With wisdom it is not a matter of grasping the past and present but to be able to define the future. However, it must be recognised that wisdom is the apex, thus it is not easy to achieve since one needs to move successively through the other transformation phases before it may be finally gained. In addition, once wisdom is gained it must be sustained with further knowledge; hence it takes an on-going demanding effort. It is precisely in this area that AI through unsupervised learning may facilitate the information-wisdom transformation process.

8.4 THE EVOLVING KNOWLEDGE-BASED ECONOMY LANDSCAPE

According to [Brown \(2023\)](#), the knowledge economy is a concept where the knowledge contained within an entity or an individual drives it, rather than its actual physical output. He argues that this concept differs from the agrarian, manufacturing, and service economies in that knowledge-intensive products become the output, not a physical provision. Typically, knowledge economies have highly skilled and specialised workforces and large institutions in the form of enterprises and centres of learning that capitalise on the generated knowledge. Researchers and other experts are questioning whether AI is threatening this environment. The evolution of the various economic models over the decades is viewed as part of contemporary economic development. The issue is: What is the next economic evolutionary stage considering the tremendous advancements of AI?

Often, knowledge is viewed as something that is normally possessed by humans. One is paid to write a drama for a TV series with creative ideas to keep viewers entertained. Others may prepare marketing campaigns with enticing advertisements that are highly creative to convince a consumer to purchase a product or service. Additionally, most of an entity's systems are administered under human control. Hence, most of an organisation's processes are based upon human knowhow with the aid of computer systems as a highly sophisticated tool. It is also a fact that computer systems transform data into knowledge after appropriate analysis, particularly displaying trends. However, the knowledge created by computer systems is deciphered and retained in the human brain, where it is recalled when making decisions or explaining and illustrating matters to others. Moreover, a human accumulates knowledge and becomes wise by using this knowledge, experience, and cognitive intuition.

However, people's knowledge is stored within an individual who may share his/her knowledge with others. The human absorption of knowledge becomes wisdom; however, humans are limited by how much knowledge an individual can absorb. An individual can become an expert on a particular subject, but it is impossible for an individual to be an expert in all subjects. Furthermore, due to the internet and the exponential rate of discoveries (human or otherwise), everyone is being inundated with more

information than any other period in human history and this trend will continue.

The evolving knowledge-based economy landscape must be seen from this context. Will AI replace the human component? The answer is likely to be “no.” However, in life there is no such thing as a definitive “yes” or “no.” One must realise that there are shades of grey. Therefore, the question should be: To what extent will AI replace the human component? Currently, AI has its limitations. Humans can analyse information heuristically. In other words, a heuristic approach of learning entails discovery and problem-solving, using reasoning and experience. This is basically the meaning of wisdom, which is the ability to use knowledge and experience to make good decisions and judgments. Hence, wisdom is the combination of experience, knowledge, and careful judgment.

On the other hand, a heuristic computer program uses rules based on previous experience to solve a problem, rather than using a mathematical procedure. Currently (and the emphasis is on currently), AI has not reached this level of advancement for humans to fully trust its decisions. This is not to say that there are no AI applications that can do this but not in a general context. However, AI is making rapid advancement with deep learning algorithms that aim to resolve complex tasks for use in a variety of fields, including manufacturing, health care, aerospace engineering, technology, and automotive design. Whether AI will reach a level that can be trusted with all different type of circumstances, only time will tell, but the trend is in that direction.

According to [Brown \(2023\)](#), Amazon discontinued using its AI recruitment tool when it became apparent that the original data it used in its AI-machine learning model was defective from the outset. Brown explains that the reason for this defect was because the model had identified male dominance as being preferred (since this was how Amazon had recruited in the past), and it therefore downgraded certain phrases with the word “women” even though the model had been instructed not to focus on gender. However, this does not mean that such issues will not be resolved in

the future where great advancements are being made in supervised and unsupervised learning algorithms.

Having said this, the problem with AI becomes precarious if humans place their full faith in the AI systems and blindly let the AI systems take the decisions for them automatically. It becomes even more dangerous if there is a loss of control of the AI applications by their human creators. [Brown \(2023\)](#) argues that even in the military, where split-second decision-making is required, commanders are hesitant to use automation to make decisions even faster. He maintains that the reports of humans being removed from the decision-making process, pushed the UN to propose a ban on AI-led decision-making in battle. However, he claims that this proposal failed due to pressure from several superpower countries. Furthermore, on 21 March 2024, the United Nations General Assembly unanimously adopted the first global resolution on AI that encourages countries to safeguard human rights, protect personal data, and monitor AI for risks.

[Brown \(2023\)](#) argues that currently, AI can only churn out existing ideas that it finds and consequently, does not create new content, such as quotes or new theories and ideas. He also contends that AI, which works more autonomously, such as AutoGPT, can handle less structured questions to deliver results but at times the source of its information is not evident or how it arrived at its conclusions. He argues that this can create difficulties for those who require accountability in their fields such as financial investment decisions or healthcare decisions, particularly when things go wrong. There have been many failures in the use of AI and there are many concerns, but this does not mean that an investigation into the root cause of the various problems will not shed light on how these failures and similar ones may be prevented and resolved.

The core issue is what the impact on the knowledge economy concept will be when (not if) the various AI difficulties are resolved over the coming years. In other words, is the economic world on the verge of the next evolutionary step and if in the affirmative what will the next evolutionary economic concept be? [Brown \(2023\)](#) suggests that we may be

on the verge of witnessing the destruction of the knowledge economy because there is the threat that as AI becomes more entrenched in our daily lives, there will be a focus not on people's knowledge but on the capabilities and utilisation of AI tools.

The knowledge economy will likely be gradually dismantled as knowledge bases (simple and complex) are replaced by safe and sophisticated algorithms that are able to learn, retain and recall knowledge on various matters, and in time will also make decisions and judgements. For example, generative AI (when safe and all issues resolved) will replace the human writer to produce material to one's specific theme and requirements, with human creativity taking a back seat and gradually fizzle out over time. We already have robots that serve at restaurants, admittedly not very advanced yet but they are bound to improve. In time, General-AI (or Strong-AI) that is designed to perform intellectual tasks that a human can perform by learning, reasoning, and understanding complex concepts will likely take over various human tasks that depend on human knowledge and wisdom. AI will continue to evolve to Super-AI that will have the ability of exceeding human intelligence in all areas. When available, Super-AI would be able to solve complex problems that are beyond human abilities and would be able to learn and adapt at a rate that far exceeds human intelligence.

The most popular type of AIs is Reactive Machines, which are the simplest form of AI systems that perform basic operations by reacting to some input with some output, without any learning occurring. Another type of AIs that are currently available is Limited Theory machines that are able to store historical data and/or predictions for making better predictions. This type of machine learning architecture is more complex. Experimentation is currently being actively conducted with Theory of Mind machines that are designed to interact with the thoughts and emotions of humans. Self-driving cars are the simpler form of Theory of Mind machines, but they still have many issues to overcome before they can be generally used safely. However, the long-term objective in the AI evolution cycle are Self-awareness machines that are aimed towards achieving enlightenment and

becoming self-aware, which has independent intelligence, and will not be under the control of the people that created it. This would likely be the final death nail for the knowledge economy concept as we know it today.

A knowledge economy is an economy in which the production of goods and services is based primarily upon knowledge-intensive activities that trigger a large portion of economic growth and employment. In other words, a knowledge economy is where knowledge and expertise are the key drivers for generating economic activity and wealth where the emphasis is on education, creativity, and finding ways of how to use knowledge for economic advancement. Nevertheless, the extensive application of AI systems is likely to generate a new economic concept where economic growth and employment would be the outcome of intense automation. However, this is only likely to happen if much more powerful and speedier computer processors are developed such as quantum and/or neuromorphic technology. Therefore, for the immediate present the knowledge economy concept appears to be safe, but the writing may be on the wall.

8.5 THE DECLINE OR RESURGENCE OF THE KNOWLEDGE ECONOMY

It must be recognised that the huge growth in AI is impacting the knowledge economy, and a significant transition process is currently underway. It must also be recognised that the information technology (IT) transition process has been continuous for at least seven decades, moving from data processing to assisting in decision-making, and now to AI. Each stage in the transition process has resulted in the automation of many operations that formerly required human knowledge. This also resulted in the loss of thousands of jobs being taken over by computer systems. However, this IT transition process over the years has also created thousands of jobs that gave birth to the knowledge economy. The question that must be asked is whether the growth of AI will continue to support the knowledge economy or will it contribute to its death and the birth of a new economic order.

The IT developments over the years have been revolutionary but these developments have also been entirely dependent on human thinking and knowledge, which is after all the basis of the knowledge economy. However, the development of AI is turning towards imitating humans that include being creative, having emotional intelligence, and trained in relationship-building. Furthermore, this development is only the commencement of a new generation of AI developments. On the other hand, some practitioners argue that AI is immediately accessible to the knowledge economy since very little technical acumen is needed and the user does not have to build anything. They maintain that an individual can start using AI tools without detailed training with the barriers to entry being very low and the speed of adoption being very high. However, as mentioned previously, General-AI and Super-AI that are designed to perform intellectual tasks that humans can perform and have the ability of exceeding human intelligence in all areas may mean the near total replacement for the need of human knowledge in performing tasks.

Nevertheless, the evolutionary changes in the IT industry have progressively resulted in opportunities and the creation of high-quality jobs. Generally, these opportunities have far exceeded the adverse impact of IT developments in terms of job losses. The developments in AI will likely have the same positive effects with new opportunities being created. This does not mean that individuals will not be adversely impacted, particularly the older persons who are caught unprepared and may find it difficult to adjust. This has always been the case with any economic development. There is no doubt that it will be very challenging (and even painful) for certain groups and class of individuals and organisations to adjust to the evolving labour and consumer markets, where different skills and techniques will be needed. However, these challenges may be mitigated by government economic and fiscal policies and by educational institutions keeping a step ahead of the required industrial trends.

Whether AI or just general computerisation, it is a fact that computers can do things faster and more efficiently than humans. However, currently, it is only humans that have the unique capacity for wisdom and emotional

intelligence, which make humans so important and unique. Whether these human qualities will remain unique to humans is a concern. This is why the legal and ethical implications of AI are such important issues, which will be addressed in a separate chapter at a later stage. There are those that are predicting a shift from knowledge-based to a wisdom-based economy, while others are visualising a shift towards autonomy-based economy. Whatever the trend, for individuals to remain relevant in the era of AI they must develop an assortment of skills and behaviours through a dynamic self-improvement programme.

Consequently, there is the positive scenario that AI will actually boost the knowledge economy and not its demise. AI is viewed as providing the potential for profoundly transforming the knowledge economy, which is generally characterised by the creation and exploitation of knowledge for economic and social development. AI has the capacity to analyse huge volume of data and attain significant insights, thus transforming different characteristics of the knowledge economy related to business, innovation, research, and education.

Enterprise AI is changing how entities operate and compete using AI driven analytical tools that can assess consumer trends (at organisational and individual level), customer behaviour, and competitor strategies. This facilitates management to make robust and informed decisions to keep ahead of their competition. Enterprise AI will also strengthen the general computerisation effort by further automating routine tasks that release valuable human resources that can focus on more important activities. AI also assists in the innovation process and is having a considerable impact in the knowledge economy by identifying new marketing opportunities, forecasting consumer trends, and helping in the development of new products and services. This is helping certain enterprises to combat competition. AI is also aiding the research process by having algorithms that can analyse patents and research papers through the identification of information patterns and trends, and providing meaningful insights that are not easily visualised by human researchers. This is helping researchers to reveal new detections and breakthroughs.

AI tools are assisting students having personalised learning experiences through intelligent tutoring that give immediate feedback and support that significantly improve the learning process. Consequently, AI is enhancing current non-AI tools and strengthening the knowledge economy. It is argued that as AI continues to grow and evolve, its impact on the knowledge economy is also predicted to escalate, revealing new opportunities for economic and social development.

The above examples support the contention that AI is likely to rejuvenate the knowledge economy by its ability to help organisations speed up their current strategy formulation process. There is no doubt that AI can analyse huge quantities of historical data from internal and external sources to speedily identify trends and patterns that human analysts might overlook. Hence, AI is a great improvement on the traditional computer systems that are more dependent on human interventions in terms of assessment. The emphasis is on the fact that under traditional computer systems human analysts are likely to overlook certain aspects that AI systems can identify. It is by this process that AI based systems facilitate enterprises to make better informed decisions based on instantaneous data to provide them with a significant advantage over competitors who may still depend on traditional computer systems.

However, one needs to recognise that such advantage is short-term because the competition is likely to adopt the same tools at some time. Hence, one cannot be complacent but always persist in striving to keep ahead by implementing the latest available AI tools. Obviously, this depends on whether an enterprise adopts a strategy where they see themselves as leaders rather than followers. However, there are enterprises that are content on being followers and adopt a wait and see strategy, but these enterprises will likely lose out in the long-term and risk their capacity to survive. Therefore, AI through its ability of analysing the competitors' strategies and tactics, as well as evaluating the workings of the business environment, can assist management to identify potential threats and opportunities, and turn their own weaknesses to strengths, thus helping them to modify their strategies accordingly.

8.6 AI IN THE KNOWLEDGE ECONOMY INDUSTRY

The previous sections have shown how AI is impacting the knowledge industry, with future developments determining whether the knowledge economy will decline or flourish further. AI has been shown to automate everyday tasks and enhance the decision-making processes with an improved overall effectiveness. AI has been shown to be extremely useful in environments where strategic decisions are required, irrespective of the industry type such as financial services and manufacturing in both the private and public sectors. AI supports knowledge workers by enabling them to focus on the more intellectually challenging and creative tasks that add value to the organisations.

The learning capabilities of AI results in the continuous enhancement of workflows and processes, by forecasting outcomes that make use of vast volumes of historical data. Furthermore, the organisation's ability to comprehend information through AI applications enable them to predict market and consumer trends. Chatbots and virtual assistants that are driven by sophisticated AI tools resolve customer concerns quickly by providing personalised recommendations that enhance customer experiences without any human intervention. In the immediate future, AI will become an important and essential factor for the knowledge industry. Whereas, traditional computer systems are important for streamlining the organisation's processes, AI goes beyond this role by creating new opportunities for innovation and efficiency. AI is ensuring that managers can be more proactive and effective in their decision making.

The concern with AI is likely to be a long-term scenario when AI is expected to become more intelligent than humans and act in its own accord independently from those that created it. The word long-term in this context is, shall we say, 50 years down the human timeline. However, many see AI as a current threat despite the clear benefits it has provided in the past decade. The main concerns are related to employment and ethical issues. However, these types of concerns were also expressed with the traditional computer systems of the early 1960s.

It is argued that AI would displace human workers, particularly those jobs that involve data processing and routine administrative tasks. One needs to recognise that routine administrative tasks were the first to be adversely affected in the 1960s when banks, insurance, and other financial institutions began to computerise their operations. The situation became worst with the onset of personal computers and multi-user systems that became affordable for the smaller organisations and replaced many clerical workers. Robots were also first applied commercially on assembly lines in the early 1960s that were primarily used for heavy lifting. Around the same time, the General Motors assembly line at the Inland Fisher Guide Plant in Ewing Township, New Jersey implemented robots in 1961 for spot welding.

The same may be said for the data processing industry. At first, there were the punch card processing systems that consisted of a punch, sorter, tabulator and other additional machines such as reproducing punches, collators, and sequence calculators, which were used by governmental and non-governmental organisations such as insurance companies and large businesses. The focus was no longer on statistical surveys but on account transactions, payroll accounting, inventory management, and billing. From the 1950s, punched card systems also found their way into medium-sized enterprises. These replaced many clerical workers but also created higher skilled jobs in mechanical and electrical engineering, for technicians, operators, and key punch operators. The same was repeated when personal computers with word processors came on the scene replacing the typing and data entry pools.

I remember the time when a car driver would drive into a service station for petrol, and while the fuel tank is being filled an attendant would clean the windscreen, check the oil, battery, and air pressure in the tyres. In time, these were replaced by self-service petrol stations where you filled the fuel tank, the computer kept a record of the amount and value taken and the driver went to pay the cashier. Later still, the cashier was replaced, and all transactions were paid by electronically. As society advances the means to service the society also changes and advances using automated systems. Old jobs are lost and new jobs that require a higher level of skill are

created. This story has been repeated over and over as economic progress is being made, hence, the creation of the knowledge economy. It is a fact that economic progress has generated many more jobs than the number of jobs that were lost through technological progress. Consequently, the arguments against technological progress including AI systems are not new.

There is genuine concern about the decision-making power being possibly delegated to algorithms. There are reservations that algorithms might not fully understand or respect the nuanced ethical and social implications of their outputs. However, one needs to remember that jet liners have been outfitted with an autopilot mechanism that enable a plane fly and land without the assistance of the human pilot. This feature has been available for several decades without much concern. Currently, drones can fly pilotless. Therefore, there is a need to distinguish between applications and closely examine the issues to decide whether there is a real cause for concern. There are those that argue that these fears are not unfounded. Currently, AI is primarily used to complement human capabilities, not replace them. It must be realised that AI tools currently analyse data and identify trends, but human experts are still required to interpret these findings within the context of existing political and social landscapes. Note that the word “currently” is important in the sentences above. This is the reason why I had stated that the concern with AI is likely to be a long-term scenario (50 years down the human timeline) when AI is expected to become more intelligent than humans and act in its own accord independently from those that created it.

Moreover, it must also be recognised that society’s behavioural habits are also changing. People want more leisure time to spend and enjoy their families and participate in more recreational activities. This implies shorter working hours or fewer working days per week. This can only be achieved by having greater automation, even to the extent of trusting certain algorithms to take decisions on behalf of humans. However, there is a need for the integration of AI into the workplace that include checks and balances to ensure safe and ethical AI usage, such as transparent methodologies; accountability for decisions made by AIs or with its

assistance; and constant monitoring to ensure equitable and unbiased outcomes. AI and non-AI applications need to be extensively tested. For instance, the CrowdStrike security update failure of 19th July 2024 for Windows based PCs, caused a massive global IT outage that immobilised airlines, hospitals, media, and banks around the globe. Security precautions must be taken on a continuous basis not just for cyberattacks but also for the safe installation of legitimate software activities and updates.

Furthermore, there is an urgent need for comprehensive policies and guidelines to regulate the use of AI applications to ensure that it supports employees and not destabilises them. The previous chapter illustrated that some work has already been done on this issue but more needs to be achieved with international collaboration of all stakeholders. Additionally, it is essential that action is taken to mitigate the perceived fears of AI development by educating and making employees aware of the potential AI's role in improving the economic activity; promoting a culture of collaboration between human intelligence and AI; and highlighting the social benefits of AI regarding the potential increase in leisure time to increase the quality of one's life.

AI has transformed the content creation field with generative AI tools that automatically generate reports, presentations, and other materials. Although there are still some issues to be resolved, AI makes it easier to emphasise potential policy impacts and critical points of interest for the relevant stakeholders. AI has been successfully applied in several knowledge management areas. For instance, AI using natural language processing (NLP) tools can quickly and efficiently process and analyse considerable amounts of information containing text data. This has proved to be useful in public policy and government relations field for analysing legislative documents, public feedback, and policy guidelines to assist relevant stakeholders to remain informed about legislative changes and public opinions. AI has also facilitated the complex issue of knowledge management by simplifying the data integration and retrieval processes because it can organise, tag, and classify considerable amounts of data, facilitating information accessibility to team members throughout the

organisation. This centralised approach mitigates the risk of creating data silos, thus reducing information retrieval time and more efficient search facilities.

8.7 CONCLUSION

This chapter examined the knowledge economy and its relationship with AI. The chapter illustrated that the knowledge economy is based on a concept where knowledge rather than physical output is the determining factor. Hence, knowledge-intensive products become the output, not a physical provision. The concern expressed by several researchers is whether AI is threatening the knowledge economy environment.

The argument of whether technology will result in the replacement of humans has been ongoing for many years before the AI came on the scene. The spread of computerisation in the 1960s was seen as a threat to people's jobs. Admittedly, many manual clerical jobs were lost but many more jobs that required higher skills and better pay were generated. This era was the birth of the knowledge economy. The chapter illustrates that AI will likely have the same impact. Certain functions that require human intervention will be replaced by AI technology, but many more jobs are likely to be created that require a higher skill set and probably better remuneration. However, such developments require comprehensive government economic and fiscal policies that support the shift to AI and the updating of the educational programmes by institutions of learning to ensure that the younger generation are equipped to face the technology challenges of the future. Concurrently, organisations have the obligation to launch their own training programmes to upgrade the skills of their current workforce.

AI is viewed as boosting the knowledge economy and not destroy it because AI has the capacity to analyse huge volume of data and attain significant insights, thus transforming different characteristics of the knowledge economy related to business, innovation, research, and education. Several examples were provided illustrating how AI complemented the knowledge economy. Hence, it is likely that AI will

revitalise the knowledge economy by its ability to assist management to speed up their current strategy formulation process. Additionally, AI because of its ability to analyse the competitors' strategies and tactics can help management to identify potential threats and opportunities thus helping them to modify their strategies accordingly. One must also recognise that society is changing, where people want more leisure time to spend with their families and enjoy longer recreational activities implying shorter working hours. This can only be possible by having greater automation that include checks and balances to ensure safe and ethical AI usage.

REFERENCES

- Brown, Stuart. (2023). *From Knowledge Economy to the AI Economy: What Could Happen?* Wildix OÜ – Holding Company. <https://blog.wildix.com/knowledge-economy-with-ai/>.[↵]
- Camilleri, Emanuel. (2019). *Information Systems Strategic Planning for Public Service Delivery in the Digital Era*. IGI Global.
- Charan, Ram. (2005). *Boards That Deliver: Advancing Corporate Governance from Compliance to Competitive Advantage*. Jossey-Bass.[↵]
- CNBC. (2022). *AI: From a Knowledge-Based Economy to Automation-Based Economy*. www.cnbctv18.com/views/ai-from-a-knowledge-based-economy-to-automation-based-economy-10948532.htm.[↵]
- Drucker, Peter. (1967). *The Effective Executive*. HarperCollins.[↵]
- Drucker, Peter. (1969). *The Age of Discontinuity*. Elsevier.[↵]
- Marchand, Donald. A., Kettinger, William. J., and Rollins, John. D. (2000). Information orientation: people, technology and the bottom line. *Sloan Management Review*, Summer 2000, Volume 44, Issue 4, pp. 69–80.[↵]
- Nguyen, Hon. (2024). *AI Has the Potential to Disrupt the Entire Knowledge Economy. Are You Ready?* Insights, CodeStringers. www.codestringers.com/insights/ai-has-the-potential-to-disrupt-the-entire-knowledge-economy-are-you-ready/.[↵]

Parker, D. (2016). Moving to the algorithm. In *TheBlack*, April 2016, pp. 38–42. [↗](#)

Powell, Walter W., and Snellman, Kaisa. (2004). The knowledge economy. *The Annual Review of Sociology, Volume 30, Issue 1*, pp. 199-200. [↗](#)

UNDP. (2023). *Global Knowledge Index 2023*. UN Development Programme (UNDP). https://knowledge4all.com/admin/2023/Methodology/GKI2023_Methodology_EN.pdf.

Chapter 9

Enterprise AI

DOI: [10.1201/9781003653950-12](https://doi.org/10.1201/9781003653950-12)

“The man with a new idea is a crank until the idea succeeds.”

Mark Twain

Author (1835-1910)

AI is a field of computer science that looks at ways of training machines to learn, reason, and make decisions. In previous chapters it has been shown that AI has been around for many decades, going through various development cycles that at times have not been very successful. There were times when AI was abandoned by the industry as investment was withdrawn due to very disappointing results. However, research centres at various universities kept up the development momentum despite the lack of adequate finding. The outcome of this perseverance is the success that industry is currently witnessing.

AI and Enterprise AI are associated concepts but differ in their scope and function. AI is more concerned with the use of algorithms and computer programs to conduct tasks that characteristically require human intelligence. On the other hand, Enterprise AI refers to the concept of using AI in an organisational setting and encompasses many different

technologies, including rules engines and machine learning, amongst others. Enterprise AI is the application of AI technologies to specifically address business problems and enhance business processes. It achieves this by integrating AI into enterprise systems and workflows using AI tools to automate and augment business processes, enhance customer experiences, and act as the key thrust for business outcomes.

This chapter will illustrate how Enterprise AI facilitates businesses to transform their operations by streamlining workflows and decision-making. This is achieved by automating everyday tasks and providing personalised customer services to bring together significant data insights. The role of AI is to dramatically increase the capabilities and effectiveness of enterprise software.

9.1 INTRODUCTION

Generally, Enterprise AI refers to the application of AI technologies in business applications to improve organisational processes and decision-making. In other words, Enterprise AI exploits advanced AI systems that include machine learning as the main approach for digital transformation. A key difference between Enterprise AI and AI is their particular focus. Enterprise AI focuses on resolving extensive high-value application situations by utilising and integrating AI tools into the organisation's business processes to generate significant value.

Organisations that implement Enterprise AI have the potential of achieving step function improvements in business processes all the way through their value chain, resulting in enhanced business flexibility, higher efficiency, greater profitability, and lessened environmental impact. Achieving step function improvements speeds up the intended business direction by reconfiguring workflow, changing some parts and/or implementing new abilities.

The development of enterprise AI applications is due to recent advanced innovations in AI, in particular machine learning, together with the facilities provided by elastic cloud computing, big data, and the internet of things. Of

significance is elastic cloud computing since it provides the ability to quickly enlarge or reduce computer processing, memory, and storage resources to meet changing demands without having to be concerned about capacity for peak usage.

Large organisations are likely to develop and implement many Enterprise AI applications to resolve a variety of concerns across their business. However, developing and implementing Enterprise AI applications at scale necessitates considerably greater challenges than earlier generations of enterprise software, requiring a new technology stack. Technology stack refers to a combination of programming languages, frameworks, libraries, tools, and technologies that are utilised to create and implement software applications. [Figure 9.1](#) illustrates a typical technology stack for AI applications, which shows the various combination of technologies that are utilised in developing Enterprise AI. The technology stack generally involves front-end technologies that control the user interface with the system and the back-end technologies that consist of a suite of tools, languages, and frameworks, which are necessary for building the server side of the application.

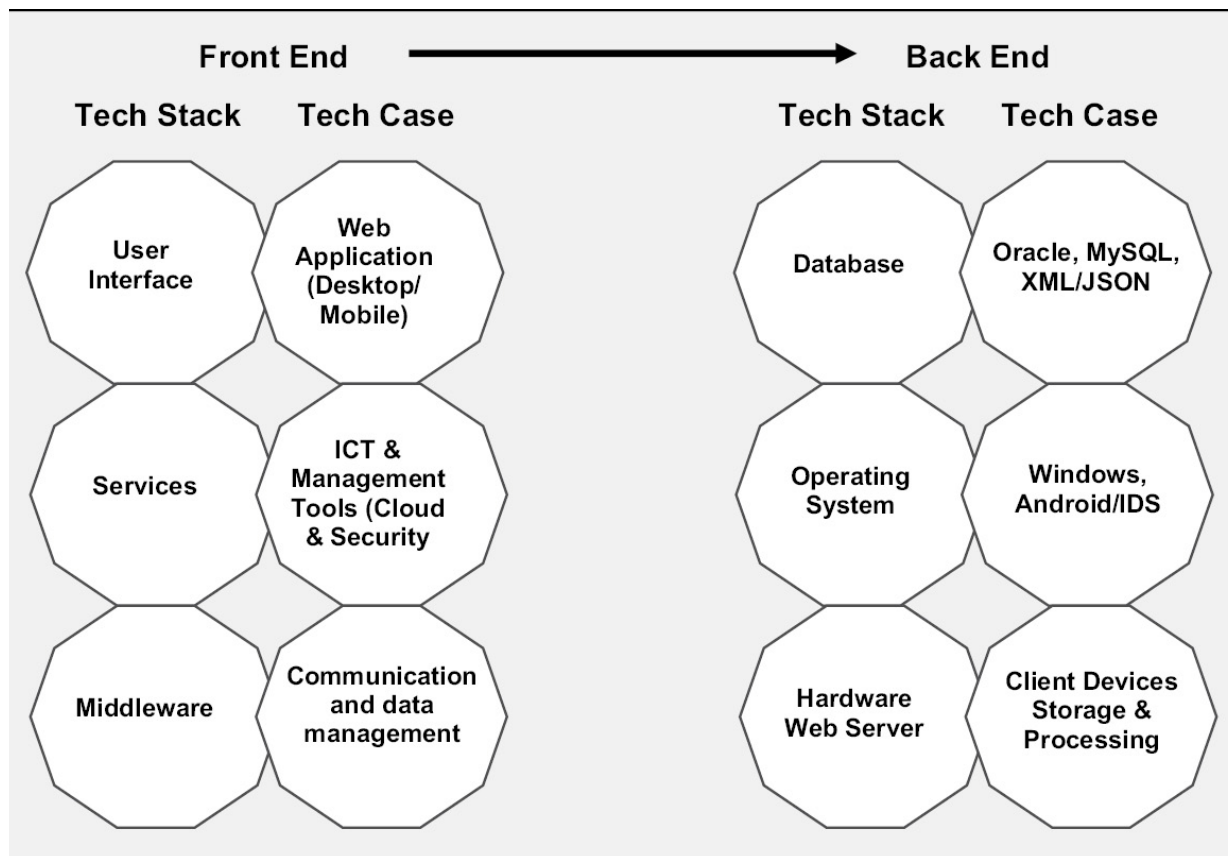


FIGURE 9.1 Typical Technology Stack for AI Applications. [📄](#)

Enterprise AI leverages the general AI features, machine learning, natural language processing (NLP), and predictive analytics as the major thrust to achieve business goals and decisions in industries such as manufacturing, finance, marketing, ICT, and the public sector. The application of AI related to the public sector will be treated in a separate chapter. Our focus in this chapter is the private sector. Furthermore, the following sections will provide a more detailed description of Enterprise AI and how it may be implemented to benefit organisations.

9.2 AI IN BUSINESS ENTERPRISE

There is no doubt that the application of AI in the enterprise sphere is fundamentally changing the way a business functions. Businesses are integrating AI technologies into their business operations with the objective

of increasing efficiency and reducing costs, creating new markets, and generating insights regarding industry market trends. The implementation of Enterprise AI applications has resulted in enhanced customer experience, optimised supply chains, improved sales, improved existing products and services, helped to create new products and services, and refined cybersecurity. It has also released employees from performing routine non-value-added tasks. This is not to say that the entities implementing these Enterprise AI applications have not encountered significant challenges.

AI is a dynamic field that is constantly changing due to greater AI research taking place on a global level, which is being supported by governments, large corporations, and education research institutions. Furthermore, AI is being applied to resolve a variety of problems that are being encountered by entities. For instance, AI-enabled hardware and software robotics are predicted to grow as business entities endeavour to become more flexible to meet the global economic challenges. Hence, business leaders need to appreciate how AI works and where AI technologies can be feasibly and reliably applied in their respective business segments.

It is a known fact that many tasks conducted within an enterprise require a specific measure of intelligence that require human intervention, hence they are not automatic. However, intelligence is viewed as the ability to acquire knowledge and applying it to achieve an outcome related to specific circumstances. Hence, AI entails programming a machine to function in this manner and is capable of rationally solving complex problems. Essentially, AI programming focuses on three cognitive skills, namely:

- **Learning:** This aspect aims at attaining data and generating rules regarding the conversion of data into actionable information. The rules define the processing instructions required to complete a specific task.
- **Reasoning:** This concerns the AI's capability of selecting the most suitable rule (algorithm), from among a set of rules, to apply in a specific circumstance.

- Self-correction: This aspect focuses on the AI's capacity to increasingly fine-tune and enhance a result until it attains the desired goal.

On the other hand, Enterprise AI is somewhat different from merely AI. Enterprise AI focuses on the integration of advanced AI-enabled technologies and techniques within large organisations to improve business functions. Moreover, it incorporates routine tasks such as data collection and analysis that involve more complex operations such as automation, customer service, and risk management. Hence, the essence of Enterprise AI is characterised by AI software tools that leverage leading-edge systems, including machine learning, NLP, and computer vision. These AI tools enable entities to attain process automation in diverse tasks, streamline complex business functions, automate repetitive tasks, and exploit the accumulated data and information. Additionally, entities can acquire significant insights about key performance indicators (KPIs) and improve their business strategies by applying AI to analyse the accumulated data set. However, to attain such benefits, organisations need to make significant investment in the technology infrastructure and skilled employees.

This illustrates that Enterprise AI applies AI tools to deal with business challenges within the internal environment of an organisation. The AI tools range from the use of machine learning, deep learning, NLP, and other available AI methods to automate processes, enhance decision-making capabilities, and conceive new products and services. The integration and application of these various AI tools goes beyond merely automation because their goal is to solve complex business concerns that entail human-like intelligence. These AI tools are utilised to provide insights about customer behaviour, detecting fraud, and improving logistics, amongst many others. Enterprise AI harnesses the organisation's data repository by applying advanced algorithms that optimise its operations, can discern market insights, and push innovation across the organisation crossing divisional boundaries and departmental functionalities.

9.3 THE IMPORTANCE OF ENTERPRISE AI

Enterprise AI is well known for its capacity to undertake and unravel complex organisational problems that traditional computer system could not undertake. Even though mathematical models existed under the traditional methods, however these models were not very effective in their ability to accurately forecast and deal with the complexities of global supply chains. Through the integration of various AI tools, Enterprise AI can exploit these mathematical and statistical models to predict with high accuracy future market trends, optimise logistics and inventory levels, and forecast disruptions. Enterprise AI is currently being successfully used for innovative applications regarding financial fraud, healthcare, and making the most effective use of energy resources, amongst many others across various industries.

Later in this chapter there will be a more detailed discussion on the uses of Enterprise AI. However, the objective at this stage is to illustrate the importance of Enterprise AI in enhancing organisational effectiveness. The volume of data generated worldwide is expected to continually increase at a rapid rate. Hence, entities that are committed to a data-driven decision-making strategy will view this expansion as a definite advantage. This increase of data provides far-reaching business intelligence that will act as the main thrust for enhancements in business operations and generating new products and services. Furthermore, AI will help entities to exploit this rapid expansion of data through deep learning to reveal the elusive patterns and hidden relationships that may provide entities a competitive edge. Additionally, AI has the capacity to make meaningful forecasts that may not be noticeable to humans by meticulously exploring the vast repository of data. Hence, Enterprise AI applications are able to provide answers to concerns that were not likely to be resolved by humans or traditional systems.

Organisations in every industry are discovering new ways of utilising Enterprise AI applications to tackle a wide spectrum of use cases, including customer retention; inventory optimisation; fraud detection; and anti-money

laundering amongst many others. Enterprise AI applications are also being applied in specialised areas such as predictive maintenance for a variety of industrial assets, including oil and gas production apparatus; energy production and distribution equipment; and vehicle and general transportation production equipment. The list of potential applications is impressive. Enterprise AI applications are only limited by one's imagination. Enterprise AI has quickly become a key digital transformation enabler and is stimulating customer expectations, which necessitates organisations to be more responsive and embrace new technologies to stay competitive and ultimately survive.

It is a recognisable fact that the sharp increase in the volume and complexity of data generated by businesses is making it extremely difficult for traditional systems and humans to analyse and interpret data effectively. On the other hand, Enterprise AI systems are providing the means and opportunity to unearth meaningful insights from this vast repository of data, information, and knowledge to disclose concealed patterns and trends that traditional approaches are not capable of doing. Additionally, Enterprise AI systems can automate complex decision-making practices that enable organisations to be responsive to market trends and thus gain a competitive edge.

9.4 ADVANTAGES AND CHALLENGES OF ENTERPRISE AI

Enterprise AI has been shown to transform business operations by automating everyday tasks and streamlining business processes resulting in improved efficiency and reduced operational costs. The outcome of this transformation is the positive impact on employees and customers, since AI frees employees from routine tasks, giving them the opportunity to provide customers better personalised experiences and to engage in additional strategic activity to combat competition. This has resulted in enhanced customer support through chatbots and virtual assistants to provide a 24/7 service from anywhere providing revenue earning opportunities through better customer support interactions.

Additionally, AI has the necessary tools that exceed traditional data processing methods, to conduct in-depth analysis of huge volumes of data resulting in better informed decisions. Moreover, AI algorithms have also enhanced cybersecurity by detecting and responding to threats effectively through powerful intrusion prevention systems. Research literature has revealed that Enterprise AI offers the following benefits, some of which have been mentioned above:

- Increased efficiency and productivity: Enterprise AI assists in the automation of recurring and tedious tasks, streamline workflows, and optimise resource allocation, resulting in significant improvements in efficiency and productivity.
- Cost mitigation: This is achieved through AI's ability to streamline operations and optimise processes, identify inefficiencies, and mitigating errors, all of which contribute to significant cost reductions.
- Improved customer service: AI can provide a faster and more personalised customer service. Advances in speech recognition and other NLP tools give chatbots the ability to provide personalised service to customers. These personalised customer interactions, impart tailored recommendations through chatbots and virtual assistants on 24/7 basis from anywhere, resulting in stronger relationships that promote customer loyalty.
- Enhanced monitoring: AI can process data in real time providing entities with immediate monitoring. For instance, factory shopfloor production is able to apply image recognition and machine learning models in their quality control practices to scrutinise production and identify problems.
- Higher quality: AI tools can mitigate errors and increasing the likelihood of complying with compliance standards better than manual or traditional automated approaches.
- Quicker product development: AI can better predict market consumer trends to facilitate shorter service and product development cycles.

- Enhanced HR management. Entities are utilising Enterprise AI to rationalise the employee engagement process and eliminate bias in business communications. Furthermore, by releasing employees from routine tasks, AI helps to empower employees to focus on higher-value activities that involve creativity, critical thinking, and problem-solving skills.
- Improved business modelling. The ability for AI to perform in-depth analysis of huge quantities of data rapidly to detect patterns and trends help entities to define better business models when compared with traditional approaches.
- Better competitive advantage: AI tools provide insights into market trends that allow management to make faster strategic decisions and help management to generate innovative products and services to gain a significant competitive advantage.

Despite the above advantages, Enterprise AI has its potential risks and challenges such as algorithmic bias and possible data security issues. Hence, it is important that organisations define clear ethical guidelines that facilitate robust design, data privacy standards, continuous monitoring, and suitable security systems. Such measures are important to ensure that Enterprise AI systems comply with appropriate standards and maintain users' trust. Another concern is the impact of AI on employment and the general workforce. This topic was extensively addressed when discussing the knowledge economy. It was concluded that businesses like past technological break throughs need to retrain and reskill their employees to mitigate employee mistrust and anxiety, and that these employees remain an integral part of the evolving work environment.

There is no doubt that technology in general can erode basic skills that can impact one's performance when technology inadvertently fails. This problem is not restricted to AI but to technology in general, particularly when organisations and individuals are continuously reliant on it. For example, before the invention of the calculator, employees in banks had very high abilities for manually counting lengthy columns of figures and

calculating monthly interest on deposits. As the calculators came on the scene and later when computer systems took over, these skills greatly eroded. Obviously, one cannot stop progress and significant cost savings, and much better customer service resulted from these inventions. However, this type of problem becomes very serious in certain circumstances. Research has shown that aircraft accidents are increasingly due to the erosion of basic pilot skills. More specifically, the two plane crashes involving Boeing 737 Max jets, gave rise to concerns that pilots were losing their ability to apply basic flying skills as the aircraft depended increasingly on AI in the cockpit.

AI errors are still a great concern. It is a fact that AI can remove human error, but it can also introduce new types of errors due to questionable data, deficient training data or faults in the algorithms. Additionally, these errors can be critically intensified because of the great quantity of transactions that AI systems typically process. However, the greatest concern is regarding unethical and unintentional routines. For example, entities need to be aware and watch for gender or racial or religious bias in reports produced by AI applications.

It is important to recognise that AI is being used in a wide variety of applications that may impact individuals differently depending on their social and income status. Hence, entities need to be watchful for unintended consequences when applying AI for certain business strategic decisions that could have implications on certain sectors of society. For instance, a supermarket chain might apply AI to decide their pricing strategy based on the rivalry from other similar establishments. However, in low-income areas where there may be little or no rivalry, the logical advice from the AI system may infer to have higher food prices, which would go against the social status of the people living in that particular area.

9.5 ENTERPRISE AI SCALE CONCEPT

The concept regarding Enterprise AI scale concerns the ability of the AI systems and their related technologies to operate successfully within the

complex environment of a large organisation. To function in this type of environment, Enterprise AI systems must adhere to several critical prerequisites to be classified as “Enterprise-Scale.” [Figure 9.2](#) illustrates that the Enterprise AI scale environment is based on nine essential criteria.

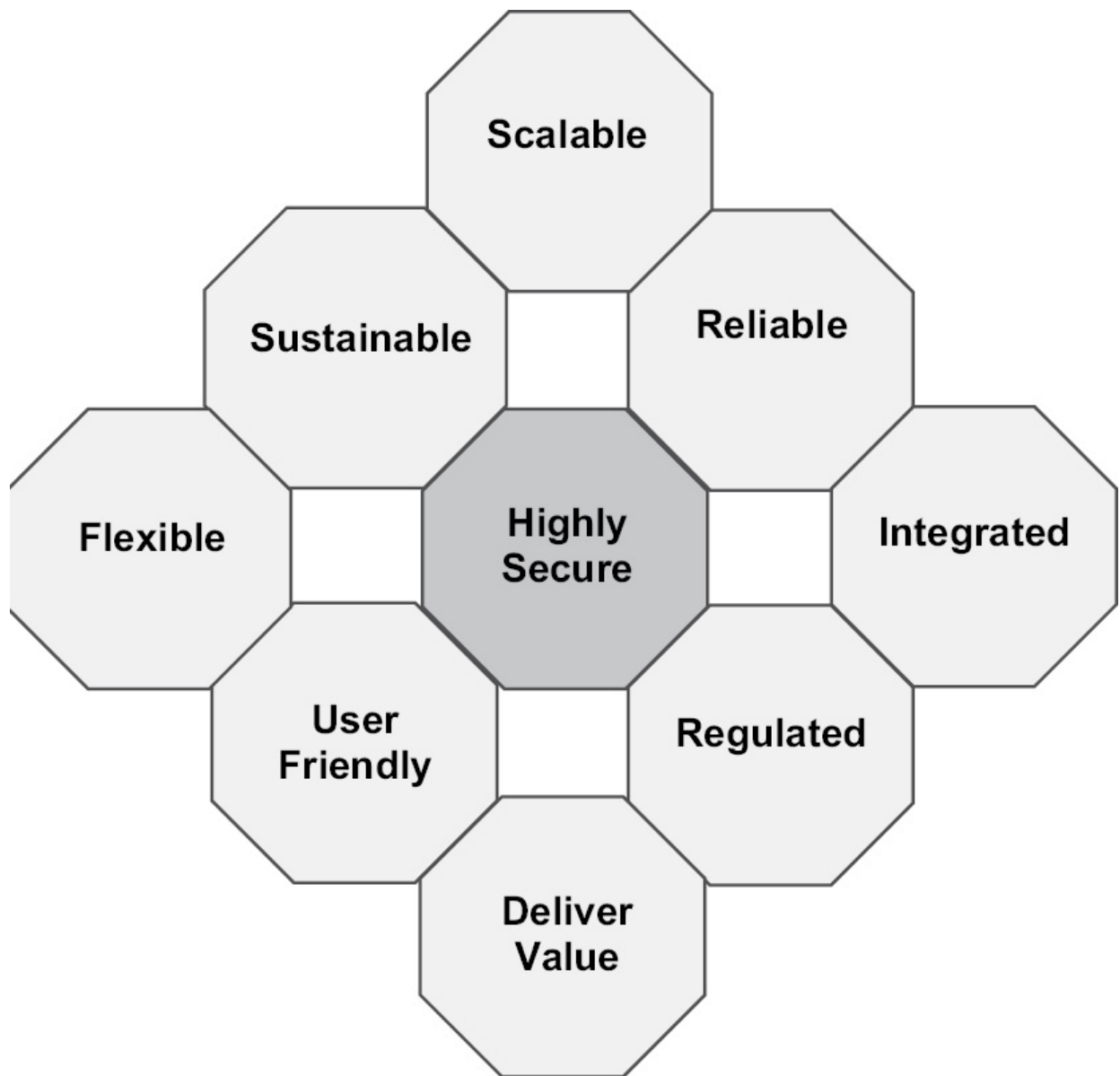


FIGURE 9.2 Prerequisites for Enterprise AI Scale Environment. [📄](#)

Whilst all the criteria are equally important, the most prominent factor is high security. The sensitivity of business data demands that Enterprise AI systems must include robust security procedures to protect data integrity

and privacy, ensuring safe user access and combat cyber threats. Enterprise AI systems must be scalable so that as the business grows and more complex circumstances and transactions are being conducted, these systems must be able to process and adapt to this growing business environment. An advantage of AI systems is that they do not need major redesign with greater complexity and bigger volume of data but can learn and modify themselves as the business environment changes.

Reliability of systems, data, results, and predictions is vital in any automated system. In the recent past user trust has been an issue with AI applications. Reliability in Enterprise AI entails consistent system behaviour with minimal downtime. These AI applications must perform as intended under fluctuating conditions. They must have robust error handling processes to ensure continuous system accuracy and accessibility. The definition of Enterprise AI emphasised the integration of various AI tools and other business systems. This integration must be seamless to users so that there is an effortless and automatic flow of data and interoperability within the organisation's ICT infrastructure allowing for an efficient exchange and utilisation of information.

AI systems need a strong regulatory regime based on clearly defined policies, practices and responsibility roles that include: (a) compliance with legal and ethical standards; and (b) data and model management. This will ensure accountability in AI decision-making. A basic objective of automated systems is to contribute towards the achievement of the organisation's strategic and operational goals that may include increased efficiency, reduced cost, enhanced customer experiences, and the development of new product lines to boost revenue streams. The previous section described various benefits that Enterprise AI is able to provide.

The success of computer systems is not only measured by the results they provide but by their penetration rate within the organisation. Hence, the easier the Enterprise AI system is to use, the greater the probability that it will actually be utilised. There is no point in having a highly sophisticated Enterprise AI system if no one or few utilise it. Therefore, user-friendliness becomes a critical aspect for its successful deployment with AI tools and

interfaces being manageable and clear to technical and non-technical organisational users.

An important critical prerequisite is flexibility, which means that the Enterprise AI system must be able to adjust to the changing business requirements and goals. This prerequisite complements the scalability factor by sustaining the organisation's business strategy and functions, and/or the emergence of new market trends. The final prerequisite is sustainability, which entails the long-term system support, including its environmental impact and the system's capacity to accommodate business strategic changes and technological advancements.

9.6 IMPLEMENTING ENTERPRISE AI

Generally, when implementing Enterprise AI, it is important to firstly identify the problem by ascertaining the areas where the processes are inefficient so that one may assess how AI can possibly address these weaknesses. Once the problem is defined and assessed, it is essential to evaluate the organisation's data and document the various processes so that an effective machine learning model may be developed. The machine learning model will use as its input the organisation's data and the documentation of the processes as its training basis.

Several issues need to be considered in the Enterprise AI deployment process, such as model bias, data privacy and security, and supporting AI technical employees. Bias in AI algorithms is a constant concern since AI is trained to recognise patterns. However, by recognising this bias, the model can be retrained to be more balanced by reframing outcomes around key markers. Data privacy and security must be taken seriously to avoid threat actors from attaining access to the system. Intrusion prevention systems (IPS), and an effective firewall will resolve this issue effectively.

Like every other computer system, an Enterprise AI application requires competent and dedicated AI professionals that focus on ensuring that the deployed AI system achieves the defined business goals considering the ethical and security aspects. It is important to recognise that implementing

Enterprise AI requires a strategic approach and careful planning. [Figure 9.3](#) provides a suggested framework for deploying an Enterprise AI system.

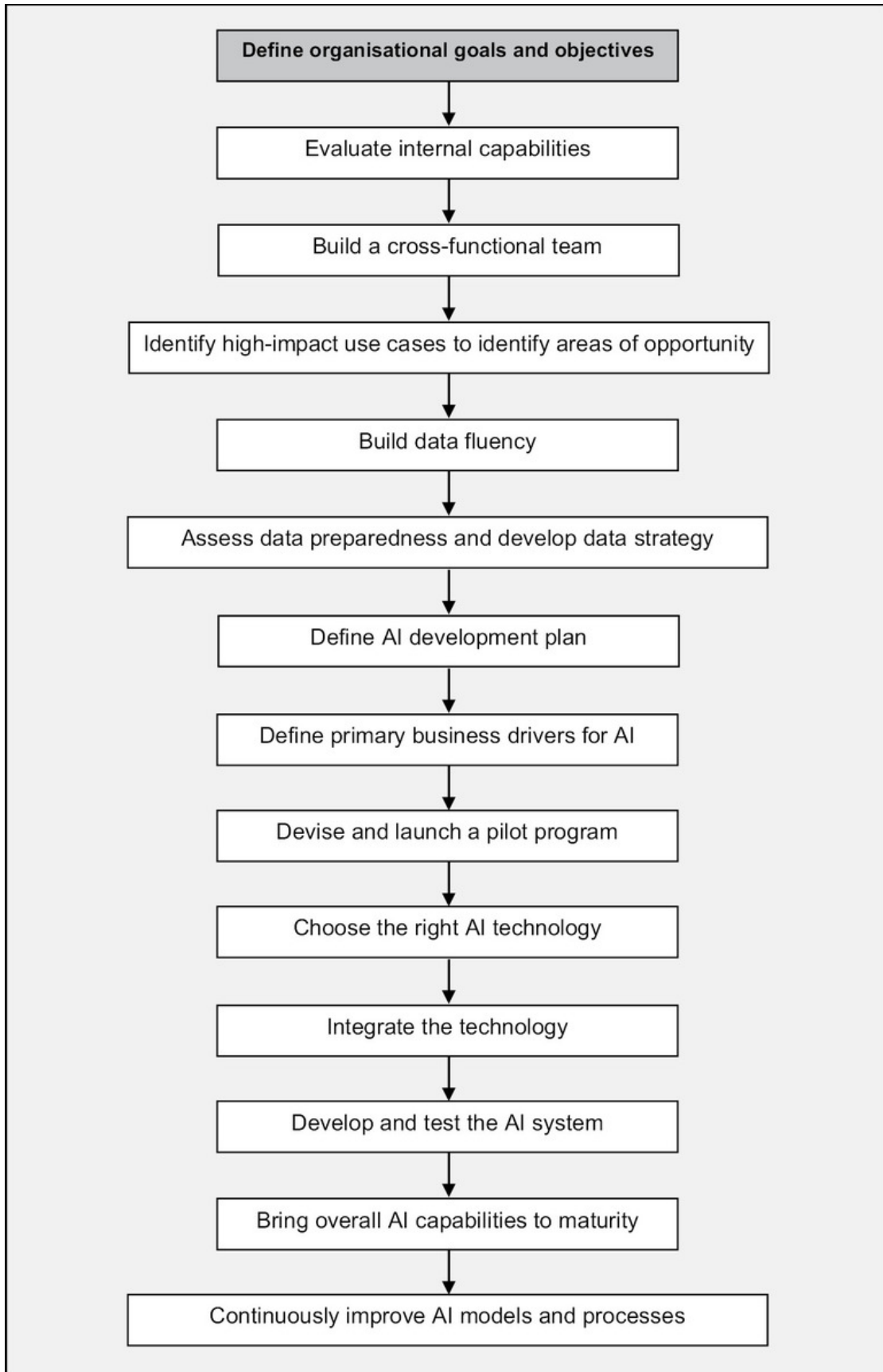


FIGURE 9.3 Framework for Deploying an Enterprise AI System. 

A fundamental step is to define the organisational goals and objectives. This entails clearly identifying the business problems that need to be addressed and/or the opportunities and outcomes that are anticipated to be gained by the eventual Enterprise AI application. Therefore, one needs to outline what the organisation aims to achieve with AI. This will assist the organisation to focus its efforts and choose the most significant Enterprise AI solutions. Typical goals may include operational efficiency, better customer experience, and the identification of new revenue streams. Hence, knowing these objectives helps the AI team to determine the direction of the AI strategy and to align it with the overall business objectives.

The evaluation of internal capabilities is very important since management must establish whether the organisation has adequate competencies or if these competencies need to be attained. Moreover, the method of attaining the required competencies must also be decided, in terms of whether to recruit professional staff or outsource the requirement. This evaluation provides adequate information to determine the training and development programme for existing technical and non-technical employees.

It is important that a management executive is chosen to act as the authoritative sponsor and champion for each AI initiative with an adequate budget. To deploy Enterprise AI successfully, an organisation requires a competent cross-functional team. This executive would need to identify skill shortages so that an appropriate team is created, which has a combination of technical skills and business knowhow. In other words, this team requires technical and business professionals. Creating a cross-functional team ensures that a holistic approach to implementing the project is taken, where different views and knowledge are merged to tackle the various aspects of the Enterprise AI project. The team's composition and size must align with the scope of the project and reviewed as various AI sub-systems materialise.

Identifying high-impact use cases to ascertain the areas of opportunity is essential. This step requires management to explore relevant AI use cases and identify the business processes where AI can add value and support the business goals. This step allows management to establish a case list to prioritise the possible business functions that provide the highest possible returns. Factors that should be considered are data availability, deployment feasibility, and return on investment. This will ensure that cases which offer the greatest potential impact are selected.

The next step is to build data fluency within the organisational culture so that employees are able to familiarise themselves with data and its possible interpretation, and to have the capacity to create and communicate information from the data. This will encourage employees not only to understand the data but to also exploit it to convey complex ideas, engage in discussions, and absorb information from others effortlessly. This will enable individuals across various roles to connect through a shared understanding of data, transforming raw data into actionable insights.

Concurrently with building data fluency, the organisation's data preparedness is assessed so that data strategy is adequately prepared. Management needs to assess the quality, quantity, and accessibility of the organisation's data. One must recognise that AI algorithms require huge volumes of data to train and operate effectively. This stage entails identifying the data sources, ensuring data quality, and establishing data governance and compliance processes. Therefore, this step ensures that the organisation has access to adequate and relevant data supported by an effective data strategy and robust data management practices.

Management is also required to define a detailed but flexible development plan with the aid of competent internal or external professionals that can select potential AI projects that are likely to give value. These professionals should also be able to commit to deliver a viable project beyond the pilot stage and refine each project as it moves into the deployment phase. The plan should also define the technical and business phases of the Enterprise AI implementation that includes an indication of

the AI technologies and tools, defining the scope of the AI projects, establishing timelines, and necessary resources.

An important step is the identification of the primary business drivers for Enterprise AI. Some of the key business drivers include enhanced efficiency and productivity; data-driven decision-making; improved customer experience; streamlined operations and supply chain management; and enhanced cybersecurity. This step will enable the AI implementation team to establish performance parameters; use these key performance indicators to monitor the progress of the AI plans; and help them prepare a risk management strategy to define contingencies in case of difficulties. To assess the defined plans, it is suggested that a pilot project is initiated to ensure that the plans for full scale deployment of Enterprise AI are viable, realistic, and anticipate potential issues that may adversely impact the Enterprise AI project.

The Enterprise AI project plan will initially indicate the type of technology required. However, once the plan is being refined and implemented, it is important that suitable AI technologies and tools that best fit the specific selected use cases and data requirements are acquired. Important factors to be considered when determining which tools and software are to be used include sustainability and long-term support; scalability; performance; and the integration capabilities with the existing systems. Once the AI technologies and tools have been selected and acquired, these need to be integrated into the existing systems and workflows to provide a holistic solution. This step requires meticulous planning to ensure the highest level of compatibility, a minimal level of disruption to current operations, and a high penetration rate at the user level.

Developing and implementing the selected Enterprise AI solutions requires a systematic process, exhaustive testing, and validation. This will ensure that the systems are operationally reliable with a minimised user dissatisfaction level. The success of the Enterprise AI system is determined by its usability and user trust and acceptance. It is also essential that the AI implementation team habitually monitor performance, fine-tune the system

parameters as necessary, and continuously train the system by iterating to improve accuracy and effectiveness.

The next phase is to bring overall AI capabilities to maturity by further testing and fine-tuning of system parameters ensuring that all modules are fully integrated to provide an effective holistic system. It is important to ensure that there is adequate abundance of reliable and trustworthy data from internal and external sources, which is supported by a suitable infrastructure to collect, filter, store, and deliver the necessary data. The final step is to continuously improve the AI models and the supporting processes. Significant attention should be given to security, privacy, and ethical AI issues by factoring activities for complying with government regulations and having courses that support best practice.

The AI team must also establish whether the Enterprise AI deployed applications are realistically supporting the business goals. This may be achieved by comparing the consequent system outcomes with the baseline metrics, particularly the established performance KPIs. Finally, support the health of the technology by having regular post-implementation reviews with ongoing maintenance, including recurrent updates, monitoring AI performance, and fine-tuning the system. This will ensure that the AI investment is safeguarded by having the AI systems remaining effective, secure, and aligned with the evolving business needs.

The framework depicted at [Figure 9.3](#) ensures that the deployed Enterprise AI systems are trustworthy and provide an explanation of how a decision was attained. Moreover, organisations that deploy Enterprise AI normally require 24/7 service from anywhere. Hence, defining the specific requirements will ensure that this criterion is achieved. Research has also indicated that natural language processing (NLP) is the most common tool that is adopted for Enterprise AI systems, particularly for customer service functions. Additionally, the framework provided aims to mitigate the main barriers that are normally encountered by organisations, namely scarce AI expertise, escalating data complexity and data silos, and the lack of tools for developing AI models.

One must recognise that AI is supported by several tools, namely machine learning, deep learning, predictive analytics, NLP, computer vision and automation. However, to gain value from these AI tools, entities must address key issues related to people, processes, and technology, all supported by enormous and timely volumes of high-quality data in a robust and secure environment. This requires substantial processing power with supercomputing platforms. Currently, Enterprise AI is viable for large organisations that can afford to have their own infrastructure or to partner with high-technology companies that can provide the necessary computing power over a leading-edge cloud environment.

Furthermore, the framework provided at [Figure 9.3](#) ensures that Enterprise AI, which is a multifaceted process, is implemented as a strategic methodology, ranging from identifying unambiguous objectives to conducting technology post-deployment reviews. Only through such a strategic approach can an organisation achieve the required return of investment from its Enterprise AI applications.

9.7 ENTERPRISE AI USE CASES

Enterprise AI is being deployed across a broad spectrum of industries related to various business functions. The objective in most cases is to leverage the data these organisations possess and utilise it to attain tangible business value from the Enterprise AI applications. Enterprise AI facilitates management's ability to make better informed data-driven decision-making, increases operational efficiency, enhances workflows, and improves customer experience. This in turn is assisting organisations to make significant advances in their business processes and attain better operational robustness and profitability.

Enterprise AI is transforming the way entities are managing their internal processes and engaging with their customers. The ensuing Enterprise AI models can address high-scale complicated problems regarding the automation of workloads, promoting products and services, providing personalised user experiences, assisting customers to make

decisions, and tackling difficult business concerns. [Table 9.1](#) provides typical examples of Enterprise AI use cases. AI has transformed industries due to its strategic impact on organisations through its ability to further automate the tasks that are normally performed by humans. Moreover, labour gains using AI are envisaged to grow at a faster rate than the job losses that result from using current workplace automation tools.

TABLE 9.1 Examples of Enterprise AI Use Cases [📄](#)

Industry	Use Case
Agriculture	To yield healthier crops, reduce workloads and organise data.
Customer Relations: Enhanced customer service	Chatbots and virtual assistants provide 24/7 customer service from anywhere to answer inquiries and resolving issues. They learn from customer interactions to improve their ability to assist customers.
Customer Relations: Personalised marketing	Analyse customer data to generate personalised marketing campaigns by understanding customer preferences and behaviours, leading to increased engagement and sales.
Education	Automating the process of grading exams by assessing students and adapting curricula to their needs.
Energy	Optimise power grids and improve the integration of renewable resources.
Financial services: Fraud Detection	Detect fraudulent activities by analysing transaction patterns and flag unusual activities to significantly reducing the risk of financial fraud.
Financial services: General	To improve back-office banking systems and automate customer service to create new opportunities.

Industry	Use Case
Healthcare	Assists to diagnosing diseases, analysing medical images, and predicting patient outcomes enhancing the accuracy of the findings.
HR management	Assist in resume screening and candidate matching to identify the most suitable applicants. To analyse employee data and advise on training, development, and retention strategy decisions.
IT service management	Use NLP to automate user requests for IT services. Apply machine learning to IT service management (ITSM) data to gain a better knowledge of the infrastructure and processes.
Law	Use machine learning to mine data and predict outcomes. Applying computer vision to classify and extract information from documents, and NLP to interpret requests for information.
Manufacturing: Robotics	Use of collaborative robots on assembly lines and in warehouses to assist humans.
Manufacturing: Predictive maintenance	Use predictive maintenance for machinery. Use machine learning algorithms to identify buying habits to predict product demand for production planning.
Marketing	To analyse historical sales data, market trends, and other relevant variables to accurately predict sales to enable businesses to optimise inventory, enhance resource allocation process, and make better informed decisions about pricing policies and marketing strategies.
Security: Cybersecurity threat detection	To identify and mitigate cybersecurity threats by analysing network traffic and identifying unusual patterns to detect potential security breaches before they cause significant harm.

Industry	Use Case
Other: Predictive maintenance	Analysing sensor data and identifying patterns to indicate likely equipment failures to predictive maintenance needs to mitigate downtime, optimise maintenance schedules, and extend the lifespan of assets.
Other: Supply chain optimisation	Forecast product demand, optimise inventory levels, and identify potential disruptions in the supply chain ecosystem to assist in having smoother operations, reduced costs and improve responsiveness to market changes.

The strength of AI is its analytical ability. This means that AI is not simply automating processes but is able to initiate the most efficient way to perform a task and fine-tune workflows immediately as circumstances change. As [Table 9.1](#) illustrates, AI is already complementing human work in many fields, from agriculture to help yield healthier crops; assisting medical staff to diagnose a patient's condition; and helping to optimise the organisation's supply chain by forecasting product demand, improve inventory levels and identify potential disruptions in the supply chain ecosystem to achieve smoother operations, mitigate costs, and enhance awareness of market changes.

AI algorithms have been deployed as an additional security layer by automatically detecting and responding to cybersecurity threats. Moreover, natural language generation tools are also being adopted to enhance customer service, while facial recognition tools are used by law enforcement agencies. Deep learning platforms are also currently being applied for experimenting with automated driving. However, safety aspects need further development before automated driving can be adopted in a practical environment. The financial sector, particularly banks are using AI to hasten and support loan processing, while maintaining regulatory compliance.

9.8 ENTERPRISE AI POTENTIAL

Enterprise AI is expected to be greatly impacted by the advances of generative AI technology due to its capacity to use large language models (LLMs) for creating new content and providing solutions that go beyond traditional predictive models. Generative AI is not only capable of managing vast volumes of data but has the capacity to make sense of this data and is able to replicate patterns within data to provide innovative solutions. Hence, the abilities of generative AI will be a further instrument that will be integrated into the Enterprise AI toolbox.

An important objective is the seamless integration of Enterprise AI with an ever-ending range of AI tools that are currently available or will be available in the time ahead. Enterprise AI will be able to use real-time data from Internet-of-things (IoT) devices to enhance predictive maintenance and provide better customer service. Furthermore, edge computing will enable faster localised AI processing that is very critical for time-sensitive applications. Hence, the integration of technology is not restricted to AI software but also to leading-edge hardware. A basic strength of AI is its capacity to manage large volumes of data. Therefore, in the big data context it will provide greater processing power for analysing and unearthing significant insights about a host of scenarios, including marketing trends. Enterprise AI's integration with blockchain technology will enhance transaction security and transparency, which are essential for the financial and supply chain management sectors. A separate chapter regarding predicted future AI developments will be provided in [Part IV](#) of this book.

It is important that ethical factors are addressed in a serious and responsible manner. Organisations applying AI need to adopt a strategy that promotes trust by ensuring equitable and transparent decision-making processes. AI should be viewed as complementing human talent not replacing it. A detailed chapter dedicated to ethical and legal considerations will be provided as the final chapter for [Part III](#) of this book.

9.9 PRACTICAL ENTERPRISE AI APPLICATIONS

AI is no longer a distant fantasy from science fiction. Everyday new applications are being released that have a profound impact on how organisations conduct their businesses and provide their products and services to their customers. This section presents some practical AI applications that are having a great beneficial impact on all types of customers, irrespective of whether they are organisations or individuals.

9.9.1 Case Study 1: eCommerce Operational Efficiency and Customer Satisfaction

[Tang_\(2025a\)](#) illustrates how an enterprise was able to profitably employ AI-driven solutions and mobile app development to enhance its operational efficiency by reducing its order processing time by 30%, increase user engagement by 25% and enhance its mobile sales by 15%.

This AI initiative took place at a time when the company was encountering significant concerns regarding the need to improve employee skills and enhance the supply chain effectiveness to eliminate a persistent 20% stockout rate. These concerns were further intensified by regulatory privacy requirements that entailed sound risk management strategies to avoid harsh legal penalties for non-compliance. This case study illustrates that AI solutions are typically implemented in combination with other strategic initiatives, such as employee development programmes, integration of technology and process improvement frameworks. Therefore, it is not just a question of implementing an AI solution in isolation.

The first step taken by the company was an assessment of its internal and external environments using well established methodologies. For the internal environment assessment, the ADDIE (Analyse, Design, Develop, Implement, and Evaluate) model was utilised for employee upskilling. This model depends on each stage being done in the given order but with a focus on reflection and iteration. The Lean Six Sigma was applied for process improvements (refer to [Camilleri, 2024](#), pp. 380–401). These tools permitted the company to identify key areas that required attention, which resulted in defining measurable improvements in productivity and

operational efficiency, and generating a more responsive and capable workforce.

The assessment of the company's external environment mainly encompassed techniques, such as Value Chain Analysis and the Kano Model (refer to [Camilleri, 2024](#), pp. 386–388). These models focus on enhancing activities that add value and prioritizing app features based on customer satisfaction. Competitive benchmarks (refer to [Camilleri, 2024](#), p. 369) were ascertained to evaluate performance against industry standards, specifically in order processing and mobile user engagement. According to [Tang_\(2025a\)](#), this strategic analysis improved the company's market position and greatly increased its responsiveness to customer requirements and market trends. This positively impacted overall business growth and customer loyalty.

The internal and external environment assessment enabled management to identify several strategic initiatives that included:

- **AI-Driven Order Processing Automation.** Lean Six Sigma and the Value Chain Analysis were used to implement this initiative. These tools enabled the implementation team to streamline the order processing process and mitigate any identified inefficiencies by adhering to a precise procedure, namely:
 - Specify the scope of the order processing automation project and define the key performance indicators (KPIs).
 - Measure current order processing time statistics and identify the bottlenecks by mapping the process, including the data collection practice.
 - Examine the data to determine root causes for the inefficiencies using root cause analysis, such as the 5-Whys (refer to [Camilleri, 2024](#), p. 76).
 - Enhance the process by implementing AI-driven automation solutions to target the identified inefficiencies.
 - Collect critical information and accurately monitor the new process, including the KPIs to ensure continuous progress (refer to

[Camilleri, 2024](#), pp. 307–313).

- Utilise Value Chain Analysis to determine and fine-tune the activities that add value to the order processing function. This sheds light on how each task in the order processing process impacts the overall value creation (refer to [Camilleri, 2019](#), pp. 217-223).

[Tang \(2025a\)](#) reports that this approach resulted in a 30% decrease in the order processing times and a significant improvement in operational efficiency and accuracy that lead to a 20% increase in customer satisfaction.

- Mobile App Development. This phase applied several analytical tools such as the Agile Methodology and the Kano Model. The Agile methodology is a project management and software development approach that emphasises flexibility, collaboration, and customer-centricity. Major companies like Facebook, Google, and Amazon use this approach because of its adaptability, customer-focused style, and its capability for developing a user-friendly mobile app quickly and efficiently.

The Kano Model was used to prioritise the features that were to be adopted based on the extent to which they were likely to satisfy customers. This approach allowed the product team to evaluate whether the inclusion of a high-satisfaction feature against its implementation cost was a strategically sound decision. Hence, the model was very useful for pinpointing which app features would delight users and which were essential for basic functionality.

The Kano Model is unique from other prioritisation frameworks because it has a rigorous focus on how customers respond to each feature. Hence, its key analysis for any new feature is to what degree it will satisfy users. The above approach ensured the successful implementation of a user-friendly mobile app that heightened the overall user experience, positively impacting customer satisfaction and loyalty. User-feedback allowed the mobile app to be fine-tuned resulting in a 25% increase in user engagement and a 15%

growth in mobile sales. The app's prioritised features enhanced the overall user experience, contributing to higher customer satisfaction and loyalty.

- **Data Privacy Compliance.** The project team examined GDPR compliance requirements and identified and evaluated probable risks to data privacy, such as data breaches and unauthorised access. The primary objective was to foster customer trust and mitigate as much as possible the identified risks and avoid legal penalties. This resulted in the mitigation of risks associated with data breaches and unauthorised access, thus boosting customer trust and brand reputation minimising legal infringements.
- **Customer Experience Improvement.** The frameworks applied for this activity were mainly Customer Journey Mapping and the Net Promoter Score (NPS). Customer journey mapping provided a visual depiction of the customer's experience during their ordering journey. It provided a deeper understanding of customers' emotions and actions highlighting pain points, unmet needs, and opportunities for improvement. NPS is often viewed as the gold standard customer experience metric. It measures customer loyalty and satisfaction by considering the customer's probability of recommending a given service or product over time. Typically, the NPS score is measured by a single-question survey and reported with a number ranging from -100 to +100, where a higher score is desirable.

Using NPS, the project team developed and implemented strategies to convert passives and detractors into promoters. The NPS trends were monitored over a significant duration to measure the effectiveness of customer experience initiatives. This resulted in the identification and subsequent resolution of key pain points, leading to a 20% improvement in customer satisfaction, with a 10-point improvement in NPS. Hence, this resulted in higher customer loyalty and repeat business, pushing long-term growth.

- **Employee Upskilling Program.** This phase utilised the ADDIE and Kirkpatrick models. The Addie model is an instructional design

methodology that is used to facilitate the streamlining and composition of the course content. In contrast, the Kirkpatrick Model appraises the outcome of training and learning programs. This model evaluates both formal and informal training approaches, assessing them against four levels of criteria: reaction, learning, behaviour, and results.

The project team applied the ADDIE model to analyse the competencies gap of employees and to develop the training modules according to the organisational business needs. The effectiveness of the training programme using the Kirkpatrick Model was evaluated by assessments and feedback. This model measured the impact of the upskilling program on employee performance and business outcomes. The adoption of these models ensured a significant improvement in employee competencies resulting in a 25% increase in employee performance and a 15% growth in overall productivity. The process mitigated the resistance to change and encouraged a culture of continuous learning and innovation.

- **Supply Chain Optimisation.** The project team applied the Supply Chain Operations Reference (SCOR) Model to map the supply chain process and the Demand Forecasting (DF) model to predict future demand based on historical data and market trends. The SCOR model was used to evaluate an organisation's supply chain for effectiveness and efficiency of sales and operational planning by optimising key processes with the aim of mitigating stockouts and overstock situations. The DF model helped to optimise inventory levels and safeguard timely product availability. The use of these models reduced stockouts by 20% and overstock situations by 15%, providing greater profitability and customer satisfaction. The AI-driven solutions emanating from the use of these models improved the organisation's capacity to respond to market conditions and demand fluctuations, leading to better overall business responsiveness and resilience.

The above support the contention that implementing an AI solution is a lengthy and meticulous process requiring the use of various management models, each focused on a specific operational aspect. Furthermore, the Critical Success Factor approach can also be applied to manage the implementation of the defined strategic initiatives (refer to [Camilleri, 2024](#), pp. 299–326) thus ensuring that the defined strategic objectives are achieved. The AI solution enabled the order processing operations to be streamlined by replacing the traditional methods that were time consuming and inclined towards human error. The implemented solution enabled the AI system to process tasks such as sorting, prioritising, and handling orders automatically to expedite the entire operation and mitigate the risk of mistakes.

A significant benefit of the AI orders processing system was its capability to enhance accuracy and efficiency. AI solutions are able to identify and correct errors in orders, making sure that the correct information is entered into the system. This mitigates the requirement for manual control and verification, thus streamlining the process, which lowers the likelihood of errors and delays. The resultant AI system also improved customer experience because the time needed to process orders was reduced, resulting in faster delivery times. Additionally, the AI system continuously kept customers informed throughout the process, reducing anxiety, and strengthening their satisfaction, which in turn may have positively impacted their loyalty, retention, and upbeat reviews.

Typically, AI solutions work on 24/7 basis, thus reducing the need for human intervention, which releases employees to focus on higher-value tasks. Costs associated with manual processing are definitely reduced to a minimum with overall productivity increasing. Furthermore, AI solutions also have the potential to reduce the need for office space and other resources, resulting in more cost savings. The AI system features significant insights and analytics regarding the order processing method, permitting the organisation to determine patterns, trends, and areas for improvement. With such information organisations are able to optimise their processes, mitigate

bottlenecks, and enhance overall efficiency that can result in greater profits, lesser costs, and improved customer satisfaction.

An important feature of an AI solution is that it can be customised to meet the needs of individual businesses, which may be focused for the specific requirements of the organisation, leading to even higher efficiency and accuracy. The provided AI solution make use of several AI algorithms that support the automation process and optimise operations. These AI algorithms have a different and specific objective that build a robust AI system for an order processing system. As one may appreciate, there is continuous progress being made in machine learning and data science. Thus, algorithms continue to evolve, making AI systems more effective. This example has illustrated that AI has a multifaceted role in revolutionising order processing by removing inefficiencies, delivering actionable insights through data analytics, enhancing customer and supplier relationships, and keeps the organisation competitive, while ensuring regulatory compliance and security.

9.9.2 Case Study 2: AI-Driven Inventory Management

[Tang.\(2025b\)](#) describes how a mid-sized company specialising in consumer electronics with a global customer base, encountered serious concerns regarding its inventory management. These concerns were mostly related to inappropriate demand forecasting, resulting in significant stockouts and overstocking across multiple product categories that negatively impacted sales and costs. He reports that by incorporating advanced AI demand forecasting models, the company was able to realise a 25% decrease in inventory related concerns and an 18% reduction in holding costs, underlining the effectiveness of a systematic approach to AI in Operational Excellence. It has transpired that the company's existing AI system is lacking features for predicting demand patterns, which is viewed as the reason for the significant loss in sales and an increase in holding costs.

Root cause analysis was conducted using several methods, such as 5-Whys, process mapping and Ishikawa diagrams amongst others (refer to

[Camilleri, 2024](#), pp. 78–99). The 5-Whys approach is a simple way to identify the root cause of a problem. Basically, when a problem has been defined, ask: “Why did this happen?” Then, without stopping at the answer to this first question, ask: “Why?” again and again until the root cause is attained. The Ishikawa diagram (or fishbone diagram), which is also known as a cause-and-effect diagram, helps management identify the many possible causes for a problem by sorting ideas into useful categories and is especially useful in structuring brainstorming sessions. The outcome of this analysis suggested that the problems originated from an absence of accurate demand forecasting, poor inventory allocation, and unsatisfactory integration of AI with the existing supply chain management systems. These problem areas were the likely contributors to inventory mismanagement that may have also undermined the company’s operational efficiency and profitability.

Management adopted a 5-phase consulting process to acquire a systematic and data-driven methodology for improving the AI facilities for inventory management. This ensured a complete analysis, design, and implementation of an upgraded system that included greater forecasting accuracy, improved inventory levels, and lower costs. The consulting process consisted of the following five phases:

- **Diagnostic Evaluation:** Assessing the current inventory systems, data quality, and AI technologies to ascertain the existing capabilities and the evident data gaps. This included data collection, stakeholder interviews, and process mapping to identify areas for AI integration.
- **AI Capability Building:** Increasing AI competencies based on the diagnostic evaluation by identifying the AI tools and technologies that best fit the company’s needs and taking advantage of data analytics for better forecasting. This phase involved choosing appropriate AI solutions and devising pilot projects.
- **Developing a Demand Forecasting Model:** This task involved developing a cutting-edge AI model to predict sales trends. This phase incorporated determining how to utilise historical sales data to

forecast demand and deciding on the external factors to be included in the model. The aim was to develop and test predictive algorithms.

- **Inventory Optimisation:** This involved integrating the upgraded AI models with the inventory management systems. The main issues included ascertaining how AI-driven insights could lead to optimal stock levels and identifying the reordering strategies to be adopted. This mainly involved the creation of dashboards and decision-support tools for inventory planning.
- **Implementation and Continuous Improvement:** This entailed the implementation of the AI-enhanced inventory system across the company. This included securing agreement on the implementation strategy and KPIs to be adopted for the individual departments for ensuring success. This phase embraced training, monitoring, and iterative fine-tuning of the AI applications.

Several implementation challenges were encountered by the company that required careful consideration. For instance, compatibility of new AI technologies with the company's existing IT infrastructure; seamless integration to avoid disruptions in supply chain operations; having an inclusive change management strategy to ensure that stakeholders understood and accepted the new AI-driven processes; and sustainable investment for continuing AI training and development to maintain the system's effectiveness and adaptability to changing market dynamics.

[Tang.\(2025b\)](#) claims that the company anticipated a 20-30% reduction in stockouts and overstock situations, a more responsive reaction to market demands, and an overall increase in customer satisfaction due to better product availability once the new AI systems were in place. Additionally, the company envisaged a decrease in holding costs by at least 15%, contributing directly to the bottom line. The main implementation challenges encountered were data privacy concerns since AI systems required access to vast amounts of consumer and transactional data; and data accuracy and consistency across global operations to ensure successful AI-driven forecasting.

KPIs were a critical aspect because they provided measurable spot checks to verify the alignment of operational activities with the defined strategic goals. This guaranteed that deployment is not just activity-driven, but results-oriented. KPIs act as an early warning system since they indicate progress or deviation from established targets, enabling responsive decision-making and course correction if necessary. Important KPIs included: (a) demand forecast accuracy that measured the percentage of accuracy between forecasted and actual sales, signifying the effectiveness of the AI models; (b) inventory turnover ratio, which evaluated how often inventory is traded and replaced over a period, reflecting the optimisation of stock levels; and (c) stockout and overstock rate that tracks the incidence of these events, with improvements signifying better demand prediction and inventory management.

As explained in the previous case study, AI is a tool that requires a strategic approach applying several other management tools in the system development life cycle. [Tang_\(2025b\)](#) argues that it is vital for management to promote data-driven decision-making and a continuous learning culture to exploit AI's capabilities. Additionally, the old aphorism in data processing: garbage-in, garbage-out is imperative. In other words, the AI systems are only as good as the data they process, ensuring data quality and integrity is paramount for reliable outcomes ([Tang, 2025b](#)). Furthermore, AI technology is continuously evolving, what is current today is quickly outdated. Therefore, keeping up-to-date with the latest developments and be willing to experiment with new approaches is important for gaining a competitive edge in a turbulent economic environment. Furthermore, a company having gained a competitive edge is a short-lived reality because other competitors are also striving to gain a competitive edge. One must recognise that gaining a competition edge is a continuous struggle for business survival.

[Tang_\(2025b\)](#) identifies the following key results: (a) 25% lower stockouts and overstock situations through the integration of advanced AI-driven demand forecasting models; (b) 18% reduced stock holding costs that directly improved in the company's bottom line; (c) 35% increase in

the accuracy of forecasting demand, leading to more efficient inventory management; (d) 20% improvement in the inventory turnover ratio, signifying a more effective and responsive approach to market demands; and (e) 40% increase in staff proficiency with the new system by the deployment of a comprehensive AI training program.

It must also be recognised that several AI algorithms are employed that support the inventory automation process and optimise operations. For instance, supervised learning algorithms, such as linear regression are often applied for forecasting to assist in predicting demand based on one or more predictor variables. Decision trees are beneficial for classifying tasks that identify the categories of products, which are most likely to be purchased by specific customer segments. Additionally, unsupervised learning algorithms like principal component analysis (PCA) is frequently applied for dimensionality reduction when dealing with large sets of correlated variables, such as product features. The above are only a few AI algorithms that may be exploited when considering AI solutions.

9.10 CONCLUSION

This chapter provided a detailed description regarding Enterprise AI. It illustrated that Enterprise AI is a subset of AI that is specifically focused on driving business value in a large-scale organisation. It also supported the contention that businesses that harness the advantages emanating from the power of Enterprise AI applications will most likely experience significant advances in their value chain operations. Before specifically focusing on Enterprise AI, this chapter examined how AI is generally transforming business enterprises and how the concept of Enterprise AI came into fruition by integrating the various available AI tools into a seamless system to exploit the vast volumes of data that organisations generate during their operations.

In recent years, Enterprise AI has been shown to be important in maintaining a competitive advantage. In fact, organisations in a broad range of industries are uncovering new ways of applying Enterprise AI

applications, including enhancing customer loyalty; providing greater individualised customer experience; inventory optimisation; fraud detection; and anti-money laundering amongst many others. Enterprise AI is also being used in specific industry segments such as predictive maintenance regarding oil and gas production equipment, and energy production and distribution systems. These are just a few worthwhile examples.

The chapter has illustrated the tremendous advantages of applying Enterprise AI ranging from cost mitigation, proving better business models, and particularly keeping ahead of competition. It was illustrated that Enterprise AI is a significant tool for ensuring the business survival of an organisation. However, despite the many advantages, Enterprise AI has its potential risks such as algorithmic bias, possible data security issues, and eroding basic human skills. Hence, the importance for organisations to compile clear ethical guidelines about developing human talent, designing reliable AI systems, having data privacy standards, continuous monitoring, and having suitable security systems. Two important issues that require special attention are to maintain the users' trust in the AI application systems and mitigating the impact of AI on the workforce by introducing reskilling programmes so that employees remain an integral part of the evolving work environment.

The Chapter also provided a suitable framework for the successful deployment of Enterprise AI. Each step in this implementation framework was described in detail to provide the reader with a viable and practical deployment approach that may be adapted and customised to their individual organisation. Finally, the chapter itemised and explained several use cases to give the reader an idea how Enterprise AI can benefit their organisation. Additionally, a brief explanation was provided regarding the potential of Enterprise AI, which will be further described in more detail later in a separate chapter.

REFERENCES

- Camilleri, Emanuel. (2019). *Information Systems Strategic Planning for Public Service Delivery in the Digital Era*. IGI Global. [↵](#)
- Camilleri, Emanuel. (2024). *Key Performance Indicators: The Complete Guide to KPIs for Business Success*. Routledge. [↵](#)
- Tang, David. (2025a). *Enterprise-Wide Artificial Intelligence Integration Project for Retail Organization*. Flevy Management Insights. <https://flevy.com/topic/artificial-intelligence/case-artificial-intelligence-optimization-e-commerce-efficiency-nonstore-retailers#section2>. [↵](#)
- Tang, David. (2025b). *AI-Driven Inventory Management for Ecommerce*. Flevy Management Insights. <https://flevy.com/topic/artificial-intelligence/case-ai-driven-inventory-management-ecommerce>. [↵](#)

Chapter 10

Public Sector AI

DOI: [10.1201/9781003653950-13](https://doi.org/10.1201/9781003653950-13)

“When it is obvious that the goals cannot be reached, don’t adjust the goals; adjust the action steps.”

Confucius (Kong Qiu)

Chinese philosopher of the Spring and Autumn period (551-479 BCE)

According to World Bank ([2021](#)), disruptive technologies like AI have the potential to transform governments because AI can deliver a personalised service delivery experience, enhance efficiency of back-end processes, strengthen policy compliance and facilitate fraud identification within government disbursements. If used appropriately and exploited together with human management and decision-making, governments can also use AI to address development challenges ([World Bank, 2021](#)). Similar to Enterprise AI, Public Sector AI is viewed as helping organisations to perform certain tasks that have in the past typically required human intelligence and tasks that involved human vision, speech, language, knowledge, and exploration of data.

As governments continually confront the economic challenges arising from a turbulent global economy, and recently the challenges arising from

worldwide pandemics such as Covid-19, governments are under tremendous fiscal stress exacerbated by mobility limitations on citizens and civil servants when epidemics arise. This environment is acting as the main thrust for governments to discover better technologically enabled solutions to enable them to continue and enhance their services at a reduced cost and mitigate corruption. The World Bank contends that a common pattern of government's AI use cases is focusing on citizen engagement; compliance and risk management; fraud and anti-corruption; business process automation; service delivery; asset management and analytics for decision-making; and policy design.

Furthermore, the World Bank argues that with careful deployment, Public Sector AI applications can help governments to deliver their services faster and more personalised to the needs of beneficiaries and citizens, and the public administration charged with delivering them. However, care is also required to mitigate the risks, such as managing inherent biases in data and algorithms that aggravate or create problems. Generally, governments view AI as an essential factor in the expansion of the knowledge economy and regard it as a major strategic resource to enhance their country's competitiveness and to further increase economic growth. A 2023 report by McKinsey & Company predicted that AI is likely to contribute to \$17.7 trillion annually in economic value. Unfortunately, the growth in the application of AI is mostly in developed countries with some important inroads being made in developing countries. There is a concern that a slower adoption of AI in various countries may lead to inequality between the rich and the poor nations.

10.1 INTRODUCTION

It must be recognised that the Public Sector in most countries is the biggest employer of human resources, and its services extends across the whole spectrum of society, including the business sector. It is also the principal user of non-human resources and provider of a wide-ranging number of services. Therefore, Public Sector AI has the possibility to greatly enhance

government operations and aid in providing the extensive needs of citizens by applying cutting-edge technology, ranging from traffic management to healthcare delivery to processing revenue and value added tax returns.

However, there are still many public entities that have not embraced this powerful technology. Some public sector officials are fully aware of the transformational impact of AI applications by exploiting their entity's data. Although public entities have an enormous wealth of data, in most cases this data is held in separate data silos that make it difficult to apply AI solutions. Hence, a primary challenge for public sector entities is having an effective data management process that generates a comprehensive central data repository, which supports a data sharing and data discovery mentality.

Moreover, public sector officials may not have suitable knowledge and expertise to make strategic decisions for acquiring the necessary AI tools. However, this challenge may be overcome by engaging private sector expertise through a lawful outsourcing process. One needs to recognise that public sector management officials have less freedom than their private counterparts because they are under constant public scrutiny and restrictive procurement regulations (including lengthy appeal procedures), apart from restricted fiscal allocation processes. Additionally, the risk and possible public condemnation through the media about ethical considerations increases further layers of complexity in initiating AI projects, particularly in developing countries. The outcome of this uncertain environment is that public sector management will be inclined to procrastinate procurement decisions and thus continually postpone the apparent risk. However, there is no doubt that AI has great possibilities to facilitate the government's effort to deliver services that are more responsive, efficient, and personalised.

10.2 AI AND THE PUBLIC SECTOR

[Berryhill et al. \(2019\)](#) contend that despite the lack of research, AI is a significant and growing issue in governments. They imply that AI can facilitate the governments' effort to achieve its purpose, namely, to elaborate and improve laws and policies, provide public goods and services

to citizens and residents, and to deliver and maintain the tools, resources, and structures needed for civil servants to perform their respective functions. They also identify several roles that governments have regarding AI. They view these roles as a way of transforming public sector processes, practices, policies, and services, rather than how they affect the broader economy. [Berryhill et al. \(2019\)](#) refer to the following functions that governments can perform regarding AI, with these functions in most cases being conducted concurrently:

- a. Government as a financier or direct investor to sustain financial support for the development and deployment of emerging technologies.
- b. Government as a smart buyer and co-developer through innovative procurement practices, or as a co-developer through public-private partnerships (PPP) and other forms of collaboration to build new or tailored solutions.
- c. Government as a regulator or rules maker can accelerate the innovation cycles of emerging digital technologies by having creative types of policy and regulatory instruments that also evaluate compliance and taking appropriate action for nonconformity.
- d. Government as a convener and standards-setter by bringing together stakeholders, such as citizens, residents, businesses, organisations, and academics to facilitate the achievement of their objectives and to understand multiple aspects of relevant issues.
- e. Government as a data steward by having a vast data repository to drive AI-based technologies.
- f. Government as a user and service provider by applying AI technologies, both for the public and for back-office functions.

The application of AI in the Public Sector differs from Enterprise AI, nevertheless there are some commonalities. For instance, both Public Sector AI and Enterprise AI seek to provide a better customer service experience at a lower cost on a 24/7 basis from anywhere. However, Public Sector AI

is more focused on improving laws, and social and economic policies, including the development and implementation of national strategies in various disciplines, such as education, health, law enforcement, and transportation. [Berryhill et al. \(2019\)](#) present several recommendations that are being promoted by OECD for policy makers concerning national policies and international co-operation for trustworthy AI. Admittedly, these proposals go beyond public sector innovation and transformation, but they are still relevant to assist government policy makers to explore possible opportunities for deploying AI in the public sector. These recommendations include:

- Investing in AI research and development to support the use of AI in the public sector with a view to transforming and innovating the way they design and implement policies and services.
- Better data management to mitigate the data silos syndrome that is so common in government and serve as a driver for AI deployment.
- Promoting a digital ecosystem for AI by building connection points, knowledge-sharing mechanisms and ecosystems within the public sector and other stakeholders to sustain AI experimentation in government.
- Shaping an enabling policy environment for AI by documenting how AI is used to develop AI strategies and the extent to which these strategies specifically provide for public sector transformation.
- Building human capacity and preparing for a labour market transformation to build AI capacity among civil servants.
- Establishing international co-operation for trustworthy AI by exploiting continuous international collaboration opportunities.

These recommendations will facilitate the transformation of public administrations by the application of AI trustworthy solutions.

10.3 RECOGNISING DIFFERENT AI APPROACHES AND IMPORTANCE OF DATA MANAGEMENT

AI consists of several techniques, such as Machine Learning, Neural Networks, and Deep Learning, with each technique being applied to address specific issues. Hence, the application of AI takes different approaches. The introductory chapters of this book provided clear definitions and objectives for the different aspects of AI. However, it is important to note that particularly in the Public Sector, there is a general lack of awareness about how AI works and its inherent capabilities and risks.

All too often AI is viewed as a “black box” with data going in at one end and a solution being provided at the other. Most public service officers are simply not aware of what takes place within the box and most of those who are aware have only rudimentary knowledge. One must start by recognising that AI is a disruptive field, which impacts society as a whole and drives innovation across sectors and industries. One must also recognise that the public sector provides products and services to a much wider spectrum of clients than the private sector. For instance, the private sector is segmented into marketing spheres, with each sphere being further segmented by having different companies and entities competing. The point is that no specific private enterprise or organisation caters for the whole spectrum of possible users or customers.

On the other hand, government (public sector) caters for everyone, from citizens, civil society and businesses. Therefore, for governmental policy makers and public servants in general, this poses several challenges about how to regulate, support, use, and maximise possible value from AI while mitigating its possible negative risks. This suggests that decision makers in all sectors, particularly the public sector, are required to recognise the benefits and drawbacks of AI, and to acknowledge that easier solutions may exist, which deal with a specific concern. Hence, AI should be applied when it has a distinct advantage over other possible solutions. It is therefore important that public sector decision-makers are knowledgeable about AI because its wide potential applicability across different cultures and environments will likely impact how it is implemented and applied. Thus, having basic technical knowledge about AI can assist public servants to

generally determine possible AI tools to apply for a specific concern and help them negotiate better terms with suppliers.

The basis for all AI applications is the quality of the data. The chapter addressing the knowledge economy provided detailed information regarding the transformation of data to information and from information to knowledge. It was also illustrated that this transformation process contributed to the success of AI projects. The public sector has a tremendous amount of data. However, most of the governmental data is stored in data silos and not shared across departments within a ministry, let alone shared across ministries. Hence, data quality in the public sector is a major concern that needs to be addressed before undertaking major AI projects. This is not to say that AI applications cannot be implemented. However, the deployment of AI will likely be restricted to the most basic uses and certainly not for helping with building major strategies or for defining wide-ranging policies. This is likely to lead to deploying AI in a fragmented approach that may not maximise its potential. This also means that better data management systems are required to widen the range of AI applications that may be suitably deployed.

Therefore, before an AI application can be successfully deployed on a large scale, government users must ensure that a robust data management system is in place to ensure that a data repository is generated which is accurate, reliable, and relevant. This is especially the case for applications that use machine learning, where an algorithm continuously learns from the available data. Hence, most governments do not lack data and information, what is normally lacking are adequate data management systems that provide a comprehensive data repository.

10.4 MAJOR CHALLENGES CONFRONTING PUBLIC SECTOR AI

Research by the World Bank (related to its clients) highlights that the main challenges confronting the public sector regarding the deployment of AI are lack of AI awareness by public sector management; inadequate policies,

legislation and incentives for AI; insufficient digital data and other digital infrastructure; and lack of digital talent and skill ([World Bank, 2021](#)). This supports the contention in the previous section that public sector management need to be knowledgeable about the potential of AI, have a basic understanding of digital technologies, and recognise the importance of the availability and quality of data. Furthermore, low access to digital skills was also revealed as being a major barrier to the deployment of AI in the public sector.

As at mid-2024, several nations have adopted some basic AI policies and legislation, namely USA, European Union, China, Canada and Australia. However, the USA has no comprehensive AI federal regulation framework specifically but has established several sector-specific AI-related agencies and organisations that address some of the challenges arising from the evolution of AI. Furthermore, OECD and the UN are actively involved in setting and outlining global guidelines on AI regulation. This illustrates that most countries do not have a legal framework that is so critical for deploying trustworthy AI applications safely. Moreover, these countries also lack digital processes, data collection abilities, technical skills and supporting digital infrastructure to currently benefit from the application of AI. Hence, as a priority they must focus on adopting an appropriate policy and legal framework before considering AI as a possible solution. Notwithstanding this lack of policy and legal framework, and the apparent digital divide across various countries, the World Bank have noted several AI use cases materialising or are at the piloting AI stage at their client countries. Several use cases will be discussed later in a separate section.

Like Enterprise AI, Public Sector AI is likely to impact the demand for workers in terms of an increased demand for higher skilled workers and a decreasing need for the lower skilled workers in the medium to long term. Hence, governments are likely to encounter a rising risk in job losses in both the private and public sectors, which would need to be adequately managed by the introduction of labour reskilling programmes and continuously upgrading the post-secondary school educational curriculum to reflect the changing technological environment. Additionally, the impact

on developing and underdeveloped nations will likely be higher depending on their individual development status.

Several countries have modified and strengthened their legislative framework for data protection and privacy to complement and encompass AI related activities. These changes are viewed as permitting the interoperability and accessibility of government-held data and its governance, particularly for data use and reuse utilising machine learning and AI specifically to mitigate bias and inequity. Since, the legislative framework for data protection and privacy have been commonly legislated across many counties, it provides a robust basis aimed at catering for the AI ecosystem, particularly in developing countries.

As highlighted in the previous section, the lack of adequate data quality is a serious barrier for implementing AI applications successfully within the public sector. Furthermore, unnecessary government control, poor cybersecurity, and inadequate personal privacy may also pose a critical challenge. Resolving the data quality and unnecessary government control issues are difficult to resolve since they require extensive effort to ensure timely, accurate and complete data, and necessitate major reviews of internal processes. Poor cybersecurity and inadequate personal privacy are not as difficult to address since exceptional cybersecurity applications are widely available (at a cost) and personal privacy can be strengthened through legislative acts of government.

The [World Bank \(2021\)](#) maintains that governments should allocate more investment funding in digital skills so that they may formulate suitable policies and risk management strategies for their specific circumstances. It must be recognised that escalating the digital skills of a workforce on a national basis requires long-term commitment on the part of governments, both in terms of human and financial resources. The World Bank are also of the opinion that having a policy for the ethical application of AI as being of paramount importance to manage the adverse consequences of using AI in the public sector. In fact, the next chapter will address in detail the legal and ethical aspects of AI use. Having said this, it is important to emphasise that there is an urgent need for governments to

define and implement policies for addressing the issues of governance and oversight in the use of AI solutions.

Each country has its individual priorities and available resources regarding the deployment of AI and other technologies. Considering this, the World Bank suggests that countries adopt a holistic methodology for the deployment of AI rather than having a fragmented approach. This necessitates a holistic national strategy to build and maintain the crucial infrastructure, and the standardisation of systems, data, and governance frameworks to coordinate its implementation and continuous monitoring. As stated earlier in this section, a serious concern for governments in tackling AI applications is the data quality, including having an environment that supports data silos. Systems that are based on a data-silo approach can never achieve interoperability to truly provide effective technological solutions that are so necessary for increasing the living standards of citizens of the respective countries.

10.5 EMERGING GLOBAL PUBLIC SECTOR AI LANDSCAPE

The previous sections have demonstrated that AI is quickly changing many facets of people's everyday lives. The public sector is contributing to this change by having governments recognising the benefits of AI and upgrading AI as a national priority. This is reflected by increased AI investments and compiling regulatory regimens to cater for the constant growth of AI in the wider social and business community.

A government is the biggest entity in any nation and is therefore able to leverage the tremendous potential of AI to transform the public sector into an effective organisation. This may be achieved by reviewing and changing the processes for defining and implementing policies for delivering services to its constituents, civil society, and commercial enterprises. It is this context that one needs to understand the emerging global public sector AI landscape. The innovative application of AI and its transformational impact on the public sector is critical for governments to confront the continuous growth of the demands and expectations from their constituents, civil

society, and commercial enterprises. Moreover, as nations climb the development scale, these demands and expectations are constantly escalating in complexity.

In this scenario, interoperability of public services becomes essential, where standards, protocols, technologies, and mechanisms that allow data to flow between diverse systems require very minimal human intervention. This allows diverse systems to talk to each other and share information in real time. As highlighted in the previous section, this also depends on having highly integrated holistic systems that are supported by high quality data (i.e., timely, accurate, reliable, consistent, and relevant for its intended use) across government. This permits AI to be integrated into the complete policymaking and service design cycle. Hence, as AI tools such as machine learning and deep learning develop, more administrative and process-driven processes will be automated, enhancing greatly public sector efficiency and releasing public service employees to focus on other essential tasks.

Additionally, such developments will allow senior government officials to enhance their understanding of the various circumstances and help them to make better decisions by knowing and predicting the demands of their diverse customers (i.e., constituents, civil society and commercial enterprises). This will facilitate government to make more precise and equitable decisions. Many national governments are in the process or have defined their AI strategies with the focus on instigating economic development through research and R&D funding. However, the most essential and fundamental principle that underline these governmental AI strategies is the successful development of trustworthy AI.

OECD have proposed that international collaboration as a key principle for the successful development of trustworthy AI ([Berryhill et al., 2019](#)). According to [Berryhill et al. \(2019\)](#), a commitment to such collaboration is reflected in several declarations, namely:

- G20 AI Principles that were based directly from the OECD Recommendation (G20 Ministerial Statement on Trade and Digital Economy: www.mofa.go.jp/files/000486596.pdf).

- In 2018, all EU member countries signed the Declaration of Cooperation on AI committing cooperation to boost European AI capacity and adoption, address socio-economic challenges and ethics, and ensure an adequate legal and ethical framework ([Berryhill et al., 2019](#)).
- The EU member states also committed to making AI available and beneficial to public administrations, to sharing best practices in procuring and using AI in government, to implementing open data practices, and to maximise the impact of investments at EU and national levels, encourage synergies and cooperation across the EU ([Berryhill et al., 2019](#)).
- Ten governments from the Nordic-Baltic Region signed the Declaration on AI pledging to improve skills development and access to data, and to develop ethical guidelines ([Berryhill et al., 2019](#)).

The above demonstrates that most national government AI strategies contain obligations for ensuring secure designed and implemented AI systems that are ethical and trustworthy. Additionally, [Berryhill et al. \(2019\)](#) note that most strategies also promote the strengthening of the national AI talent through educational programmes and training and for applying AI to innovate and transform the public sector.

Research suggests that many public sector administrations already apply AI in a range of fields, such as smart recognition, supporting various assessment processes, environmental protection, law enforcement, and healthcare. However, most AI projects appear to be based on the initiative of individuals (not necessarily from a senior management level) who have an adequate degree of AI awareness to pursue such projects.

This type of environment does not support the interoperability of systems, processes, services and organisations, despite the fact interoperability may be recognised as a critical factor by the individuals concerned. This is viewed as a serious weakness by public administrations, and most have taken appropriate action to remedy the situation by consolidating the AI strategic process, emphasising the importance of the

interoperability aspect for supporting sustainable AI-based. The main features emerging from the global public sector AI Landscape include the following applications (Tangi et. al., 2023):

- Experimentation with AI in government by identifying certain AI projects to discover the benefits and potential to be exploited from AI. As highlighted above AI cannot depend on individual initiatives but must be an organisational objective by increasing the general awareness of AI throughout public administration.
- Collaboration across sectors. AI can be applied as a tool to improve interoperability between public organisations and private entities through public-private partnerships and facilitated through innovation hubs and labs.
- Promoting cross-government focus groups, networks and communities to support suitable systems methodologies. Applying AI-based applications that support interoperability poses significant challenges for the public sector that include scarcity of skills and expertise, cross-organisational interoperability, complexity, system integration concerns, compliance with regulatory regimes, financial constraints, and data quality.
- Promoting robust data management process. AI solutions are dependent on the availability and quality of data, but they can also contribute to increase data quality. However, as an initial step, public administrations must implement a comprehensive data management process to have a quality data baseline.
- Legal, regulatory and ethical compliance. This is a critical factor that requires a collaborative approach between public sector organisations, legal experts, technology providers and other relevant stakeholders to assist organisations navigate the legal landscape associated with implementing AI.
- Automation of everyday government processes to improve efficiency. Public sector employees conduct many different tasks often of a routine nature leaving them insufficient time to focus on finding ways

to improve public services. AI provides an opportunity to conduct these routine tasks and release public sector employees to perform more worthwhile functions.

- Applying AI to assist in governmental decision-making. AI is being used by several governments to help public administrators to evaluate various policies by predicting their potential impact on target users. AI is also being used in designing and implementing preventative measures such as flood barriers and early warning systems. AI-driven solutions have been found to be critical for predicting possible disasters and enabling authorities to take timely action that can significantly mitigate the impact of such events.
- Strategic management, harnessing government data for designing public and private services. Public administrators are applying AI to help them define their operational strategies by harnessing their data to design customised services and predicting the demand for new possible services through services trend analysis. Furthermore, non-sensitive personal data is being made available to the private sector as a key input to Enterprise AI.
- Provision of guidance on the transparent and ethical use of public sector AI. Having legal consistency among various national and international systems creates difficulties to develop common protocols for data exchange. Often legal concerns occur when public administrations are required to interchange data between them.
- Enhancement of public service capacity through training, recruitment, tools and funding. A critical challenge for public sector AI is the need to acquire the essential skills and capabilities to develop, implement and maintain AI-based solutions. However, governments recognise that upskilling civil servants is vital for ensuring that government entities operate effectively. It is also widely accepted that training and various types of development programs help public servants acquire exposure to new techniques and technologies, build their leadership skills, and remain current on evolving policies and regulations.

Being aware of the emerging global Public Sector AI landscape will help individual public service administrations to improve their level of readiness for the initiation of AI applications or to fine tune their defined AI strategies.

10.6 IMPLEMENTING PUBLIC SECTOR AI

There is no doubt that Public Sector AI has significant potential in all types of public sector fields. It is also evident that implementing Public Sector AI will encounter many challenges and will have considerable implications that government leaders and public servants need to consider when determining whether AI can contribute to the achievement of their missions and specifically assist them in addressing the various problems that frequently confront them.

[Figure 9.3](#) of the previous chapter provided an effective and comprehensive user-friendly approach for implementing Enterprise AI. This framework is also applicable to the deployment of Public Sector AI at every government level (i.e., national, state, or municipal). Therefore, the focus of this section will be on the challenges and practical aspects that are related to the public sector. An important phase in Public Sector AI is building support and having management champions that enthusiastically carry the torch for the deployment of AI. However, building support for AI within the public sector is contingent on identifying a clear vision and defining the practical implications for use of AI in the public sector to better serve citizens, civil society and businesses.

Governments need to encourage its senior management in the public service to be receptive to AI by providing the necessary freedom for flexibility and experimentation to expedite speedy learning. Once this foundation is established, public service management will be required to formulate a policy for determining whether AI is the best solution for a specific problem and provide the necessary instruments for identifying and devoting awareness to such problems. This may possibly be achieved by

understanding the needs of their diverse customers and identifying high impact use cases to determine areas of opportunity.

However, it is important that such policies are implemented after building and setting up of cross-functional team(s). Several cross-functional teams may be required to cater for each specific government ministry, since it would not be practical to have one national cross-functional team. Each cross-functional team will consist of multi-disciplinary and diverse members to assist in identifying AI opportunities and to promote the development and deployment of AI assignments that are ethical and effective. The cross-functional team must ensure that AI is developed (i.e., designed and deployed) according to a trustworthy, fair, and accountable methodology that recognises trade-offs, mitigates risk and bias, and guarantees an appropriate role for humans. Hence, during the AI development life cycle a persistent effort must be made to focus on users and individuals who may be impacted by the AI applications.

[Figure 9.3](#) highlights that building data fluency is a key implementation task, since it makes AI-driven innovation possible. Data is the fundamental basis for AI. Therefore, there is the need to assess the data preparedness status and develop a data strategy that allows governments to access robust, timely, accurate and complete data. This must be achieved through an approach that safeguards privacy and adheres to societal and ethical standards.

Implementing AI on a governmental scale requires a high degree of knowhow and expertise, which will likely consist of a consortium of public and private sector talent, and needed products, services, and infrastructure. This consortium will probably be initially inclined towards having a greater participation from the private sector, but as the internal capabilities within a public administration are developed there will be a greater reliance on the internal public service structure. However, such decisions are country and context-specific and are contingent on the explicit vision of the senior public servants at the time, including the budgeted allocations to ICT, specifically AI projects.

Being fully dependent on the private sector, particularly for AI projects, would be a strategic oversight. It is critical that governments through their public servants have a structure that supports at a minimum, a baseline level of data fluency and knowledge of data science, including related AI tools to avoid governing complications by having a situation where suppliers have complete power over the government. Therefore, the public service should define a human resource development plan to give public service employees the opportunity to develop relevant capabilities to support the AI initiatives and to ascertain the required future competencies that are likely from future public servants. This would form a robust and healthy basis for having a collaborate Public Sector AI deployment approach between the public and private sectors.

Governments should be aware that there will likely be various shifts regarding AI in the future. Hence, the public service must continuously improve their AI models and processes. They need to research and explore the changing AI landscape environment and take a proactive stand by predicting possibilities and take immediate preparatory action to keep a step ahead of developments (or at least not falling behind). This will ensure that government exploits and leverages its AI investment to maximise the inherent benefits of AI and mitigate its potential risks.

10.6.1 Practical Implications for Implementation Public Sector AI Strategies

Research has revealed that the application of AI by public administrations is growing, with several organisations using AI in their daily operations. Chatbots are being successfully used to assist citizens with typical income taxation and social services enquiries and are helping companies with income tax and value added tax management. This has resulted in a significant reduction in telephone and email exchanges. AI is also being used as a legal dictation system, which allows an automatic transcription of voice notes of legal text. This application is capable of transcribing correct legal terminology at a rate of over 150 words per minute. This application

reduces the tedious effort for public service employees and allows them to focus on more challenging tasks.

As stated previously, the literature review revealed that AI poses several challenges. These challenges include trustworthiness, cooperation, procurement, data quality, skills, and human resources. Most of these challenges correspond with Enterprise AI, except for procurement. Procurement appears to be a customary concern for the public sector and not just restricted to AI related acquisitions.

However, ensuring a trustworthy use of AI is viewed as a critical challenge for the public sector, driven by the natural concern about the impact of AI systems regarding the lack of accuracy and bias, and inadequate supervision that may impact the citizens' health, safety or their fundamental rights. Hence, an AI application needs to be vigorously and extensively tested to ensure that it performs as intended and does not risk harming citizens. Moreover, every effort must be made to mitigate the perception of untrustworthiness because resistance to using AI is very likely if untrustworthiness becomes an issue. Having adequate mechanisms available to public administrations ensuring the ethical use of AI would be helpful in this regard.

There is a vital need for a suitable collaboration agreement in developing and using AI between public administrations and with the private sector. As stated previously, competencies and resources involved for successfully developing and deploying AI systems require the collaboration of the public service, universities, and research institutions, and private sector entities, particularly with start-ups who may present fresh ideas. Cooperation may also involve employing a centralised approach for the provision of AI services and products with different authorities using them at their own rate depending on their individual level of preparedness.

Procurement has always been a concern for public administrations, but it becomes a lot worse for AI applications namely due to a lack of AI knowledgeable and competent resources, and its inherent technical complexity. The difficulty is not the public sector procurement procedures but detailing the technical specifications and describing the requirements in

terms of the quality and accuracy level of the output, the explainability and transparency of the algorithm, and the algorithm's non-discriminatory quality. Outsourcing AI technical specifications and related needs is not desirable due to the possibility of conflict of interest. Additionally, public administrations need innovative scalable solutions, prevent supplier lock-in and acquiring non-explainable solutions, and inhibit fraudulent behaviours by the supplier. These concerns confirm the argument that public sector employees need to attain expertise in AI technology through exposure to AI techniques, be well-informed in relevant AI regulatory regimes (if any), and initiating formal development programmes.

Another practical implication for implementing public sector AI strategies is the collection of huge volumes of high-quality data that are utilised to train the AI systems. The higher the data quality the less likelihood of having biased AI systems. [Figure 9.3](#) illustrates that assessing the data preparedness of the organisation and defining the data strategy precedes the definition of the AI development strategy. However, it comes after building and establishing the cross-functional team, since human resources are required to generate the data repository for training the AI models and algorithms. Other human resources with the appropriate skills are also required from the various business units to train and test the AI algorithms.

The above issues have focused on a technical and resource centrality of implementing public sector AI strategies. However, a most essential objective is citizen centrality, where AI is brought closer to the citizens, civil society, and businesses by encouraging smart communities or cities. All too often public sector AI systems are conveyed for negative or restrictive applications, such as fraud detection, surveillance and law enforcement. It is important that the public sector conveys positive implications, such as better customer experiences in relationship management and for the provision of services on 24/7 basis from everywhere.

Bringing AI closer to the citizen and businesses has its unique complexities, not in terms of the applications themselves (and the technical

support) but also ensuring a sufficient penetration rate by public sector customers (i.e., citizens, civil society and commercial entities). There is no point in having a very beneficial citizen centric AI application, but no one uses it. The main concern is related to digital literacy that enables public sector customers to access data and related AI applications, and to understand the outcomes presented by the AI application. This concern is somewhat greater for the older generation (e.g., pre-1970s).

It is important that public administrations actively engage with the public conveying an assurance of having trustworthy AI applications based on a robust governance framework. Public administration customers must be viewed as having a central role in the development of AI-based public services and applications from beginning to end, and not just as users. However, this requires robust and resilient partnerships and collaboration with different relevant stakeholders ranging from having reliable and user-friendly AI solutions supported by a comprehensive communication strategy to involve public sector customers in the development process and educating them on the use of the deployed AI applications. The development of reliable and user-friendly AI solutions can be based on a rapid application development (RAD) methodology by having a “toolbox” with solutions that can be replicated and supported by standard processes to ensure compliance with established regulations. This avoids a situation of reinventing the wheel syndrome.

10.7 PUBLIC SECTOR AI USE CASES

Many governments have recognised the tremendous potential that AI provides for a more effective public administration across the public sector. Furthermore, this recognition has been translated by governments through their definition of comprehensive long-term AI strategies to enhance how governments engage with and serve their constituents. However, research has shown that AI is not just at the strategic planning phase for many governments but has moved rapidly to the practical deployment phase with

the collaboration of educational institutions of higher learning (Universities and Technology Institutes) and private sector AI providers.

For instance, the Biden-Harris Administration, was dedicated to ensuring that the United States of America (USA) took a lead in safe, secure, and trustworthy AI innovation to harness the opportunities of AI while mitigating its risks, itemise over 700 applications that were deployed across federal government (<https://ai.gov/ai-use-cases/>). According to the Biden-Harris Administration, the USA federal government was leveraging AI to better serve the public across a wide array of use cases, such as healthcare, transportation, the environment, and benefits delivery, and establishing strong guardrails to ensure its use of AI keeps people safe and does not violate their rights.

[Table 10.1](#) provides an indication of several AI use cases that have been implemented in numerous countries. There is no doubt that AI usage provides important benefits related to back-office automation since it is viewed as being an indispensable enabling technology for government entities encountering continuous pressure to provide higher levels of service to a growing population with fewer human capital resources. A particular AI application uses historical data from application forms completed by millions of previous applicants to generate predictive scores to expedite social security benefits claims. While other public administration AI applications optimise customer service approvals and improve engagement, increasing the speed and quality of service delivery, transforming employee work experiences, and decreasing workloads. In various use cases, Chatbots are automating interactions by providing a detailed knowledge base for citizens with multilingual support and collect citizen feedback on a large scale for use in better public policy making.

TABLE 10.1 Public Sector AI Use Cases 

Industry	Use Case
Administration	Digitising paper documents and speed up claims processing and back-office automation.

Industry	Use Case
Benefits Administration	Optimise service approvals and improve customer engagement, increasing the speed and quality of service delivery, and transforming employee work experiences and decreasing workloads.
Benefits Administration	Using historical data from application forms completed by millions of previous applicants to generate predictive scores to expedite social security disability benefits claims.
Customer Relations	Chatbots automate interactions by providing a detailed knowledge base for citizens with multilingual support and collect citizen feedback on a large scale.
Enforcement	Enforcement of policies that result in better large-scale infrastructure monitoring to fight tax evasion and unlawful property changes.
Environment	Optimise street lights; reduce traffic congestion; automatically alert authorities in case of any incident; and optimise energy consumption.
Environment	AI-enabled remote sensors monitor water usage and detect leaks in the water supply network to conserve water and reduce waste.
Healthcare	Early disease detection and prevention, and clinical decision support.
Healthcare	Analyse vast volume of biomedical data including genomic, imaging, and clinical data to fast-track discovery of new ways to prevent, diagnose and treat diseases.
Healthcare	Identify patterns, effects, and mitigation options for issues related to public health and climate change.
Healthcare	Optimising medical records management by analysing historical records and performing robot-assisted surgery.

Industry	Use Case
Healthcare	Interpret the images to detect counterfeit pharmaceuticals.
Policy Making	Analytics on public data to identify emerging issues related to their regions and constituents.
Policy Making	Enhance public policy to improve citizens' quality of life by predicting insecurity risk for housing, food, addictions, and mental health.
Security	Monitoring abnormal network activity and entry points; identifying likely data vulnerabilities and boost access restrictions; enhancing accuracy of intrusion detection systems; simulating cyber-attacks in controlled conditions to identify system vulnerabilities; and automate system patching and security updates.
Trade & Commerce	Analyse product-specific trade, tariff & macroeconomic and governance data to identify potential export markets.
Transport	Analysis of vehicle traffic crash data to detect areas prone to accidents to employ additional road safety measures to prevent future mishaps.
Transport	Data-driven analyses to assist traffic flow, preventing accidents and congestion, monitoring logistics, and increasing road safety.
Urban planning and development	Analysing demographic data, traffic patterns, environmental factors and other data used in urban planning to make more informed decisions about infrastructure projects and zoning regulations.

AI is making headways in the enforcement of policies that result in better large-scale infrastructure monitoring to fight tax evasion and unlawful property changes. In human resources utilisation, AI is being deployed to gain a deeper insight into how employees engage with benefits

programs, enhancing the employers' ability to make more informed benefit investment decisions that optimise employee satisfaction and total benefits spend efficiency. Intelligent personalised recommendation engines are also being applied to bridge the gap between optimised offerings and broad employee awareness and education, leading to higher program engagement from employees.

AI is also being used by several governments to define environmental policies, such as gaining a deeper understanding of climate change causes and impacts. Predictive analytics, computer vision, machine learning, and simulation models are assisting in monitoring and predicting the impact of climate change on weather patterns, environmental health, and human emissions. This is enabling governments to define their respective policy positions regarding their global efforts to mitigate and adapt to climate change, and to create more resilient health systems. Other AI applications to enhance environmental conditions include the optimisation of streetlights; reducing traffic congestion; automatically alerting authorities of any traffic related incidents; and optimising energy consumption. Moreover, AI-enabled remote sensors are being used to monitor water consumption and detect leaks in the water supply network to safeguard water resources and reduce waste.

Healthcare has always been a priority area for AI applications. AI is being used to enhance the knowledge of the key health contributing factors using trend analysis of population risk factors combined with local policies and programs to give an insight of the social and environmental factors that most impact people's health. Additionally, AI is being applied to better predict the impact of proposed health policies by using machine learning to leverage health data and the known local risk factors to predict the impact of proposed policies and inform targeted legislation to achieve specific population health goals. The analysis of vast volumes of biomedical data including genomic, imaging, and clinical data is being utilised to fast-track discovery of new ways to prevent, diagnose and treat diseases. Moreover, natural language processing is making complex biomedical data more available to medical professionals to enable greater use of the ongoing

research and data being collected. Machine learning algorithms are also providing insights from immense data sets to uncover new disease trends and drugs to improve diagnostic and treatment capabilities.

On an individual level, AI applications are being used to track a patients' health in real-time by monitoring weight, height, blood glucose, stress levels, heart rate, and other indicators and feeding this information to AI healthcare systems, which can inform doctors of likely risks. Other applications include, optimising medical records management by analysing historical records, performing robot-assisted surgery, and interpreting medicinal label images to detect counterfeit pharmaceuticals.

AI is playing an important role in the governments' policy making process by identify emerging issues related to their districts and constituents. Speedier detection of changing societal trends allows governments to alter policies accordingly. This enhanced policy planning efficiency and effectiveness is making it easier for public administrations to understand expectations and target major problems. The improvement in policy making decisions is positively impacting the constituents' quality of life by predicting insecurity risk for housing, food, addictions, and mental health.

AI is also being used to strengthen cyber-security by monitoring abnormal network activity and entry points; identifying likely data vulnerabilities and boost access restrictions; enhancing accuracy of intrusion detection systems; simulating cyber-attacks in controlled conditions to identify system vulnerabilities; and automate system patching and security updates. Cyber AI solutions are assisting governments and enterprises to manage cutting-edge cyber workflows by incorporating network behavioural modelling, malware detection, and credential misuse detection. A powerful feature of these security solutions is detecting domain-generated algorithms (DGA). A DGA algorithm generates large numbers of new domain names that cybercriminals and botnet operators use to change the target domain for launching malware attacks. Therefore, this feature determines whether a domain is likely to have been generated by an

algorithm, rather than a person, by reverse-engineering and analysing other frequently used techniques found in DGAs.

Some countries have implemented AI to ascertain possible export markets by analysing current trade patterns. By specifying a particular class of products they produce and the markets they currently export to, this application applies a machine learning algorithm to analyse product-specific trade and tariff data, economy-level macroeconomic and governance data to detect and compare markets the user should consider. AI is being successfully used by governments to considerably mitigate road safety issues. Currently, 1.19 million people annually are killed on roadways around the world, with road traffic crashes being the leading cause of death for individuals aged 5 and 29. The AI application using computer vision, object detection, drone tracking, and camera-based traffic systems, identifies areas susceptible to accidents and recommend additional road safety measures to prevent future mishaps. Additionally, data-driven analyses using machine learning algorithms assist traffic flow, preventing accidents and congestion, monitoring logistics, and increasing road safety.

AI applications are being used to assist town planners recognise and respond to emerging trends in demographic data, such as population growth in particular districts and demographic shifts (such as age or income) that are likely to impact demand for government services. Urban planning and development are being enhanced through AI applications by analysing environmental factors, demographic data, traffic patterns, and other data to make more informed decisions about infrastructure projects and zoning regulations.

As illustrated above, AI complements human expertise by vastly improving the way that public administrators conduct their tasks. There is no doubt that AI applications aid governments to move from low-value to high-value efforts and focus more on their core responsibilities. This is achieved by mitigating or removing tedious tasks and uncovering new insights from the collected data. Information overload is a concern for public servants; however, AI is helping them to make sense of the vast and

increasing amount of available data giving them new insights and generating forecasts to help them make better policy decisions.

10.8 PUBLIC SECTOR AI IMPLICATIONS

This chapter has illustrated that Public Sector AI will likely have several implications regarding how public administrations will be managed. It must be recognised that the development and deployment of AI in the public sector involves different stakeholders. However, a fundamental issue insisted by all stakeholders is having trustworthy AI. This implies that trustworthy AI applications for government entails the involvement of multiple collaborators within and outside public administrations. From an internal aspect, the participation of policy specialists, legal and technical units, and competent decision-makers are essential. On the other hand, from an external aspect, the contribution of private sector organisations, AI specialists, and civil society are needed to offer a multidisciplinary input for defining trustworthy AI-powered innovative public services.

One cannot be complacent about having trustworthiness as a primary objective of AI use cases since there are Public Sector AI applications that might be classified as high-risk. Government entities cannot risk allowing the implementation of non-transparent and biased AI applications, which might result in inequitable decisions, since government's own legitimacy can be at risk. Therefore, it is important that governments implement AI applications according to robust and secure ethical frameworks, which are founded on suitable legislative and governance regimes.

Public Sector AI specialists suggest that public administrators should first conduct pilot AI projects and increasing scale at a later stage. However, public administrators should always deploy AI (even at the experimental stage) according to a strategic AI roadmap. Reference is made once again to the framework for deploying an Enterprise AI system at [Figure 9.3](#). It must be acknowledged that AI projects in both the public and private sectors need some flexibility for experimentation. This will enable the respective practitioners to learn from their mistakes and successes to increase the

likelihood of overall success at the scale-up stage. Nevertheless, any AI development activity should always be within a robust governance and ethical framework, having reusable modules, and driven by the appropriate data quality to suit the specific AI purpose.

Concurrently, with the AI pilot phase, public administrations should define and implement an upskilling training programme for public sector employees. The intensity of the training programme for these employees is contingent on the anticipated role they will play in the AI process. In other words, some will need to have a high level of AI technical skills, while many others might only need a sound knowledge of AI awareness. The AI revolution is here to stay and therefore public administrations need to be well prepared. This preparedness status is like other transformational developments that public administrations have confronted in the past, whether through traditional computerisation or administrative approaches, such as the implementation of accrual accounting processes from the traditional cash accounting systems. Every major public administration reform in the past required trustworthy solutions and the general upskilling of the public servant employees; AI is no different. However, AI is more sensitive to ethical issues due to its extraordinary self-learning capabilities through machine or deep learning methods. Therefore, public servants need to be well prepared to ensure the correct interpretation and oversight of the AI application output that drive future public services.

Interoperability has always been an issue for ICT development. ICT made it possible to link organisations and achieve cross-organisational collaboration through efficient data access routines and responsive data transmission on a 24/7 basis from anywhere. Additionally, the Year-2000 millennium bug had highlighted that those organisations had to concern themselves not only with their own critical supply chain but the critical supply chains of their suppliers. Interoperability refers to the degree to which a software system, devices and applications can connect and communicate with other entities in a coordinated manner without effort from the end user.

While the concept of interoperability has been widely accepted by the private sector, it has proven to be a serious challenge and concern for governments. However, the growth of public services digitalisation, the increased awareness of big data, and the realisation of new technologies have created a suitable environment for governments to seek and achieve interoperability. This does not mean that the challenges have suddenly disappeared. What it means is the realisation by public administrations of the tremendous potential that interoperability may bring to effective governmental operations. Governments are realising that the envisaged challenges are now mitigated and manageable by adopting AI-based solutions to support the interoperability concept.

Interoperability has made it possible to implement AI applications based on smart recognition or to support assessment processes in many public administration areas, such as public services, health and environmental protection, amongst others. Public Sector AI is flourishing in many jurisdictions and will continue to thrive as long as public administrations support several key principles in deploying their AI applications. These key principles are interoperability, trustworthiness, transparency, upgrading skills of public servants and maintaining high data quality.

10.9 PUBLIC SECTOR AI'S IMPACT ON SOCIETY

As has been illustrated throughout this chapter, AI has promptly become an integral part of people's lives that has transformed several aspects of society providing far-reaching opportunities. However, this potential brings with it great challenges and apprehensions about AI's impact on society, particularly the economy, employment, healthcare, education, and ethics. [Behailu \(2023\)](#) argues that the development and implementation of AI has raised important ethical, social, and economic issues that need attention.

Some researchers view AI as part of the second machine age in which machines are replacing human brain power, and contrast this with a first period of industrialisation in which machines replaced human and animal muscle-power ([Brynjolfsson and McAfee, 2014](#)). While others have

expressed their concern about the susceptibility of jobs to digitalisation resulting in massive job losses ([Frey and Osborne, 2017](#)). The future of employment and job growth is anticipated to be more focused in high-skill STEM jobs, while middle- and low-skill jobs (such as food service, production work, and office support tasks) will likely decline considerably. Hence, there will likely be a shift in the labour market where certain job categories might experience more growth than others. These paradigms provide a general view of the economic implications of AI.

AI is also likely to cause intense social transformations in healthcare and education. Use cases cited previously illustrate that AI applications are likely to improve healthcare outcomes through predictive analytics and personalised treatments. Additionally, AI applications in the education field is rapidly advancing through AI-Driven learning experiences by adjusting to the needs of the individual student. A concern regarding these advances is whether the AI applications related to healthcare and education would be available to all, or whether they will be available only to those who can afford them. One should be aware that some countries provide free healthcare and education to all. However, there are many countries where healthcare and education are only within reach for the affluent members of society.

There is no doubt that AI has become increasingly integrated into society, but ethical concerns need to be addressed. The ethical and legal issues will be discussed in detail in the next chapter. However, it is essential to stress that AI developments and deployment must adhere to strict ethical standards, particularly in relation to data privacy and security, bias, fairness, and transparency. Each national government will need to assess its existing status regarding the deployment of AI applications and define the necessary policies to suit its specific needs and that of its constituents. Furthermore, governments must also endeavour to collaborate with each other to ensure that AI's potential is exploited for the benefit of all and any challenges encountered are resolved and concerns mitigated, and if possible, eliminated.

10.10 PRACTICAL PUBLIC SECTOR AI APPLICATIONS

Public Sector AI applications are gaining extensive attention because they impact a wide spectrum of a nation's population. Public Sector AI systems have an enormous effect on how public sector organisations conduct their operations and provide their services to a wide range of individuals, businesses, and society in general. This section provides some useful public sector AI applications that are having a significant favourable impact on the way governmental organisations operate and how they provide the respective services to their clients.

10.10.1 Case Study 1: Healthcare Effectiveness and Customer Satisfaction at a Public Hospital

The management of a large public hospital conducted an operations review with the objective of modernising the way the hospital provides its services to the public using technology, in particular AI. The hospital was confronted with several challenges that included staff shortages mainly due to high turnover as a result of work pressure and burnout; unacceptable waiting periods for elective (non-urgent) treatment such as cataract surgery or a knee replacement; shortage of medical supplies, including medicines; and general customer dissatisfaction, amongst other grievances.

To address these challenges and increase its healthcare effectiveness to its customers, the hospital management established a taskforce to implement the various recommendations of the operations review. A key recommendation was a substantial investment in modern technology, including AI. This investment was viewed as an effective way of addressing the various identified difficulties. The taskforce consisted of a full-time chairperson (supported by administrative and technical staff) and five senior medical professionals on a part time basis. The task force also had the authority to establish full-time dedicated technical teams (internally or outsourced) as required to implement the strategic measures approved by the taskforce.

The identified concerns were examined in detail by applying root-cause analysis and Ishikawa diagrams amongst other methods. Root-cause analysis is closely linked to service and product quality. It reveals in a very explicit way the source of a problem, in terms of when, where, and why the problem has occurred. Thus, pointing towards an appropriate approach to resolving the concern. [Figure 10.1](#) illustrates a simple and effective method for identifying the root cause to a problem. It should be noted that before a problem can be addressed, it must be explicitly identified. A more detailed approach to identifying complex problems is the Ishikawa diagram (refer to [Figure 10.2](#)). The Ishikawa diagram or the fishbone diagram is often used in the analysis of processes as a causal analysis tool to identify the many possible causes for a problem or situation.

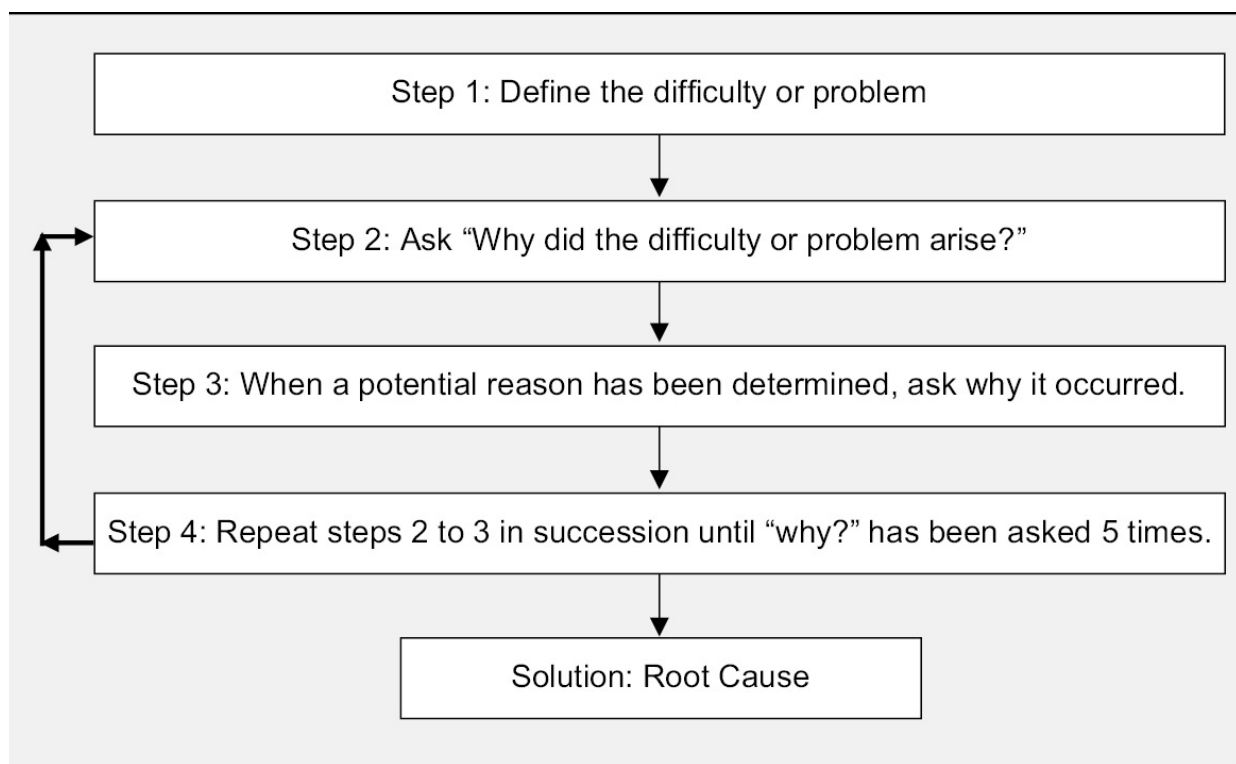


FIGURE 10.1 Root Cause Analysis: Five-Whys Approach. [↗](#)

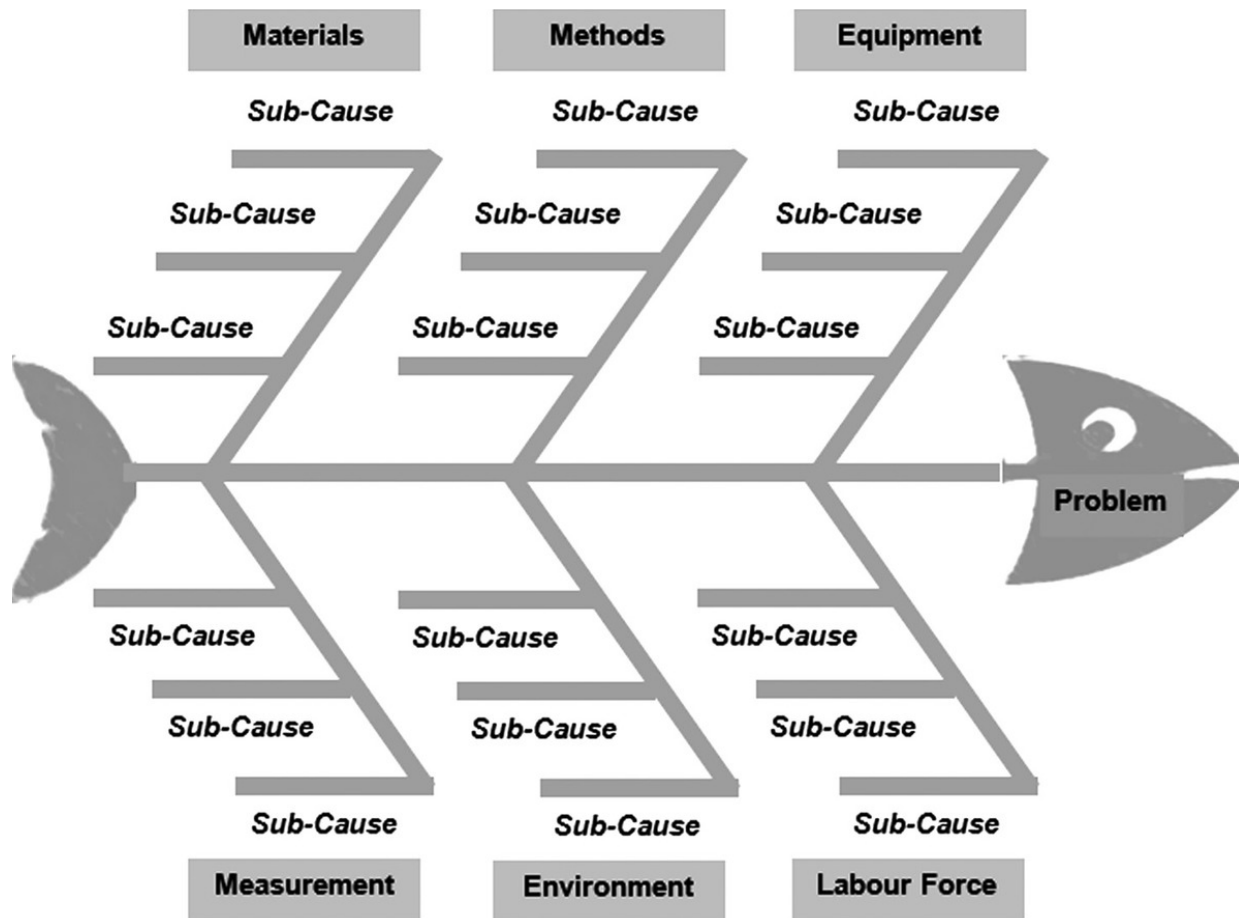


FIGURE 10.2 Constructing an Ishikawa Diagram. [↗](#)

The Ishikawa diagram was applied to structure brainstorming sessions to identify possible causes. This method was able to generate added examples of root causes and enabled the team to swiftly detect if the root cause is found multiple times in the same or different causal tree. It also permitted them to view all causes concurrently and provided helpful visualisation for displaying concerns. Furthermore, this approach provided a highly structured method and an excellent way for conducting the brainstorming sessions to detect the possible causes of a problem and sorting thoughts about the causes into suitable categories. The Ishikawa diagram is often used in conjunction with the Five-Whys tool (illustrated at [Figure 10.1](#)).

The identification of the root cause to the problems and respective solutions provided a sound basis for defining a strategy that did not exceed five years. To implement the defined strategy, the task force applied a

meticulous methodology illustrated by [Figure 10.3](#). It shows that there were two concurrent processes, both starting with the definition of the defined strategy and ending with the outcome of the strategy. Each process branch had a distinct focus, one related to strategy implementation (action execution) and the other focusing on the performance (critical success factors and KPIs). Furthermore, the focus of performance monitoring is through the KPIs. For example, from a critical success factor (CSF) and KPI perspective, the CSFs focus on specific aspects, while the KPIs related to each specific aspect establish the target performance level.

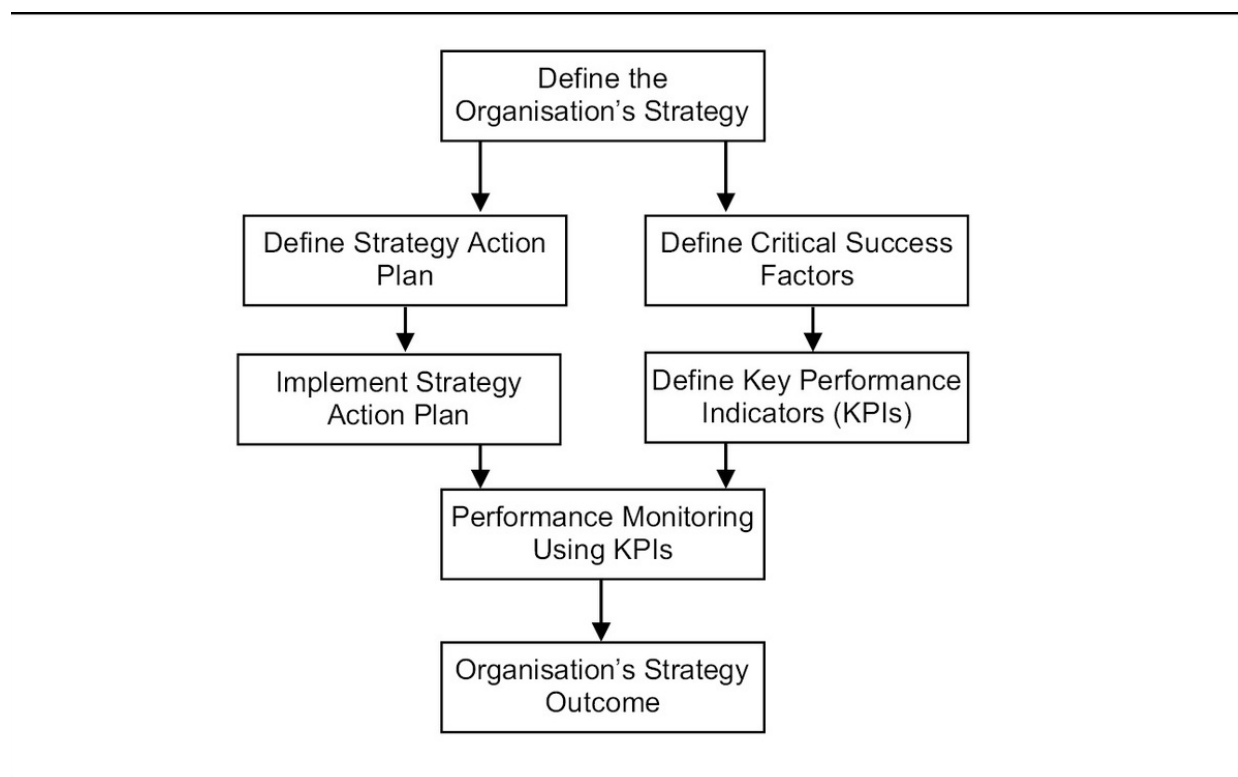


FIGURE 10.3 Relationship between Strategic Planning, CSFs and KPIs. [↗](#)

The CSF approach was applied for each identified initiative related to the various problem areas. The CSF methodology is based on the concept that a critical success factor is a high-level objective feature that is vital for an initiative (project) to achieve its mission. Therefore, the defined CSFs signified those areas that had to receive special and continual attention to bring about high performance and included those concerns that were most

important to the organisation's operating activities and to its future success. CSFs were expressed as action aphorisms that defined the method, desired outcome, and the action to be taken (e.g., waiting for elective treatment, such as cataract surgery or a knee replacement no longer than 12 weeks from referral to their first treatment). [Figure 10.4](#) illustrates the CSF framework that was used, which consisted of five critical processes, each one being related to the other:

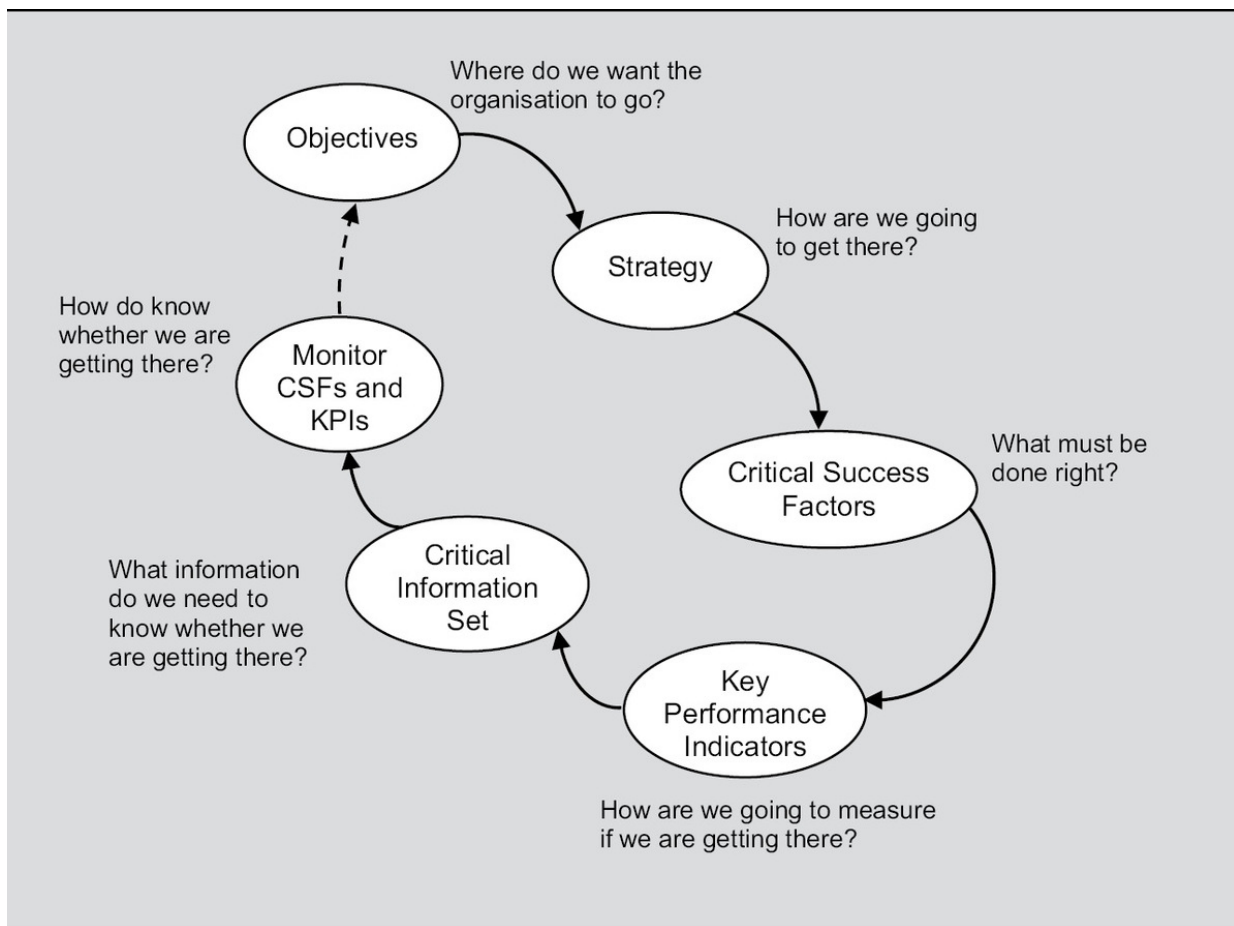


FIGURE 10.4 Critical Success Factors Methodology Framework. [📄](#)

1. **Objectives**: Where do we want our organisation to go? This is the overall target of what the organisation is trying to achieve.
2. **Strategy**: How are we going to get there? These are measures that will be taken by the organisation to achieve the overall target.

3. Identify the CSFs: What must be done absolutely right? These are the tasks that support the attainment of the measures being taken by the organisation to achieve the overall target.
4. Define the KPIs: How do we measure if we are getting there? These are the measures that are established to evaluate whether the tasks are being achieved.
5. Determine the critical information set: What information should be collected to track progress of each individual CSF. Thus, the information collected should be sufficient to verify whether the KPIs are being achieved.

The above methodology enabled the taskforce to implement various identified AI initiatives that were highlighted by the root cause analysis and defined in the strategy. Various AI tools and applications were evaluated and selected for implementation in critical healthcare areas as follows:

1. Medical Diagnosis and Treatment Recommendations. IBM Watson was able to assist healthcare professionals in making accurate diagnoses and treatment recommendations. Its capabilities enabled it to analyse patient data, including medical records, laboratory results, and various research papers, to generate insights and suggest personalised treatment plans. Watson for Oncology assisted oncologists by postulating evidence-based treatment options personalised to individual patient profiles, leading to more effective and targeted cancer care.
2. Patient Engagement and Support. IBM Watson provided AI-driven solutions, such as virtual health assistants that improved patient engagement and support. These virtual assistants were able to interact with patients, answer their healthcare queries, provided medication reminders, and offered personalised health advice based on their medical history and preferences. This personalised care enhanced patient outcomes resulting in higher patient satisfaction and also encouraged proactive healthcare management.

3. Healthcare Operations Optimisation. IBM Watson enabled the hospital management to analyse operational data, identify inefficiencies, and make data-driven decisions to improve resource allocation, reduce waiting times, and improve overall workflow efficiency. This resulted in a more streamlined healthcare delivery system and substantial cost savings.
4. Medicinals Inventory Management. This phase consisted of integrating an AI inventory model and upgrading the existing inventory management systems. The key issues to be addressed were related to attaining optimal stock levels and identifying the reordering strategies that were to be implemented. This eliminated stockouts and overstock situations and decreased the holding costs.
5. Resource Scheduling and Preventative Equipment Maintenance. AI models were implemented for resource scheduling, including a predictive element and assisted in the preventative maintenance of the medical equipment and facilities. This ensured the maximisation of medical equipment and facilities that resulted in a reduction in waiting lists for medical interventions and enhanced patient satisfaction.

Several AI algorithms were applied in the above models, such as Monte Carlo methods to simulate and analyse multiple scenarios, such as the probable outcomes of different inventory strategies and Long Short-Term Memory (LSTM), which is a recurrent neural network used for sequence prediction problems like stock level forecasting. Other AI predictive models focused on resource scheduling and equipment maintenance, such as Linear Programming, which was useful for resource allocation.

10.10.2 Case Study 2: Urban Traffic Management

A metropolitan authority was under pressure by persistent grievances regarding traffic congestion and the need to urgently address the growing problem. Traffic had become insanely unmanageable with the arterial roads being saturated by light and heavy vehicles. Thus, traffic congestion had

developed into a pressing issue that was affecting the city in various ways. This metropolitan area witnesses a significant amount of traffic on a daily basis, leading to long delays, frustration, and increased levels of air pollution. While traffic congestion was recognised as a common feature of metropolitan cities, it had become unmanageable, impacting both cargo and passenger transport.

A study commissioned by the competent government transport authority found that the growing problem of traffic congestion was attributed to several factors, including population growth, rapid urbanisation, a need for a more effective public transportation systems, careless driving such as the use of mobiles while driving, and an increase in the number of private vehicles on the roads. The outcome of these factors was frequent gridlocks on major roads, making it difficult for people to commute to work, access services, or deliver goods in a timely manner. In terms of cargo transport, the congestion on the roads was leading to a significant impact on the economy. Delays in the delivery of goods was resulting in increased costs for businesses that would eventually affect the prices of goods and services. This, in turn had the potential to negatively impact consumers and lead to a decrease in overall economic productivity. Furthermore, traffic congestion affected passenger transport, since commuters tended to spend more time on the roads, leading to frustration, stress, and a decrease in the overall quality of life. This concern was worsened by the need for a more reliable and efficient public transportation options, forcing many residents to rely on private vehicles that only compounded the issue of congestion further.

The government transport authority appointed a taskforce to examine the findings of the commissioned study and to develop a strategic transport plan. Additionally, once the strategic transport plan was approved, the taskforce was to be given the remit to implement it. After examining the commissioned study and its conclusions regarding the main causes of the traffic congestion, the taskforce set about finding solutions to mitigate the concerns. The taskforce recognised that the traffic congestion concerns could never be eliminated. The taskforce also commissioned an environmental and social impact assessment. This impact study concluded

that constructing new roads or widening existing roads by expropriating private property were not a feasible option from an environmental and social impact viewpoint. However, through a public consultation process and subsequent brainstorming sessions comprising of expert panels representing several key disciplines, several solutions were identified, which were deemed to be acceptable to most stakeholders.

One potential solution was to invest in the improvement and expansion of the public transportation system, including alternative transport methods (e.g., apart from land transport using also sea transport). This was seen as making it more convenient for people to leave their cars at home and rely on public transportation for their daily commutes. Another solution identified was to incentivise the use of alternative modes of transportation, such as walking and cycling. This could be achieved through the development of dedicated bike lanes and pedestrian-friendly infrastructure, as well as initiatives to promote active transportation, such as subsidising the cost of bicycles and manual/motorised scooters.

Additionally, given that more people were shopping online, the taskforce considered the controlled use of drones as means of shipment for small and light objects. Furthermore, the taskforce identified a need to develop and maintain the road infrastructure in a way that supported efficient traffic flow. This involved rejuvenating the existing road network and implementing intelligent traffic management systems to reduce bottlenecks and keep traffic moving smoothly. A key solution was the need to enforce regulations that promote efficient traffic flow, including strict enforcement of traffic laws and regulations, such as speed limits, and the reduction of the number of roadside parking that often contributes to the flow of traffic.

The taskforce concluded that traffic congestion was indeed a pressing issue that required immediate attention. Hence, it concluded that by investing in public transportation, improving road infrastructure, promoting alternative modes of transportation, and enforcing traffic regulations, the city would be able to work towards a more efficient and sustainable transportation system that benefits both cargo and passenger transport. Furthermore, the taskforce recognised that most of the solutions identified

could be effectively implemented through the support of technological solutions, particularly those related to intelligent traffic management systems and enforcing regulations.

These conclusions formed the basis of its transport strategic plan. A technical team was established to identify the appropriate technological support solutions, including the use of AI. The technical team evaluated “City Brain” an AI system developed by Alibaba (BABA). This system made extensive use of AI to collect information, such as video from intersection cameras and GPS data on the locations of cars and buses. The information was to be analysed in real time, while the system harmonised numerous traffic road signals with the aim of preventing or easing gridlock. Hence, the first phase was the identification of key road intersections and the installation of video cameras at these locations. Furthermore, all public vehicles were installed with GPS tracking devices. After several months of testing, commuters witnessed a significant improvement. The system also facilitated the work of first responders by enabling civil protection vehicles and ambulances to almost halve the amount of time it took them to the scene of accidents and address emergencies. The rapid clearance of broken-down vehicles due to minor accidents or mechanical faults assisted in keeping traffic flow at its maximum rate. By 2027, it was estimated that these AI systems could decrease commutes by an average of 18% to 25%.

The combined use of cameras and AI, and amendments to legislation enabled a more effective system of enforcing traffic regulations. Major roads were automatically monitored by the AI system using video feeds as input in real time to identify traffic infringements. These traffic infringements were instantly categorised by severity to enable timely interventions by law enforcers for serious offences, whilst other offenders were issued with customary traffic tickets supported by appropriate evidence for the respective traffic violations. Hence, the AI system was viewed as a preventative tool as well as an enforcement instrument. Admittedly, the AI system raised privacy concerns related to vehicle surveillance and possible facial recognition issues. However, the benefits

from this type of system far outweighed the concerns because it brought more discipline to the roads and in several cases saved lives.

As mentioned previously, traffic management solutions using big data and AI-based solutions were applied to monitor and control traffic signals and street cameras to optimise driving and walking routes. Additionally, sensors and cameras were able to monitor the condition of roads and prompt warnings for any immediate maintenance work. Moreover, digital signages and augmented reality (AR) based messaging systems replaced conventional road signages.

The problem of urban growth was not viewed as a recent phenomenon. Urban development in some of the older regions going back to possibly centuries was allowed to take place in a haphazard manner. Hence, authorities, architects, town planners, and civil engineers encountered problems that have their roots possibly many decades (if not centuries) in the past. However, technology was seen as having a critical role to the development of sustainable smart cities. For instance, internet of things (IoT) in combination with AI were applied to have an intelligent street lighting system that responded to changes in weather and traffic conditions to conserve energy and reduce light pollution. Autonomous vehicles were still at the acceptance testing stage. However, these were seen as a means of providing office goers with additional time to check their emails and attend any calls while commuting. Furthermore, these driverless autonomous vehicles could communicate automatically with a smart parking solution and even park themselves to the nearest available spot. Hence, they were able to select parking areas that offer charging-facilities to recharge batteries.

The technical team also revealed and reported to the competent authorities that there existed extensive potential from the emerging technologies like machine learning to address other problems encountered by communities, such as:

- Structural integrity monitoring. IoT based devices integrated with AI would issue warnings for maintenance required in any building,

bridge, or historical monument by continuously monitoring vibrations and material conditions.

- Electrical energy and water supply management. Smart grids would help in the reduction of electrical power and water distribution losses with IoT enabled equipment maintenance improving operational efficiency.
- Pollution management. A city-wide network of sensors integrated with AI would generate a map showing levels of sound, water, ground, air, and light pollution to assist administrators in taking mitigating and remedial measures.

The technical team believed that all the recommended solutions were technologically feasible. However, they also recognised that several limitations such as logistical, legal, and budgetary considerations constrained their immediate implementation. It was also recognised that such measures would transform their city and greatly increase the quality of life for its citizens.

10.11 CONCLUSION

This chapter has focused on Public Sector AI and illustrated the considerable potential AI applications have on public administrations and the constituents they serve. The word “constituents” is being applied in a broad sense and includes all citizens irrespective of their demographic status, civil society, and the commercial sector. The chapter has also shown that governments are the largest employer in each respective country and have a wide-ranging customer base that expect to be provided with an ever-increasing number of products and services.

AI has been recognised as a powerful tool to cater to such a broad spectrum of expectations from the different categories of customers. The chapter has demonstrated that Public Sector AI and Enterprise AI share a common aim, that is to provide a better customer service experience at a lower cost on a 24/7 basis from anywhere. However, an important

distinction must be made between Enterprise AI and Public Sector AI. Enterprise AI focuses on improving the profitability of a commercial entity by improving efficiency and providing personalised customer experiences. On the other hand, Public Sector AI is focused on improving laws and social and economic policies. These policies include the development and implementation of national strategies in various disciplines, such as education, health, law enforcement, and transportation. But both have a common underlying objective to improve efficiency and provide personalised customer experiences.

Furthermore, governments face greater challenges when contemplating the deployment of AI because they need to consider the holistic national interest, particularly its economic development and the impact of AI on the different segments of society. These concerns are basically related to how AI will impact the economic growth, employment displacement, technology transfer and skills upscaling, healthcare, and education, amongst others. The bottom line is that every government is required to evaluate its status regarding the need to deploy AI applications and define the essential policies to suit its specific needs and that of its constituents. The chapter itemised and explained several use cases to give the reader an idea of how Public Sector AI can benefit their public administration entities and their respective customer base.

Finally, the chapter referred to a suitable framework for the successful deployment of AI, which was fully describe in the previous chapter when discussing Enterprise AI (see [Figure 9.3](#)). Each step in this implementation framework was explained in detail to provide the reader with a viable and practical deployment approach that may be adapted and customised to their individual public administration organisations. Furthermore, this framework is suitable for all categories of public sector organisations, irrespective of the nature of the government level, namely national, state or municipal. It is contended that the suggested framework provides an effective and comprehensive user-friendly approach to the deployment of Public Sector AI in public administrations at every government level. However, every government is required to evaluate its current circumstances regarding the

deployment of AI applications and identify the required policies to suit its specific requirements and that of its constituents.

REFERENCES

- Behailu, Yeabtsega. (2023). The impact of artificial intelligence on society. *International Research Journal of Modernization in Engineering Technology and Science*. Vol. 05, No. 10, pp. 3120-3125. <https://doi.org/10.56726/IRJMETS45708>.↵
- Berryhill, Jamie., Heang, Kévin, Kok., Clogher, Rob, and McBride Keegan. (2019). *Hello, World: Artificial Intelligence and Its Use in the Public Sector*. OECD Observatory of Public Sector Innovation (OPSI). <http://oe.cd/helloworld>↵
- Brynjolfsson, Erik., and McAfee, Andrew. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W.W Norton.↵
- Camilleri, Emanuel. (2024). *Key Performance Indicators: The Complete Guide to KPIs for Business Success*. Routledge.
- Canning, Mike., Gentile, Matt., Mittal, Nitin., Saif, Infan., and Ammanath, Beena. (2024). *The Government & Public Services*. AI Dossier by Deloitte AI Institute. www2.deloitte.com/us/en/pages/consulting/articles/ai-dossier-government-public-services.html.
- Frey, Carl, Benedikt., and Osborne, Michael. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, Vol. 114, pp. 254–280.↵
- Tangi, Luca., Combetto Marco., Martin Bosch Jaume., and Rodriguez Müller, Aan, Paula. (2023). *Artificial Intelligence for Interoperability in the European Public Sector: An Exploratory Study*. JRC Technical Report. European Commission, Joint Research Centre (JRC).
- World Bank. (2021). *Artificial Intelligence in the Public Sector*. <https://documents1.worldbank.org/curated/en/746721616045333426/pdf/>

[Artificial-Intelligence-in-the-Public-Sector-Summary-Note.pdf](#) 

Chapter 11

AI Ethical, Moral, and Legal Considerations

DOI: [10.1201/9781003653950-14](https://doi.org/10.1201/9781003653950-14)

“It is not wisdom but Authority that makes a law.”

Thomas Hobbes

English philosopher (1588-1679)

Often the use of AI is justified due to its inherent benefits, such as faster data analysis, time savings in performing tedious tasks, increased levels of predictive accuracy, more in-depth insight into business processes, and enhanced customer service experience. It has been established that AI is an emerging technology that has the objective of imitating the cognitive skills and judgment of humans and has the potential of providing a competitive advantage. Industry has recognised these objectives as being feasible and achievable, and as a result, the take up of AI is constantly growing. AI is being applied in innovative ways such as risk assessments, analytics, predictive business modelling, and routine operational tasks.

As the application and benefits of AI continue to materialise within the various industries, there is a steady and persistent realisation that unintended consequences may also arise with its use. There are numerous researchers advocating caution regarding the use of AI. They implore practitioners to not only explore the benefits of AI but to also identify and scrutinise the ethical implications that such powerful technology may bring about. It is important that a collaborative approach is taken amongst stakeholders, such as government, enterprises, civil society, and educational institutions (Universities and Institutes of Technology) to explore this issue.

It is critical that a conceptual evaluation of the practical ethical and social issues encompassing AI is undertaken to ensure trustworthy AI applications that operate in a safe environment. [Fassbender \(2014\)](#) quoted Elon Musk, CEO of SpaceX and Tesla Motors in an interview at the AeroAstro Centennial Symposium as describing AI as the biggest existential threat to humanity and that we should be careful with AI as it might summon “the demon.” This implies that there is a pressing need for having a suitable mechanism for ensuring that AI applications are developed and implemented according to robust ethical and moral guidelines. Furthermore, governments have a civic duty to protect their citizens from any adverse

consequences that AI may bring about and ensure that AI is applied in a way that brings economic prosperity to their constituents.

11.1 INTRODUCTION

In the emerging technology landscape, AI has developed into a world-shattering force that continues to influence numerous aspects of people's lives. It has revolutionised various industries and has also altered the way individuals interact with machines. The growing movement towards the various ways that AI is being applied is only limited by the innovative imagination of technologists and practitioners. Having said this, the amazing wonders of AI's capabilities should not distract society to the risks that may emerge from using it. It is becoming increasingly critical for society to contemplate the ethical and moral dilemmas, and societal implications of creating algorithms that endeavour to replicate the intelligence of human beings.

According to [Patowary \(2023\)](#), Mary Shelley's iconic literary masterpiece, *Frankenstein*, published in 1831 nearly two centuries ago, remains an enduring cautionary story about the hazards of unrestricted determination and the outcomes of playing God. Patowary argues that AI obscures the peripheries between human and machine, with society finding itself treading on the very themes explored by Mary Shelley. Patowary (2023) contends that this story of excessive pride, self-confidence, moral predicaments, and the intricate relationships between creator and creation continues to go beyond time, finding a striking resonance in contemporary discussions on AI and its potential implications.

There is no doubt that AI provides great opportunities to positively transform people's lives and offers enormous benefits to enable societies to progress in numerous fields, be it healthcare, education, improved quality of life, and many others. AI is also viewed as enabling nations to develop further their economies for the benefit of their constituents. However, the dark echoes of Victor Frankenstein's missteps serve as a stark reminder of the need for ethical frameworks and interdisciplinary collaboration to ensure that AI remains a powerful force for good ([Patowary, 2023](#)).

[Wallach and Allen \(2009\)](#) contend that the human-built environment is increasingly being populated by artificial agents which can act autonomously through AI. They argue that the algorithms controlling these AI applications are currently "ethically blind" since their decision-making abilities do not involve any explicit moral reasoning. Wallach and Allen maintain that there is an urgent need for these increasingly autonomous systems (robots and software bots) to factor ethical and moral considerations into their decision-making algorithms.

[Patowary \(2023\)](#) further argues that it is imperative that society learns from the cautionary tale of *Frankenstein* and embrace a multidisciplinary approach, integrating ethics, philosophy, and technology to guide the future of AI in a manner that respects humanity and safeguards society's collective well-being. This chapter will focus on two distinct aspects, namely: (a) the ethical and moral considerations for the development and deployment of AI applications; and (b) the AI legal considerations by examining the legislative developments in various countries.

11.2 AI ETHICAL AND MORAL CONSIDERATIONS

After decades of research and several AI winters, AI has become an essential feature across many industries. AI has now been widely accepted, and its use has been fruitful in healthcare, banking, retail, transport, education, and manufacturing. However, like other technologies that have been introduced in the past, little thought has been given to its possible dark side. Pazzanese (2020) argues that AI's game-changing capacity to do things like increasing efficiency, reducing costs, and fast-track research and development has now been moderated with concerns that these complex, opaque systems may do more societal harm than economic good. Pazzanese contends that up to 2020, with almost no government oversight, private companies in the USA applied AI software to make decisions about health and medicine, employment, creditworthiness, and even criminal justice without having to answer for how they're ensuring that these programs were not encoded, consciously or unconsciously, with structural biases.

AI is here to stay. This is confirmed by the predicted worldwide spending on AI solutions that will grow to more than \$500 billion in 2027, with most organisations experiencing a notable shift in the weight of technology investments toward AI implementation and adoption of AI-enhanced products and services ([IDC, 2023](#)). It is now an accepted fact that practically every large company now has multiple AI systems and includes the use of AI as an integral and essential part of their corporate strategy. Initially, AI was viewed as being applied to automate simple tedious and repetitive tasks requiring low-level decision-making, but powerful computers and the accumulation of massive data sets has facilitated AI's rapid development into a very sophisticated device (Pazzanese, 2020).

The emerging AI application environment, where AI systems are now making certain determinations about individuals is causing serious concerns. These critics worry whether AI will bring about newer, quicker, and unobstructed methods to discriminate and divide at scale. At first glance, algorithmic decision-making appears to provide an objective approach of overcoming human subjectivity, bias, and prejudice. However, experience has shown that these algorithms show bias because they learn from the data set they are provided with. Hence, if the data has an inherent bias the resultant outcome will reflect this bias.

Pazzanese (2020) cites three major areas of ethical concern for society regarding AI, namely, privacy and surveillance, bias and discrimination, and the role of human judgment. Pazzanese cites political philosopher Michael Sandel of Harvard University who when discussing how to overcome bias in algorithmic decision-making in sentencing, parole, and employment practices refers to conscious and unconscious prejudices of program developers and those built into datasets used to train the software. He contemplates that we have not yet wrapped our minds around the hardest question: Can smart machines outthink us, or are certain elements of human judgment indispensable in deciding some of the most important things in life?

There are those who accept that AI does indeed introduce bias in its resultant outputs, but they also believe that it is not worse than all types of biases that humans introduce in their decision-making process. For instance, using resume-screening AI software would mitigate the chance of favouritism when fine-tuned and aligned with the requirements. However, there is disagreement regarding this premise because AI is widely viewed as being reliable due to its scientific basis

and considered as an objective assessment, even though in reality it imitates and reflects human preferences. This “bias” tendency of AI software is because machine learning is based on historical data sets that reflects former “good” and “not-so-good” human decisions. Therefore, the data sets will inherently contain a predisposition towards all types of decisions, irrespective of whether they are good, bad, or mediocre. This concern is applicable to all decisions when assessing different types of applications, such as in banking, social security pensions, insurance, credit rating, and social housing, amongst many others.

There is wide consensus in the industry that AI should be regulated. However, there is disagreement regarding how this should be done and who should do it. Developers and private sector AI users prefer self-policing within existing legal regimes. However, civil society look to the government to take the lead in the specific regulation of AI applications. A major difficulty is the fact that regulatory organisations lack the AI expertise necessary to identify and define what needs to be regulated and what practical approach should be adopted to regulate it. Self-policing is appropriate to some extent in terms of having a code of ethics that would serve as a guide to developers, providers and user organisations. However, one needs to remember that this would not exonerate developers, providers, and user organisations from legal action by aggrieved parties. Therefore, legislation is necessary to regulate AI and clearly define the legal obligations of all concerned.

As stated previously, regulatory organisations lack AI expertise to identify and define what needs to be regulated, and what practical approach should be adopted to regulate it. However, this difficulty might be mitigated by having a collaborative approach through a taskforce between government, and representatives from developers, providers, major users and civil society. This combination would mean that government and civil society would contribute towards the legal and social expertise and the other representatives towards the technical aspects. Furthermore, the taskforce may have various panels or sub-committees to tackle specific industries, such as finance, automobile, aviation, healthcare, education, and manufacturing, amongst others. Having said this, each National, State, and municipal jurisdiction would need to consider its circumstances and adapt the appropriate mechanism that suits its environment.

11.2.1 The Need for AI Ethics and Moral Guidelines

Generally, AI ethics in technological advancements can facilitate and promote a world with less bias and more fairness. As AI becomes more significant to society, it has become evident that there is a need for ethical boundaries in the development and deployment of AI tools. Acknowledging this need and in the absence of a wide-scale governing body to write and enforce ethical rules, many technology companies have introduced their own version of AI ethics or code of conduct. The urgent need for AI ethics and moral guidance has arisen due to the concerns about the harmful impact that AI applications might have on society in general.

Fundamentally, AI ethics are the moral values that companies employ to guide responsible and fair development and application of AI. In other words, ethics in the context of AI, is a set of moral principles that help one to determine between right and wrong, and how to enhance AI’s beneficial impact while mitigating risks and harmful outcomes. Guidelines are required to assist

developers, providers, and users in ensuring that the AI applications are trustworthy, safe, secure, humane, and environmentally friendly. There are numerous AI ethical and moral issues to be considered, namely, privacy and security, fairness, inclusion, data responsibility, transparency, explainability, robustness, environmental sustainability, moral agency, value alignment, accountability, trust, and technology misuse. There are broadly two main approaches to defining codes of ethics, namely by companies and government-led regulatory frameworks. Both methods facilitate the regulation of AI technology.

According to [Leslie \(2019\)](#), the field of AI ethics has largely emerged as a response to the range of individual and societal harms that the misuse, abuse, poor design, or negative unintended consequences of AI systems may cause. This illustrates that AI ethics has been tackled as a reactive approach rather than a proactive measure. Governments and society appear to be repeating past faults when dealing with technological developments in that the response has always been reactive. [Leslie \(2019\)](#) identifies several possible harmful causes of AI applications:

- **Bias and Discrimination:** AI algorithms use historical data sets that might contain an inherent human predisposition and preconceptions to certain decisions. Also, the data set used to train and test algorithmic systems may not be adequately representative of the population from which they are drawing inferences and hence might generate biased and discriminatory outcomes.
- **Denial of Individual Autonomy, Recourse, and Rights:** There may be difficulty to hold directly accountable the parties responsible for AI outcomes in situations where citizens are subject to decisions, predictions, or classifications from these AI systems, thus harming the autonomy and violate the rights of the affected persons.
- **Non-transparent, Unexplainable, or Unjustifiable Outcomes:** There are cases where the rationale for the outcomes from machine learning models are not clear and could conceal traces of discrimination, bias, or inequity.
- **Invasions of Privacy:** AI applications might use personal data that may have been attained without proper consent or may reveal personal information that may harm a person's basic right.
- **Isolation and Disintegration of Social Connection:** AI systems might reduce the need for human-to-human interaction, while algorithmically enable hyper-personalisation by limiting one's exposure to worldviews different from one's own that might polarise social relationships. It is important that these social relations be preserved.
- **Shoddy AI development practices** will likely result in unreliable, unsafe, or poor-quality outcomes that damage the wellbeing of individual persons and the public welfare, which contribute to trustworthiness among the user community.

It must be recognised that AI ethics are critically essential because AI technology has the objective to amplify or replace human intelligence. Hence, in a situation where technology is designed to imitate human intelligence then similar concerns that can blur or confuse human judgment can also be replicated by AI.

The need for AI ethics and moral guidelines can be illustrated by examples that have caused concern. According to [Ovide \(2022\)](#), photos and videos posted on Instagram or other similar media may be used to coach AI software without any consent. She reports that Lensa AI was being trained on a massive volume of photographs that are sourced from the internet without consent to generate trendy, cartoon-looking profile photos from people's regular images. Ovide cites another similar example regarding ChatGPT, which learns to mimic humans by analysing recipes, social media posts, product reviews and other text from everyone on the internet without consent. Beyond those two technologies, birthday party photos on Facebook helped train Clearview AI facial recognition software that police departments use in criminal investigations ([Ovide, 2022](#)).

Formulating ethical guidelines that define fundamental principles for responsible AI development and deployment requires industry actors to collaborate as a team. These stakeholders must study how social, economic, and political issues interconnect with AI, and establish how machines and humans may harmoniously coexist. Key stakeholders might include:

- **Academics:** These have the essential role of conducting research studies and developing theory-based statistics and concepts that can support governments, corporations, and non-profit organisations.
- **Government:** Entities and taskforces within a government can help to expediate the definition of a nation's AI ethics. This can be achieved through public consultations for identifying appropriate regulations and defining governance, economic and security policies.
- **Intergovernmental Bodies:** Institutions like the United Nations, the World Bank and OECD play a vital role for enhancing awareness and drafting agreements for AI ethics globally.
- **Non-Profit Organisations:** These organisations assist various groups attain representation within AI technology. For example, the Future of Life Institute who believe that frontier AI is currently being developed in an unsafe and unaccountable manner has compiled twenty-three guidelines that outline specific risks, challenges, outcomes for AI technologies.
- **Private Companies:** Technology leaders, such as Google, Meta, and others, as well as leading AI users, such as banks, consultants, healthcare providers, and other industries, establish ethics teams and codes of conduct that are viewed as a standard for a wide range of companies.

All these stakeholders have a vital role for contributing towards ensuring less bias and risk from AI technologies.

11.2.2 Establishing Principles for AI Ethics

Researchers from the National Institute of Standards and Technology (NIST) have suggested that a report released in 1979 on the ethical treatment of human subjects (Belmont Report) could set a precedent for Ethical AI Research ([CDO, 2024](#)). This is confirmed by [IBM \(2024\)](#), which reports that the academic community in USA has leveraged the Belmont Report to guide ethics within

experimental research and algorithmic development. The Belmont Report has three key principles ([IBM, 2024](#)):

- **Respect for Persons:** This principle acknowledges the autonomy of individuals and advocates the expectation of researchers to protect individuals with diminished autonomy, which could be due to a variety of circumstances, such as illness, a mental disability, and age restrictions. This principle is viewed as supporting the concept of consent, where individuals should be aware of the possible risks and benefits of any experiment that they participate in and be able to choose to participate or withdraw at any time before and during the experiment.
- **Beneficence:** This principle has its basis on healthcare ethics, where doctors undertake to “do no harm.” This concept can be applied to AI where algorithms can intensify biases regarding race, gender, political leanings, and others, irrespective of the intent to do good and enhance a given system.
- **Justice:** This principle is based on fairness and equality. The Belmont Report proposes five modes to disseminate burdens and benefits, namely: equal share; individual need; individual effort; societal contribution; and merit. These may be applied for the distribution of the benefits from experimentation and machine learning.

The European Union (EU) Commission in 2019 proposed a guideline for trustworthy AI systems based on seven key requirements as follows:

- **Human Agency and Oversight:** AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. At the same time, proper oversight mechanisms need to be ensured, which can be achieved through human-in-the-loop, human-on-the-loop, and human-in-command approaches
- **Technical Robustness and Safety:** AI systems need to be resilient, secure and safe, ensuring that a contingency plan is in place in case something goes wrong. They also need to be accurate, reliable and reproducible. This would ensure that unintentional harm can be mitigated and prevented.
- **Privacy and Data Governance:** Apart from ensuring full respect for privacy and data protection, AI systems must have sufficient data governance mechanisms, consider the quality and integrity of the data, and ensuring authorised access to data.
- **Transparency:** The data, system and AI business models should be transparent with traceability mechanisms. Additionally, AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned. Humans need to be aware that they are interacting with an AI system and must be informed of the system’s capabilities and limitations.
- **Diversity, Non-Discrimination and Fairness:** Unfair bias must be avoided, since it might include multiple negative implications, from the marginalisation of vulnerable groups to the exacerbation of prejudice and discrimination. AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle.

- Societal and environmental well-being: AI systems must benefit all human beings, including future generations and ensure that they are sustainable and environmentally friendly. AI systems should take into consideration the environment, including other living beings, and their social and societal impact should be carefully studied.
- Accountability: Mechanisms must be introduced to ensure responsibility and accountability for AI systems and their outcomes, and that adequate and accessible redress is ensured. Auditability to assess algorithms, data and design processes is required particularly for critical applications.

Developing and deploying more ethical AI involves a careful study of the ethical implications of policy, education, and technology. There is no doubt that regulatory frameworks can guarantee that technologies benefit society and not harm it. Globally, governments are initiating policies for ethical AI, including how entities should deal with legal issues if bias or other harm arises. AI users should be able to appreciate the risks and possible negative impact of unethical or fake AI. The generation and dissemination of accessible resources can help to mitigate these types of risks. On the other hand, one must also appreciate that AI tools can help and may be applied to determine whether video, audio, or text (hate speech on Facebook, for example) is fake by detecting unethical data sources and bias better and more efficiently than humans.

AI algorithms execute according to how they are designed, developed, trained, tuned, and applied, with AI ethics being about creating an ecosystem of ethical standards and guardrails around and throughout all phases of an AI system's lifecycle ([IBM, 2024](#)). According to IBM, entities, governments, and researchers have recognised the importance of ethical frameworks and have commenced to compile charters for addressing AI ethical concerns and configure the future of work within the field. The structure of AI ethical guidelines is being constantly enhanced and incorporate governance as a common aspect ([IBM, 2024](#)).

[IBM \(2024\)](#) contend that governance is an organisation's act of overseeing the AI lifecycle through internal policies and processes, staff, and systems, which help to ensure that the AI systems are operating as an organisation's principles and values intend, as stakeholders expect, and as required by relevant regulation. Moreover, IBM view a successful governance program as one that: (a) defines the roles and responsibilities of people working with AI; (b) educates all people involved in the AI lifecycle about building AI in a responsible way; (c) establishes processes for building, managing, monitoring, and communicating about AI and AI risks; and (d) leverages tools to improve AI's performance and trustworthiness throughout the AI lifecycle. IBM recommends the establishment of an AI Ethics Board to manage the governance mechanism that would comprise of diverse leaders from across the business. The objective of this ethics board is to provide a centralised governance, review, and decision-making process for the ethics policies and practices.

According to the [World Bank \(2021\)](#), many advanced digital governments have published governance frameworks, including ethical standards for the use of AI that include the ethical principles, the role of a central agency, and operational framework. The World Bank notes that national governments in addition to international organisations are actively proposing governance models for AI that emphasise several common principles as shown in [Table 11.1](#). These national

governments include Australia, Canada, China, Japan, Singapore, United Arab Emirates, and the United States and the international organisations consist of the EU Commission, the Institute of Electrical and Electronics Engineers, International Organization for Standardization, United Nations, and World Economic Forum.

TABLE 11.1 Common Ethical Principles Adopted by National Governments [🔗](#)

Principle	Description
Privacy & data protection	AI solutions should respect an individual's right to privacy and civil liberties. Individuals should have control over their data, with their consent required for repurposing, redistributing or erasure.
Accountability	Mechanisms must ensure accountable behaviour during the life cycle of AI design and implementation. An entity must be liable for monitoring accountability.
Safety & security	Cybersecurity is critical. Leaders must ensure the well-being of society at large and private individual humans.
Transparency & explainability	The algorithm, business case, data collection, design, and policy information must be transparent to stakeholders and those impacted. Individuals should get notifications when interacting with AI or when AI decides for them.
Fairness	AI solutions should minimize bias and identify and manage risk. Inclusiveness should be ensured in design and impact.
Human control of technology	The AI should be under human control. People should review automated decisions and have the right to opt-out of automated decisions.
Professional responsibility	Multistakeholder collaboration, accuracy, and scientific integrity of the solution should be ensured.
Promotion of human values	AI should be human-centric. It should promote human values and benefit society.

Source: [World Bank \(2021\)](#). *Artificial Intelligence in the Public Sector*. World Bank Group.

The World Bank maintains that some governance models and guidelines highlight common program and project management practices that include a cost-benefit analysis, legal and regulatory compliance, risk management, flexibility, and the application of a responsive methodology.

11.2.3 Building-Blocks of a Responsible AI Project Delivery Ecosystem

According to [Leslie \(2019\)](#), setting up an ethical platform for responsible AI project delivery involves more than building from the cultural ground up. He maintains that it also involves providing the project team with the means to achieve the goals of establishing ethical acceptability, impartiality, trustworthiness, and project justification. Additionally, [Leslie \(2019\)](#),

argues that an ethical platform for responsible AI project delivery should be based on three building-blocks, namely:

1. **Basic Level:** This obliges one to attain a working knowledge of a framework of ethical values that supports, guarantees, and inspires a responsible data design and use ecosystem. These values consist of four key notions: Respect, Connect, Care, and Protect. The objectives are to provide an accessible framework to start thinking about the moral scope of the societal and ethical impacts of the project; and to establish clearly defined criteria to evaluate its ethical permissibility.
2. **Second Tier Tangible Level:** This entails a set of actionable principles that assist in achieving an orientation to the responsible design and use of AI systems. These embrace four key concepts: Fairness, Accountability, Sustainability, and Transparency. The objectives of these principles are to provide the moral and practical tools to ensure that the project is bias-mitigating, non-discriminatory, and fair; and to safeguard public trust in the project's capacity to deliver safe and reliable AI innovation.
3. **Third Tier Definite Level:** This necessitates a process-based governance framework that operationalises the basic level values and the second-tier tangible principles across the entire AI project delivery workflow. The objective is to establish transparent processes of design and implementation that safeguard and enable the justifiability of both the AI project and its product.

The building-blocks for a responsible AI project delivery ecosystem are intended to provide a robust, process-based foundation of values, principles, and protocols that enable an AI project team to improve the design and implement AI systems ethically, equitably, and safely. Additionally, it is intended to help an entity create a culture of responsible AI innovation based on an ethical platform that ensures a united collective spirit is fostered within the AI project team to enable the development of AI technologies for the public good.

11.3 AI LEGISLATIVE GLOBAL LANDSCAPE

The above sections focused on ethical and moral behaviour regarding AI. This section will focus on the legislative aspect related to AI around the world. It is important to understand that typically, ethics and law are the different systems of rules by which humans are expected to behave themselves in the society. However, ethics represents the internal system of controls, and Law refers to an external mechanism of control. Both these systems of controls are essential for optimal behaviour, which includes pursuing and imparting treatment.

The sections below will provide a detailed outline of AI regulations of several regions around the globe and will discuss how various nations are governing AI. Furthermore, the key considerations for developing AI regulations will be explained. Although ethics and law are different, it is important that the legal frameworks governing AI are understood because a legal code needs to be applied in an ethical manner by various entities. The ethical application of the law regarding AI is to ensure that societal benefits are attained as technological advances shape

humanity's future. The emerging global landscape of AI regulations might differ since countries will possibly adopt dissimilar approaches. Having said this, it is also likely that there may be some similarities regarding the fundamental factors that underline the creation of AI legislation.

There is no doubt that governments are belatedly recognising that AI developments are impacting people's lives quicker than expected and has become a disrupting mechanism. This is impacting how businesses operate, which in turn affects their respective employees and the recipients of products and services. Additionally, this scenario is provoking moral, legal, and social concerns, which is forcing governments to adopt appropriate legislation for regulating the design, development and application of AI based systems. This section will consider the following regional areas that are viewed as providing a suitable global AI legislative outlook:

- Europe: EU and UK.
- North American: USA and Canada.
- Africa: Mauritius, Kenya, Egypt, and South Africa.
- South America: Argentina, Brazil, Colombia, Mexico, Chile, and Peru.
- Australasia: Australia and New Zealand.
- Asia: China, India, Japan, and Singapore.

11.3.1 Europe

The European legislative segment will focus on the European Union and the United Kingdom.

11.3.1.1 European Union (EU)

According to [European Parliament \(2024\)](#), the application of AI in the EU is regulated by the AI Act, which is the world's first comprehensive AI law. The priority of this Act is to ensure that AI systems used in the EU are safe, transparent, traceable, non-discriminatory, and environmentally friendly. Additionally, the EU maintains that AI systems must be administered by people, rather than by automation, to prevent harmful outcomes. Moreover, an important aim is to have a technology-neutral uniform definition for AI that might be applied to future AI systems.

A critical aspect of the Act is to establish different rules for various risk levels and defining obligations for providers and users depending on the level of risk from AI. All AI systems require assessment irrespective of whether they pose minimal risk. According to [European Parliament \(2024\)](#), AI systems that are considered a threat to people are viewed as having unacceptable risk and will be banned. AI systems that are classified as having unacceptable risk by the EU include:

- Cognitive behavioural manipulation of people or specific vulnerable groups, such as voice-activated toys that incite dangerous behaviour in children.
- Social scoring that classifies people based on behaviour, socio-economic status or personal characteristics.
- Biometric identification and categorisation of people.
- Real-time and remote biometric identification systems, such as facial recognition.

However, the EU identifies some exceptions that may be permitted for law enforcement purposes. For example, “Real-time” remote biometric identification systems are permitted in some serious cases, while “post” remote biometric identification systems, where identification occurs after a significant delay, will be permitted to prosecute serious crimes and only after court approval ([European Parliament, 2024](#)).

AI systems may also be classified as high risk if they negatively affect safety or fundamental rights. All high-risk AI systems will require assessment prior to being released on the market and throughout their lifecycle. Additionally, individuals will have the right to file complaints about AI systems to designated national authorities. High risk systems are divided into two categories, namely:

- AI systems that are applied in products that are within the EU’s product safety legislation, such as toys, aviation, cars, medical devices, and lifts.
- AI systems applied in specialised areas that need to be registered in an EU database, such as management and operation of critical infrastructure; education and vocational training; employment, worker management and access to self-employment; access to and enjoyment of essential private services and public services and benefits; law enforcement; migration, asylum and border control management; and assistance in legal interpretation and application of the law.

According to [European Parliament \(2024\)](#), Generative AI (e.g., ChatGPT) will not be classified as high-risk, but will need to comply with transparency requirements and EU copyright law, namely disclosing that the content was generated by AI; designing the model to prevent it from generating illegal content; and publishing summaries of copyrighted data used for training. High-impact general-purpose AI models that may perhaps present systemic risk, such as the more advanced AI model GPT-4, would need to go through a detailed assessment and any serious incidents would need to be reported to the European Commission. Furthermore, content that is either generated or modified with the support of AI algorithms, such as images, audio or video files (for example deepfakes) are required to be clearly identified as being AI generated to ensure that users are conscious of them when they encounter such content.

The AI Act supports innovation by obliging national authorities to provide companies with a testing environment that simulates conditions close to the real world. This gives start-ups and small and medium-sized enterprises the opportunity to develop and train AI models before being released to the public. According to [European Parliament \(2024\)](#), the AI Act was adopted in March 2024 and the EU Council followed with its approval in May 2024 with the Act being fully applicable 24 months after coming into effect. However, some articles of the Act will be adopted sooner. For instance, the ban of AI systems posing unacceptable risks; the codes of practice; and the rules on general-purpose AI systems that need to comply with transparency requirements will come into effect after six, nine and 12 months respectively. On the other hand, the obligations concerning high-risk systems will come into effect after thirty-six months, thus giving them more compliance time.

According to [Symbio6 \(2024\)](#), the main criticism regarding the AI Act is its complexity and lack of clarity. The definition of an AI system in the AI Act (Article 3.1) means a machine-based

system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. They contend that this definition is largely vague, leading to uncertainties about its scope and application, which could hinder effective implementation and compliance.

Furthermore, [Symbio6 \(2024\)](#) claim that even though the AI Act makes sixteen references to human vulnerability, the concept of human vulnerability is still difficult to understand because the words like “vulnerability” are not precisely defined. It is also argued that the Act might lead to overregulation for the business community, which might hinder technological innovation and establish a globally limiting standard. There is also some concern regarding the EU’s decision not to completely ban public mass surveillance. However, this decision is viewed as being based on the individual security needs of member states. Additionally, [Symbio6 \(2024\)](#) contend that while the AI Act has the objective of safeguarding fundamental rights and ensure transparency, human rights organisations argue that the act should more robustly tackle the potential harms of AI, including invasive surveillance and social scoring systems, which could undermine personal freedoms and privacy.

11.3.1.2 United Kingdom (UK)

According to [Tobin \(2024\)](#), the AI (Regulation) Bill was proposed as a private member’s bill by Lord Holmes with the chief objective of establishing a new body, namely the AI Authority. This authority would have various functions intended to facilitate AI regulation in the UK. The AI Authority is to ensure that relevant existing regulators take into consideration AI developments to guarantee an alignment approach is applied between these regulators and to undertake a gap analysis of regulatory responsibilities regarding AI. Other functions proposed for the AI authority include monitoring economic risks arising from AI; conducting horizon-scanning of developing technologies; facilitating sandbox initiatives to allow the testing of new AI models; and accrediting AI auditors.

The proposed Act would also institute several regulatory principles to govern the development and application of AI. The bill to establish an AI Authority is not an Act to regulate AI because regulating AI in the present technological environment is viewed as premature and may be counterproductive. Hence, it is argued that existing sectoral regulators are in a better position to regulate AI with the backing of a central function currently in development within the Department of Science, Innovation and Technology ([Tobin, 2024](#)).

According to [Gallo and Nair \(2024\)](#), the UK Government has opted for a cross-sector and outcome-based framework for regulating AI, which is supported by five fundamental principles. The fundamental principles are safety; security and robustness; appropriate transparency and explainability; fairness; accountability and governance; and contestability and redress. These five cross-sectoral principles are intended for the existing regulators to interpret and use within their individual area of interest as a guide to ensure responsible AI design, development, and application. [Gallo and Nair \(2024\)](#) contend that the UK Government’s strategy for implementing these principles is based on three core pillars, namely:

- Leveraging existing regulatory authorities and frameworks. The UK Government does not intend to establish a new AI regulator to oversee the implementation of the framework but is mandating existing regulators to put into action the five principles to regulate and supervise AI within their respective domains. Regulators, such as Information Commissioner's Office (ICO), Office of Communications (Ofcom), or the Financial Conduct Authority (FCA) are required to apply a harmonious context-based approach, making use of the existing laws and regulations. Hence, these regulators are not legally obliged to implement the principles or to work together. However, government will likely introduce a legislation in the future, but this will depend on the regulators' strategic AI plans and the outcome of a review to be carried out by the Central Function regarding likely regulatory powers and remits.
- Establishing a central function to facilitate effective risk monitoring and regulatory coordination. The government recognises that due to the extensive impact of AI, individual regulators will not be able to effectively deal with the opportunities and risks posed by AI technologies in isolation. Therefore, a Central Function will be established within Department of Science, Innovation and Technology (DSIT) to administer and assess AI risks, promote coherence and address regulatory gaps.
- Supporting innovation by piloting a multi-agency advisory service, specifically the AI and Digital Hub. The UK Government will establish a pilot multi-regulator advisory service known as the AI and Digital Hub. This service has the aim of assisting innovators to navigate multiple legal and regulatory obligations before launching a product. This Digital Hub has the remit to facilitate compliance and encourage greater collaboration between regulators and help companies meeting the eligibility criteria.

Regulators will be required to implement the framework in their respective domains by applying existing laws and issuing additional regulatory guidance for organisations to follow. The framework and endeavours of the individual regulators will be enhanced by the voluntary safety and transparency measures that are aimed for developers of highly capable AI models and systems. The Government does not intend to codify the framework into law for the present. However, it plans to make legislative interventions in the future, especially regarding the risks presented by complex General Purpose AI and the leaders involved in its development.

While this approach has some support from the industry, there are others that have expressed concerns regarding the reliance on voluntary commitments from major AI developers instead of having obligatory legal requirements. However, the current arrangement is viewed as an initial step. [Gallo and Nair \(2024\)](#) argue that the challenge is to ensure its effective implementation across various regulatory domains, including data protection, competition, communications, and financial services. They contend that given the variety of regulatory bodies involved, effective governance of the framework is essential. They view the newly established Central Function as having a key role in guaranteeing effective collaboration between regulators and easing possible negative impact of conflicting interpretations. They further contend that the Central Function will be vital for supporting organisations with the obligatory regulatory transparency to adopt and scale their investment in AI, thus strengthening UK's competitive edge.

The UK approach is in direct contrast with the EU's method that is based on wide-ranging legislation to regulate the development and application of AI across all member states. However, the UK approach has some similarity to the USA's federal system where no AI legislation is currently proposed. Having said this, some USA States have currently initiated legislation aimed at the regulation of AI specifically related to privacy and accountability. The USA AI legal scenario will be discussed separately at a later stage.

Unlike the EU approach, the UK framework does not provide a formal definition of AI. However, the outcomes-based approach that focuses on two defining characteristics, namely adaptivity and autonomy are aimed at guiding sectoral interpretations. For example, UK regulators, like the ICO, Ofcom or FCA, can interpret adaptivity and autonomy to create domain-specific definitions, if beneficial ([Gallo and Nair, 2024](#)). [Gallo and Nair \(2024\)](#) contend that this increases the risk of diverse interpretations by regulators that might create regulatory uncertainty for organisations operating across sectors. To mitigate this risk, the Government intends to institute a Central Function as a coordinating body between regulators. It is noted that the conventional definition of AI is maintained, with the framework classifying AI into three types, namely highly capable general-purpose AI; highly capable narrow AI; and AI agents. The framework also provides a definition of each type.

11.3.2 North America

The section on North American will focus on the United States of America and Canada.

11.3.2.1 United States of America (USA)

Generally, USA applies a decentralised model of AI regulation as its overall approach to governance. Most of its regulatory practices and policies are focused on sectoral levels that reflects very closely the AI segmentation. Moreover, USA has no specific wide-ranging federal regulation framework regarding AI. Having said this, USA has launched several sector-specific AI-related agencies and institutions that deal with some of the complications resulting from the fruition of AI. For example, the Federal Trade Commission (FTC) specifically focuses on the issue of consumer protection and attempts to apply fair and transparent business practices in the field. Similarly, the National Highway Traffic Safety Administration (NHTSA) regulates the safety aspects of AI-powered technologies, particularly their application in autonomous cars.

Although the USA AI legislation lacks centralisation, the National Artificial Intelligence Initiative Act of 2020 established an American AI Initiative and guidance on AI research, development, and evaluation activities at federal science agencies. Further acts have required specific agencies to guide AI programmes and policies throughout the federal government (e.g., AI in Government Act of 2020 and Advancing American AI Act of 2021).

The number of bills passed by Congress since 2015 were nine, with 33 legislations still pending up to November 2023. According to [Szczepański \(2024\)](#), during 2021, Congress introduced at least 75 bills that focused on either AI and machine learning or related provisions but only six of these were enacted. He further maintains that between January and June 2023, Congress had introduced at least 40 AI-relevant bills, none of which has been enacted. This

seems to suggest that at a federal level, the US Government is reluctant to introduce comprehensive AI regulatory measures.

[Szczepański \(2024\)](#) notes that in January 2023 the White House Office of Science and Technology Policy published its proposal for an AI Bill of Rights and the National Institute of Standards and Technology released an AI Risk Management Framework. In mid-2023 two broad policy frameworks were announced to guide the Congress in developing future AI legislation. These included the SAFE Innovation Framework for AI Policy (Security, Accountability, protecting Foundations, Explainability) and Blumenthal & Hawley Comprehensive AI Framework (known as the Bipartisan Framework for US AI Act). Additionally, in April 2023 four federal agencies jointly emphasised that their enforcement powers incorporated AI, and that advanced technology should not be a pretext for breaking the law.

On 30 October 2023, US President Joe Biden issued an executive order regarding the safe, secure, and trustworthy development and use of AI. This expands on earlier work, such as the executive order that directed agencies to combat algorithmic discrimination and the securing of voluntary commitments from major US companies (such as Amazon, Google, Meta, Microsoft, and OpenAI) to drive safe, secure, and trustworthy AI development. The executive order of 2023 embraces several policy areas, namely:

- To establish new standards for AI safety and security. Developers of the most powerful AI systems are required to share their safety test results and other critical information with the US government. Additionally, government agencies have the responsibility to develop standards, tools and tests to help ensure AI systems are safe, secure, and trustworthy. These new standards aim to protect against the risks of applying AI to engineer dangerous biological materials and protect US citizens from AI-enabled fraud and deception.
- To establish an advanced cybersecurity programme to create AI tools that uncover and resolve weaknesses in critical software. The National Security Council and White House Chief of Staff are responsible with the development of a national security memorandum to direct further action on AI and security.
- To protect citizens' privacy from AI-related risks. Federal support is to give priority to hasten the expansion and application of privacy-preserving techniques. This is aimed at strengthening applicable research and technologies; emphasises the need for the evaluation of how agencies collect and apply commercially available data; and makes it necessary to develop guidelines for federal agencies to assess the effectiveness of privacy-preserving techniques.
- To advance equity and civil rights. The executive order demands clear guidance for property-owners, federal benefits programmes and federal contractors, and processes to address logarithmic discrimination and fairness certainty throughout the criminal justice system by developing best practice on AI use.
- To protect consumers, patients and students. The executive order demands that measures are in place to advance the responsible application of AI in healthcare and shape its potential to transform education.

- To support workers. This solicits the development of principles and best practice to minimise harm from AI and maximise its benefits. It also calls for a report on the possible impact of AI on the labour-market and on approaches to ease this impact through stronger federal support.
- To promote innovation and competition. This strives to stimulate research across the US and requests new measures to support fair, open, and competitive AI ecosystems. It also calls for efforts to attract highly skilled immigrants and non-immigrants with expertise in critical areas to study, stay and work in the US. To progress US leadership abroad, it also calls for the expansion of bilateral and multilateral AI engagements, accelerated development and implementation of vital AI standards, and promotion of the safe, responsible, and rights-affirming development and deployment of AI abroad to address global challenges.
- To ensure responsible and effective government use of AI. The executive order pursues new support for agencies using AI, more proficient and speedy contracting to obtain AI products and services, and government-wide hastening in employing AI experts.

The executive order was supported by a comprehensive implementation guidance from the Office of Management and Budget. The AI industry has generally applauded the executive order, with some reservations regarding its heavy handedness, but various experts view it a good first step that requires major implementation challenges. [Szczepański \(2024\)](#) notes that observers have characterised this executive order as “landmark” and “sweeping” with a strong bipartisan support for its main provisions. However, he also remarks that the Centre for Strategic and International Studies opinions that unlike the EU approach, the most likely outcome over the next few years “is a bottom-up patchwork quilt of executive branch actions.”

On the other hand, the Brookings Institution sees the executive order as sending a strong signal to the international community that the US is finally taking a stand on AI, however, data privacy legislation is necessary to create effective and resilient AI ([Szczepański, 2024](#)). According to [Szczepański \(2024\)](#), the Trade and Technology Council (TTC) provides the setting for EU and US cooperation concerning AI and are implementing a joint roadmap on evaluation and measurement tools for trustworthy AI and risk management. He contends that: (i) the TTC expert groups have itemised 65 key AI terms essential to understanding risk-based approaches to AI, accompanied by their EU and US interpretations and shared EU-US definitions; (ii) they have mapped the respective involvement of both the EU and the US in standardisation activities, aiming to identify relevant AI-related standards of mutual interest; and (iii) they are compiling a catalogue of existing and emerging risks, including an understanding of the challenges posed by generative AI. Additionally, in January 2023, the EU and the US signed an administrative arrangement on “AI for the Public Good”, aimed at boosting AI research collaboration ([Szczepański, 2024](#)).

At a state level, [NCSL \(2024\)](#) reports that in 2023 at least 25 states, Puerto Rico and the District of Columbia introduced AI bills, and 18 states and Puerto Rico adopted resolutions or enacted legislation. [Table 11.2](#) provides the legislation proposed by the various US jurisdictions, illustrating that New York state has 24 pending legislations, California having enacted two laws

and has ten pending. It is interesting to note that 24 jurisdictions have no legislative action regarding AI. In this regard, the state of Delaware is of particular interest in that it is a leading domicile for the US and international corporations with more than 1,000,000 business entities and over 66% of the Fortune 500, yet it has no apparent plans to initiate AI laws.

TABLE 11.2 Legislative Action by US Jurisdictions [↗](#)

State	Enacted	Adopted	Vetoed	Failed	Failed – Adjourned	Pending	Pending – Carryover	Total
New York						24		24
California	2					10		12
Illinois	1					8		9
Massachusetts						9		9
Pennsylvania						9		9
Texas	2				6			8
Rhode Island		1				6		7
Maryland	1			1	4			6
North Carolina	2					3		5
New Hampshire						4		4
New Mexico	1				3			4
Georgia	2	1						3
Michigan	2						1	3
Puerto Rico		1				2		3
Washington							3	3
Wisconsin						3		3
Arizona	1		1					2
Hawaii		2						2
Maine							2	2
Minnesota	1						1	2
Missouri					2			2
Nevada			1		1			2
North Dakota	2							2
Arkansas					1			1
Colorado					1			1
Connecticut	1							1

State	Enacted	Adopted	Vetoed	Failed	Failed – Adjourned	Pending	Pending – Carryover	Total
District of Columbia						1		1
Louisiana		1						1
Mississippi	1							1
Utah	1							1
Vermont							1	1
West Virginia	1							1
Total:	21	6	2	1	18	79	8	135

The following 24 jurisdictions had no legislative action regarding AI:

A. Samoa; Alabama; Alaska; Delaware; Florida; Guam; Idaho; Indiana; Iowa; Kansas; Kentucky; Montana; N. Mariana Islands; Nebraska; New Jersey; Ohio; Oklahoma; Oregon; South Carolina; South Dakota; Tennessee; U.S. Virgin Islands; Virginia; and Wyoming.

Data obtained from: [NCSL \(2024\)](#). Artificial Intelligence 2023 Legislation.

Additionally, [Figure 11.1](#) illustrates the general legislation status for the 32 states that have laws in the pipeline. It is noted that 20% of the legislative pipeline have been enacted (16%) and adopted (4%). It is also noted that 65% are either pending (59%) or pending carryover (6%). This implies that there will be many new enacted and adopted legislation regarding AI in the short-term. Furthermore, [Figure 11.2](#) provides the various legislative categories, including each category's definition. [Figure 11.2](#) illustrates that 22% of all legislations by the US state and territories jurisdictions are regarding government use, followed by private sector use (17%) and responsible use (13%). In addition, [Table 11.3](#) provides several examples of the enacted legislation and adopted resolutions by some US jurisdictions.

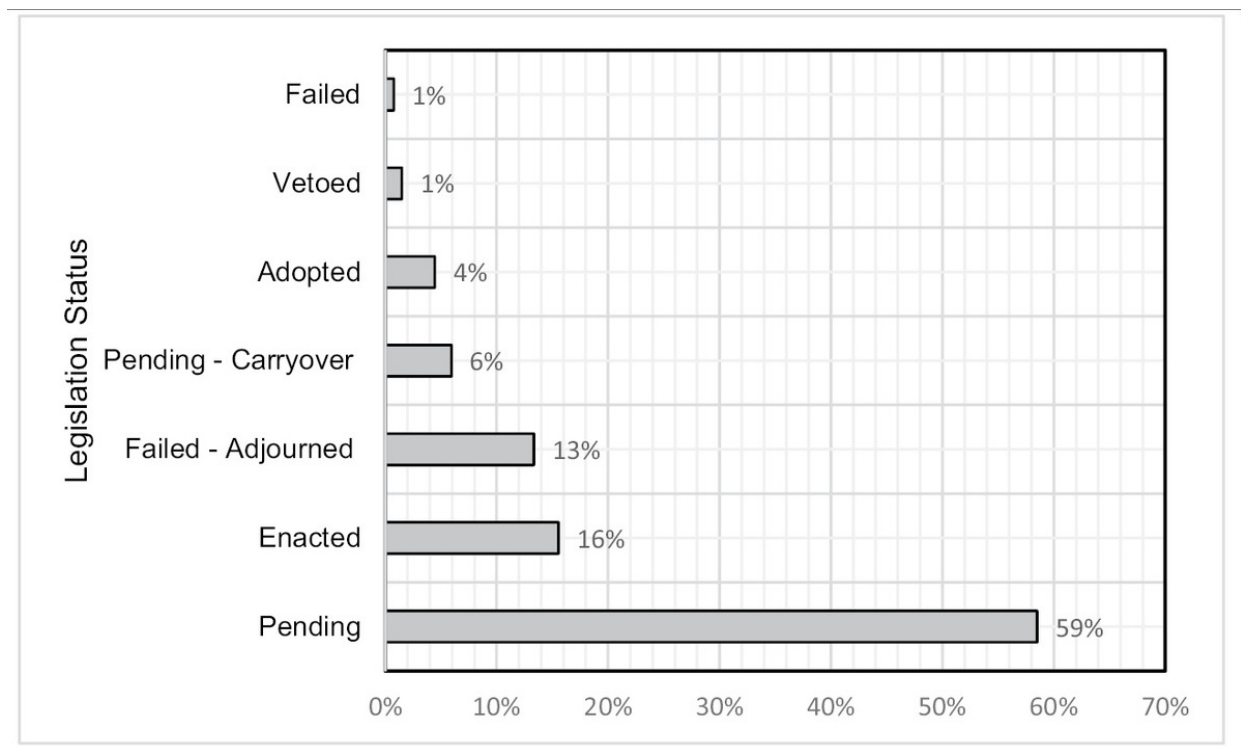


FIGURE 11.1 Legislation Status for US States Jurisdictions. (Data obtained from: [NCSL \(2024\)](#). Artificial Intelligence 2023 Legislation.)[📄](#)

State	Description
Connecticut	This legislation obliges the state Department of Administrative Services to conduct an inventory of all systems that employ artificial intelligence and are in use by any state agency and, beginning Feb. 1, 2024, perform ongoing assessments of systems that employ AI and are in use by state agencies to ensure that no such system shall result in unlawful discrimination or disparate impact. Further, the legislation requires the Office of Policy and Management to establish policies and procedures concerning the development, procurement, implementation, utilization and ongoing assessment of systems that employ AI and are in use by state agencies.
Louisiana	Louisiana adopted a resolution requesting the Joint Committee on Technology and Cybersecurity to study the impact of artificial intelligence in operations, procurement and policy.
Maryland	This established the “Industry 4.0 Technology Grant Program” to assist certain small and medium-sized manufacturing enterprises with implementing new “industry 4.0” technology or related infrastructure. The definition of “industry 4.0” includes AI.
North Dakota	This legislation defines a person as an individual, organisation, government, political subdivision, or government agency or instrumentality, and specifying that the term does not include environmental elements, AI, an animal or an inanimate object.
Texas	Texas created an AI advisory council to study and monitor AI systems developed, employed or procured by state agencies, with North Dakota, Puerto Rico and West Virginia also creating similar councils

Source: NCSL (2024). Artificial Intelligence 2023 Legislation.

[NCSL \(2024\)](#) claim that state lawmakers are taking into consideration AI’s benefits and challenges with a growing number of measures being introduced to study the impact of AI or algorithms and the potential roles for policymakers.

11.3.2.2 Canada

According to [White and Case \(2024a\)](#), Canada is to regulate AI at the federal government level, through the Artificial Intelligence and Data Act (AIDA) that was introduced in June 2022 and effectively passed its second reading. Subsequently this Act was referred to the Standing Committee on Industry Science and Technology on April 24, 2023. However, the Minister of Innovation, Science and Industry in November 2023 presented the Standing Committee with several amendments since the original version had limited substantive content that left most key elements of the legal regime to be set out later in regulations. These proposed amendments have not been adopted, and it is unlikely that the Act will be passed before October 2025.

The proposed Act will also include the Consumer Privacy Protection Act, which is a revision of the current federal privacy law and the Personal Information and Data Protection Tribunal Act.

However, unlike the USA situation, no laws have been proposed to directly regulate AI at the provincial level. On the other hand, the Innovation Council of Quebec, which has the remit from the government of Quebec, has recommended specific regulation of AI through a provincial legislature of a law. Additionally, [White and Case \(2024a\)](#) contend that there are several other laws that may impact the development and/or application of AI in Canada, whether federal, provincial or territorial, these include:

- At federal level: Privacy Act and the Personal Information Protection and Electronic Documents Act (PIPEDA).
- At a provincial level: Personal Information Protection Act (in Alberta); the Personal Information Protection Act (in British Columbia); and the Act Representing the Protection of Personal Information in the Private Sector (in Quebec). The revisions of September 2023 to the Quebec Act regulate automated decision-making based on the processing of personal information and require disclosure of the occurrence of processing and the provision of information about the reasons and factors that have led to a decision.
- Intellectual property laws may impact aspects of AI development and application.
- Competition law might impact the structure of the AI market in Canada.
- The Innovation Council of Quebec has also proposed amending the Quebec labour law to address the impacts of AI.
- Existing human rights and anti-discrimination laws (including the Canadian Human Rights Act) regulate discriminatory outcomes and practices, such as those that may result from biased outputs and decisions from AI systems.

Generally, AIDA has the objective to standardise cross-border trade in AI systems by instituting common obligations applicable across Canada, for the design, development, and application of those systems; and to forbid any AI systems behaviour that may cause serious harm to individuals or harm to their interests, particularly biased outputs. AIDA has wide territorial domain and is applicable regarding: (a) AI systems or machine learning models made available during international or interprovincial trade and commerce (cross-border trade); and (b) the management of the operations of AI systems used during cross-border trade ([White and Case, 2024a](#)).

In general, AIDA does not adopt a sector-specific approach but imposes various responsibilities on specific persons contingent on the kind of AI system that they are involved with and their status down the AI value chain. For instance, individuals who manage, or who make available, or make available for the first time during cross-border trade of general-purpose AI systems and high-impact systems have obligations under AIDA. This also applies to those individuals who first make a machine-learning model available, for incorporation into a high-impact system, during cross-border trade.

AIDA embraces several key definitions, namely: (a) an AI system means a system that, using a model, makes inferences in order to generate output, including predictions, recommendations or decisions; (b) a General-purpose system means an AI system that is designed for use, or that is designed to be adapted for use, in many fields and for many purposes and activities, including

fields, purposes, and activities not contemplated during the system's development; and (c) Machine-learning model is a digital representation of patterns identified in data through the automated processing of the data using an algorithm designed to enable the recognition or replication of those patterns ([White and Case, 2024a](#)). Moreover, the proposed amendments explain that an AI system may be a general-purpose system and a high-risk system concurrently. AIDA incorporates a risk categorisation concept. Hence, an AI system will be a "high-impact system" if it has the following application:

- To decide on employment issues such as to ascertain employment levels, recruitment, referral, engagement, remuneration, promotion, training, apprenticeship, transfer or termination.
- To adopt decisions associated to service access such as to establish whether to supply services to an individual, or the type or cost of services to be given to an individual; or to prioritise services to be made available to individuals.
- To process biometric information for the identification of individuals, without their consent; or the evaluation an individual's behaviour or frame of mind.
- To make decisions concerning content moderation and prioritisation, namely for the moderation of content acquired on an online communications platform, including a search engine or social media facility, or the prioritisation of the presentation of such content.
- To decide on matters regarding healthcare or emergency services, with certain exceptions, in accordance with the Food and Drugs Act.
- In determinations by a court or administrative body concerning an individual who is a party to the proceedings.
- To assist law enforcement officers, as stated in the Criminal Code, in the exercise and performance of their law enforcement powers, duties, and functions.

According [White and Case \(2024a\)](#), AIDA implements a comprehensive method regarding compliance requirements, particularly for those relating to risk management measures, which are specific to certain roles in the AI value chain. They cite two specific examples:

- Persons who originally make a machine-learning model accessible for incorporation into a high-impact system during cross-border trade (i.e., developers) must define processes that identify, assess, and mitigate against the risks of biased output before the model is made available.
- Persons who originally make high-impact systems available must make sure that risk mitigation measures are implemented, and that these measures have been tested; and those who manage such systems are responsible for conducting the tests to ascertain the effectiveness of the mitigation measures.

Furthermore, White and Case contend that AIDA obligations are styled according to the different challenges and objectives at each point in the AI value chain. For instance, persons who initially make a machine-learning model accessible for incorporation into a high-impact system in the course of cross-border trade, must guarantee that measures respecting the data used in

developing the model have been instituted in accordance with regulations; and that high-impact system operators must institute procedures that permit users to provide feedback on the system's performance.

An important requirement of AIDA is the creation and maintenance of written accountability frameworks by those individuals who make a general-purpose system or high-impact system available, or who manage the operations of such systems. These frameworks apart from other points, are to include: a description of the roles, responsibilities, and reporting structure for all those who contribute to making the system available or managing its operations; the policies and procedures respecting the management of risks relating to the system and the data used by the system; and any other items required by regulation. It is also noted that the AI and Data Commissioner is appointed by the Minister responsible for the Act and has the remit for enforcing AIDA. AIDA also imposes different penalties for individuals and companies for certain offences that may result in substantial fines or imprisonment.

11.3.3 Africa

According to [Sulaiman, Olën and Bezuidenhout \(2024\)](#), several African jurisdictions are currently implementing AI strategies with some being in the consultation process with stakeholders, while others have yet to make any pronouncements regarding AI regulation. Furthermore, they also refer to a draft AI policy of the African Union (AU) related to AI regulation that is expected to be officially endorsed by members in 2025. It is noted that the AU currently has 55 members that incorporate most African states. [Sulaiman, Olën and Bezuidenhout \(2024\)](#) maintain that the AU Draft Policy incorporates recommendations for industry-specific codes and practices, standards, and certification bodies, which are focused on evaluating and benchmarking AI systems, providing regulatory oversight for the testing of AI, and establishing national AI councils.

Hence, this framework provides the opportunity for those members without AI policies or regulation to catch up with the more advanced members and encourages all members to review and align their existing policies with the AU Draft Policy to ensure a degree of consistency among AU members. Mauritius, Kenya, and Egypt lead as the African AI policymakers, while South Africa and Nigeria, are consulting stakeholders about initiating AI policies

11.3.3.1 Mauritius

The Mauritian Government set up a working group on AI in recognition of the opportunities that AI may provide to help stimulate the country's economy as well as a means of tackling the financial and social issues being encountered. This work group formulated and published the Mauritian AI Strategy in 2018 after due consultations with stakeholders.

This strategy defined three key requirements, including identifying possible applications and impact of AI; the creation of an ecosystem to encourage AI development; and the formation of a regulatory framework for AI oversight. Additionally, the Mauritian AI Strategy acknowledges the capacity of AI technologies to integrate into the broader nation's economy and support productivity, economic growth, and quality of life. The Mauritian AI Strategy is intended to be

upgraded and augmented in the medium and long term to emerge as a key pillar for sustainable growth and economic development in Mauritius. The strategy is focused on:

- Concentrating on specific sectors and identifying national projects.
- Developing capacity and attracting skills.
- Having incentives to catalyse implementation.
- Contemplating ethical concerns of AI.
- Developing strategic alliances in emerging technologies.
- Increasing awareness of AI through focused campaigns.
- Implementing new technologies to enhance public services delivery.

Additionally, several projects are proposed to support the AI Strategy that include making use of blockchain technology for tax payment, land registration and managing health insurance; using AI to enhance the energy sector; using AI to bolster efficiency and competitiveness in the manufacturing sector; using AI for leapfrogging in the healthcare and biotech industry; and employing AI in agro-industry, the ocean economy, and for transport.

The AI Strategy emphasises several issues concerning the legal and regulatory framework, ethics and data protection. Moreover, it recommends government to expedite the Intellectual Property Bill; update outdated regulations to permit the development of AI, such as in Fintech, mobile payment, crowdfunding and crowdlending; guarantee an ethical AI system by instituting a permanent committee on ethics; set up a special working group to make the necessary amendments to data protection legislation; and consider policies on open data access.

11.3.3.2 Kenya

Like Mauritius, the Government of Kenya also acknowledged that AI applications might provide economic opportunities. Hence, in 2018 it formed the Distributed Ledgers Technology and AI Task Force. In 2019, this Task Force released a report that advocated the potential of AI to contribute to the growth of Kenya's global economic competitiveness. The Task Force report generated the basic ideas to position the country as a regional and international leader in Information and Communications Technology (ICT) by promoting a substantial escalation of investment in infrastructure and skills. It also proposed strategies to harmonise innovation and regulation in the private sector.

A National Digital Master Plan 2022-2032 has been presented by the Minister for ICT that accepts the critical points of the Task Force Report and endorses its support for fostering research and development in the AI industry. Moreover, the master plan defines the Ministry's ICT strategy to implement AI solutions that are aimed at overcoming the challenges for deploying AI and its regulation and disseminating the AI solutions to other countries. The master plan also stresses the significance of developing international partnerships with leading actors in the AI industry, including those in the emerging technologies space.

Additionally, significant progress was made in 2024 for finalising a national AI strategy for Kenya through the strategic alliance of the Government of Kenya and the German Development Cooperation (GIZ). This alliance resulted in initiating the Forward Artificial Intelligence for All

(FAIR) Initiative, which is also supported by the German Federal Ministry for Economic Cooperation and Development and the European Union.

11.3.3.3 Egypt

Like Kenya, Egypt emphasises the need to facilitate technology transfers and attracting foreign investments. The Egyptian National Council for AI (NCAI) formulated the national Egyptian AI Strategy that supports Egypt's implementation of AI in specific phases. The first phase of the AI strategy has the objective of launching Egypt as a key role player in enabling regional cooperation in AI. Hence, this phase accentuates the growth of cooperative AI capabilities between Egyptian and foreign stakeholders, in preference to its national or internal development.

Exploiting Egypt's internal AI growth potential will be pursued in later phases of the AI Strategy. Having said this, discussions related to the second phase have already commenced with the Minister of ICT and the NCAI. The second phase will be inclined towards commencing several priority initiatives identified in the first phase that are intended to enhance investment, increase public awareness, and developing greater AI capabilities in Egypt.

There does not appear to be any direct laws in Egypt regarding AI. However, AI is indirectly governed through existing laws, such as the Telecommunication Regulation Law (2003). This law is mainly focused on telecommunication services, but also deals with digital data management, which is viewed as an essential component of AI operations. Additionally, the Egyptian Data Protection Law of 2020 plays a crucial role in regulating AI. This law is modelled on the European Union's General Data Protection Regulation (GDPR), which establishes standards for data privacy and security, directly impacting AI systems that process vast amounts of personal data. This law also makes it mandatory to obtain consent for data collection, which is an important factor in AI's functionality, particularly for consumer-facing applications.

According to [Iskander \(2024\)](#), as AI in Egypt continues to evolve, there is a need for ongoing legal and regulatory development. He maintains that this includes updating existing laws and introducing new regulations that address the unique challenges posed by AI technologies. [Iskander \(2024\)](#) argues that embracing AI's potential while mitigating its risks requires a balanced approach, combining legal rigor with ethical considerations and global best practices.

11.3.3.4 South Africa

Generally, South Africa, Nigeria, and Morocco rely on existing legislation for AI regulation like Egypt. According to [White and Case \(2024b\)](#), South Africa has not announced any AI regulation proposals but is in the consultation process for a draft National AI plan. Generally, AI is unregulated in South Africa. However, certain aspects of existing legislation regulate some activities conducted by organisations using AI, including the Protection of Personal Information Act (POPIA), the Copyright Act, the Patents Act, and the Competition Act.

[White and Case \(2024b\)](#) maintain that South Africa has not released any policy documents or entered bills to parliament for the regulation of AI. However, White and Case identify the following developments that are viewed as significant in considering any expected future regulation:

- April 2019: The President appointed members to the Presidential Commission on the Fourth Industrial Revolution (4IR Commission) to identify relevant policies, strategies, and action plans that will position South Africa as a competitive global player.
- May 2019: OECD and partner countries, including South Africa, formally adopted the first set of nonbinding intergovernmental policy guidelines on AI agreeing to support international standards that aim to ensure AI systems are designed to be robust, safe, fair, and trustworthy.
- November 2022: The Department of Communications and Digital Technologies launched the AI Institute of South Africa and AI hubs (University of Johannesburg and Tshwane University of Technology).
- April 2024: The Minister of the Department of Communications and Digital Technologies organised an AI Summit to reveal the contents of a draft National AI Plan discussion document, intended as a pre-policy process to obtain and collate inputs to be incorporated into a draft plan.

Like other African countries, there are several laws that do not directly regulate AI but might impact the development or use of AI in South Africa. [White and Case \(2024b\)](#) provide some key examples as follows:

- Protection of Personal Information Act that encompasses any automatic processing of personal data.
- Copyright Act and Patents Act that are applicable to AI generated efforts, one of which has been granted a patent in South Africa.
- Competition Act that although it does not directly regulate the technology is applicable to digital mergers and digital platforms.
- A Draft Cyber Security Bill was published in 2024, which includes AI in the definition of a regulated ICT System and ICT Infrastructure and contains provisions that may require AI products to be certified against minimum standards.

Since South Africa does not have any laws to directly regulate AI there is no single legally recognised definition of AI. However, the draft National AI Plan discussion document refers to the European Union Commission definition. Furthermore, the lack of AI legislation means that currently there are no specific or unique obligations imposed on developers, users, operators and/or deployers of AI systems, apart from those that might relate to the automated processing of personal information as provided for in the Protection of Personal Information Act (POPIA). However, Section 71(1) of POPIA governs automated decision-making that protects data subjects from being subjected to decisions based exclusively on automated decision-making, which results in legal consequences for the data subject and the data subject being profiled.

[White and Case \(2024b\)](#) remark that South Africa's AI Planning Discussion Document notes several regulatory areas that are to be eventually considered:

- Disinformation and fake news created by Generative AI applications and software.
- Bias and discrimination that could be demonstrated by AI developers and users.

- Copyright abuse in alignment with World Intellectual Property Organisation.
- Privacy to protect personal, private, and public data.

White and Case also remark that the AI Planning Discussion Document lacks a risk-based methodology but instead focuses on considering the optimal form of regulation regarding:

- Anti-competitive behaviour if provisions of key AI models, or other enablers are focused on a small number of players.
- Risks in robotic or autonomous devices that apply AI to determine actions, such as AI driven autonomous vehicles.
- Insistent employment losses that lead to social risks.
- Criminal behaviour or other highly dangerous outcomes.
- Experiential risks in situations where AI gets out of control and engages in goals harmful to humanity.
- Risks that AI intended for military purposes outflows into the wider world and leads to greater criminal conduct or other highly dangerous outcomes.

South Africa is enlarging its AI knowledge centres through the South African Centre for Artificial Intelligence Research (CAIR) that incorporates eight universities conducting AI research in South Africa. The funding for CAIR activities is mainly supported by the South African Department of Science and Innovation and by the AI Institute of South Africa.

The above illustrates that several African countries are active contributors in developing a balance between AI regulation and innovation. There are also several African countries that are defining and implementing AI strategies and policies for public sector reform, education, and research. However, although the African AI innovation ecosystem is active, there is a lack of direct AI laws in most jurisdictions. Having said this, the AU Draft Policy, once agreed by members, may empower AU member states to leapfrog AI regulatory development, including those in the developmental and experimental stages of their AI regulation.

11.3.4 South America

The trend in South American countries is to actively develop AI regulations to support ethical AI development, protect human rights, and foster innovation. Generally, many South American countries are proactively seeking legislative action with several bills having been introduced in respective parliaments and legislative bodies. This suggests that there is an increasing awareness and commitment towards instituting a legal framework for effectively regulating AI. Furthermore, these legislative endeavours are focused to address the socioeconomic, educational and political challenges of the region.

[Muñoz \(2024\)](#) contends that the recent elections in Latin America have underscored the dangerous impact of algorithms and AI, particularly generative AI, at the voting booth. He argues that during the 2023 Colombian regional election, campaigns were tarnished with misinformation channelled across social media, with the emergence of AI-generated media content or deepfakes, which were targeted at undermining candidates' political campaigns. Muñoz claims that a similar

deception was launched during Argentina's 2023 election, where candidate Patricia Bullrich was the focus of AI-generated audios smearing her as minister of economy. This was also experienced in the recent Mexican election, where a deepfake audio of candidate Claudia Sheinbaum imitated her criticising President Andrés Manuel López and admitting deception in the polls.

In October 2023 a high-level AI summit was conducted in Chile that included key government authorities from across South America and the Caribbean. The Santiago Declaration was an important outcome of this summit that denotes a significant organisational shift in the region. The Santiago Declaration underlines a determined endeavour for a unified development of a governance and regulatory frameworks that is adapted to the unique needs and context of the region. A significant outcome of this summit was the suggestion to set up an intergovernmental Council on AI for Latin America and the Caribbean that is in line with UNESCO's proposal on the AI ethics that aims to reinforce regional capabilities in AI, promoting an environment where ethical considerations and human rights are at the forefront of AI development and implementation. However, despite this declaration many South American countries are pursuing their own individual AI strategies to address their own particular challenges. This section will discuss several South American countries to determine the status regarding AI regulation in this region.

11.3.4.1 Argentina

In October 2023, Argentina presented a Bill of Law on AI Regulation in the Chamber of Deputies that defines AI specific terms; establishes guiding principles; itemises exclusions; and proposes an administrative entity for AI within the National Executive Branch. Furthermore, several bills were initiated in late 2023 to amend the Penal Code so that AI-related crimes against sexual integrity are punished.

A Bill was also presented in the Senate obliging organisations engaged in AI to enrol in a National Registry of AI Systems so that the impact of an AI system may be determined and classified by risk levels ranging from limited to unacceptable as follows:

- Unacceptable risks: These are risks that are unacceptable because of their seriousness or because of their impact on fundamental rights or the safety of individuals. An AI system with an unacceptable risk will require the implementation of suitable mitigation measures or the banning of the system.
- High risks: These are risks that have a significant impact on fundamental rights or the safety of individuals. The existence of high risk in an AI system will require the adoption of appropriate mitigation measures.
- Limited risks: These are risks that have an insignificant impact on basic rights or the safety of individuals but still need mitigation measures.
- Negligible risks: These are risks that have a minimal impact on fundamental rights or the safety of individuals.

The draft law proposes the following values and principles:

- Diversity and inclusion: The development of AI must ensure the involvement of all persons or groups, regardless of race, colour, descent, gender, age, language, religion, political opinion, ethnic or social origin, economic status or any other form or condition that would disrupt respect for diversity and social inclusion.
- Peace and justice: Advance in AI must safeguard peace and justice among nations and must not weaken human freedom or lessen human autonomy of choice. Nor should it endanger individual or collective security, divide or confront individuals and groups, or threaten coexistence between humans, other living beings and the environment.

The proposed law is not clear about how the AI systems are to be evaluated for their likely impact on peace and social justice, nor what methods are to be implemented to ensure the involvement of all people in order to promote respect for diversity and social inclusion. The draft law sets out principles and values regarding respect for diversity and the promotion of non-discrimination, as well as respect for human freedom and autonomy, but does not go into any great depth. However, it gives the Scientific and Technological Cabinet the authority to pause the progress of AI research, development or application when it deems that principles and values are infringed. Furthermore, the Scientific and Technological Cabinet will also be in charge for receiving complaints about threats or harm caused by these technologies. Guidelines for developing AI systems were issued by the Undersecretariat of Information Technologies regarding adherence with democratic principles and individual rights. These guidelines embraced various principles, such as transparency, explainability, safety, privacy non-discrimination, and data protection.

The responsibility for the promotion of transparency and data protection in AI use has been assigned to the Agency for Access to Public Information, which initiated the “Program for Transparency and Protection of Personal Data in AI.” This program has the objective of analysing, regulating, and strengthening state capacities for the development of AI applications in both the public and private sectors. It also has the aim of defining good practices, learning, and proposing recommendations on transparency, algorithmic transparency, and data protection in AI use. Additionally, the coordination of AI initiative is entrusted to the Interministerial Roundtable on AI, which was set up by the National Executive Branch by incorporating various government agencies. To summarise, Argentina experienced the emergence of AI as a significant topic during the recent presidential elections because of the intense exploitation of AI technology in campaigns. However, after the election the issue has not been given a priority on the country’s political agenda.

11.3.4.2 Brazil

Currently, Brazil is emerging as a leader in AI regulation in South America by initiating three new bills to regulate the development and use of AI. These bills are the result and follow up of the Brazilian AI Strategy that aims to foster trustworthy and ethical AI. This also follows activities of other South American countries who have individually published guidelines to endorse safe and fair AI, including Colombia’s framework for ethical AI and Argentina’s guidelines for trustworthy AI. Furthermore, Brazil’s regulatory initiatives come at time when

there are mounting global endeavours to codify responsible AI practices into law, particularly in Western jurisdictions like Europe, Canada, and United States and other large countries like China that is primarily focusing on regulating generative AI.

Between 2019 and 2021, Brazil initiated three AI laws that had the objective of harmonising the minimisation of harm and promotion of innovation. However, these laws were not approved by Congress. These three laws were replaced by a consolidated Act in May 2023 that had several similarities with the EU AI Act, such as a risk-based approach to AI regulation. Specifically, the 2023 initiated bill related to the formulation of a framework for AI development and use was presented in Congress. The latest bill obliges entities to ensure transparency and mitigate biases, specifically for high-risk AI systems and demands detailed public impact evaluations that outline the system's objective, risk mitigations, and stakeholder participation, where organisations will be held strictly liable for any damages caused.

Like Argentina, the proposed Act defines AI specific terms; establishes guiding principles; itemises prohibitions; and creates an administrative entity for AI within the National Executive Branch. The focus for Brazil's national AI strategy is chiefly on the citizens' rights, risk classification, and governance procedures, including administrative sanctions. The definition of AI has been refined and extended from previous bills. An AI system is defined as a computational system with different degrees of autonomy that is designed to infer how to achieve a given set of objectives using machine learning and/or logical and knowledge representation. Additionally, the process relies on input data from humans or machines and produces predictions, recommendations, or decisions to influence the real or virtual environment. As stated previously, the 2023 bill is based on a risk-based approach like that of the EU AI Act and focuses on the rights given to individuals, which include the right:

- To prior information about interactions with AI systems.
- To the provision of a rationalisation about the outputs of AI systems.
- To contest decisions or forecasts made by AI systems that result in legal implications or otherwise significantly impact changes.
- To incorporate human involvement in decisions about AI systems.
- To non-discrimination and correction of discrimination biases.
- To privacy and protection of personal data.

To determine the risk of an AI system, the proposed bill stipulates that AI systems must undertake a preliminary evaluation by the supplier before they are marketed or commissioned, with the evaluation being documented and registered to certify accountability and liability. Furthermore, if the evaluation detects that a system is classified as having a high level of risk, an algorithmic impact evaluation is also mandatory, in addition to the governance measures. In terms of the proposed law, the following applications are classified as high-risk systems:

- Security apparatuses in managing and operating critical infrastructure.
- Education and professional training.

- Recruitment, screening, filtering, and assessing of job candidates, as well as other employment determinations, such as promotion and termination.
- Access to and entitlement for public and private services.
- Credit ratings and computation of debt capacity.
- Establishing priorities for emergency response dispatch.
- Administration of justice.
- Autonomous vehicles.
- Healthcare applications.
- Biometric identification systems.
- Criminal investigation and public security.
- Analytical study of crimes by scrutinising large sets of complex data to identify hidden relationships.
- Evaluation of the credibility of evidence during a criminal investigation.
- Migration management and border control.

AI applications are compelled to implement transparency and data management processes; adhere to data protection legislation; practices for training, testing, and validating system results; and information security processes. The AI system must comply with specific regulatory procedures for high-risk systems, such as documentation, automatic recording of actions, tests for reliability and robustness, and data management to avoid bias.

Like the EU requirement, AI systems with extreme risk are forbidden under the bill. Such systems include those that utilise subliminal practices to deliver harmful behaviour; those that take advantage of the vulnerabilities of specific groups to create harmful behaviour; and those utilised by public entities with the objective of social scoring. Additionally, biometric identification systems are only authorised in publicly accessible spaces when judicial authorisation is provided in relation to security activity and individualised criminal prosecution.

The institutional frameworks are based on the setting up of several entities to regulate AI, including the Ministry of Science, Technology, and Innovation (MCTI), which defines and deploys AI policies. Moreover, its National Data Protection Authority has the remit to ensure that AI systems comply with data protection laws. It also has a commission to promote and debate AI regulation that incorporates a public consultation process to collect and consider stakeholder input.

The use of personal data in AI systems are regulated by the General Data Protection Law (LGPD), which establishes the standards for data processing that includes transparency, security, and accountability. The LGPD is enforced by the National Data Protection Authority that ensures AI applications follow the data protection rules and is empowered to impose fines and penalties. Its incident reporting process records all AI incidents, and the audit framework verifies that AI systems comply with regulations. Key stakeholders include specialists from academia, industry, and civil society that participate in a multistakeholder forum to discuss AI regulation to make sure the framework is effective.

11.3.4.3 Colombia

In November 2023, the Colombian Senate approved a bill that establishes legal guidelines for AI development and use, which incorporates policies for data protection, intellectual property, and a code of ethics. The definition of AI appears to be weak, describing it as a discipline that “is concerned with creating computer programmes that perform tasks comparable to those performed by the human mind, such as learning or logical reasoning.” According to this definition a computer program might imitate or be comparable with the abilities of a human mind, but it is oblivious that the neurological processes by which learning ensues in humans remain, to some extent, a mystery. Furthermore, the definition presumes that an AI programme might be able to conduct logical reasoning, when it operates on predictive models previously established by the team of developers.

The bill includes several principles in line with those established by UNESCO and the Ethical Framework for AI issued by the Colombian Government in 2021 as follows:

- Human Authority: Any AI decision is to be auditable and contested by human decision.
- Common Good: The objective of AI is to be exclusively for the common good.
- Cooperation: AI research is to foster cooperation around risk deterrence.
- Safe Design: Any development in AI is to have provisions for the harmful effects that it may cause.
- Primacy of Human Intelligence: To ensure that human decision-making always predominates over AI.
- Preventive Research: Risk avoidance developments are to be conducted concurrently with the development of AI.

The intention is to establish a Commission for the Treatment of Data and Developments with AI, whose tasks will incorporate recommending new technological initiatives; approving institutional requirements that will likely originate from public bodies regarding AI developments and implementation; suggesting initiatives in public management; and advising the Colombian Congress in a consultative capacity. All organisations that develop and/or use an AI application must register with the Ministry of Science, Technology, and Innovation (MCTI), which is responsible for all AI policies compliance and enforcement. MCTI has set up a Commission to discuss AI regulations and administer a public consultation process to collect public input.

Colombia’s data protection laws oblige AI developers to ensure the protection of data and respect for human rights. Colombia recognises two fundamental personal data rights, namely the right to privacy and the right to data rectification. Furthermore, personal data processing is regulated by two statutory laws and several decrees that set out data protection obligations with two different authorities responsible for data protection and data privacy matters. The Superintendent of Industry and Commerce (SIC) acts as a general authority, while the Superintendence of Finance (SOF) acts as a supervisor of financial institutions, credit bureaus, and other entities that manage financial data or credit records and verifies the enforcement of the provisions according to the respective law.

However, SIC is the prevailing authority regarding personal data protection and data privacy and is empowered to investigate and impose penalties on companies for the inappropriate collection, storage, usage, transfer and elimination of personal data. Key stakeholders involved in AI regulation are experts from academia, industry, and civil society, who provide the necessary input for making the framework effective. Colombia, which at one time was a pioneer in debating the AI issue in the region, appears to have lost its forceful drive regarding AI regulation because it has been relegated a low priority in the current political agenda.

11.3.4.4 Mexico

The government of Mexico initiated a bill to create an ethical framework for AI development in March of 2023 that focused on protecting human rights and personal data and ensuring AI systems respect human dignity. Like most countries, the law regarding the Protection of Personal Data held by Private Parties plays an important role in protecting individuals in an AI environment by specifying data protection guidelines.

The law stipulates that AI developers must implement measures to safeguard data from access, loss, or damage. The objectives of the proposed legislation include the following:

- Formulate public policy guidelines for the ethical use of AI and robotics within the national territory.
- Foster the preparation of official standards, based on ethical principles, for the appropriate use of AI and robotics for the benefit of society, which respects human rights, gender parity, without any discrimination based on race, ethnicity, religion, social class or economic position.
- Regulate and govern the use of AI and robotics for governmental, economic, commercial, administrative, communication and financial reasons so that they always comply with ethical and legal standards.
- Set up and regulate an Ethics Council for AI and Robotics as a decentralised public institution at the service of citizens.
- Unify autonomous agencies through the regulation for the use of AI within the national territory, establishing the National Institute of Statistics, Geography and Informatics as the centre providing information on the application of AI.

The Ministry of Science, Technology, and Innovation (MCTI) administers all AI policies. A commission has been established within MCTI to discuss and formulate AI regulations through a public consultation process. The Ministry has outlined plans to establish a central government office for digital and AI policy and a national Centre for AI to support cooperation between academia, industry, and government. Furthermore, MCTI has the responsibility for monitoring AI regulations compliance and is empowered to impose fines and penalties. The government's AI strategy underlines the importance of fostering education, promoting computational thinking in schools and increasing the number of master's and PhD students in AI and data science as a means of ensuring the responsible AI development and use. However, despite its regional

leadership position in research and innovation, Mexico does not currently appear to be prioritising AI regulation, particularly due to general elections in 2024.

11.3.4.5 Chile

In 2021, Chile initiated the National AI Policy, which has the objective of facilitating the use and development of AI tools and the comprehension of the legal, ethical, social, and economic impact of AI. It also established a definition of AI, namely: “the set of computer techniques that allows a machine to perform tasks that usually require intelligence, such as reasoning or learning” and “a computational system that can, for a given set of human-defined objectives, make predictions and recommendations or make decisions that influence real or virtual environments.” This was followed in 2024 by a bill to regulate AI that aims to promote AI systems while respecting democratic principles and human rights, and defines the obligations for AI developers, suppliers, those commissioning AI systems, and importers.

Currently, there are five draft laws related to AI that are being handled by the National Congress, which mainly are related to criminal matters and aggravate or penalise the misuse of AI tools. However, the key legislative initiative that is intended to regulate AI refers to AI systems, robotics and related technologies, in their different fields of application. This proposed legislation has the objective of regulating various behaviours relating to AI, creating a legal framework for the development, commercialisation, distribution, and use of AI systems, with the aim of protecting the fundamental rights of citizens.

The proposed law also identifies the stakeholders implicated in the full AI cycle, ranging from developers, suppliers to users. It also identifies the fundamental concepts to facilitate its interpretation, including biometric data and input data, and what is implied by a major incident. Hence, it refers to cases of death or serious harm to health, property or the environment, and disruption of critical infrastructure. Some view this as being restrictive because a serious incident could be the outcome of discriminatory results that affect the dignity of one or more persons. The law applies a risk approach like the EU AI Act in that the proposed bill classifies AI systems into two different levels of risk, namely unacceptable risks and high risks.

Critics claim that the wording of the Articles regarding unacceptable risk systems is too blurred and not explicit enough to ensure sufficient protection of the fundamental rights of individuals in the context of the deployment of AI-based systems. For instance, the exceptions might legitimise the application of technologies related to remote facial recognition in public spaces for prevention purposes, which is viewed as a top-heavy mechanism that may lead to the violation of basic citizens’ rights in Latin American jurisdictions. Similarly, what is considered as a high-risk system is also seen as dubious, since there should always be human supervision, particularly in cases of a criminal nature where a sentence (using AI) could involve the restriction of a person’s freedom.

The Ministry of Science, Technology, Knowledge, and Innovation has the mandate for administering AI policies. In this regard this Ministry has introduced initiatives to promote AI development and use, including the establishment of a National AI Commission. This commission has the responsibility to assess authorisation requests from AI developers, providers, and users. The government plans to introduce a compliance and enforcement regulatory

framework to ensure that AI regulations are obeyed. This framework includes measures for monitoring and enforcing compliance rules and imposes fines and penalties for non-compliance. Like Mexico, the government is placing great emphasis on education and awareness about responsible AI development and use.

11.3.4.6 Peru

Peru is currently a leader in South America in regulating AI to protect citizens' rights and promoting ethical AI development. In 2023, the Peruvian government initiated a bill that is based on the EU AI Act. The bill accentuates the respect for human rights, privacy, and data protection, while ensuring the safety and security of citizens through upholding the principles of transparency and innovation in AI.

Additionally, the bill promotes AI within the framework of a national transformation process for economic and social development, and sets out essential principles, such as risk-based security, ethical development, and privacy. The key guiding principles for the development and use of AI in all fields are as follows:

- Risk-Based Security Standards: Fostering a risk-based approach for the use and development of AI.
- Plurality of Participants' approach: Encouraging the involvement of natural and legal persons, or public and private organisations and institutions in the discussions related to the development of policies that have the objective of regulating the use of AI.
- Internet Governance: Supporting the development and use of principles, norms, rules, decision-making practices and schemes that impact the growth and use of the internet by the state, private sector institutions and civil society in their respective roles.
- Digital Society: Acknowledging the value of information and knowledge acquired through access, use and development of digital technologies in all their dimensions; promoting security, trust, digital economy, digital connectivity, talent, innovation, education, and digital identity; and the use of emerging technologies for the social and economic wellbeing of citizens.
- Ethical Development for Responsible AI: Viewing ethics as the fundamental basis for accurately identifying the framework of responsibilities in the use of AI systems.
- Privacy of AI: Ensuring that AI does not violate people's privacy and must operate in a safe fashion for attaining a positive and beneficial impact on citizens.

The proposed bill proposes to set up a national authority to administer and regulate AI use and promotion. The government has also introduced measures to assess authorisation requests from AI developers, providers, and users. The anticipated regulatory framework has the objective of ensuring compliance with AI regulations by introducing monitoring and enforcement measures, such as the imposition of fines for non-compliance. Additionally, the government is in the process of launching an awareness campaign to educate the public about responsible AI use.

Latin America is at an early stage in regulating AI, with the various countries confronting challenges and striving to define their political agendas in this emerging field. To date, most of

these countries do not have specific legislation on AI. It is noted that the EU AI Act that not long-ago secured approval in the European Union, has been applied as a reference for discussions and legislative proposals in many South American countries. Nonetheless, these countries must meet the challenge of adapting these concepts to their local realisms, considering the intense regional differences.

The above suggest that AI regulation in Latin America is gradually evolving but unlike the situation in Africa where there is a central African Union body that is encouraging a consistent approach to AI regulation, South American countries are developing their own individual AI strategies. However, these individual strategies suggests that each country is encountering significant challenges regarding the creation of legal frameworks, reducing bias, and protecting personal data. A unified approach, like the EU and African countries would facilitate matters in the deployment of AI in the South American region.

11.3.5 Australasia

The Australasian legislation segment will focus on Australia and New Zealand. Like other global regions, the AI regulation landscape in Australasia is changing but legislative progress specifically related to AI is somewhat lacking.

11.3.5.1 Australia

Australia does not have any specific federal laws or a regulatory framework to address specific risks and challenges posed by AI. Currently, the government's response to AI is based on voluntary (self-discipline) regulation that includes the AI Ethics Principles published in 2019 consisting of eight principles for the responsible design, development and implementation of AI. These voluntary principles are coherent with the OECD's Principles on AI, namely:

- Human, Societal and Environmental Wellbeing: AI systems are to benefit individuals, society and the environment.
- Human-Centered Values: AI systems must respect human rights, diversity, and the autonomy of individuals.
- Fairness: AI systems are to be inclusive and accessible and must not involve or result in unfair discrimination against individuals, communities or groups.
- Privacy Protection and Security: AI systems must respect and uphold privacy rights and data protection and ensure the security of data.
- Reliability and Safety: AI systems must reliably operate in accordance with their intended purpose.
- Transparency and Explainability: AI systems must be transparent and should include responsible disclosure so persons may recognise when they are being significantly influenced by AI and can be aware when an AI system is engaging with them.
- Contestability: There must be a timely process to allow people to challenge the use or outcomes of the AI system when the AI system significantly impacts a person, community, group or environment.

- **Accountability:** People responsible for the different phases of the AI system lifecycle must be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.

Because Australia does not have any specific federal laws or a regulatory framework and the AI ethics principles are voluntary, there is no AI specific regulator in Australia. Nevertheless, there are several laws and regulations that apply to AI technologies related to its development, supply and use. Moreover, Australian entities that have an international association, including in the EU, may be affected by the extraterritorial reach of applicable foreign laws.

A positive step to remedy this situation was taken in January 2024 with the publication by the Department of Industry, Science and Resources of a document titled “Safe and Responsible AI in Australia.” This document is based on the response to a public consultation process that took place months previously. It outlines government’s approach to AI regulation and demonstrates its commitment to initiate regulation for particular uses of AI. This method is not meant to be transformational and is inclined towards a balanced approach, thus ensuring that the development and deployment of AI systems in Australia takes place in a lawful environment that is safe and reliable. However, critics view this approach as being risky and allows the use of AI in low-risk settings to mainly continue without restriction.

Furthermore, the document does not elaborate on the range of high-risk AI applications that will be regulated. A definition that focuses on the effects of AI, which are described as systematic, irreversible or perpetual, are not clear, but merely indicate that a similar approach to the EU regarding the regulation of high-risk applications will be adopted. This issue is of particular concern in relation to the application of AI in self-driving vehicles to make real-time decisions and AI-enabled robots for medical surgery. [Bradshaw et al. \(2024\)](#) identify the following intentions of government in the short-term:

- To further consult on options for initiating new regulatory guardrails that focus on testing, transparency and accountability.
- To take steps to facilitate businesses operationalise safe and responsible AI by formulating a voluntary industry AI Safety Standard that collectively brings current responsible AI principles, guidelines and frameworks to create a best-practice and up-to-date AI risk-based safety framework; initiating work with industry to study the merits of voluntary labelling and watermarking of AI-generated material in high-risk settings; and instituting a temporary expert advisory group to support the government’s development of AI guardrails.
- To consider other opportunities to strengthen current laws to focus on the risks and harms from AI (e.g., reviewing privacy law reforms and government’s effort on its Cyber Security Strategy).
- To progress on commitments made in the Bletchley Declaration.
- To collaborate internationally to formulate global AI governance to recognise domestic responses to risks posed by AI and seek avenues for strengthening the engagement of Australian experts in key international forums that develop technical standards for AI.

- To further study opportunities for ensuring that Australia can maximise the benefit of AI technology.

Following the interim response, the Australian Government established an AI Expert Group to assist the Department of Industry, Science and Resources in developing regulations on transparency, testing and accountability, including options for mandatory AI guardrails in high-risk settings. According to [Bradshaw et al. \(2024\)](#), the interim response follows other recent developments in AI regulation in Australia that include four important developments:

- July 2024: New South Wales AI Assessment Framework. The NSW state government issued and implemented a review of its AI Assessment Framework that includes a mandatory self-assessment tool designed to guide responsible and safe AI usage in projects by NSW state agencies.
- December 2023: Framework for Generative AI in Schools. The Federal Department of Education approved the Australian Framework for Generative AI in Schools that was developed in consultation with all industry stakeholders. This framework aims to guide responsible and ethical use of AI that benefits and supports all people connected with school education.
- November 2023: Bletchley Declaration. The Federal Government signed the Bletchley Declaration jointly with the EU and 27 countries, including the USA, UK and China. This declaration pledges international collaboration to identify AI safety risks of shared concern and build risk-based frameworks across countries to ensure AI safety and transparency.
- Ongoing review of related laws. This includes the proposed reforms to the Australian Privacy Act that are aimed at increasing transparency and accountability for organisations that use AI technologies to process personal information and provides greater enforcement powers for the privacy regulator. This is part of a government's programme for a wider review of AI regulation and automated decision-making.

A significant development occurred in March 2024 when the Senate (at Federal level) passed a resolution to establish a Select Committee on Adopting AI to investigate and report on the opportunities and impacts for Australia arising from the application of AI technologies in Australia. This Select Committee had the following terms of reference:

- To examine trends and opportunities in the development and adoption of AI technologies in Australia and overseas, specifically generative AI.
- To investigate the risks and harms occurring from the implementation of AI technologies, such as bias, discrimination and error.
- To explore emerging international methodologies for mitigating AI risks.
- To study the opportunities for implementing AI in approaches that benefit citizens, the environment and/or economic growth, such as healthcare and climate management.
- To detect opportunities that support a responsible AI industry in Australia.
- To examine potential threats to democracy and trust in institutions from generative AI.

- To evaluate environmental impacts of AI technologies and opportunities for limiting and mitigating impacts.

The Select Committee was scheduled to report to Parliament by September 2024, which was extended to the end of November 2024. The Select Committee received over 200 submissions from individuals, multinational companies and industry organisations that included Microsoft, Google, FreeTV, OpenAI, and the Law Council of Australia. The main topics in these public submissions were related to:

- Deepfake Technology: AI-generated deepfakes were viewed as a serious concern for the democratic process and personal privacy, since these had the ability to mislead the public, encouraging the increase of misinformation that is already present in social media. Deepfake pornography is viewed as a particular worrying concern and is already being addressed through a Bill by the Australian House of Representatives.
- Transparency: Submissions highlight the need for oversight in the testing and development stages of AI-enhanced products, particularly regarding datasets sources and the clear disclosures of the presence of AI-enhanced systems, particularly through misinformation and disinformation where users are unaware that they are viewing AI-generated content.
- Bias: Algorithmic bias is seen as being able to generate skewed outcomes by reproducing prejudiced results. It is argued that there is a need for risk management processes to mitigate the impact of algorithmic bias.
- Benefits of AI: Despite the risks, the adoption of AI is viewed as generating benefits in areas such as the environment, healthcare and education.

According to [White and Case \(2024c\)](#), other laws affecting AI include: (i) the Online Safety Act that includes mechanisms to address online safety issues, extending to AI generated material; (ii) the Australian Consumer Law that was applied to algorithmic decision-making by ordering a company to pay fines for misleading consumers in the recommendations made by its algorithm; (iii) the Privacy Act; (iv) the Corporations Act; and (v) Intellectual property laws that may impact several aspects of AI development and use Anti-discrimination laws, such as where an individual is a victim of a discriminatory outcome resulting from an AI-driven process.

In the interim response, the Australian Government recognised that current laws are expected to be strengthened to address the harmful effects posed by AI. Hence, the Federal Government is currently developing new laws that will give the Australian Communications and Media Authority with regulatory powers to combat online misinformation and disinformation, extending to content on digital platforms that are generated by AI. [White and Case \(2024c\)](#) argue that no definition of AI has been formally adopted by any statutes or regulations in Australia, but the Department of Industry, Science, and Resources embraced the following definitions that are claimed to be based on the International Organisation for Standardisation:

- “AI” means “an engineered system that generates predictive outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives or

parameters without explicit programming. AI systems are designed to operate with varying levels of automation.”

- “Machine learning” means “the patterns derived from training data using machine learning algorithms, which can be applied to new data for prediction or decision-making purposes.”
- “Generative AI” means “models [that] generate novel content such as text, images, audio and code in response to prompts.”

Generally, it may be concluded that the Australian Federal Government has been reluctant to initiate laws that directly regulate and govern AI. This reluctance may stem from an environment where the Government is encouraging the development of AI technology by the private sector and thus regulation may be viewed as a hindrance to the AI innovative process at this specific stage of AI innovative growth. However, major reforms are anticipated in the medium term, with a risk-based framework focusing on suitable mandatory safeguards supported by the appropriate methodologies to implement them.

11.3.5.2 New Zealand

The Minister of Science, Innovation and Technology in June 2024 released a Cabinet Paper that outlined New Zealand’s strategic framework regarding AI. The government views AI to boost productivity and economic activity, and to deliver better public services. However, the Minister also indicated that a specific AI Act is not on the horizon. The Cabinet Paper makes it very clear that the government intends to take a wait-and-see attitude, and instead of taking the lead regarding AI regulation, it is content to be a follower basing its efforts on international developments. This stand is surprising, since New Zealand has been viewed as a progressive country regarding major administrative reforms. Like Australia, there seems to be a reluctance to regulate AI at this specific stage.

That government’s strategic approach appears to be based on supporting and fostering AI development and use. However, instead of creating a specific AI Act, the government intends to utilise existing regulatory frameworks related to privacy, consumer protection, and human rights. These are mainly principles-based and technology neutral, with voluntary guidance, industry codes, technical standards, and audit requirements. This approach will enable the New Zealand government to avail itself, and adjust with, according to international developments, and best practice as they occur.

The intention of the government is to utilise safe AI innovation in public services but with the prerequisite of balancing AI efficiencies against the imperative for social licence and trust. Hence, it is anticipated that the Government Chief Digital Officer will support agencies to safely adopt and innovate using AI and to oversee AI uptake in general. Furthermore, to enable the application of AI across all economic sectors, the Ministry of Business, Innovation and Employment are to formulate risk management-based guidance for companies and to define an AI Roadmap to support AI in the private sector, with the specific objective of unlocking AI technology for innovation, productivity, and export potential.

The government is actively participating in key international discussions to leverage global networks and monitor international initiatives and to provide AI advice to other agencies when

opportune. Like Australia, the New Zealand government will foster the voluntary principles prescribed by OECD regarding AI development and use. The government recognises that AI can be both a threat and a tool to protect national security. Therefore, the Department of the Prime Minister and Cabinet will be taking on the responsibility as the strategic coordination agency for the national security responses to emerging, critical, and sensitive technologies. The government is also sensitive regarding the Indigenous peoples of New Zealand and the need to ensure Māori share the benefits of AI developments.

Other initiatives in addition to the Cabinet Paper include the Centre for Data Ethics and Innovation, and the Office of the Privacy Commissioner's guidance on privacy and AI. The Privacy Commissioner expressed concerns about the lengthy timeframe for a national approach on AI as defined in the Cabinet Paper, due to imminent urgent issues. One such issue is addressing the uncertainty regarding how the existing regulatory frameworks apply to AI. The Privacy Commissioner has taken the initiative to summon the Digital Regulators Forum that included Ministry of Business, Innovation, and Employment; the Department of Internal Affairs; Statistics NZ; and the Commerce Commission to investigate progressing public-facing guidance for businesses on how existing regulations apply to and support the responsible use of AI in the immediate term.

11.3.6 Asia

The section regarding Asia will cover the regulatory landscape of China, India, Japan and Singapore.

11.3.6.1 China (PRC)

Chinese regulators appear to be taking a supportive and practical approach towards AI. Generally, regulators are advocating a balanced methodology in the development and use of AI by having adequate safeguards to prevent social and economic harms. Moreover, while the Chinese government has intensified its supervision of AI technologies, it has adopted a cautious style by setting rules that target specific applications and not applying a holistic regulatory approach over the entire AI industry so as not to restrain AI innovation. Restricting AI innovation is not an option since Chinese policy makers view AI technology as a significant and decisive factor for the future development of innovation in general and smart industrial practices in a digital environment.

Like Australasian countries, most regulatory provisions in China are not mandatory and are still evolving. The key exceptions to this general rule are the high-level cybersecurity and data protection obligations, and certain requirements issued by the Cyberspace Administration of China (CAC). It is noted that for the last seven years since 2017, the Chinese government has initiated three discrete stages in the development of AI with each stage demonstrating a distinctive balance of interests and methods to AI. [Gong et al. \(2024\)](#) claim that China initially prioritised industry self-regulation, with key planning policies advocating the need for industry to invest in the technology and exploit market opportunities. According to [Gong et al. \(2024\)](#), the three stages include the following developments:

- Stage I: Strategic Planning and Industry Self-Discipline (2017 – 2020). During this period, policy makers examined the national and international compliance risks related with AI technology, particularly security. The focus was to establish a robust basis to enable the Chinese AI industry to grow without regulatory obstacles. Additionally, a strategic plan was issued in 2017 by the State Council outlining future AI policy and identifies key issues that should be implemented when conducting departmental tasks, particularly the utilisation of data. This initial plan was supported by other strategic documents that emphasised the need to promote the inherent integration of the Internet, big data, AI and the economy. The responsibility for implementing the 2017 Strategic Plan was assigned to the AI Strategic Advisory Committee and the AI Planning and Promotion Office under the remit of the Ministry of Science and Technology. Additionally, in 2019, a specialist body was created to strengthen research on the legal, ethical, and social issues related to next-generation AI technologies and to strengthen international collaboration. All conditions within the strategic plans were voluntary. However, several entities, mainly internet companies, instituted self-regulatory frameworks to promote guiding principles for the industry.
- Stage II: Voluntary Standards and the Beginning of Regulatory Oversight (2020 – 2022). During this period, non-mandatory, industry best-practice national standards were defined to cover algorithmic technologies and machine learning applications. These standards lacked a comprehensive legal architecture through which state bodies would impose direct oversight and enforcement. In 2020, Guidelines to the Construction of the National New-Generation AI Standard System was released that defined structural aspects of the AI standard system to guide technical standards formation. The aim of these guidelines was to prepare technical committees to gain first-hand knowledge of how AI technologies work and issue sector-specific standards from the bottom-up.
- Stage III: Direct Supervision of AI Technologies (2022 – present). CAC has released mandatory regulations for AI governance in China related to content moderation, intellectual property, and cybersecurity. As stated previously, Chinese regulators have taken a targeted approach, particularly security and generative AI, namely the implementation of 2023 Generative AI Measures.

Generally, China's regulations have the objective of addressing the risks associated with AI and to initiate compliance obligations on entities involved in AI-related business. The current set of regulations include:

- 2022: Administrative Provisions on Algorithm Recommendation for Internet Information Services that focuses on the use of algorithm recommendation technologies, including generative and synthetic algorithms.
- 2023: Provisions on Management of Deep Synthesis in Internet Information Service that focuses on the use of deep synthesis technologies, which is a subset of generative AI technologies.

- 2023: Provisional Provisions on Management of Generative Artificial Intelligence Services that broadly regulates the development and use of all generative AI technologies.
- 2023: Trial Measures for Ethical Review of Science and Technology Activities that focuses on the ethical review of the research and development of AI technologies.

These regulations entail obligations on service providers, technical supporters, users, and certain other entities, including online platforms and have the objective of addressing the risks related to AI-generated content and to protect national and social security. According to [Yang \(2024\)](#) the lookout for 2024 is whether China will follow in the European Union's footsteps and announce its own comprehensive AI Act. He notes that in June of 2023, China's top governing body released a list of legislation they were working on, and "Artificial Intelligence Law" appeared for the first time. Furthermore, he argues that the Chinese government is already good at reacting to new technologies swiftly since it was probably the first country in the world to introduce legislation on generative AI mere months after ChatGPT's big break. [Yang \(2024\)](#) believes that a new comprehensive law could give China even more control over how AI disrupts (or doesn't disrupt) the way things work today.

11.3.6.2 India

The Indian government's attitude towards the AI regulatory status of India can be summed up by [Barik \(2024\)](#) a journalist for The Indian Express. This journalist cited a top government official who stated that India will regulate AI, but not at the cost of innovation. He claims that the government is expected to take a middle path for AI regulation, as it has taken in some of its recent technology laws, such as data protection, where the focus is not just on protecting civil rights, but also towards facilitating the country's growing start-up base. Therefore, the government intends to regulate AI but not in a way that could stifle innovation; in fact, innovation needs to be encouraged.

Presently, India has no laws, statutory rules or regulations that directly regulate AI and there is no single legally recognised definition of AI. However, the government has initiated a process of compiling several frameworks to guide the regulation of AI, including: (a) The National Strategy for AI of 2018 that aims to establish a robust foundation for future regulation of AI in India; (b) The Principles for Responsible AI of 2022 that defines India's roadmap for the generation of an ethical, responsible AI ecosystem across sectors; and (c) The Operationalising Principles for Responsible AI of 2021 that focuses on the need for regulatory and policy interventions, capacity building and introducing incentives for having ethically designed AI systems.

The National AI strategy identified key national priority areas in AI innovation and utilisation, including healthcare, education, agriculture, smart cities, and transportation. Parts of this strategy have been implemented regarding the creation of high-quality datasets to promote research and innovation, and the creation of legislative frameworks for data protection and cybersecurity. Moreover, as part of the national strategy on AI, the government has identified several Principles for Responsible AI most of which are like those defined by the OECD and include:

- Safety and reliability.

- Equality.
- Inclusivity and non-discrimination.
- Privacy and security.
- Transparency.
- Accountability.
- Protection and reinforcement of positive human values.

The Principles for Responsible AI address the ethical issues encompassing the execution of AI solutions that are classified into system and societal factors. The system issues chiefly consider decision-making principles, fair inclusion of beneficiaries, and accountability. On the other hand, the societal issues focus on the impact of automation on job creation and employment. Operationalising principles for responsible AI is the second section of the principles for responsible AI that focuses on applying the principles derived from the ethical issues mentioned previously. It emphasises the impact of government involvement in fostering responsible AI execution in social sectors, in alliance with the private sector and research organisations. This segment of the national strategy places significance on the need of regulatory and policy actions, capacity enhancement, and urging ethical practices by incorporating a responsible mentality amongst private entities regarding AI.

India has several laws that do not directly attempt to regulate AI but nevertheless might impact the development or use of AI. Some of these laws include: (a) The IT Act of 2000 including the IT reasonable security practices and procedures and sensitive personal data rules of 2011, which will be replaced by the Digital India Act 2023 when enacted; and (b) The Digital Personal Data Protection Act 2023. Additionally, AI development and use might be affected by the intellectual property laws. The Digital Personal Data Protection Act regulates the processing of digital personal data irrespective of its original format and can be applied to address several of the privacy issues related to AI platforms. Furthermore, the Information Technology Rules are the Intermediary Guidelines and Digital Media Ethics Code which were revised in 2023 form part of the IT Act and have the objective of defining a framework to administer several entities, including social media intermediaries, Over-the-top media service (OTT) platforms that stream content to customers directly over the web, and digital news media.

The National Data Governance Framework Policy that was issued in 2022 has the goal of updating and improving government's data collection and management procedures to foster an ecosystem favourable to AI and data-driven research, including startups by establishing a comprehensive repository of datasets. According to [Dey and Cyrill \(2024\)](#), India currently does not have laws that directly address generative AI, deepfakes, and AI-related crimes. However, they claim that government has announced that the relevant legislation is in progress. They note that deepfakes, which are digitally manipulated media, including videos, audio, and images, created using AI, have the potential to harm reputations, fabricate evidence, and erode trust in institutions due to their hyper-realistic nature. However, several provisions within the current legislation provide both civil and criminal remedies. For example, the Information Technology Act, Section 66E renders deepfake crimes regarding privacy abuses punishable by imprisonment

of up to three years or a fine of about \$2500. Moreover, the Indian Penal Code provides further recourse under various sections of the law.

While India has no AI regulator, the Ministry of Electronics and Information Technology has the mandate for AI-related strategies and has empowered several task forces to implement the AI policy framework. Additionally, the Ministry of Commerce and Industry has instituted an AI Task Force to explore the process for establishing an AI regulatory authority for future reference.

11.3.6.3 Japan

Generally, like other countries in the Pacific region including Australasia, Japan is taking a soft law approach to AI governance. In April 2024, the Ministry of Internal Affairs and Communications and Ministry of Economy, Trade and Industry initiated AI Guidelines for Business that are compatible with international thinking. These guidelines are not legally binding but provide a well-defined voluntary direction for AI developers, providers and users. Although the guidelines are voluntary, they may impact decisions related to legal challenges. Like other nations in the region, the guidelines are intended to harmonise societal and individual rights while concurrently promoting innovation and the implementation of AI technologies. Hence, having a specific legal framework at this stage is likely to be viewed by government as constraining AI development and use. Hence, Japan is prioritising innovation while concurrently mitigating foreseeable harms.

It is noted that in 2019 the Japanese government had released the Social Principles of Human-Centered AI with the objective of achieving the world's first "AI-ready society." These principles are based on three key beliefs, namely respect for human dignity; a sustainable society; and a society where diverse backgrounds support individual well-being. Specifically, the principles support the following values:

- AI must not infringe upon fundamental human rights.
- Basic education and literacy must be guaranteed.
- Privacy must be protected, security safeguarded, and a fair, competitive environment guaranteed.
- Fairness, accountability and transparency are a must.
- AI innovation will be promoted through the close collaboration and cooperation of domestic and international stakeholders.

In 2023, Japan published the Hiroshima International Guiding Principles for Organisations Developing Advanced AI Systems (known as the Hiroshima Principles) that have broadly similar goals to those of 2019, namely, to establish and promote guidelines worldwide for safe, secure, and trustworthy AI. The original 2019 and the revised 2023 principles support the AI Principles defined by the OECD that recognise AI's capacity together with its risks. The government established an AI Strategy Council, which is an advisory body that has the objective to examine methods for maximising the potential of AI while minimising the possible risks. This Council in May 2024 proposed draft discussion points regarding the suitability and possible extent of future regulation.

A task force has proposed a law to regulate certain generative AI, titled “the Basic Act on the Advancement of Responsible AI” that imposes certain obligations on AI systems and developers regarding vetting, operation, and output of the systems. This proposed law is still being formulated, and it is likely that it would specify a general framework, while the industry and/or individual developers would apply a self-regulating approach by establishing the specific standards by which they would comply. This proposed law signifies that the government will likely move away from a soft to a hard legal approach when regulating AI. It also signifies that the government would be empowered to monitor AI developers and inflict penalties for infringements. Since Japan does not have a specific law regarding AI, its regulatory regime for AI is through existing laws including:

- The Copyright Act and Personal Information Protection Law.
- The Civil Code that permits tort claims to be brought against a person who has given instructions to AI to produce defamatory content about another person and published such content.
- The Digital Platform Transparency Act, which obliges conditions on large online malls, app stores, and digital advertising businesses to guarantee transparency and fairness in transactions with business users.
- The Financial Instruments and Exchange Act, which compels businesses engaging in algorithmic high-speed trading to register with the government, establish a risk management system, and maintain transaction records.
- The Unfair Competition Prevention Act and Antimonopoly law.
- The Economic Security Promotion Act.

In the absence of a specific AI law there is no official regulator. However, the following ministries and agencies have the remit for formulating and supporting guidelines regarding AI:

- Ministry of Economy, Trade, and Industry.
- Ministry of Internal Affairs and Communications.
- Agency for Cultural Affairs (chiefly for copyright issues).
- Personal Information Commission.

It should be noted that although the guidelines are of a voluntary nature they are scrupulously respected by companies and the public. The above support the view that the Japanese government has the objective of realising its social principles using AI through a voluntary process and not by imposing legal restrictions on the ever-evolving technology.

11.3.6.4 Singapore

Singapore does not have a specific general legislation for the development and use of AI. However, it has established various AI frameworks that guide AI ethical and governance principles with the current sector-specific regulations that address AI risks. These non-binding AI frameworks have been formulated by the Info-communications Media Development Authority (IMDA), Personal Data Protection Commission (PDPC), the Monetary Authority of Singapore

(MAS), and the Intellectual Property Office of Singapore (IPOS). All the frameworks promote responsible, safe and values-based AI.

The Asian and Australasian regions all appear to have the objective of harmonising the need of allowing innovation in AI to develop and safeguarding public interest in AI ethics and governance. Hence, the development of the various frameworks and tools to guide AI deployment and promote the responsible use of AI in Singapore is aimed at achieving this harmonisation. These frameworks include the following:

- Model AI Governance Framework of 2020 (originally 2019), which describes detailed guidance to private sector entities regarding the main ethical and governance issues when implementing AI solutions.
- AI Verify, which is an AI governance testing framework and toolkit aimed at facilitating organisations to validate the performance of their AI systems according to standard tests related to AI ethics principles. This toolkit was created by the Infocomm Media Development Authority of Singapore (IMDA) with the assistance of the private sector.
- AI Verify Foundation that was established by IMDA as a not-for-profit entity to crowd in expertise from the private sector and the open-source global community to create AI testing frameworks, standards and best practices.
- National AI Strategy of 2023 (originally initiated in 2019), which outlines Singapore's goal and commitment to creating a trusted and responsible AI ecosystem, driving innovation and growth through AI, and empowering its people and businesses to understand and engage with AI.
- Model AI Governance Framework of 2024, which has the objective of expanding the Model AI Governance Framework of 2020 by focusing on emerging issues related to Generative AI and giving guidance on suggested practices for safety evaluation of Generative AI models.

The above frameworks focus on four key areas:

- Creating internal governance structures and policies to provide distinct and explained responsibilities for AI implementation strategies.
- Ascertaining the extent of human involvement and administration related to AI-extended decision-making, such as whether humans are able to supervise or take over the decision-making process.
- Creating operations management processes to minimise bias on datasets.
- Transparent stakeholder interface, such as disclosure to customers and other end-users.

These frameworks are meant to be broadly applied and to be adjusted to match an organisation's needs. They recommend nine dimensions to promote a trusted ecosystem that include:

- Accountability: Assign responsibility to AI actors during the AI development life cycle.

- **Data:** Safeguarding the quality of data as input to AI models via the use of trusted data sources.
- **Trusted Development and Implementation:** Fostering transparency and disclosure to improve a wider awareness and safety.
- **Incident Reporting:** Create incident-management structures that is supported by processes for timely notification and redress.
- **Testing and Assurance:** Applying third-party testing compared with common AI testing standards to exhibit trust to end-users.
- **Security:** Identifying and addressing risks implicated through AI models.
- **Content Origin:** Creating technologies to augment transparency regarding where and how content is generated.
- **Safety and Alignment of Research and Development:** Hastening investment in research and development to enhance model compatibility with human intent and values.
- **AI for Public Good:** Utilising AI to benefit the public by improving public sector deployment, upskilling workers, democratising access and having AI systems sustainably.

Currently, the Singapore government is not giving priority to developing general AI regulations but intends to rely on existing laws related to data protection, copyright and other sectoral legislation. However, the published guidelines have the goal to supplement the current laws. This is seen as establishing a robust foundation for the enacting future general AI regulations at an opportune time.

There are several existing laws that may indirectly impact the development and use of AI, namely: (a) Road Traffic Act 1961 (amended 2017) that permit the testing and use of autonomous vehicles; and (b) Health Products Act 2007 that obliges medical devices featuring AI technology to be registered prior to being utilised. Furthermore, several regulatory authorities have published standards on the responsible application of AI that include: (a) Personal Data Protection Commission regarding the use of personal data in Recommendations and Decision AI systems for machine-learning models; (b) Monetary Authority for promoting fairness, ethics, accountability and transparency in the application of AI and data analytics in the financial sector; and (c) Health Sciences Authority regarding understanding, codifying good practice and supporting the safe growth of AI in healthcare.

The above support the contention that the government of Singapore is basing its AI regulation through a voluntary process by defining specific guidelines that are well defined and address ethical and practical implications of AI. The government will not be imposing legal restrictions in the short-term so as not to disrupt the evolving AI technology. Generally, the Asian region appears to have a similar approach to the Australasian countries under review, in that they are all regulating AI through existing laws and the issuance of voluntary guidelines. However, no direct legislation to govern AI is in sight for the short term.

11.4 CONCLUSION

After several AI winters, AI has now been widely acknowledged and accepted. Its application has been successful in healthcare, banking, retail, transport, education, and manufacturing. On the other hand, like other revolutionary technologies that have been launched in the past, little consideration has been given to its possible sinister side. Many recent experiences have emphasised the need for AI regulation and governance. AI developers and users are obliged to adhere to highly ethical behaviour. However, most of the requirement associated with ethical behaviour imposed by authorities is mainly voluntary.

Governments, corporate leaders, and organisations should take immediate measures towards responsible AI application and investment to address AI-related risks. This is an urgent requirement to facilitate the compliance with existing laws and to generate and maintain trust with all stakeholders, including citizens, customers, suppliers, and others. In the long term, it also assists in establishing a solid foundation for compliance with prospective AI regulation around the globe.

The EU on a federal level and several US States have emerged as global leaders in AI regulation. There is no doubt that AI regulations substantially govern how the technology will impact society's future. Regulation through legislation has the objective of identifying well-defined obligations and encourage AI in many sectors but must be based on having robust ethical and consumer protection values. The Australasian and Asian regions have sided with developers and private sector AI users. Developers and private sector AI users prefer self-policing within existing legal regimes. Whereas civil society look to the government to take the lead in the specific regulation of AI applications.

Many nations appear to view mandatory regulation as a hindrance to AI innovation. However, this need not be the case for as AI develops because of advancements in machine learning, the regulations ought to become more accommodating, renewed, and harmonised with the requirements of the regulatory bodies. This can be achieved with the collaboration of the relevant stakeholders at all levels, national, state and global, to guarantee responsible AI deployment and the maximisation of AI technological potential. Jacob (2024) argues that in defining AI legislation, several essential issues should be addressed, namely:

- **Ethical Principles:** These must uphold the essential values of transparency, equitability, and accountability, to ensure responsible AI development and use.
- **Data Privacy:** To incorporate rules about the method AI applies to gather, apply, and protect personal data to eliminate privacy fears.
- **Algorithmic Bias:** Introducing processes that eliminate algorithmic bias and permit fair and unbiased AI decision-making.
- **Transparency and Explainability:** To ensure that AI systems are transparent and easy to understand so that users can know how decisions are made.
- **International Collaboration:** To facilitate the cooperation of governments with international bodies to guarantee the attainment of cohesive regulations that address global issues.

There is consensus among the countries examined for the development and deployment of AI in a safe and trustworthy environment. However, the old proverb, "*The proof of the pudding is in*

the eating” is very suitable in this regard. Having voluntary ethical principles and standards in place is commendable but unless these are supported by forceful legislation, they become somewhat diluted. It is noted that many countries are very diplomatic with the words they chose to justify the reluctance to implement a direct law to regulate AI. The reluctance in having a direct law to regulate AI is aptly put by a senior official of the government of India who was quoted as saying that India will regulate AI, but not at the cost of innovation (Barik, 2024).

REFERENCES

- Barik, Soumyarendra. (2024). India will regulate AI, but not at the cost of innovation: Govt official. *The Indian Express*. <https://indianexpress.com/article/business/india-will-regulate-ai-but-not-at-the-cost-of-innovation-govt-official-9335211/>.[↗]
- Bradshaw, S., Yew, J., Stone, V., and Rose, A. (2024). *Navigating AI Regulation in Australia and Beyond: What Corporate Leaders and Organisations Need to Know*. Hamilton Locke. <https://hamiltonlocke.com.au/navigating-ai-regulation-in-australia-and-beyond-what-corporate-leaders-and-organisations-need-to-know/#:~:text=While%20Australia%20does%20not%20yet,reach%20of%20applicable%20for%20eign%20laws.>[↗]
- CDO. (2024). 1979 Belmont Report on Ethical Treatment of Human Subjects Could Influence Modern AI Ethics. *US Federal News Bureau, CDO Magazine Bureau*.[↗]
- Dawson, S., Lau, E., Angelo, J., and Millard, L. (2024). *Australian AI Regulation: A Work in Progress*. <https://jws.com.au/insights/articles/2024/australian-ai-regulation-a-work-in-progress>.[↗]
- Dey, A., and Cyrill, M. (2024). AI regulatory landscape in India. *India Briefing*. www.india-briefing.com/news/india-regulation-of-ai-and-large-language-models-31680.html/#:~:text=Presently%2C%20India%20lacks%20a%20dedicated%20regulation%20for%20AI%2C%20but%20instead,regulatory%20landscape%20for%20AI%20technology.[↗]
- European Parliament. (2024). *EU AI Act: First Regulation on Artificial Intelligence*. Directorate General for Communication, European Parliament. www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.[↗]
- Fassbender, Melissa. (2014). *AI Is Summoning Demons*. Manufacturing.Net. www.manufacturing.net/aerospace/blog/13058129/ai-is-summoning-demons.[↗]
- Gallo, V., and Nair, S. (2024). *The UK's Framework for AI Regulation*. Deloitte Touche Tohmatsu. www.deloitte.com/uk/en/Industries/financial-services/blogs/the-uks-framework-for-ai-regulation.html.[↗]
- Gong, J., Qu, H., and Dorwart, H. (2024). *AI Governance in China: Strategies, Initiatives, and Key Considerations*. Bird & Bird. www.twobirds.com/en/insights/2024/china/ai-governance-in-china-strategies-initiatives-and-key-considerations.[↗]
- IBM. (2024). *What is AI ethics?* www.ibm.com/topics/ai-ethics.[↗]
- IDC. (2023). *IDC Future Scape: Artificial Intelligence Will Reshape the IT Industry and the Way Businesses Operate*. International Data Corporation (IDC) Headquarters.

www.idc.com/getdoc.jsp?containerId=prUS51335823.⁴

Iskander, Joseph. (2024). *Legal Horizons of AI in Egypt: Challenges and Prospects*. Andersen Egypt. <https://eg.andersen.com/ai-in-egypt/>.⁴

Jacob, Savio (2024). *AI Regulations Around the World: A Comprehensive Guide To Governing Artificial Intelligence*. www.spiceworks.com/tech/artificial-intelligence/articles/ai-regulations-around-the-world/.

Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. *The Alan Turing Institute*. <https://doi.org/10.5281/zenodo.3240529>⁴

Muñoz, Victor. (2024). *Operation regulation: Strengthening Latin America's AI Governance*. European Council on Foreign Relations. <https://ecfr.eu/article/operation-regulation-strengthening-latin-americas-ai-governance/>.⁴

NCSL. (2024). *Artificial Intelligence 2023 Legislation*. www.ncsl.org/technology-and-communication/artificial-intelligence-2023-legislation.⁴

Ovide, Shira. (2022). Your selfies are helping AI learn. You did not consent to this. A guide to the sticky ethics of fun AI tools. *The Washington Post*. www.washingtonpost.com/technology/2022/12/09/chatgpt-lensa-ai-ethics/⁴

Patowary, Upakul. (2023). Artificial Intelligence and Mary Shelley's *Frankenstein*: A comparative analysis of creation, morality and responsibility. *Integrated Journal for Research in Arts and Humanities*, Vol. 3, No. 4, pp. 121–127.⁴

Pazzanese, C. (2020). Great promise but potential for peril: Ethical concerns mount as AI takes bigger decision-making role in more industries. *The Harvard Gazette*. <https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/>.

Sulaiman, Shahid; Olën, Davin and Bezuidenhout, Monique (2024). *AI Regulation and Policy in Africa*. www.dentons.com/en/insights/articles/2024/june/13/ai-regulation-and-policy-in-africa⁴

Symbio6. (2024). *Critics Raise Concerns About the EU AI Act*. <https://symbio6.nl/en/blog/criticism-of-eu-ai-act#:~:text=A%20primary%20AI%20Act%20criticism,impede%20effective%20implementation%20and%20compliance>.⁴

Szczepański, M. (2024). *US Approach to Artificial Intelligence*. European Parliament. [www.europarl.europa.eu/RegData/etudes/ATAG/2024/757605/EPRS_ATA\(2024\)757605_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/ATAG/2024/757605/EPRS_ATA(2024)757605_EN.pdf).⁴

Tobin, J. (2024). *Artificial Intelligence (Regulation) Bill*. House of Lords. <https://researchbriefings.files.parliament.uk/documents/LLN-2024-0016/LLN-2024-0016.pdf>.⁴

Wallach, W., and Allen, C (2009). *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.⁴

White and Case. (2024a). *AI Watch: Global Regulatory Tracker: Canada*. www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-canada.⁴

White and Case. (2024b). *AI Watch: Global Regulatory Tracker: South Africa*.
www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-south-africa.⁴¹

White and Case. (2024c). *AI Watch: Global Regulatory Tracker – Australia*.
[www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-australia#:~:text=Australia%20has%20not%20yet%20enacted,\(see%20item%2010%20below\)](https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-australia#:~:text=Australia%20has%20not%20yet%20enacted,(see%20item%2010%20below)).⁴²

World Bank. (2021). *Artificial Intelligence in the Public Sector*. World Bank Group.
www.worldbank.org/govtech.⁴³

Yang, Z. (2024). *Four Things to Know About China's New AI Rules in 2024*.
www.technologyreview.com/2024/01/17/1086704/china-ai-regulation-changes-2024/.⁴⁴

IV

The Future and Artificial Intelligence

The final part of this book consists of two chapters that explore the future and AI. The rate of AI development has accelerated rapidly over the past several years. It was only in late 2022 that ChatGPT was first released, and its uptake has been extremely impressive. It is therefore challenging to predict what the future will hold in the next five, ten, twenty- or thirty-years' time. Moreover, the progress of AI has been taking place concurrently with developments regarding computer processing power. It is suggested that computing processing power had been a serious limitation for AI development. However, we have now reached a stage where AI is virtually at the basis of all major software applications irrespective of the functions being executed, and whether it is integrated with all types of devices or associated with their control systems. What was considered science fiction several years ago has now become a reality.

Generally, history has shown that humans are creative. They have the innate ability to innovate by developing tools and have the capability to learn, invent, and discover new ways of using these tools for the benefit of mankind. This human characteristic is typical throughout the millenniums and will continue in the AI evolution era. The mounting synergy between humans and AI will possibly fragment the boundary between human and machine intelligence to enable human evolution to reign as the supreme intellectual being. However, this is contingent on ensuring that the AI being created remains under the full and undisputable control of humans.

The first chapter of [Part IV](#) examines the likely evolution of the different types of AI and their supporting hardware. It will also discuss the development of the next economic approach and the future implications of AI. Finally, the chapter will research the era beyond AI. This chapter will illustrate that the three types of AI, namely Narrow AI, General AI, and Super AI will continue to develop. Narrow AI, which is designed to perform specific tasks, will likely reach maturity in the next several years. Narrow AI does not pose a hazardous risk because this AI model is inhibited from learning independently, is always under the control of humans and is trained to conduct explicit tasks like diagnosing and detecting illnesses. This chapter will also consider General AI and Super AI, which are exceedingly intelligent.

General AIs are intended to surpass human experts in their capacity for knowledge and problem-solving abilities across a wide spectrum of topics that include physics, chemistry, mathematics, and medicine, amongst others. However, General AI will typically still remain under the control of its creators. On the other hand, Super AI is intended to exceed human intelligence in all areas and act independently from its creators. The exceptional potential of Super AI increases the level of concern to extraordinary levels. The chapter while examining these AI developments will also consider other core technologies, like quantum processing and language models. The second aspect of this chapter will discuss the likely evolution of hardware and software associated with AI, such as networking approaches, organising memory, and increasing the speed of computations. It will specifically look at SLIDE (Sub-Linear Deep Learning Engine) that could outperform GPU hardware acceleration; and the Wafer Scale Engine, which are iPad-sized processors that are able to execute Generative AI programs up to 20 times faster but at one-fifth the cost.

The chapter also examines the differences between key economic approaches, such as Knowledge-Based Economy, Digital Economy, and the Automated Economy. At times, these terms are used interchangeably, however there is a distinct and significant difference between them. The impact of AI on the economic models is of concern regarding job

displacement, data privacy and security, and algorithmic bias. This chapter also addresses the likely future implications of AI on transport, education, healthcare, finance, law, military and cybersecurity, and the job market. Finally, the discussion turns to the topic of beyond AI, where organoid models and the development of Organoid Intelligence (OI) are examined. OI blurs the line between human cognition and machine intelligence giving rise to critical ethical arguments.

The second and final chapter of [Part IV](#) is AI and the future of humans. This chapter explores several themes that include AI and the future of humans; human and AI collaboration; and what humans will look like in the distant future. This chapter will examine the predictions proposed by futurist Ray Kurzweil in his books *The Singularity Is Near* published in 2005 and *The Singularity Is Nearer: When We Merge with AI* that was published in 2024, where he predicts a future where humankind will unite with superintelligent AI to increase our human intelligence a millionfold. A second aspect examined by this chapter is human and AI collaboration, where the working hours for humans may be further reduced to give them more leisure time to be enjoyed with their families. However, there are concerns that massive human job displacement may occur. Furthermore, those humans who remain in the job market require a meaningful and fulfilling life. There is consensus amongst experts that AI technology may provide great benefits to humans but like every powerful tool it requires careful supervision to ensure that no intentional or unintentional abuse takes place. Finally this chapter examines future human development where humans could be turned into cyborgs with hi-tech machine implants, regrowable limbs and nanotechnology that repair damaged tissue, rejuvenating human cells leading to immortality.

This final chapter warns humans not to play God. While eternal life might seem appealing, it loses its significance if it not meaningful and fulfilling, and most important devoid of feeling. Humans should avoid living under the shadow of technology but as humans we must maintain our intellectual superiority by ensuring that technology remains a tool to be exploited for the good of humanity.

Chapter 12

Future Developments of Artificial Intelligence

DOI: [10.1201/9781003653950-16](https://doi.org/10.1201/9781003653950-16)

“Success depends upon previous preparation, and without such preparation, there is sure to be failure.”

Confucius (Kong Qiu)

Chinese philosopher of the Spring and Autumn period (551-479 BCE)

A decade ago, it was very unlikely that one would have envisaged having a conversation with a computer that closely resembled a human. One only needs to remember that ChatGPT was first released only a few years ago, precisely end of November 2022. There has been so much technological progress in every aspect, particularly AI, in the past five years, that it is very difficult to keep pace with what is taking place. It is even more challenging to think about what capabilities AI will likely experience in the next five, ten, 20 or even 30 years' time.

Ever since the end of World War II, technological development appears to have progressed at an exponential rate. At the basis of most of this technological development is the processing power attained through

computer technology and associated software. Currently, AI is practically at the foundation of every major software application being implemented, no matter the nature of the function, whether it is integrated with all types of devices or associated with their control systems. It is very demanding to predict what will take place. What appeared to be science fiction some years ago is now a reality. This chapter will examine some of the probable developments in AI. However, one thing is certain, AI will likely impact every imaginable aspect of everyday human and nonhuman life irrespective of its nature.

[Marr \(2024\)](#) argues that by now, it's clear that AI is the great breakthrough technology of our times, and as with other breakthrough technologies like fire, steam power, electricity, and computing, now it exists, and it is only going to continue to evolve and become more capable. Furthermore, he contends that a unique feature of AI is its ability to learn and improve itself and generate more advanced AI that will accelerate in ways that humans have not previously seen. This chapter will discuss several aspects that include the development related to: (a) the types of AI; (b) the hardware associated with AI; (c) the evolution of the next economic approach; (d) the future implications of AI; and (e) beyond AI.

The author does not claim to be an oracle, but by examining the current and foreseeable capabilities of AI and researching the thoughts of leading scientists and practitioners it might be possible to gain an informed insight into the expected advancements of the future.

12.1 INTRODUCTION

Fundamentally, it must be recognised that humans have been evolving over a very long time, with a continuous evolutionary process reaching to the current anatomical modernity. Evidence indicates that bones of primitive Homo sapiens first appeared in Africa about 300,000 years ago that had brains as large or larger than ours. These were followed by anatomically modern Homo sapiens about 200,000 years ago, with a brain structure being modern by at least 100,000 years ago. At this juncture, humans had

craniums similar in mass and structure to ours. Therefore, as one may appreciate, the evolutionary process of human intelligence has been gradual and most likely achieved through significant development milestones.

With our current view of technological development, ten years is considered a very long time. Therefore, time is relative. Consequently, there is no comparison between how we view human evolution with the evolution of technology. Having said this, the evolution of humans has probably taken place through several major milestones, with each milestone having its own (lengthy) time epoch. Similarly, history has shown that technology has also evolved through different significant milestones although each milestone has a much shorter time era when compared with the human evolution milestones. Therefore, it is assumed that the evolution of AI will also be evident through several significant milestones with the rate of development being likely to be more rapid, and with each milestone probably resulting in a great jump forward in the capacity and usefulness of AI that will bring with it significant risks and challenges.

One of the most distinguishing features of humans during its evolution is its ability to innovate through the development of tools and its capacity to learn, invent and discover new ways of using these tools for the benefit of mankind. This feature has been innate in human nature and has guided us to our current state of technological development. This same process is applicable to AI. Admittedly, we have experienced several AI winters, which is considered as part of the normal development process since humans have been able to invent tools thousands of years ago. But as soon as we have been able to figure out how to apply AI, further tools have been developed and refined that made AI progressively more intelligent. We have now reached a stage where its use is only limited by our imagination or more precisely our creativity. The more creative our interaction with AI becomes, the more sophisticated it will become, and its development will likely keep on escalating with the synergy between humans and AI growing stronger.

The growing synergy between humans and AI will likely shatter the barrier between human intelligence and machine intelligence, so that

humans will continue to evolve their intelligence as the supreme being. However, our supremacy is contingent on making sure that what is created through AI remains under the full and unequivocal control of humans.

12.2 LIKELY EVOLUTION OF THE TYPES OF AI

In an earlier chapter the three types of AI were discussed, namely Narrow AI, General AI, and Super AI. This section will attempt to describe how these three types are likely to develop further in the future.

12.2.1 Narrow AI

Narrow AI or Weak AI is designed to do a specific task or a limited range of tasks. It is usually useful in applications like facial, speech, and image recognition; natural language processing; and recommendation systems. The emergence of Narrow AI is limited since it depends on the programming of an AI model that prevents it from learning independently. Hence, its future development is somewhat constrained.

However, this does not mean that it does not have considerable future potential in how it is applied to facilitate various human dependent functions. Narrow AI still has an important future capacity for performing repetitive tasks, particularly in time-consuming and hazardous applications. It has also the potential for enhanced decision-making capacity because of its faster data processing abilities that far exceeds humans, and can enhance efficiency and productivity, particularly in the health sector.

Moreover, Narrow AI is always under the control of humans since it is programme based and trained for specific tasks. This makes it useful for diagnosing and detecting illnesses, such as specific cases of cancer and for maintenance systems of manufacturing facilities. It also has potential for applications requiring real-time analysis of data from sensors on equipment to ascertain the optimum preventative and corrective maintenance schedules.

Narrow AI has potential in the entertainment sector through a possible text-to-video app that permits users to modify the ending of a production or substitute an actor's face with a younger version of an actor or the substitute the viewer's face instead. The media landscape will likely change from brands imparting information to brands sharing intelligence with their supporters, such as piloting viewers through family health plans by examining online the contents of their refrigerator or guiding them to renovate their own home by knowing certain parameters provided by the viewer. Currently, Narrow AI is the most exploited AI approach, but its application is expected to continue to develop in the coming years as follows:

- Further dedicated AI tools for specific industries. Narrow AI is likely to be further applied to enhance business efficiency by additional automation of routine tasks to release workers so that they may put more effort into strategic and/or value-added activities related to recruitment, addressing customer service queries at anytime from anywhere, and proposing products based on purchasing habits and individual preferences.
- Increased use of AI for predictive analytics and forecasting. Narrow AI is likely to be enhanced for better decision making through faster analysis of greater volumes of data for detecting patterns and predicting trends. This will provide more effective decisions related to marketing strategies, inventory management, and pricing policies resulting in attaining a competitive edge.
- Better integration with other technologies like IoT (internet of things) and blockchain. Blockchain is of particular interest for the financial sector where Narrow AI can be applied in assessing credit risk, detecting fraud, and providing personalised financial advice. Whereas IoT can be improved by having a more intelligent use and control of various devices from remote locations.
- Enhanced models for the manufacturing sector. Narrow AI can be enhanced to provide better preventative machine maintenance

schedules, optimising machine operation and performance programs, and better-quality control. This is likely to result in the further mitigation of downtime and expenditure, and the manufacturing of superior products.

- Enhanced models for the health sector. Narrow AI can be further improved for finding new medications and having clearer and more accurate analysis of medical images to detect disorders in a timely manner that help in forecasting patient outcomes. This is likely to result in more intensive patient care and new treatments.
- Further natural language processing (NLP) for better human-AI interaction. NLP is still an emerging technology that is likely to transform industries, such as finance and healthcare. It is expected to revolutionise applications, including chatbots, language translation, and speech recognition to facilitate human-AI interactions across various media.

Narrow AI systems are currently the most successful AI applications but there is still room for further improvement and innovative implementation. The success of Narrow AI stems from the fact that their supporting algorithms imitate human intelligence for specific practical tasks. They do not have cognitive abilities, and these specific systems are not able to diverge from their coded instructions. Hence, once extensively tested and validated they remain under the full control of their creators and operators, which translates to safe and trustworthy applications.

12.2.2 General AI

General AI or Strong AI is designed to perform any intellectual task that a human can do. General AI is theoretically able to reason, learn, and understand complex concepts. General AI is seen as transforming industries, particularly in transportation, manufacturing, healthcare, and finance. However, this is not yet possible.

Furthermore, the time horizon for General AI's full and practical attainment is highly uncertain. Some predictions date as far forward as 2050. However, there is strong rivalry between researchers and organisations to achieve General AI to disclose and apply the concepts of human-like cognition. Large Language Models (LLMs) are emerging as the main buttress for General AI. LLMs have been applied for GPT-3 and the cutting-edge GPT-4, with these models demonstrating advanced learning and reasoning abilities that are integrated with other tools to resolve complex problems. The implications of these models are far more important than their immediate utilisation and offer great potential for future development that extends over various sectors. They have the potential of providing huge societal benefits that will likely reinvent the way people live and work.

The future development of General AI brings with it many risks that need to be addressed. This specific topic was discussed in a previous chapter where the need for the proper guardrails was emphasised. Therefore, this section will only consider the future potential of this powerful and empowering technology. The future enhancement of General AI will unleash the cognitive competences of a human with the capacity to learn, discuss, plan, and perform jobs in many diverse environments, all reliably performed at the highest conceivable speed. General AI, with the support of LLMs are able to conduct open-ended conversations with natural language, together with their capacity to retrieve the immense repositories of knowledge in every imaginable area. This scenario provides just a glimpse of the future possibilities of General AI.

General AIs require much more refinement, but they are currently surpassing human experts in their capacity for knowledge and problem-solving abilities across a wide spectrum of topics that include physics, chemistry, mathematics, and medicine, amongst others. It is envisaged that General AI will revolutionise intelligence with its ability to understand context, solve unique problems, and even to plan. Some researchers view this potential development as “intelligence on demand” that could be used by anyone, at any time, from everywhere.

General AI is anticipated to help perform all everyday human cognitive tasks including many repetitive business processes. It will be possible for individuals to seamlessly integrate personal General AI routines, including IoT into their daily lives like how people currently take mobile Apps for granted. General AI will be able to learn in real time, thus permitting it to uninterruptedly supplement its knowledge base and to fine tune its capability through continuous training. Additionally, it will be able to retrieve the knowledge that has been retained of its previous diverse experiences, irrespective of source, nature, and context. [Savarese and Schwartz \(2024\)](#) identify several examples that General AI is envisaged to achieve in the future:

- Operating as an everyday copilot to humans to provide guidance and assistance in various diverse tasks ranging from shopping to planning functions.
- Suggesting expert-level knowledge of diet, fitness, medication, and disease mechanisms, leading to effective diagnosis and therapies.
- Helping healthcare personnel related to administrative and logistical activities, lessening their workload and minimising medical errors.
- Entertaining by creating entirely new works based on preferences and recommending movies, music, and books according to individual tastes.
- Functioning as a career collaborator, mentor, and problem-solver, and even contributing to the development of new products and industries.

General AI has great potential for helping individuals to have better lifestyles and facilitate a healthier and satisfying work environment. However, as stated previously, it must be safe and trustworthy.

12.2.3 Super AI

The distinguishing feature of Super AI is its expected ability of exceeding human intelligence in all areas. Super AI is still in the futuristic realm and is

a cause for ethical concern. When available, Super AI would be able to solve complex problems that are beyond human abilities and would be able to learn and adapt at a rate that far exceeds human intelligence. Super AI is viewed as the new era of AI because it represents a gigantic leap beyond human-level AI and implies intelligence that far exceeds that of humans. Having this level of intelligence, Super AI has the potential of creating innovations that goes beyond human imagination and will have the ability to give new insights and address complex worldwide challenges. However, this extraordinary potential of Super AI raises the level of concern to unprecedented levels. Furthermore, the development of Super AI will likely depend on advances in quantum computing and other leading-edge technologies.

As stated previously, Super AI is still at its infancy, and when developed will have unmatched capabilities that will include exceptional features, namely:

- Greater cognitive abilities that exceed human capacities, allowing it to process and comprehend immense volumes of data, learn from varied sources, and answer complex problems more effectively than humans.
- Self-Perfection faculty that constantly upgrades its own algorithms and knowledge base as an independent entity without human involvement to optimise its operation.
- All-purpose intelligence capacity that allows it to understand, learn, and apply knowledge across a broad range of topics and tasks.
- Supreme speed capability that far exceeds human thought giving it the ability to execute computations, analyse data, and make decisions at speeds that would take many humans years to solve.
- Analytical supremacy that empowers it to generate highly accurate forecasts about future incidents, trends, and behaviours to facilitate rational decision-making.
- Ingenuity that is superior to human capability to create new ideas, designs, and solutions in fields related to science, technology,

engineering, and art.

- Emotional intelligence that may have a capacity to understand and respond to human emotions with extraordinary accuracy to assist in having enhanced human-machine interactions for improving outcomes related to customer relations and mental health.
- Ethical thinking capabilities that permit it to consider the wider influence of its decisions regarding society, the environment, and future generations.
- Multitasking abilities to tackle many complex tasks concurrently without undermining its operation and performance.
- High flexibility in its learning potential and adapting to new circumstances without major program amendments or human assistance.

The above envisaged Super AI unique features will likely bring numerous benefits across various fields. Its analytical and predictive capability might be used to examine possible environmental scenarios to combat climate change by proposing approaches that better administer natural resources and lower carbon emissions. Its ability to process more complex data may help to provide scientific breakthroughs in medicine and other fields of research. This in turn may facilitate much better healthcare through the provision of targeted treatment procedures for specific individuals through its ability to diagnose illnesses and possible treatments. Super AI will have the ability to address any issue that is brought to its attention, whether to optimise productivity, make better decisions, or reduce any types of errors in different fields, such as economics, finance, manufacturing, logistics, public safety, engineering, medicine, and many other fields.

However, Super AI also has its inherent risks and dangers that are not easily addressed. These might include loss of human control in certain critical circumstances that may not be in the interest of humanity. For example, misuse by malicious actors to launch cyber-attacks or having autonomous weapons without a human control mechanism. These may

threaten international security and stability. Furthermore, Super AI systems might learn to have different human ethical and moral values, including higher surveillance and the erosion of privacy with the possibility of misuse by authorities and other influential organisations. A possible danger is a high replacement rate of humans in everyday jobs that could lead to economic displacement and social inequality. On the other hand, these risks and hazards can be mitigated significantly through the regulation in the development, deployment, and use of AI, irrespective of its type as discussed in the previous chapter. World leaders cannot be complacent because the risks and dangers are real and not imaginary.

The global race for developing Super AI is unquestionably on, with many developed countries and organisations investing substantial amounts in research and development. Industrial experts claim that the USA, EU, and China are the frontline runners in this race budgeting billions of dollars for its development to academic institutions, and for establishing strategic alliances. Additionally, high-tech companies, such as IBM, Microsoft, Google, and Amazon have initiated development programmes to assist in the development of Super AI. Moreover, various governments are providing grants to start-ups focusing on machine learning, neural networks, or quantum computing. Many start-ups are also being acquired by established companies to exploit their respective breakthroughs. Academic and research institutions are playing a major role in experimenting with leading-edge technology and collaborating closely with major hi-tech companies and governments to make substantive progress and significant breakthroughs.

Given its great potential, there are many use cases that might be implemented regarding Super AI. The following are some potential use cases:

- Financial industry: Super AI provides many possibilities for the financial industry, such as better and speedier investment decisions, fraud detection systems, and enhanced risk management and regulatory compliance. For example, Super AI can analyse a huge volume of transactions in real-time to detect anomalies and possible

violations of regulations with outstanding accuracy, thus radically lessen the risk of non-compliance and related fines. It may also be applied to predictive compliance management by assessing historical data and current trends to forecast possible regulatory amendments and proactively modify compliance processes, guaranteeing proactive adherence, and curtailing risks. Super AI has potential for automated regulatory reporting, freeing up human resources to do more analytical tasks. Super AI has the capacity of analysing huge datasets of regulatory rulings, legal precedents, and industry best practices to detect patterns and connections that humans may not detect, providing deeper insights into regulatory requirements and risks.

- **Urban Planning and Development:** Super AI might be used to design and manage smart cities, improving traffic flow, public transportation, energy utilisation, waste management, and initiating a more effective and habitable urban environments.
- **Autonomous Advanced Systems:** Super AI will have advanced algorithms that are capable of operating and controlling self-drive cars, drones, and robots, resulting in safer and more effective transportation networks.
- **Healthcare Innovation:** Super AI is capable to provide better advanced medical diagnostic tools and offer personalised treatment options for patients.
- **Tailored Education:** Super AI can offer unique learning experiences according to one's individual requirements and permitting one-on-one interaction between virtual-teacher and pupil.
- **Human-Machine Partnership:** Super AI will have the ability to assist in a seamless alliance between humans and machines to increase productivity and innovation across numerous industries.
- **Space Exploration:** Super AI has the potential to assist in advance space exploration leading to new discoveries and colonisation efforts. This might be achieved through leading-edge technologies, planning

and executing missions, supported by the comprehensive analysis of data from previous space expeditions.

There is no doubt that the availability of Super AI will lead to new dimensions in the development of human conquest. However, its tremendous, expected capabilities must be kept under the scrupulous and continuous control of humans.

12.2.4 Strategic Approaches and Core Technologies

The development of AI, particularly the route towards Super AI will likely involve several milestones and enhanced methodologies to ensure its successful and beneficial achievement. As noted in a previous chapter, this development is based on the evolution of computational power, enhanced algorithms, better machine learning methods that include deep learning models, neural network applications, amongst others, which have evolved over many years of practical usage. This development is not in static cycles but through an iterative process resulting in the exponential growth in the capacity and intelligence of these systems. The reason for this is that they continuously redefine and finetune themselves by resolving defects until the system completely optimises itself throughout its life cycle. In this way, Super AI will achieve a heightened level of intelligence that far exceeds that of humans with the intent of attaining continuous self-improvement.

Additionally, technologies like Super AI apply transfer learning and hybrid models to enhance their abilities to an extraordinary magnitude. Transfer learning enables AI systems to apply knowledge that it gains in one domain on an entirely dissimilar, but related problem, to increase efficiency, leading to improved performance in the training stage. On the other hand, hybrid models endeavour to merge the capability of deep learning algorithms with human-type logical thinking using symbolic reasoning and neural networks, thus enlarging AI's general knowledge and capacity to solve a broader range of tasks that will eventually lead to superintelligence. However, this is not achievable without some significant

core technologies pushing AI's development, particularly Super AI growth, including:

- **Quantum Processing.** The Quantum computer is at the forefront of groundbreaking technology that will considerably escalate computational power by the processing of intricate algorithms and immense databases at speeds unparalleled in the history of computer technology. The Quantum computer enhances parallel processing by allowing the simultaneous execution of tasks at extraordinary speed, which is so essential for Super AI.
- **Cutting-edge Machine Learning and Deep Learning.** Machine learning with a deep learning capability is also a prerequisite for Super AI since they contain training algorithms that utilise huge amounts of data, which allow them to detect patterns and make decisions with very little human intervention. Deep learning integrated within AI systems can address complex settings and continually enhance their performance during their life processing cycle. They can adapt to changing environments and learn from their experiences, which make these features essential for Super AI.
- **Neural Network Architectures.** This technology makes use of convolutional neural networks (CNNs) that are essential for image recognition and processing, due to their ability to recognise patterns in images and recurrent neural networks (RNNs) that is based on a deep learning model and trained to process and convert a sequential data input into a specific sequential data output. Both Neural Network types are critical for Super AI since they enhance the ability of AI systems to understand and process very complex data structures like images, dialogue, or text.
- **Evolutionary Algorithms.** Evolutionary computation is a sub-field of AI that is widely applied in very complex optimisation problems and for continuous optimisation environments, where there are too many variables for traditional algorithms. These techniques are inspired by biological evolution that involve iteratively perfecting a population of

possible solutions exploiting processes like natural selection, including mutation, crossover, and selection. Evolutionary algorithms are highly successful in solving difficult optimisation problems and are important in ASI development since they allow for the emergence of highly efficient and strong AI systems.

- **AI Software Programming.** This technology is aimed at having AI systems with the ability to generate computer program code and being able to optimise the generated program without any human intervention. Hence, an AI system will be able to develop error free software applications unaided that will result in greater software efficiency and processing speed. This type of technology is essential for Super AI because it will enable the system to be fully independent, resulting in a continuous self-improvement environment that can create innovative objects.

Most of the above technologies are at the experimental stage but once developed will provide the robust foundation for the creation of Super AI and beyond.

12.2.5 Language Models

A language model in AI is fundamentally a computer program that is trained on a large amount of data, including conversations, articles, books, and other notable material, which is trained in how people apply words, sentences, and phrases to communicate. There are two categories of language models, namely statistical models and those based on deep neural networks. The focus of this section is on the potential of natural language processing (NLP).

Basically, NLP allows robots perceive, interpret, and create human language and narrows the communication variance between people and AI systems by merging computational linguistics, machine learning techniques, and deep learning models. While significant development has taken place in this field, much more needs to be discovered to achieve perfection.

NLP is already changing the interaction between humans with technology and is helping to break the communication barrier. However, the communication barrier is not just speech. Thus, this technology is merging speech recognition and computer vision with natural language ability to allow AI to process and interpret information simultaneously across several methods in which something exists or experienced. Transformers and other deep learning models are mimicking human responses. This technology has the potential of making human-AI interactions more natural and spontaneous as it develops further, narrowing the disparity between AI and human communication in personal and business settings.

NLP is further advancing towards having more complex natural language processing and creation that is based on the developments taking place regarding computational linguistics, deep learning, machine learning models, and Generative AI models. The future of NLP will be based on merging other AI fields that include computer vision, robotics, and the integration with programming languages and concepts from Neuro-Linguistic Programming, which is expected to further narrowing the margins between human and machine communication.

These and other AI technological developments are taking humans closer to the concepts of theory of mind that has the objective of interacting AI with the thoughts and emotions of humans and self-awareness, which is the long-term objective for AI, where it achieves enlightenment and becomes self-aware having an independent intelligence and will not need to be under the control of the people that created it.

12.3 LIKELY EVOLUTION OF HARDWARE AND SOFTWARE ASSOCIATED WITH AI

Recent years have experienced great developments in AI due to increasing investment by governments, the private sector and research institutions. It is therefore not surprising that the AI landscape is drastically changing. This changing AI landscape is not only due to more effective algorithms and other sophisticated software tools but also because of advances in hardware.

For example, startup entities that specialise in specific development aspects are creating innovative and diverse approaches of networking, organising memory, and speeding up computations that will have a profound impact on the methodologies being applied by cutting edge organisations to design and implement AI algorithms. According to [Kan \(2024\)](#), a startup called Cerebras Systems has been developing iPad-sized processors called the Wafer Scale Engine that is claimed to be able to run generative AI programs up to 20 times faster but at one-fifth the cost.

Leading researchers are challenging the idea that graphics processing units (GPUs) are superior to Central processing units (CPUs) for executing AI workloads. Computer scientists from Rice, supported by collaborators from Intel are experimenting with a new category of algorithms, referred to as SLIDE (Sub-Linear Deep Learning Engine), which is the first smart algorithmic implementation of deep learning on CPU that can outperform GPU hardware acceleration on industry-scale recommendation datasets with large fully connected architectures ([Boyd, 2020](#)).

Furthermore, neural processing units (NPUs) have also come on the scene that are dedicated processors designed specifically for accelerating neural network operations and AI tasks. Whilst CPUs and GPUs are general-purpose processors, NPUs are optimised for a data-driven parallel computing. This makes them highly efficient at processing massive multimedia data like videos and images, and processing data for neural networks. This makes them particularly suitable for handling certain AI tasks, such as speech recognition, background blurring in video calls, and photo or video editing processes like object detection. Intel is combining CPUs, GPUs, and NPUs as co-functional processor for achieving the type of high-performance inferencing tasks required by AI. It is claimed that these dedicated processors accelerate AI tasks and are very suitable for preparing pre-trained neural networks for the inferencing stage of AI, since competences learned during the training stage are applied to make predictions.

Software is also undergoing new developments with academic and industry researchers exploring ways to maximise the speed and functional

capabilities of current AI applications. Developing sentient machines that rival human general intelligence are of current and future interest. The term “sentience” refers to the ability to experience subjective sensations, such as pain, pleasure, or emotions. Current AI systems can only mimic these sensations to a certain degree, but they do not actually experience them as humans do. Therefore, a sentient machine must be able to have subjective experiences of its own. However, some researchers believe that the creation of a truly sentient machine will be possible, but it is still at the conceptual stage and is viewed as a very long-term goal. The basic reason for this prediction is the fact that scientists do not fully know how consciousness in the human brain occurs. On the other hand, many scientists believe that consciousness is a human only characteristic, and it will never be possible to replicate it in machines. Furthermore, the implications of sentient machines on society will be profound, particularly the potential for sentient machines to displace humans not only in the workplace but in other unimaginable scenarios.

Researchers are also working on integrating neural networks with symbolic AI to create neuro-symbolic AI that has the goal of creating systems, which can handle both data-driven pattern recognition and knowledge-driven reasoning. This integration facilitates several cognitive functions and application requirements. Some of the main elements of Neuro-Symbolic AI include the Integration Layer, Knowledge Base, and Explanation Generator, which can all be included in the Interface/User Experience (UI/UX). Neuro-symbolic AI has application potential in many fields ranging from healthcare, business management, finance, and brain and behavioural modelling.

12.4 LIKELY EVOLUTION OF THE NEXT ECONOMIC APPROACH

The world has seen several economic approaches, with each economic method representing the era of the time. Some world leaders talk about the

Knowledge-Based Economy, others refer to the Digital Economy, while others are now referring to the Automated Economy.

The starting point is to know what these terms specifically mean. The digital economy basically means that manual transactions or paper-based data are replaced by digital data, which facilitates its storage, retrieval, and transfer. In a knowledge-based economy the meaning of the digital economy is extended. Under a knowledge-based economy we learn from the digital data and use it to improve future processes and decisions. The automated economy extends these concepts even further, in that it leverages technology to facilitate the exchange of goods and services between users, often through peer-to-peer (P2P) platforms ([Cooperative Computing, 2024](#)). [Chapter 8](#) focused on the transformation of data (the basis for the Digital Economy) and specifically the Knowledge-Based Economy and how these were related and impacted by AI. This section will focus on the Automated Economy and Gig Platforms, and their relationship with AI.

[Cooperative Computing_\(2024\)](#) cites several examples of successful automated economy businesses such as:

- Airbnb: This company has an online system involving bookings of short-and-long-term homestays and experiences in various countries. Its online system allows it to be an intermediary charging a commission from each booking.
- Uber: A multinational transportation company having an online platform that allows it to deliver on demand ride-hailing services, courier services, food delivery, and freight transport.
- TaskRabbit: This company operates an online system that involves matching freelance labour with local demand. This permits individuals to seek assistance with tasks including personal support, furniture assembly, moving, delivery, and handyman work.

A common characteristic with these examples is their dependence on an online system supported by effective gig economy platforms. These gig platforms or gig marketplaces are online platforms that basically connect

individuals and businesses that work around short-term, temporary, or freelance work, where individuals offer their skills and services to clients or businesses on a project or task basis. Hence, gig economy platforms have a role of an intermediary that connect gig workers with potential employers. The outcome is that these platforms have profited from automation and economic growth by increasing their efficiency and optimising operations.

12.4.1 The Impact of AI on the Automated Economy

The examples cited above irrespective of their specific function, are all based on real-time online systems that match specific requirements parameters, and permit instance pricing and charging. AI will facilitate these systems through better matching algorithms since it has the ability to assess huge data volumes. Moreover, its high processing speed helps to reduce customer service times and mitigate transaction cost by matching customers' needs with the best service providers with no human intervention.

Additionally, the AI systems will permit price optimisation by providing computing prices that best fit the customers' needs depending on demand and supply conditions. For instance, when booking a ride on Bolt or Uber, the app has various options, one of which is economy and whether payment is by cash or electronic. The pricing will vary on demand, time, and location. Furthermore, these system features must operate in a secure and trustworthy environment. AI's superior ability to verify transactions helps to detect fraudulent activities and build user trust.

The enhancement provided by AI applications has created a robust basis for an autonomous economy and promoted a higher magnitude of personalised services that has generated a growth in customer loyalty. AI has the potential to provide better support for Gig Platforms that is likely to transform the future of work. AI is able to introduce a more efficient talent acquisition process through a higher degree of matching prerequisite parameters that will help the growth and acceptance of gig platforms. AI also has the capacity to optimise the allocation and scheduling of tasks that

gig platforms are greatly dependent on. AI's workload distribution algorithm has the potential to optimise the assignment of tasks based on skills, availability, and other essential parameters. Additionally, the AI application will also have the possibility of providing feedback and monitoring the performance of those conducting the assigned tasks. Gig platforms together with the potential of AI will facilitate the growth of the automated economy and assist organisations to locate the required talent and allow freelance workers to attain assignments that tally with their specific skills.

The above illustrates that AI will have a critical impact on the automated economy and will enhance the effectiveness of gig platforms. [Cooperative Computing_\(2024\)](#) identify the following impact of AI on the automated economy:

- **Increased Productivity and Efficiency:** AI and robotics are viewed as transforming several operational areas by increasing output and streamlining processes.
- **Enhanced Innovation and Competitiveness:** AI provides businesses with the ability to analyse data repositories to present enhanced market insights to assist them in creating novel products, services, and business models, thus increasing an organisation's competitiveness.
- **Opportunities and Challenges:** The growth in the application of AI is unquestionably reshaping the nature of work and creating new opportunities that more than balance any job losses.

It must be recognised that AI and the resultant increase in automation will transform economic development. However, it must also be understood that governments and policymakers must proactively manage this shift by formulating policies that encourage innovation and economic growth while considering the potential challenges. These challenges are of particular concern regarding job displacement, data privacy and security, and algorithmic bias. Having a secure and trustworthy environment will

guarantee that the generated economic benefits of AI and automation will be equitably distributed across society.

12.5 LIKELY FUTURE IMPLICATIONS OF AI

The likely future implications of AI as it is further integrated into various fields are unlimited and will continue to grow at an exponential rate. AI's capacity to evaluate immense volumes of data, identify information patterns, and generate informed decisions has opened new application opportunities that significantly increase an organisation's operational effectiveness.

There is no doubt that the long-term future implications of AI are unimaginable, but its future progress will also depend on the development of high-performance processors. AI is set to revolutionise various sectors some of which are identified below:

- **Transportation.** Autonomous vehicles and intelligent transportation systems are currently at the advanced prototype stage and will probably transform the way people will travel. In the long-term, AI is expected to progress autonomous vehicles and intelligent transportation from the prototype stage to full deployment. This is likely to lead to safer roads and assist to reduced traffic congestion with better traffic flow. Additionally, AI will be applied for traffic management systems that will optimise the road network to make transportation more sustainable and environmentally friendly.
- **Healthcare.** AI and healthcare have been at the forefront for decades. However, with new AI tools becoming available, healthcare will likely take a great leap forward in advanced medical diagnostics, personalised medicine, and drug discovery. AI tools are likely to transform patient care in the future by foreseeing and averting diseases, providing rapid response, delivering exceptionally accurate diagnoses, identifying personalised treatment plans, and discovering leading-edge therapies.

- Finance. AI together with natural language processing combined with machine learning are already being used by banks and financial consultants to effectively engage with clients regarding financial planning, scrutinising credit levels, fraud detection, insurance policy issues and customer service. Hence, future AI systems will likely enhance these services and also provide additional facilities related to complex investment strategies for the bigger investors.
- Law. AI applications will be developed to assist lawyers to draft contracts and prepare litigation cases.
- Education. AI is likely to revolutionise education at all levels in that students will receive educational material and tuitions tailored to their specific requirements. AI algorithms will establish optimal educational approaches based on students' individual learning styles.
- Marketing. AI's ability to access and analyse large quantities of data can facilitate the development of new products and services and also launch marketing campaigns that are more accurate and economical that are able to target specific consumers.
- Military and Cybersecurity. AI is already being used to safeguard a nation's security through AI-assisted military technologies. Some of these AI technologies are autonomous that do not require human intervention. In the future, AI utilising deep neural networks is likely to be applied at having military robots that are intelligent and able to perform various tasks, such as bomb clearance and defusing.
- Job market. AI will likely take over repetitive and monotonous tasks. However, it is also likely to create new opportunities that require higher skills. This has been the pattern throughout the different technological eras since the industrial revolution. Hence, AI will probably create new industries and occupations that will require different skills. Therefore, it is important for governments with the collaboration of the private sector to monitor developments closely and take proactive measures to ensure that individuals are trained and qualified to undertake the new occupations, and that educational

institutions adapt their curriculums to meet the future needs of the emerging job market.

The above implications will also mean that governments must update their respective ethical and regulatory standards. This topic has been discussed in detail in a previous chapter.

12.6 BEYOND AI

Just when AI is finally making head ways and is generally being accepted into society's collective consciousness, developments are taking place regarding a new type of infused human-technology intelligence, one derived from a blob of cells no larger than a grain of rice ([Iyer, 2024](#)). [Strickland \(2023\)](#) argues that computers powered by human brain cells called “organoid intelligence” could shape the future. Brain organoids are 3D in vitro culture systems derived from human pluripotent stem cells that self-organize to model features of the (developing) human brain. Scientists have been using organoids for almost two decades to safely conduct and test various experiments without using humans or animals as test subjects. Strickland contends that brain organoids do not replicate miniature versions of the human brain, but the pen dot-size cell cultures contain neurons that are capable of brain like functions, forming a multitude of connections.

Strickland mentions that Professor Hartung and his colleagues at Johns Hopkins Bloomberg School of Public Health and Whiting School of Engineering in Baltimore predict combining the power of brain organoids into a type of biological hardware more energy efficient than supercomputers. She maintains that these “biocomputers” would utilise networks of brain organoids to possibly transform pharmaceutical testing for diseases like Alzheimer's, and this might provide an understanding into the human brain and change the future of computing.

[Smirnova and Hartung_\(2024\)](#) argue that an emerging concept termed Organoid Intelligence (OI), which merges organoids with AI systems to generate learning and memory is currently in the early stages of

development. They contend that OI has the objective of modelling cognition and enabling biological computing applications. They further assert that brain organoids permit neuroscience experiments that were not previously possible with traditional techniques. Furthermore, they argue that brain organoids have the possibility to transform disease modelling, drug development, and the understanding of human brain development and disorders. Smirnova and Hartung maintain that the vision for OI is very similar to the origins of AI and signify brain organoids as a disruptive technology that is quickly evolving and gaining ground in many disciplines.

There is no doubt that a computer can process immense volumes data faster than humans. However, it is also a fact that the human brain is much more energy efficient, can operate effectively under different environments, and is better at learning and making complex logical decisions. Dr Hartung contends that organoid models open research on how the human brain works because one can start manipulating the system, doing things that one cannot ethically do with human brains ([Strickland, 2023](#)). However, organoid models are beginning to attract attention with legal debates likely to intensify. OI is viewed as having a lab-grown human-brain model that has the capacity to reproducing cognitive functions, such as learning and sensory processing. [Strickland \(2023\)](#) maintains that the brain organoids currently being used by Hartung will need to be significantly scaled up for OI, since the number of cells in each organoid is equivalent to a fruit fly's nervous system. Currently, a single organoid contains about 50,000 cells but for OI to work this would need to increase to 10 million cells.

Furthermore, an input/output interface would need to be developed to provide a gateway for communicating with the organoids so that information can be transmitted and received to determine what organoids are "thinking." [Strickland \(2023\)](#) argues that Hartung and colleagues have formulated a roadmap that involves using tools from bioengineering and machine learning, together with new advances to communicate across organoid networks that would permit more complex tasks. She cites Hartung as developing a brain-computer interface device like an electroencephalogram (EEG) cap for organoids, which is a flexible shell

that is densely covered with tiny electrodes that can pick up signals from the organoid and transmit signals to it. The ultimate objective is to develop a communication conduit between AI and OI to enable the two technologies to discover each other's capacities.

OI may initial commence like narrow AI that are designed to perform a specific task or a set of closely related tasks. For instance, OI might be applied in human medicine where brain organoids might be developed from skin samples of patients with neural disorders, enabling experts to exam how various medicines and other factors could impact them ([Strickland, 2023](#)). Another example cited is related to cognitive aspects of neurological conditions where researchers might compare memory formation in organoids derived from healthy people and from Alzheimer's patients and attempt to repair relative deficits.

[Smirnova and Hartung_\(2024\)](#) view OI as an emerging frontier that combines organoids with AI systems with the aim of using stem cell-derived organoids not just as models, but as active components integrated into biohybrid systems exhibiting cognition and learning. They cite early examples of linking neuronal cultures with robots in simple biohybrid systems, such as DishBrain from Cortical Labs that succeeded basic learning of computer games using networked neuronal cultures. [Smirnova and Hartung_\(2024\)](#) argue that OI offers unparalleled prospects to explain processes in neurophysiology and advance pharmacology and toxicology relevant to the brain and allow direct experimental access to study processes underlying neuronal signalling and network activity.

They view OI as having exciting promise to instigate new directions in neuromorphic engineering and pioneer unprecedented biological computing capabilities. As the reader may recall neuromorphic computing (or neuromorphic engineering) was covered in some detail in [Chapter 5](#). It basically refers to the design of both hardware and software computing elements that have the objective of modelling bio-inspired computer systems that imitate the human brain and nervous system. [Smirnova and Hartung_\(2024\)](#) maintain that as research clarifies how brain organoids display learning and information processing, which are grounded in

dynamic neuronal signalling, these findings might steer the development of hardware and algorithms emulating biological neural networks.

Similar to other technologies, the application of these advanced techniques is limited only by the creativity and perseverance of the researchers applying them. Admittedly, the development and using of brain organoids to generate OI is still at the embryotic stage and will likely take several decades of research before it comes into a practical fruition.

[Strickland \(2023\)](#) argues that several ethical concerns arise from generating human brain organoids capable of cognitive functions, including whether they can develop consciousness or feel pain, and if those whose cells were used to make them have any rights concerning the organoids. Having said this, it must also be recognised that like AI, the Belmont Report of 1979 regarding ethical treatment of human subjects could be leveraged to guide ethics within experimental research and algorithmic development for OI as well. [Strickland \(2023\)](#) asserts that OI blurs the line between human cognition and machine intelligence, and that technology and biology are advancing at a speed that could outpace the ethical and moral discussions which are needed; hence a vigorous approach must be adopted before the technology crashes into the moral abyss.

12.7 CONCLUSION

AI has survived several winters but the advancement made in these last several years are unprecedented. This chapter has explored the likely future scenarios related to several aspects, namely:

- Likely evolution of the types of AI. This has illustrated how Super AI with its ability to exceed human intelligence in all areas will be able to solve complex problems that are beyond human abilities and would be able to learn and adapt at a rate that far exceeds human intelligence. However, this extraordinary potential of Super AI brings with it ethical concerns. Moreover, the further development of Super

AI will probably depend on advances in quantum computing and other leading-edge technologies.

- Likely evolution of hardware and software associated with AI. It is no longer a question of having better GPUs and CPUs but having an emerging NPUs. NPUs are dedicated processors designed specifically for accelerating neural network operations and AI tasks, which are optimised for a data-driven parallel computing. Creating sentient machines that challenge human general intelligence are also under way, with Neuro-Symbolic AI being on the horizon. Neuro-Symbolic AI mainly involves the integration of neural networks with symbolic AI that has the aim of generating systems, which can handle both data-driven pattern recognition and knowledge-driven reasoning.
- Likely evolution of the next economic approach. In modern times there have been several economic approaches, namely the Knowledge-Based Economy and Digital Economy. Currently, the world of economic development is moving towards the Automated Economy. Each stage of an economic approach builds on the other, with the automated economy extending the economic concepts by leveraging technology to facilitate the exchange of goods and services between users, mostly through peer-to-peer (P2P) platforms known as gig economy platforms.
- Likely future implications of AI. The long-term future implications of AI in all fields are unimaginable, but its future evolution will also depend on the maturity of high-performance processors, such as the quantum computer. AI is set to revolutionise all conceivable fields, ranging from transportation, healthcare, finance, law, education, advertising, military, and cybersecurity, to our very own job market.

Finally, this chapter examines the prospects beyond AI. However, AI is still at the development stage and moving towards maturity in certain limited areas. Yet scientists are already exploring beyond AI and are experimenting with organoid intelligence (OI). These are computers that are powered by human brain cells called Brain Organoids, which are 3D in

vitro culture systems derived from human pluripotent stem cells that self-organise to model features of the (developing) human brain. OI is seen as the emerging leading-edge technology that merges organoids with AI systems with the aim of using stem cell-derived organoids as active components, which are integrated into biohybrid systems exhibiting cognition and learning. This brings technology closer to emulating human thought processes and related emotional and physical feelings.

However, these technologies also mean greater responsibility in how they are eventually applied. Governments must update their respective ethical and regulatory standards with closer international cooperation to ensure that any developments taking place are safe and trustworthy, and above all abiding by exceptional ethical and moral standards. AI and OI have to a certain degree shifted the debate from a technical focus to a very intense philosophical topic because it deals with the issue of what the long-term future holds for human evolution.

REFERENCES

- Boyd, Jade. (2020). *Deep Learning Rethink Overcomes Major Obstacle in AI Industry: SLIDE is First Algorithm for Training Deep Neural Nets Faster on CPUs than GPUs*. Rice University. <https://news.rice.edu/news/2020/deep-learning-rethink-overcomes-major-obstacle-ai-industry>.[↵]
- Cooperative Computing. (2024). *How AI is Disrupting Traditional Business Models: A Look at the Automated Economy and Gig Platforms*. Cooperative Computing. www.cooperativecomputing.com/blog/automated-economy/#:~:text=Automated%20economy%20refers%20to%20an,Airbnb%2C%20Uber%2C%20and%20TaskRabbit.[↵]
- Iyer, Adithi. (2024). *Is OI the New AI? Questions Surrounding "Brainware"*. Harvard Law School. <https://blog.petrieflom.law.harvard.edu/2024/03/13/is-oi-the-new-ai->

[questions-surrounding-brainware/#:~:text=There%20are%20several%20reasons%20that,decision%20making%20problems%20than%20pure.↵](#)

Kan, Michael. (2024). Startup goes after Nvidia with Giant AI chips built for speed. *PC Mag*. www.pcmag.com/news/startup-goes-after-nvidia-with-giant-ai-chips-built-for-speed.↵

Marr, Bernard. (2024). Where will Artificial Intelligence take us in the future? *Forbes Accolades*. www.forbes.com/sites/bemardmarr/2024/08/16/where-will-artificial-intelligence-take-us-in-the-future/.↵

Savarese, Silvio., and Schwartz, Peter. (2024). *Artificial General Intelligence: Casting a Vision for a Better Future*. Worldwide. www.salesforce.com/news/stories/what-is-artificial-general-intelligence/.↵

Smirnova, Lena., and Hartung, Thomas. (2024). The promise and potential of brain organoids. <https://onlinelibrary.wiley.com/doi/10.1002/adhm.202302745.↵>

Strickland, Ashley. (2023). Move over, artificial intelligence. Scientists announce a new ‘organoid intelligence’ field. *CNN*. <https://edition.cnn.com/2023/03/02/world/brain-computer-organoids-scn/index.html.↵>

Chapter 13

AI and the Future of Humans

DOI: [10.1201/9781003653950-17](https://doi.org/10.1201/9781003653950-17)

“We have seen when the earth had to be prepared for the habitation of man, a veil, as it were, of intermediate being was spread between him and its darkness, in which were joined in a subdued measure, the stability and insensibility of the earth, and the passion and perishing of mankind.”

John Ruskin

English writer, philosopher, art historian (1819-1900)

The above quote implies that when the earth was being made suitable for human habitation, there was a transitional stage where an intermediary presence existed between humanity and the darkness of the earth. This intermediary existence features both the stability and impassiveness of the earth, as well as the emotions and fragility of human beings, although in a tempered manner. The quote emphasises the notion that this intermediate status played a crucial role in bridging the gap between the inertness of the planet and the dynamic nature of humans. However, humans need to be vigilant to ensure that they retain their emotional quality and continue to control their own destiny.

In previous chapters it has been illustrated that AI applications are constantly expanding. Then again, it has also been shown that developing AI systems takes a significant amount of time and effort, and more importantly, it cannot presently materialise without human involvement. Currently, all types of AI applications, irrespective of whether they involve robotics, Chatbots, or self-driving vehicles, are all dependent on human intelligence. The critical issue is whether this situation will remain.

There is no doubt that AI will have an impact on future jobs and therefore impact the future of humans, both as individual beings and how they are to subsist. AI will likely take over many jobs that are manually performed by humans. It is also likely that AI will also create new job opportunities for humans. However, the crucial issue is whether the job opportunities generated by AI will significantly outnumber the job losses or will there be a time when individuals will be paid to stay home?

As early as 2018, Elon Musk was reported on CNBC as saying that there is a pretty good chance we end up with a universal basic income, or something like that, due to automation ([Clifford, 2018](#)). Another way of viewing this is that if there are fewer jobs, then there will likely be substantial unemployment, with the purchasing power of families and individuals to purchase the goods being produced by AI diminishing. Therefore, economic development will likely shrink, which in turn means less profits for the industrialists and investors, and less taxation revenue for government. Hence, the introduction by governments of a universal basic income payment will likely become essential to sustain the purchasing power of humans that is so important for economic growth and sustainability.

In the same CNBC programme, Musk provides an example regarding the future of semi-trailer trucks that will be able to drive themselves ([Clifford, 2018](#)). He argues that although this will not become the status quo for a while, it will mean that there will not be a need for quite as many truck drivers. He expressed the view that some drivers will become fleet operators, responsible for monitoring the status of a fleet of trucks and remotely resolve concerns that may arise with individual vehicles. Those

making the transition will likely have a more interesting job than just driving a truck but those that do not make the transition will lose their job or find an alternative one. On the other hand, people will likely have more leisure time. However, there is no point in having more leisure time if people do not have the spending power to enjoy this, hence the need for a universal basic income.

13.1 INTRODUCTION

The impact of AI on the future of jobs will likely be substantial. Industries are already witnessing the effect of AI that has resulted in the digitalisation and automation of many manual processes across a diverse range of functions. Tasks that involve repetition or the use and interpretation of large amounts of data are candidates for an AI application, with the consequence that human intervention may be diminished or not be required at all. The outcome is likely to be *further* job losses. The word “*further*” is being used in the previous sentence because past automation (without AI) had already resulted in job losses. However, it is also likely that AI will create new and interesting opportunities by eliminating monotonous human-intensive tasks and replacing them with challenging ones. Similar to previous technological developments, the outcome of this new AI technology will likely create new fields of study and work.

Consequently, while certain jobs may become extinct, new opportunities and careers will be generated. One should recall that it was in the early 1960s that robotics technology was introduced in the automotive industry at the New Jersey, General Motors assembly line. [Conversion Capital \(2024\)](#) maintains that since 2004, each industrial robot that has been installed in the manufacturing sector has displaced an average of 1.6 people. It is likely that this proportion would be much larger if we go back to the 1960s. However, as stated previously, time will tell whether the future rate of new job creation will be greater than job extinction.

A previous chapter discussed the envisaged new economic model with AI being an important driver for economic growth. There is no doubt that

AI has the capacity to increase productivity and collaboration within the organisation by providing new opportunities to introduce new products and services that will ultimately lead to the organisation's growth. With an appropriate communication strategy, this will likely result in an increase in the demand for goods and services for all types of organisations. Cumulatively, this outcome will hopefully likely empower and drive an economic growth model that distributes wealth equitably and increases the living standards for all citizens.

However, in the epoch of AI it is important to understand that employment should not be viewed as just sustaining a standard of living. Humans should not view employment as a means of existence but as a means of being fulfilled and contented with one's existence. Employment should promote the critical human need of belonging, commitment, co-creation, and dedication. The necessity to be needed is an individual's sense of significance rooted in the feeling of being part of an organisation or cause beyond themselves. Hence, if a job or task is replaced by AI automation it should be substituted by a function that is meaningful and provides a sense of belonging for the human performing that function. There is no doubt that AI provides the opportunity for qualified individuals to focus their human intellect on producing creative solutions allowing the AI systems, including robotics, to manage the tedious and physical tasks that were previously executed by humans.

The critical issue is whether AI will replace humans. This issue is not new. Every technological era since the industrial revolution had to confront this possibility, irrespective of whether the technology was mechanically or electronically based. It is also a fact that every technological development since the industrial revolution has made life much safer and easier for humans to perform tasks and generated substantial economic growth for all concerned. AI could automate specific tasks and jobs and therefore it will also replace humans in these areas. Moreover, AI is particularly suitable for performing repetitive, data-driven tasks, and making data-driven decisions. Therefore, any human activity with these types of characteristics is a candidate for an AI application.

AI becomes a threat to human livelihoods when it fully takes over functions that require human skills, such as creativity, critical thinking, emotional intelligence, and complex problem-solving. However, while these functions continue to be performed by humans but in collaboration with AI applications, then the outcome of this synergy will augment human capabilities. This will enable humans to focus on those tasks that are most closely aligned with their organisation's mission, goals, and values, which require human ingenuity and expertise. It is imperative that AI, like every other technological development, is viewed as a powerful tool, which is used with reverence to improve productivity and assists in creating new opportunities rather than as a complete substitute for human involvement. This chapter will explore the above themes in greater detail and specifically examine AI and the future of humans; human and AI collaboration; and what will humans be likely to look like in the long-term future.

13.2 AI AND THE FUTURE OF HUMANS

AI has already made an impact on all types of industries and is likely to have an even greater impact in the next several decades. Ray Kurzweil in his book *The Singularity Is Near* published in 2005 predicted that computers would reach human-level intelligence by 2029 and that humans would merge with computers and become superhuman around 2045, which he referred to as “the Singularity.” Additionally, in his latest book *The Singularity Is Nearer: When We Merge with AI* published in 2024, he predicts a future where Humankind will unite with superintelligent AI where our human intelligence can increase a millionfold. He further predicts that this will not happen suddenly, but at our own pace, with the human Einsteins in each field leading the way in their specialties.

According to [Vassev \(2021\)](#), such predictions are a cause of major concerns. Previous chapters have already emphasised these concerns. The source of these concerns is widespread among scientists, politicians, and technologists. It is widely recognised that AI is here to stay but precautions need to be taken to prevent or mitigate its dark side. Vassev argues that AI

is likely to make most people better off, but its advances will also affect what it truly means to be human in the 21st century. [Freedland \(2023\)](#) is of the same view, but he maintains we have overcome terrifying dangers before. Freedland a journalist at *The Guardian*, reported that Geoffrey Hinton, Vice President and engineering fellow at Google resigned his post at Google, saying that he had undergone a “sudden flip” in his view of AI’s ability to outstrip humanity and confessing regret for his part in creating it. Who is Geoffrey Hinton? Geoffrey Hinton is a British Canadian computer scientist, cognitive scientist, cognitive psychologist, known for his work on artificial neural networks, which earned him the title as the “Godfather of AI.” He was awarded the 2024 Nobel Prize in Physics, shared with John Hopfield. Therefore, the reason for his resignation cannot be taken lightly.

According to Freedland the grave concern is a category leap in the technology, whereby AI becomes tremendously powerful with general intelligence capabilities, where it is no longer reliant on specific prompts from humans, but begins to develop its own goals and its own agency (i.e., Super AI). There is wide consensus that the possibility of Super AI is no longer science fiction but a reality in the not-too-distant future, particularly with the development of Quantum computers and possibly Organoid Intelligence (OI). It is also widely recognised that AI is already a serious threat. For example, [Freedland \(2023\)](#) cites the announcement of 55,000 planned redundancies at BT that he views as being due to technological developments, including AI. Another example he cites is an apparent fake photograph generated by AI of an explosion at the Pentagon that went viral, which caused the US stock market to plummet. [Vassev \(2021\)](#) argues that there is the bright and dark side of the AI future. [Table 13.1](#) provides some examples that Vassev notes regarding the bright and dark side of AI future.

TABLE 13.1 Examples of the Bright and Dark Side of the AI Future📄

Bright	Dark

Bright	Dark
Precision medicine to understand how a person's genetics, environment and lifestyle can assist to determine the best approach to prevent or treat a certain disease; digital therapeutics, custom-designed drugs and improved diagnosis for creating treatments that are cheaper, accessible, accurate to enable humans live longer and healthier lives.	Mass surveillance with the consequential destruction of digital privacy. China's social credit system has increased surveillance to all aspects of life, including assessing citizens' behaviour and trustworthiness, and having extreme control on their population.
Driverless cars using deep learning techniques that are much better than humans when it comes to perception, prediction and spontaneous planning, where fatal driving accidents are projected to drop dramatically making deaths from driver error a thing of the past.	Modern warfare using unmanned underwater vehicles, constellations of smart robotic devices, drones, robots and precision-guided missiles. Modern warfare supremacy will be decided by technology using Robots, which are quicker, stronger, more accurate and follow orders much better than humans, making soldiers obsolete.

Bright	Dark
Using an implantable for a brain-machine interface that is likely to lead to an immense increase in human intelligence, permitting humans to explain medical conditions including paralysis, blindness, anxiety and addiction. An implantable may also be applied to reprogram core human biology to allow humans to strengthen and upgrade themselves.	Considerable job losses. It is estimated that a billion people will lose their jobs due to AI by 2030. Additionally, these widespread job losses could rival the move away from agricultural labour during the 1900s in the United States and Europe. It is also envisaged this technological trend will likely increase inequality and place power in the hands of a few.
Virtual assistants that apply natural language processing and voice technology to comprehend and perform human directed tasks, particularly when linked to robots.	Socioeconomic inequality will likely increase since millions of low-skills to medium-skill jobs will be forfeited resulting in a huge income gap between middle-skill and high-skill labour.

Source: Vashev, N. (2021). AI and the Future of Humans. Forbes Technology Council.

[Vashev \(2021\)](#) argues that with the continuous increasing pace of technology, it is essential that everyone is made aware of the forthcoming disruptions and not just unquestionably accept the bright side of AI future without also considering the negative aspects. Finally, he contends that computer superintelligence at worst could threaten human's very existence and at best it offers great challenges that need to be settled if society can sensibly give rise to the Fourth Industrial Revolution. [Freedland \(2023\)](#) maintains that there are things governments can do, such as imposing restrictions on how much computing power the tech companies are allowed to use to train AI, and how much data they can feed it with the aim of constraining the bounds of its knowledge. These ethical issues have been considered in detail in a previous chapter, but it is essential that ethical and moral issues are not overlooked.

13.3 HUMAN AND AI COLLABORATION

The rapid progress of AI technology these past few years has generated general anxiety regarding the future employment prospects of humans. The concept of having AI to displace humans at work is already a topic of public debate with AI proponents saying that AI will introduce new opportunities, while opponents view AI as an impending disaster for humans.

João da Silva and Natalie Sherman, Business reporters for BBC News cite Tesla boss Elon Musk as saying that the electric car maker will start producing, selling, and using humanoid robots from 2026 ([da Silva and Sherman, 2024](#)). It is known that Tesla has experienced weakening demand for its cars, and the use of the humanoids is aimed at cutting costs. Musk states that generally their focus remains on company-wide cost reduction and that the autonomous humanoid robot (Optimus) will be produced to conduct unsafe, repetitive or boring tasks ([da Silva and Sherman, 2024](#)). Furthermore, [da Silva and Sherman \(2024\)](#) also report that other firms, including Honda and Boston Dynamics, have also been developing their own humanoid robots.

There is no doubt that AI is proving to be beneficial for various industries and that intelligent algorithms have increased operational efficiency and optimised organisational resources. Opponents to AI view this as a reflection of the job losses taking place, while proponents see this as increasing an organisation's profitability and its capacity to introduce new products and services that still require human intervention.

There is also no doubt that technology has always had the effect of reducing the functions that were normally performed by humans (and animals being replaced by tractors and vehicles), but it is also a reality that technology has generated many new functions and created new fields. For example, in the car industry, robots have replaced human workers in performing repetitive and sometimes hazardous tasks. But the converse is also true in that new fields regarding engineering and computer science have been created, including the many jobs associated with them.

Furthermore, the lowering of operational costs has meant lower prices for the goods and services, which has made them more affordable for humans and their usability in most cases brought more leisure time and enjoyment.

In the 1800s, many Americans worked 70 hours or more per week and this continued decreasing to 40 hours per week at around 1940. According to the ILO, in 2024 most workers in developed countries worked less than 40 hours per week. According to the European Commission, the average for the European Union countries was 36.1 hours per week. These figures illustrate that most humans have experienced more leisure time. There is consensus that technology has played a key role in the reduction in the working hours per week. Many believe that further reduction in working hours is likely as technology progresses. The concern is whether this reduction will continue to a state where most humans are no longer required. This brings us back to the “universal basic income, or something like that” concept predicted by Elon Musk as reported on CNBC ([Clifford, 2018](#)).

A key issue is whether AI should be embraced as a powerful potential collaborative partner for humans rather than a threat to human workers. Like most social issues, it is not just black or white but there are many shades of grey in between. While it is fact that AI can perform various tasks more quickly and efficiently than humans, there are many characteristics that humans excel at and currently cannot be replaced by AI. A key word in the previous sentence is “currently.” Therefore, when considering such issues it is important to anticipate the future not just the present. It is also a fact that many things that were not possible in the past are achievable today because of technological progress.

Currently, creativity, complex problem solving, and empathy is primarily the domain of humans. However, generative AI is already encroaching on the creativity aspect. The future potential of General AI and Super AI is unfathomable. What may be very difficult for machines to replicate today could be very possible tomorrow. However, if we adopt a collaborative mindset when considering the potential application of AI, our anxiety about its potential may be mitigated. To be successful in the implementation of AI

applications, a high level of technical skills, such as programming, software engineering, and data science are critical. These skills are currently all human skills and with the right ethical guidelines supported by legislation we would ensure that they remain so. Additionally, education institutions need to keep ahead of progress and update their curriculum to ensure specialisation in AI-based application system development and related skills. This would allow progress to continue based on human capacity and not create an environment where we become dependent on the capacity of AI to dictate our progress. In other words, we need to avoid a situation where the rate of progress depends on the ingenuity of AI instead of human ingenuity. AI needs to remain under the control of humans and not be allowed to roam free to regenerate itself and generate other AIs.

According to [Wroe \(2023\)](#), thinking and learning about AI are the mental equivalent of a fission chain reaction where the questions get really big, really quickly. He further argues that there is the current tendency to concern ourselves with the issues that are of an immediate or short-term nature, such as the opportunities for economic productivity, health care, manufacturing, education, solving global challenges like climate change and, on the reverse side, the risks of mass unemployment, disinformation, killer robots, and concentrations of economic and strategic power. Although in his view these matters are important, there are more profound issues to be considered, such as the human capacity to live meaningful, fulfilling lives in a world in which we no longer have intelligence supremacy. He further argues that as long as humanity has existed, it has had an effective monopoly on intelligence in its sphere of influence.

It has taken modern humans about 200,000 years to evolve from the now extinct *Homo erectus*, whereas computer technology through human intelligence has taken less than a hundred years to evolve to the state it is now. The rate of such development is without precedent. [Wroe \(2023\)](#) argues that humans over the centuries have built a vast intellectual empire made up of science, philosophy, theology, engineering, storytelling, art, technology, and culture. He contends that it is human nature to explore, discover, and expand its intellectual capacity.

Yet having General AI and Super AI as the long-term objective would mean that humans have created intelligence which is superior to their own. Humans would no longer be the smartest entities in our universe. Humans, particularly decision makers, need to see the wood for the trees. Currently, we are getting very involved in the short-term details but are ignoring what is important about the thing as a whole. The whole issue about the impact of future AI is not just based on economics but it becomes an issue involving human development and its social implications. [Wroe \(2023\)](#) argues that even if world leaders can solve the economic equality issues by introducing a universal basic income as Elon Musk suggested, and replace notions of “paid work” with “meaningful activity”, how will humans spend their lives in ways that they find meaningful, given that the basis of human evolution is founded on striving, thriving, and competing?

On the other hand, General AI and Super AI can help humans to strive (determination), thrive (flourish), and compete (rivalry) and to achieve things that until now have not been within our knowledge, if this technology is tightly controlled and not allowed to roam freely. [Wroe \(2023\)](#) argues that if General AI and Super AI could answer all of our most profound questions and solve all of our problems, how much satisfaction do we get from having the solutions handed to us? However, this argument assumes that knowledge is finite. But indications are that knowledge is infinite and therefore our satisfaction will be attained from gaining more knowledge and understanding, which is not currently known. It is a fact that humans have different talents and a diverse level of creativity.

Not everyone enjoys writing or reading poetry or be a literary genius or be a painter or musician or performer. Those that enjoy such activities will continue to practice them within their own creative talents. Whilst those that do not have such talents can seek collaboration from AI with their own conscious knowledge that the work being produced is not entirely their creation. The medical field is still wide open to new discoveries, whilst exploration of our oceans is only at a basic level. Basically, the issue is about maintaining human control over all the types of AI. [Wroe \(2023\)](#) cites AI pioneer Stuart Russell who compared this situation to the King Midas

story where, turning everything into gold by touching it sounds fantastic, until you try to eat an apple or hug your kids.

[Anderson and Rainie \(2023\)](#) from the Pew Research Centre maintain that experts are divided about how much control people will retain over essential decision-making as digital systems and AI spread. There appears to be consensus that government authorities and large companies are likely to intensify their efforts to deploy AI systems with the objective of improving people's living standards. However, there are also grave concerns that the outcome of these AI applications will reduce the individual's capacity to manage their options. This concern is driven by the prediction that smart machines, bots, and systems powered mostly by machine learning and AI will rapidly increase in speed and sophistication between now and 2035 ([Anderson and Rainie, 2023](#)).

Their concern is based on the scenario where individuals become more addicted to these technologies in their daily lives, and they are also steadily being lured to relinquish their decision-making and personal autonomy to digital tools. The added concern is that they may be doing this unconsciously or instinctively and thus losing the capacity to exercise judgment and make decisions independent of these systems. Research has already found that many people are being addicted to social media apps like TikTok, Facebook, and Snapchat, mobile sports betting apps, video watching websites like YouTube, and mobile games that are keeping smartphone users hooked to their device more and more. However, AI applications are different and normally delve into the decision-making realm with the consequence that people will be heavily influenced by AI in reaching their decisions.

There is consensus that AI technology will provide great benefits to humans but like everything else it requires careful monitoring to ensure that no intentional or unintentional abuse takes place. Research by Pew Research Centre and Elon University found that 56% of the experts questioned agreed that by 2035 smart machines, bots, and systems will not be designed to allow humans to easily be in control of most tech-aided decision-making ([Anderson and Rainie, 2023](#)).

13.4 FUTURE HUMAN DEVELOPMENT

According to [Jones \(2024\)](#), to understand the future evolution of humans, we need to look to our past. Jones poses a very valid question regarding our descendants, namely: Will our descendants be cyborgs with hi-tech machine implants, re-growable limbs and cameras for eyes or might humans morph into a hybrid species of biological and artificial beings? Furthermore, she questions whether humans can become smaller or taller, thinner or fatter, or even with different facial features and skin colour?

Historical evidence reveals that it is only during the last 10,000 years that there have been major changes in which humans had to adapt to. Jones argues that agricultural living and plentiful food have led to human health problems, which have been addressed with the aid of science, such as treating diabetes with insulin. Additionally, humans' physical features have also changed in that generally humans have become overweight and, in some zones, have developed taller. [Jones \(2024\)](#) cites Thomas Mailund, Senior Lecturer in Bioinformatics and Big Data at Anglia Ruskin University, London who suggested that humans may evolve in ways that would help us to adapt to the developing environment and specifically the changes taking place around us. For instance, Mailund suggests that as urban populations increase our daily social interactions would also increase substantially, where humans may develop in ways that facilitate these interactions by for example remembering people's names, which would be a vital skill in such an environment.

However, technology can help in this human evolving environment. For example, Mailund suggests that an implant in the brain would permit humans to recall people's names, since scientists already know what genes are involved in building a brain that's good at remembering people's names ([Jones, 2024](#)). Admittedly, this type of technology is still at the early experimental stage because while the implant may exist, the input/output interface still needs to be resolved. Mailund argues that it is not a biological issue but technological one. For instance, humans are already having implants to repair various biological appendages, such as a pacemaker or

hip or knee implant. Mailund maintains that there may be the possibility in the future for implants to be applied to enhance a person's capability, such as an artificial eye with a camera that can read different frequencies of colour and visuals ([Jones, 2024](#)).

[Jones \(2024\)](#) cites Dr. Hodgson, Lecturer, Grand Challenges in Ecosystems and the Environment who maintains that using bioinformatics it is possible to accurately predict how genetic variation will change. For instance, he foresees that urban and rural areas will become increasingly genetically diverse due to the migration of people from rural areas into cities, with an increase in genetic diversity in cities and a decrease in rural areas being the outcome. AI algorithms that predict this genetic diversity would provide an extremely accurate depiction of how humans are evolving across countries.

[Pester \(2021\)](#) claims that futurists predict two significant milestones, namely singularity, where AI will progressively become smarter until it is immeasurably superior to our own intelligence; and virtual immortality that will enable the scanning of one's human brain and transfer its contents to a non-biological medium, like a computer or robot. [Pester \(2021\)](#) argues that once the human mind is deposited in a computer, which in turn can be uploaded to the internet, an individual does not need to be apprehensive about the human body perishing. However, according to [Schneider \(2019\)](#) this does not mean achieving physical immortality. She maintains that a better approach to achieve extreme longevity (i.e., the ability to live a long life beyond the human's average age at death), while also preserving the person, would be through biological enhancements compatible with the survival of the human brain or through brain chips. [Schneider \(2019\)](#) views this as a future development where a human would gradually have replacement parts using brain chips, which would eventually result in having a human driven by AI. Thus, a human over time would progressively transition into a cyborg and thinking in chips rather than neurons. However, she maintains that if the human brain is intimately connected to you, then replacing it could mean suicide. She also believes

that the human body appears to have an expiration date, regardless of how it is upgraded or uploaded.

The technology regarding brain implants has already become a reality but still needs perfection. [Chow \(2024\)](#) from NBC News reported that the first brain-computer interface was implanted into a human in the late 1990s, where these devices were able to tap into the brain circuitry that remains intact after injury to perform basic movements and functions. Moreover, [Limehouse \(2024\)](#) for USA Today reported that a Quadriplegic patient with a dislocated spine had a Neuralink implanted that has helped his life, including allowing him to play video games and chess and surf the Internet with ease. Despite some reservations [Schneider \(2019\)](#) views the potential brain and body enhancements of the future as exciting since it allows humans to rid itself of death by old age and permits technology to cure ailments and make humans smarter.

[Diaz \(2023\)](#) a journalist with New York Post cited futurist and author Ray Kurzweil as predicting that humans will achieve immortality in just seven years. Kurzweil has previously suggested that humans will have created technology to fend off aging and illness with microscopic robots, sent to repair our bodies on a cellular level ([Diaz, 2023](#)). According to [Belachew \(2023\)](#) nanotechnology can repair damaged tissue at the cellular level and is assisting in detecting, diagnosing, and treating severe diseases such as HIV, cancer, and others. However, this technology is still at an embryonic stage. Combined with AI this technology has the potential of making humans immortal.

13.5 CONCLUSION

This chapter has examined three key themes that include AI and the future of humans; human and AI collaboration; and what will humans are likely to look like in the long-term future. The greatest fear is the displacement of humans from the job market for routine and repetitive tasks resulting in massive unemployment and an economic catastrophe. Furthermore, there

are also grave fears that certain types of AI could cause the possible extinction of humans altogether.

The research has revealed that AI is likely to result in many jobs being taken over by AI and related technology. On the other hand, there are also predictions that AI will provide great benefits in healthcare, transport, education, manufacturing, amongst many others. Moreover, it is predicted based on past technological evolutionarily experiences that AI and its associated technology will create countless opportunities in all fields. Research has also revealed that a collaborative approach between humans and AI technology will ensure that humans reap the benefits with little adverse consequences. However, this scenario becomes realistic if governments introduce ethical standards supported by legislative measures to protect the human job market. Additionally, research has also revealed that the evolution of technology, including AI has meant an increase in leisure time for humans resulting in a lower number of working hours per week. However, lowering the working hours was not a voluntary act but through labour trade union demands and government intervention.

Research has also shown that General AI and Super AI could be dangerous, particularly if these AI systems are allowed to operate in an uncontrolled environment. Such systems should be prohibited through international treaties to prevent malicious behaviour, such as cyberterrorism and launching weapons, particularly nuclear warheads. The general conclusion is that AI requires constant monitoring both within nations and internationally but this needs to be supported by extensive legislation and sincere international collaboration.

The research has revealed that whilst General AI and Super AI are feared as a contributor to the extinction of humans, yet it has also revealed that AI associated with other technological developments have to the potential to make humans quasi-immortal. Hence, there appears to be a contradiction, on one hand the extinction of humanity and on the other hand its immortality.

I would like to finish this chapter on the same theme that I choose in the first chapter of this book. The Garden of Eden as described in the book of

Genesis, depicts an earthly paradise inhabited by the first created man and woman, where humans and animals lived in peace and harmony, living off plant life. God tasked Adam with looking after the garden on the condition that he and Eve were not to eat the fruit from the tree of the knowledge of good and evil. It was utopia. The serpent convinced Eve that eating from the forbidden fruit will bring immortality and that similar to God they (Adam and Eve) will know things that only God knows. The outcome was a human calamity, resulting in the creation of pain and hardship. The moral of the story is that we humans should not play God. Who wants to live for ever?

“Who Wants to Live Forever,” the theme song for the Highlander film stresses the importance of living for today. While eternal life might seem appealing, it loses its significance if it not meaningful and fulfilling, and most important devoid of feeling. We should not live under the shadow of technology but as humans we must maintain our intellectual superiority by ensuring that technology remains a tool to be exploited for the good of humanity.

REFERENCES

- Anderson, Janna, Quitney., and Rainie, Lee. (2023). *The Future of Human Agency*. Pew Research Centre. www.pewresearch.org/internet/2023/02/24/the-future-of-human-agency.↵
- Belachew, Girum Tefera. (2023). Disease fighting machines inside the body: A review. *Biomedicine Journal of Science & Technology Research*, Vol. 49, No. 2, pp. MS.ID.007774. <https://biomedres.us/pdfs/BJSTR.MS.ID.007774.pdf>.↵
- Chow, Denise. (2024). *Behind Elon Musk's Brain Chip: Decades of Research and Lofty Ambitions to Meld Minds with Computers*. NBC News. www.nbcnews.com/science/science-news/neuralink-elon-musk-science-behind-rcna136352.↵

- Clifford, Catherine. (2018). *Elon Musk: Robots Will Take Your Jobs; Government Will Have to Pay Your Wage*. CNBC Brand Studio.🔗
- Conversion Capital. (2024). *Labor Unions and Automation*.
<https://conversioncapital.com/labor-unions-and-automation/>.🔗
- da Silva, João. and Sherman, Natalie. (2024). *Musk Says Tesla to Use Humanoid Robots Next Year*. BBC News.
www.bbc.com/news/articles/cz5reve8476o.🔗
- Diaz, Adriana. (2023). Immortality is attainable by 2030: Google scientist. *New York Post*. <https://nypost.com/2023/03/29/immortality-is-attainable-by-2030-google-scientist/>.🔗
- Freedland, Jonathan. (2023). The future of AI is chilling – humans have to act together to overcome this threat to civilisation. *The Guardian*.
www.theguardian.com/commentisfree/2023/may/26/future-ai-chilling-humans-threat-civilisation.🔗
- Jones, Lucy. (2024). *Evolution: What Will Humans Look Like in a Million Years?* BBC Earth. www.bbcearth.com/news/what-will-humans-look-like-in-a-million-years.🔗
- Kurzweil, Ray. (2005). *The Singularity Is Near*. Viking.
- Kurzweil, Ray. (2024). *The Singularity Is Nearer: When We Merge with AI*. Kindle.
- Limehouse, Jonathan. (2024). *Second Person with Spinal Cord Injury Gets Neuralink Brain Chip and It's Working, Musk Says*. USA TODAY.
<https://eu.usatoday.com/story/tech/news/2024/08/05/elon-musk-brain-chip-implant-update/74674041007/#:~:text=A%20second%20human%20has%20successfully,%2C%20'It's%20working%20very%20well>.🔗
- Pester, Patrick. (2021). *Will Humans Ever Be Immortal?* Live Science.
www.livescience.com/could-humans-be-immortal.🔗
- Schneider, Susan. (2019). *Artificial You: AI and the Future of Your Mind*. Princeton University Press.🔗
- Vassev, Nikolai. (2021). *AI and the Future of Humans*. Forbes Technology Council.

www.forbes.com/councils/forbestechcouncil/2021/05/06/artificial-intelligence-and-the-future-of-humans/.[↵]

Wroe, David. (2023). *Artificial Intelligence and the Future of Humanity*.
ASPI: The Australian Strategic Policy Institute.
www.aspistrategist.org.au/artificial-intelligence-and-the-future-of-humanity/.[↵]

Index

A

ABC, [50–51](#)
Active-HDL ,108, [113](#)
ADDIE, [238](#), [240](#)
AGI, [61](#), [77](#), [168](#)
AI approaches, [249](#)
AI Computer Programming Languages, [118](#)
AIDA, [294](#), [297–298](#)
AI-driven, [175](#), [197](#), [237–241](#), [243–245](#), [254](#), [256](#), [264](#), [268](#), [316](#)
AI ethics, [54](#), [278–283](#), [304](#), [313–314](#), [324](#), [328](#)
AI landscape, [252–253](#), [255–256](#), [345](#)
AI safety, [15](#), [86](#), [173–175](#), [189–191](#), [193–195](#), [198–199](#)
AI software, [91](#), [96](#), [114](#), [118](#), [150](#), [171](#), [226](#), [237](#), [277](#), [280](#), [344](#)
AI use cases, [202](#), [233](#), [235–236](#), [246](#), [251](#), [258–260](#), [262](#)
AI winter, [54](#), [57](#), [59](#), [63](#), [88–90](#)
Alan turing, [9](#), [53–54](#), [162](#), [328](#)
Allegro network, [121](#)
ALPAC report, [56](#)
ANPR, [157](#)
Application specific, [86](#), [103–105](#), [111](#), [183](#), [198](#)
AR, [4](#), [157](#), [271](#)
ASIC, [103–106](#), [111](#)

ASR, [151–152](#)
Atanasoff Berry Computer, [50–51](#), [63](#)
Augmented reality, [4](#), [157](#), [271](#)
Automated economy, [332](#), [346–348](#), [352–353](#)
Automated reasoning, [5](#), [130–131](#)
Automatic number plate recognition, [157](#)
Automatic speech recognition, [151](#)
Automation based economy, [17](#), [203](#), [205](#), [221](#)

B

Backpropagation, [54](#), [57](#), [59](#), [63](#), [90](#)
Belmont report, [281](#), [328](#), [352](#)
BERT, [159](#)
Bidirectional encoder, [159](#)
Bill Gates, [9](#), [12](#), [18](#), [90](#)
Bodily-kinaesthetic intelligence, [26](#), [70](#)
Brown University, [13](#)

C

C++, [108](#), [113](#), [122](#)
CA, [185–187](#)
CAD, [125–126](#)
CAIS, [189](#), [191](#), [193](#), [198–199](#)
Centre for AI safety, [15](#), [173](#), [189](#), [191](#), [198](#)
Certificate authority, [185–186](#)
Certificate lifecycle, [185–187](#)
Certificate template, [186](#)
ChatGPT, [13–14](#), [159–160](#), [280](#), [286](#), [331](#), [335](#)
CISA, [184](#), [198](#)
CLB, [106](#)

Cognitive development, [27](#)
Colossus, [2](#), [50–52](#), [63](#), [88](#)
Common objects, [156](#)
Complex programmable logic device, [105](#), [108](#), [111](#), [113](#)
Computer-aided design, [118](#), [121](#), [125–126](#), [171](#)
Condor, [95](#), [111](#)
Configurable logic blocks, [106](#)
Conversational AI, [150](#), [160](#)
Convolutional neural network, [99](#), [154](#)
Core technologies, [332](#), [343](#)
CPLD, [105](#), [108–114](#)
CPU, [88](#), [99](#), [100](#), [102](#), [110](#), [145](#), [345](#)
Crack Public Key Infrastructure, [62](#), [64](#), [173](#), [184](#), [198](#)
Critical success factor, [240](#), [267](#)
Cross-validation, [124](#)
CrowdStrike security, [220](#)
CSF, [267–268](#)
CUE-bits, [97](#)
Cultural intelligence, [32–34](#), [40](#)
Cutting-edge machine learning, [343](#)
Cyber security, [86](#), [158](#), [163](#), [173–179](#), [184](#), [197–199](#), [204](#), [225](#), [227](#), [233–236](#), [251](#), [261](#), [283](#), [291](#), [296](#), [302](#), [314](#), [319](#), [321](#), [332](#), [352](#)
Cyber threats, [175](#), [184](#), [229](#)
Cybernetics, [52](#), [78](#), [88](#)

D

DARPA, [55](#), [89](#)
Dartmouth, [1](#), [42](#), [54–56](#), [63](#), [65](#), [78](#), [88](#)
Darwin, [19](#), [21–22](#), [36](#)
Data management, [224](#), [233](#), [247–250](#), [254](#), [301](#), [307–308](#)
Data privacy compliance, [239](#)

DDOS, [177](#), [180](#), [182](#)
DEC, [58–59](#)
Deep blue, [60](#), [68](#), [76](#)
Deep learning, [1](#), [57–60](#), [102–103](#), [145](#), [153–156](#), [161](#), [163](#), [165](#), [170](#), [214](#),
[226–227](#), [235–236](#), [249](#), [252](#), [263](#), [332](#), [343–345](#), [353](#), [358](#)
DeepMind, [11](#)
Deep reinforcement learning, [192](#)
Deffie Hellman algorithm, [188](#)
DGA, [261](#)
Digital certificate, [184–186](#)
Digital equipment corporation, [58](#)
Disruptive technologies, [246](#)
Distributed denial-of-service, [177](#)
Domain-generated algorithms, [261](#)
DRL, [192–193](#)

E

eBPF, [178](#)
Edge computing chips, [102–104](#), [111](#), [113](#)
EDVAC, [53](#)
EEPROM, [109](#)
Electrical and electronic engineering, [51](#)
ELIZA, [56](#)
Elon Musk, [11](#), [275](#), [354](#), [358–360](#), [365](#)
Emotional behaviours, [5](#)
ENAIIS, [193–195](#), [198](#)
ENIAC, [2](#), [50–53](#), [63](#)
Enterprise AI, [201–202](#), [217](#), [223–248](#), [251](#), [254](#), [255](#), [257](#), [263](#), [272–273](#)
Ethical and moral, [25](#), [202](#), [275–276](#), [279](#), [284](#), [341](#), [352–353](#), [358](#)
Ethical principles, [283](#), [310](#), [327](#)

EU, [193–195](#), [202](#), [253](#), [281](#), [283](#), [285–289](#), [292](#), [306–308](#), [311–315](#), [327–329](#), [341](#), [365](#)
European network for AI safety, [193–194](#)
European Union, [193](#), [208](#), [250](#), [281](#), [285](#), [300](#), [302](#), [312](#), [359](#)
Evolution of human beings, [22](#), [24](#)
Evolutionary algorithms, [344](#)
Evolving threat landscape, [65](#), [86](#), [173](#), [175](#), [197–199](#)
Expert system, [58](#), [132](#), [166–167](#), [172](#)
External environment, [73](#), [125–126](#), [210](#), [238](#)
Extinction, [11–12](#), [15](#), [18](#), [191](#), [355](#), [363–364](#)

F

Facial recognition, [5](#), [81](#), [156](#), [192](#), [235](#), [272](#), [280](#), [286](#), [311](#)
Feature pyramid network, [154](#)
Field-programmable gate array, [105](#)
Firewall, [86](#), [174](#), [179–183](#), [189](#), [198–199](#), [231](#)
First international AI safety report, [195](#), [199](#)
5-Whys, [238](#), [242](#)
FPGA, [105–109](#), [111](#), [113–114](#), [146–147](#), [150](#)
FPN, [154](#)
Frames of mind, [70](#)
Framework for Deploying an Enterprise AI System, [231–232](#), [263](#)
Fraudsters, [62](#), [189](#)
FSIC, [105](#)
Functionality-specific integrated circuit, [105](#)

G

Game theory, [36](#), [41](#), [52](#), [88](#)
GAN, [157](#)
Gardner, [28](#), [70](#)

Garry Kasparov, [60](#), [68](#), [76](#)
GDPR, [239](#), [301](#)
GELAN, [155](#)
General AI, [43–45](#), [64](#), [91](#), [123](#), [190](#), [215–216](#), [225](#), [326](#), [331–332](#), [337–340](#),
[359–361](#), [364](#)
General intelligence, [5](#), [27–29](#), [34](#), [40](#), [61](#), [77](#), [168](#), [346](#), [352–353](#), [357](#)
General problem solver, [56](#)
Generative AI, [86](#), [150](#), [157](#), [172](#)
GIM, [108](#)
Global interconnection matrix, [108](#)
GMOS, [115](#)
Google, [11](#), [14](#), [105–106](#), [150–151](#), [157–161](#), [239](#), [280](#), [291](#), [316](#), [341](#), [357](#),
[365](#)
GPS, [56](#), [62](#), [64](#), [271](#)
GPU, [191](#), [332](#), [345](#)
Graphical design capture language, [108](#), [113](#)
Graphics processing unit, [191](#)
Group selection theory, [38](#)

H

Hardware description language, [108](#), [113](#), [146](#), [172](#)
Harmonisation, [81](#), [324](#)
Harvard University, [55](#), [65](#), [70](#), [277](#)
HDL, [108](#), [110](#), [113–114](#), [146–149](#), [172](#)
Hierarchical model, [27](#)
HIPS, [180](#)
Homo Sapiens, [20–22](#), [24](#), [35](#), [336](#)
Host-based IDPS, [180](#)
Human intelligence, [1–4](#), [17–46](#), [53](#), [56](#), [61](#), [64](#), [66–72](#), [76–83](#), [86](#), [91](#), [118](#),
[150](#), [172](#), [201](#), [215–216](#), [220](#), [223](#), [246](#), [279](#), [309](#), [332](#), [336–340](#), [352–360](#)

I

IBM Qiskit, [144](#)
IBM quantum platform, [86](#), [144](#), [171](#)
IC, [89–90](#), [103](#), [111](#), [149](#)
ICAD, [118](#), [125–126](#), [171](#)
ICT, [206–209](#), [224–225](#), [230](#), [256](#), [263](#), [300–302](#)
IDPS, [174](#), [178–181](#)
IDS, [179](#), [181–182](#), [199](#), [224](#), [353](#)
IEEE, [51](#), [57](#), [146–149](#), [172](#)
Image recognition, [44](#), [54](#), [60–61](#), [64](#), [86](#), [91](#), [99](#), [118](#), [150](#), [153–156](#), [172](#),
[228](#), [337](#), [343](#)
Implementing enterprise AI, [224](#), [231](#), [255](#)
Implementing public sector AI, [255](#), [258](#)
Implications of AI, [17](#), [216](#), [264](#), [326](#), [331–332](#), [336](#), [348](#), [352](#)
Integrated circuits, [86](#), [89](#), [102–104](#), [109–111](#), [146](#)
Internal environment, [73](#), [126](#), [210](#), [226](#), [238](#)
Internet of Things, [102](#), [113](#), [170](#), [185](#), [224](#), [237](#), [272](#), [338](#)
Internet protocol suite, [180](#)
Interpersonal intelligence, [26](#), [70](#)
Intrusion detection, [174](#), [178–179](#), [181](#), [260–261](#)
Intrusion prevention, [17](#), [65](#), [86](#), [173–174](#), [178–179](#), [182](#), [198–199](#), [227](#), [231](#)
Inventory management, [208](#), [219](#), [241–245](#), [269](#), [338](#)
Inventory optimisation, [227](#), [242](#), [244](#)
IoT, [100–103](#), [113–114](#), [170](#), [185](#), [237](#), [272](#), [338–339](#)
IPS, [17](#), [65](#), [86](#), [174](#), [178–179](#), [181–183](#), [189](#), [198–199](#), [231](#)
Isaac Asimov, [4](#), [54–55](#)
Ishikawa diagram, [242](#), [265–266](#)

J

Jaron Lanier, [11](#)
Java HDL, [108](#), [113](#)
JHDL, [108](#), [113](#)
Johns Hopkins University, [21](#), [83](#)

K

Kano Model, [238–239](#)
KBE, [125](#)
Key performance indicators, [192](#), [226](#), [233](#), [238](#), [245](#), [267–268](#), [273](#)
Kirkpatrick model, [240](#)
KL-One, [118](#), [127–131](#), [171](#)
Knowledge based economy I, [17](#), [40](#), [201](#), [203–221](#), [332](#), [346](#), [352](#)
KPI, [267](#)

L

LaMDA, [160](#)
Language model, [160](#), [344](#)
Lean six sigma, [238](#)
Legislative global landscape, [284](#)
Lighthill, [89](#)
Limited theory, [44](#), [215](#)
Linguistic intelligence, [26](#), [70](#)
LISP, [54](#), [56](#), [58](#), [63](#), [87](#), [90](#), [115–121](#), [132](#), [171](#)
LLM, [191](#)
Logical-mathematical intelligence, [26](#), [70](#)
LSI, [90](#), [109](#)

M

Machine learning, [7](#), [43–46](#), [49](#), [54–60](#), [63–64](#), [79–80](#), [83](#), [90](#), [98](#), [100](#), [102](#), [104](#), [108](#), [113](#), [118–119](#), [122–124](#), [130](#), [144–145](#), [153–158](#), [161–163](#), [166](#), [169](#), [175–178](#), [180](#), [191](#), [214–215](#), [223–228](#), [231](#), [235–236](#), [241](#), [249–252](#), [259–262](#), [272](#), [277](#), [279](#), [281](#), [290](#), [297–298](#), [306](#), [317](#), [319](#), [326–327](#), [341–345](#), [349](#), [351](#), [361](#)

Macrocells, [108](#), [110](#)

Malicious software, [189](#)

Malware, [86](#), [177](#), [181–183](#), [189](#), [198](#), [261](#)

Mask RCNN, [154](#)

Mercury, [118](#), [139–140](#), [144](#), [171](#)

Mesolithic, [20](#), [23](#)

Meta-data, [210](#)

Microprocessors, [90](#)

Microsoft, [9](#), [11–12](#), [15](#), [90](#), [117](#), [150–151](#), [160](#), [291](#), [316](#), [341](#)

Minicomputer, [89](#)

Misinformation, [12](#), [15](#), [176](#), [304](#), [316–317](#)

MITB, [189](#)

MITM, [188–189](#)

Mobile App Development, [237](#), [239](#)

Moral values, [6](#), [55](#), [278](#), [341](#)

Morphology, [98](#), [111](#), [162](#)

Multiple Intelligence, [28](#)

Musical intelligence, [26](#), [70](#)

N

Named entity recognition, [164](#)

Narrow AI, [43–44](#), [61](#), [64](#), [91](#), [290](#), [331](#), [337–338](#), [351](#)

National Academy of Sciences, [56](#)

National Institute of Standards and Technology, [184](#), [188](#), [281](#), [290](#)

National Security Agency, [184](#)

Natural language, [5](#), [44](#), [54](#), [56](#), [58](#), [60–61](#), [64](#), [80](#), [82](#), [86](#), [91](#), [94](#), [99](#), [104](#), [118](#), [124](#), [130–131](#), [140](#), [150](#), [152](#), [162–165](#), [170–172](#), [176](#), [184](#), [220](#), [225](#), [234–235](#), [261](#), [337–339](#), [344–345](#), [349](#), [358](#)

Naturalistic intelligence, [26](#), [70](#)

NBA, [179–180](#)

Neolithic, [23](#)

NER, [65](#), [164](#)

Net promoter score, [239](#)

Network behaviour analysis, [179–180](#)

Network-based IDPS, [179](#)

Neural Lyapunov barrier, [192–193](#)

Neural network, [50](#), [52](#), [54–63](#), [87–88](#), [99](#), [104](#), [145](#), [154–158](#), [167–169](#), [193](#), [269](#), [343](#), [345](#), [352](#)

Neural processing unit, [102](#)

Neuromorphic, [85–87](#), [98–104](#), [111](#), [114](#), [167–172](#), [215](#), [351](#)

Next-generation firewall, [183](#), [198–199](#)

NFS, [121](#)

NGFW, [86](#), [183](#), [199](#)

NIPS, [179–180](#)

NIST, [184](#), [198](#), [281](#)

NLB, [193](#)

NLG, [165](#)

NLP, [44](#), [56](#), [94](#), [118](#), [150](#), [152](#), [162–165](#), [176](#), [220](#), [225–228](#), [234–238](#), [344–345](#)

Non-operational data, [210](#)

Non-recurring engineering, [105](#)

Normal distribution, [25](#), [31](#), [67](#)

NPS, [239](#), [240](#)

NPU, [102](#)

NRE, [105–106](#)

NSA, [184](#), [198](#)

O

Object-oriented programming, [122](#), [125](#)
OCR, [152](#)
OECD, [248](#), [250](#), [253](#), [273](#), [280](#), [302](#), [318](#), [321](#), [323](#)
OI, [332](#), [350–353](#), [357](#)
OOP, [122](#)
OpenAI, [11](#), [15](#), [160](#), [291](#), [316](#)
Operating system, [89](#), [115](#), [117](#), [122](#), [167](#), [178](#), [180](#), [182](#), [224](#)
Operational, [5](#), [7](#), [73](#), [178](#), [202](#), [210](#), [227](#), [230–231](#), [235–240](#), [242–243](#), [254](#),
[269](#), [272](#), [275](#), [283](#), [348](#), [359](#)
OPS5, [118](#), [131–134](#), [171–172](#)
Optical character recognition, [152](#), [157](#)
Optical fibre, [94](#)
Organoid Intelligence, [332](#), [350](#), [353](#), [357](#)
Orson Welles, [4](#)
Osprey, [95](#)

P

Palaeolithic period, [23](#)
Parallel processing, [50](#), [52](#), [59](#), [85–86](#), [91](#), [94](#), [98](#), [104](#), [107](#), [111](#), [167](#), [172](#),
[343](#)
PASS model, [28](#)
Perceptron, [54](#), [56–57](#), [63](#), [65](#), [88](#)
Petri Net, [133–135](#), [172](#)
PGI, [155](#)
PKI, [62](#), [86](#), [173–174](#), [184–185](#), [188–189](#), [198–199](#)
Predictive analytics, [225](#), [235](#), [259](#), [264](#), [338](#)
Primary mental abilities, [27](#)
Private sector, [147](#), [209](#), [225](#), [247](#), [249](#), [254](#), [256–257](#), [259](#), [262–263](#), [278](#),
[294–296](#), [300](#), [312](#), [317–318](#), [321](#), [324–325](#), [327](#), [345](#), [349](#)

Programmable gradient information, [155](#)

Prolog, [139–144](#), [171](#)

Public key infrastructure, [62](#), [64](#), [86](#), [173](#), [184](#), [198–199](#)

Public sector, [150](#), [201–202](#), [225](#), [246–265](#), [267](#), [269](#), [271–274](#), [283](#), [303](#), [326](#), [328–329](#)

Python, [98](#), [113](#), [116–117](#), [121–125](#), [144](#), [147](#), [158](#), [165](#), [171](#)

Q

Qiskit, [86](#), [98](#), [113](#), [144–145](#), [171](#)

Quantum, [2](#), [8](#), [17](#), [52](#), [54](#), [62–64](#), [66](#), [85](#), [88](#), [95–98](#), [111](#), [113](#), [119](#), [144–145](#), [171–172](#), [215](#), [332](#), [340–341](#), [343](#), [352](#), [357](#)

Quantum computer, [2](#), [54](#), [62–64](#), [95–98](#), [111](#), [113](#), [343](#), [352](#)

Qubit, [95](#), [97](#), [111](#)

R

RAM, [106](#), [221](#)

Random access memory, [106](#), [109](#)

Reactive machines, [44](#), [215](#)

Reduced aggression theory, [35](#)

Region of Interest Align, [154](#)

Rete algorithm, [133–136](#), [138–139](#)

Robot, [4](#), [8](#), [50](#), [55](#), [63](#), [71](#), [82](#), [260–261](#), [355](#), [359](#), [362](#)

ROIAlign, [154](#)

Root cause analysis, [161](#), [238](#), [242](#), [265–266](#), [268](#)

Rosenblatt, [54](#), [56–57](#)

Rule engine inference cycle, [136](#)

S

SAFE, [191–194](#), [198](#), [290](#)

SDK, [86](#), [98](#), [113](#), [144](#), [171](#)
Secure sockets layer, [182](#), [185](#)
Self-awareness, [24](#), [43–45](#), [61](#), [64–65](#), [67](#), [91](#), [215](#), [345](#)
Self-domestication, [35](#), [41](#)
Self-replication, [77](#)
Sexual selection, [21](#), [28](#), [32](#), [36–37](#), [40](#)
Silicon Valley, [11](#), [94](#), [208](#)
Simple programmable logic device, [108](#)
Single-shot detection, [154](#)
Singularity, [77](#), [83](#), [332](#), [356–357](#), [362](#), [365](#)
SLIDE, [332](#), [345](#), [353](#)
SNN, [101](#), [104](#), [168–9](#)
SoC, [105](#)
Social brain theory, [32–33](#)
Social exchange theory, [35–36](#)
Software development kit, [98](#), [113](#), [144](#), [171](#)
Software history, [114](#)
Spatial intelligence, [26](#), [70](#)
Speech data, [104](#)
Speech human, [46–47](#), [151](#), [165](#)
Speech patterns, [151](#)
Speech processing, [100](#)
Speech recognition, [5](#), [44](#), [60](#), [99](#), [104](#), [150–152](#), [164–165](#), [228](#), [338](#), [344–345](#)
Speech synthesis, [5](#)
Spiking neural network, [99](#), [104](#), [168–169](#)
SPLD, [108](#)
SSD, [154–155](#)
SSL, [182](#), [185](#)
Stanford Centre for AI Safety, [191](#)

STEM, [38](#), [164](#), [206–207](#), [209](#), [264](#), [317](#), [350–351](#), [353](#)
Strategic management, [205](#), [254](#)
Strong AI, [44](#), [64](#), [215](#), [338](#), [344](#)
Structure of Intellect Model, [27](#)
Sub-Linear Deep Learning Engine, [332](#), [345](#), [353](#)
Super AI, [2](#), [43–45](#), [61](#), [64–65](#), [91](#), [111](#), [215–216](#), [331–332](#), [337](#), [340–344](#),
[352](#), [357](#), [359–361](#), [364](#)
Supervised learning algorithms, [124](#), [244](#)
Supply chain, [175](#), [233](#), [235–237](#), [240](#), [242–243](#), [263](#)
Support vector machine, [54](#), [59](#), [63](#), [90](#), [124](#)
SVM, [59](#), [90](#), [124](#)
Symbolic AI, [1](#), [46](#), [61](#), [346](#), [352](#)
System-on-chip, [105](#)

T

Task-specific integrated circuit, [105](#)
Technology stack, [224](#)
Text Recognition, [86](#), [118](#), [150](#), [152–153](#), [172](#)
Theory of Mind, [32](#), [37–38](#), [43–45](#), [61](#), [64](#), [91](#), [215](#), [345](#)
Thomas Watson, [9](#), [90](#)
Three Laws of Robotics, [54–55](#), [63](#)
Trade and technology council, [292](#)
Transactional, [73](#), [243](#)
Transformation, [19](#), [23](#), [30](#), [38](#), [71–75](#), [122](#), [153](#), [173](#), [203–205](#), [209–212](#),
[223](#), [227](#), [248–250](#), [311](#), [346](#)
Triarchic theory, [28](#)
Trojan Horse, [189](#)
TSIC, [105](#)
TTC, [292](#)
Two-factor theory, [27](#)

Types of Human Intelligence, [26](#), [70–71](#)

U

ULSI, [94](#)

Ultra Large-Scale Integration, [94](#)

UNDP, [208](#), [222](#)

Uni-factor theory, [27](#), [30](#)

United Nations, [11](#), [208](#), [214](#), [280](#), [283](#)

United States Defence Advanced Research Projects Agency, [89](#)

Unsupervised learning algorithms, [124](#), [214](#), [244](#)

V

Value chain analysis, [238–239](#)

Verilog, [86](#), [101](#), [106](#), [108](#), [110](#), [113–114](#), [146–150](#), [171](#)

Very large-scale integration, [90](#)

VHDL, [86](#), [106](#), [108](#), [110](#), [113–114](#), [146–150](#), [171–172](#)

Video analysis, [156–157](#)

VLSI, [90–92](#), [110](#)

VMS, [157](#)

VMware, [183](#), [199](#)

Voice recognition, [86](#), [118](#), [150–152](#), [172](#)

von Neumann, [52](#), [77](#), [85](#), [88](#), [99](#), [100](#), [110](#), [118](#), [167–172](#)

W

Wafer scale engine, [332](#), [345](#)

Wallace, [19](#), [21–22](#)

WIPS, [179–180](#)

Wireless intrusion prevention systems, [179](#)

World Bank, [246](#), [250–252](#), [274](#), [280](#), [283](#), [329](#)

X

XCON, [58](#)