KEMENTERIAN SAINS,
TEKNOLOGI DAN INOVASI
MINISTRY OF SCIENCE, TECHNOLOGY AND INNOVATION

THE NATIONAL GUIDELINES ON AI GOVERNANCE & ETHICS

"AI FOR MALAYSIA, AI FOR ALL"
To Enhance The Development and Deployment of AI Technology

Published by:

Ministry of Science, Technology and Innovation
Floors 1-7, Block C4 & C5, Complex C,
Federal Government Administrative Centre
62662 Putrajaya, MALAYSIA

603 8885 8084 | enquiry@mosti.gov.my | mosti.gov.my

©Publisher's Copyright
All rights reserved. No part of this publication may be reproduced, copied, stored in any retrieval system or transmitted in any form or by any means-electronic, mechanical, photocopying, recording or otherwise, without prior permission.

Table of Contents

Foreword

Foreword By The Deputy Prime Minister III

Preface By Minister of MOSTI IV

Executive Summary V

Preamble VI

Section 1: Introduction

1.1 Introduction 2

1.2 Key Considerations for National Guidelines Development 4

1.3 Objectives of The National Guidelines on AIGE 5

Section 2: Understanding AI

2.1 Understanding AI 8

2.2 A Case for Change - AI in Modern Transformational World 11

2.3 Ethical Issues in AI 13

2.4 Building Trust 14

2.5 Global Calls for Ethics of AI 16

2.6 National Considerations for Responsible AI 18

2.7 Responsible AI 21

2.8 Human-Friendly AI (HF-AI) Platform 28

Section 3: National Guidelines For Stakeholders

Introduction: Summary of Part A, Part B and Part C 32

Part A: National Guidelines For End Users of AI

A1.0 Introduction 36

A2.0 Responsible AI for the End Users 38

A3.0 What You Can and Cannot do as End Users 42

A4.0 Aligning End Users Contributions with Responsible AI and Sustainability 43

Table of Contents

Part B: National Guidelines For Policy Makers In Government, Agencies, Organizations And Institutions

B1.0 Introduction 46
B2.0 Responsible AI for Policy Makers 51
B3.0 What You Can Do as Policy Makers 61
B4.0 The Independent Advisory Body. 69
B5.0 Aligning Government Initiative with Responsible AI and SDG & ESG Principles 70

Part C: National Guidelines For Developers, Designers, Technology Providers And Suppliers

C1.0 Introduction 72
C2.0 Responsible Principles for Sector Players 74
C3.0 What You Can Do as Sector Players 78
C4.0 Demonstration and Sectors Cases 93
C5.0 Code of Ethics 102
C6.0 Aligning Industrial Economic Contributions with Responsible AI and SDG & ESG Principles 103

Section 4: Benchmarking And International Smart Partnership

4.1 Benchmarking and x International Smart Partnership 106

Section 5: Conclusion

5.1 Conclusion 110
5.2 Moving Forward 111

List of Abbreviations 114
List of Figures & Tables 115
List of Appendices 116
Acknowledgement 135

YAB Dato' Sri Haji Fadillah bin Haji Yusof
Deputy Prime Minister

Foreword By

Deputy Prime Minister

Artificial Intelligence (AI) is revolutionizing various aspects of society, including governance, work, and problem-solving. However, it is crucial to uphold ethical principles such as fairness, privacy, security, and transparency. To develop and use AI responsibly, we must work towards technological, societal, institutional, and legal approaches that align our AI systems with our values by actively prioritizing transparency, accountability, and fairness.

The Malaysian government is committed to ensuring AI technologies are deployed responsibly, respecting human rights, privacy, and societal well-being. Hence, the National Guidelines of AI Governance & Ethics have been prepared to meet international standards for sustainable development and corporate responsibility.

Consumer rights must be at the center of AI development and implementation. Policymakers must establish policies and pass laws and regulations that provide safe and consumer-centered technology, utilizing principles such as the right to object and receive an explanation, the right to have personal data deleted, and the right to collective redress. As we navigate a sustainable future, it is essential to align our approach to AI with global sustainability goals and environmental, social, and governance requirements.

On behalf of the Government of Malaysia, I would like to thank all ministries and government agencies, particularly the Ministry of Science, Technology & Innovation (MOSTI), the Working Committee of the National Blockchain and Artificial Intelligence Committee (NBAIC), academia, namely, Universiti Teknologi Malaysia, industry partners, including associations, civil society organisations, and professional bodies, both local and international, who have shared valuable input and experiences. All have worked hand in hand to make these National Guidelines of AI Governance & Ethics possible. Congratulations!

YB Tuan Chang Lih Kang
Minister of Science, Technology and Innovation

Preface By

Minister, MOSTI

The Ministry of Science, Technology and Innovation (MOSTI) is implementing the National AI Roadmap 2021-2025 to harness AI's power for economic growth and society's betterment. One of the Roadmap strategies include the establishment of AI Governance which emphasis on AI leadership, governance, ethics, cybersecurity, standard development, risk management, and assessing the impact of responsible AI. We also recognize the importance of establishing clear national guidelines for the responsible use of AI technologies and ethical risks. These ~~plan~~ guidelines promotes seven AI throughout the AI lifecycle. The Ministry and Ministry of Investment, Trade and Industry (MITI) are actively participating in ISO/IEC 42000 AI international standardization efforts to develop technical standards for AI technologies internationally.

Currently MOSTI aggressively working to create a thriving national AI innovation ecosystem that allows everyone (government, business and society) to capitalize on the benefits of AI in a secured and safe manner for economic prosperity and social well-being. Implementing our National AI Roadmap Action Plans represents a significant step forward in harnessing AI's transformative power for society's benefit.

Together, let us embark on this journey towards a more prosperous, equitable, and sustainable future powered by AI innovation and embrace the transformative power of AI while ensuring its benefits are enjoyed by all and no one is left behind. I hope these National Guidelines on AIGE will provide a reference to build trust, help to mitigate ethical risk as well as enhance AI adoption while simultaneously accelerating the country's competitiveness.

Executive Summary

Artificial Intelligence (AI) is the intelligence of machines or software as opposed to the intelligence of humans. The ultimate goal of AI is to create technology to enable machines to function in an intelligent manner, which includes learning, reasoning, understanding natural language, and perception, together with helping to enhance human capabilities and improve decision making. Generative AI (GenAI) has recently become a focus of excitement, and is a technology which involves training models to learn linguistic patterns and structures from large datasets and then using those learned patterns to generate new content.

The tremendous contribution of AI to sustainable economic development in different industries is rapidly becoming evident, making it an instant focus of attention at industrial, academic and even government levels. In August 2022, the Malaysian Ministry of Science, Technology and Innovation (MOSTI) launched the Malaysian National Artificial Intelligence Roadmap 2021-2025 (AI-RMAP), which set out frameworks for the integration of Artificial Intelligence (AI) into the different sectors of the economy through policy initiatives. Malaysia will be considered a high technology nation when cutting edge technologies like AI become a critical driver of productivity and competitiveness for the whole economy, not just the technology sector.

For this reason, Malaysia is actively participating in ISO Artificial Intelligence Standards development (ISO/IEC 42000). The Malaysian government has proposed seven principles to guide the development of Trusted and Responsible AI, in alignment with principles established by UNESCO, the OECD, the European Commission, and others. These are:

- Fairness
- Reliability, Safety and Control
- Privacy and Security
- Inclusiveness
- Transparency
- Accountability
- Pursuit of human benefit and happiness

The intention of the National Guidelines on AI Governance & Ethics (AIGE) are to develop and deploy AI in a safe, trustworthy, and ethical manner, taking an approach to developing and deploying AI from both an ethical and a legal point of view in line with AI-RMAP.

Preamble

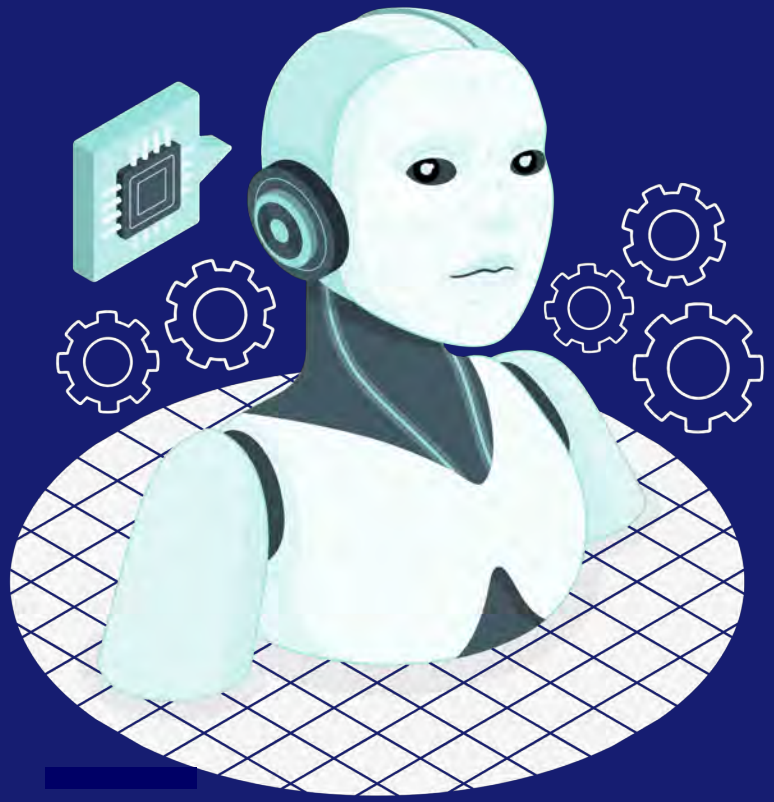
These National Guidelines AIGE (“the National Guidelines”) are intended to establish Responsible AI practices as the government aspires to enhance the development and deployment of Artificial intelligence (AI) through the Malaysian National Artificial Intelligence Roadmap 2021-2025 (AI-RMAP). The government hopes that the responsible use of AI will benefit the three main stakeholders, namely (1) the society, consisting of users, workers and consumers in general, (2) government agencies and policy makers, and (3) industry, including developers and designers.

AI is basically the application of technology to carry out intelligent behaviour using data and input on a computer system. The ultimate goal of AI is to create technology to enable machines to function in an intelligent manner which includes learning, reasoning, understanding natural language, and perception, which together help to enhance human capabilities and improve decision making. We can also define AI as the simulation of human intelligence in machines programmed to think like humans and imitate their actions. The field of AI brings together several approaches and techniques, including machine learning, natural language processing, robotics, and computer vision.

In August 2022, the Malaysian Ministry of Science, Technology and Innovation (MOSTI) launched the Malaysian National Artificial Intelligence Roadmap 2021-2025 (AI-RMAP), which set out frameworks for the integration of AI into the different sectors of the economy through policy initiatives. In order to achieve the objectives of the Roadmap, two action plans have been initiated: the development of National Guidelines for AI Governance and Ethics and the Institutionalization of Responsible AI. These documents intend to develop and employ AI in a safe, trustworthy, and ethical manner, taking an approach to developing and deploying AI from both an ethical and a legal point of view.

The Malaysian government recognizes that legal and cultural requirements on the use of AI may vary from one culture to another, especially in Malaysia. By means of the National Guidelines, it aspire to promote the beneficial use of AI aligned with the national ethical principles embodied in our Code of Conduct. Acknowledging that AI is a dynamic field, the National Guidelines will may be amended from time to time to reflect technological progress and changing ethical norms.

Section 1
INTRODUCTION



1.1 Introduction 2

1.2 Key Considerations for National Guidelines Development 4

1.3 Objectives of The National Guidelines AIGE 5

1.1 INTRODUCTION

AI plays a major and increasing role in modern society, and AI technology is appearing in various forms at all levels of our contact with society, from AI-powered devices in our pockets and handbags and small daily chatting intelligent robots to large industrial and government-level assisted offices where influential decisions are made at a national government or organizational level (Qin et al., 2023). AI is described as a system that uses the power of science and technology to create a system which can think and act like humans, and which can even surpass some precise algorithms of humans, having been applied and put into practice in many fields. AI-RMAP has defined AI as “a suite of technologies that enables machines to demonstrate intelligence, to adapt to new circumstances, and is used to amplify human ingenuity and intellectual capabilities across a broad range of challenges” (AI-RMAP, 2021). Generative AI (GenAI) has recently become a focus of excitement. GenAI is an AI technology which involves training models to learn patterns and structures from large datasets and then using those learned patterns to generate new content. This content can be manifested in various forms, ranging from textual representations in natural language, and images (including photographs, digital paintings, and cartoons) to videos, music, and software code, thus covering the full spectrum of symbolic expressions of human thought.

AI is basically designed to enable humans to make the best decisions. Its tremendous contribution to sustainable economic development in different industries is rapidly becoming evident, making it an instant focus of attention at the industrial, academic and even government levels

- Heylighen, 2017 -

70%

of major transnational companies will will have adopted at least one type of AI technology in 5 major sectors or areas, namely **finance, education, retail, transport and automation** by 2030.

- McKinsey Global Institute -

As in blockchain, 9 out of 10 leading businesses have invested heavily in AI technologies, and **37%**

of businesses and organizations in **education, manufacturing, retail, tourism, pharmaceuticals, IT, communication and automobiles** in developed countries are employing AI.

These are among the biggest employers; and they intend to increase that number to **90%** by 2030.

- Bughin et al. 2018 -

In comparison, less than **15%**

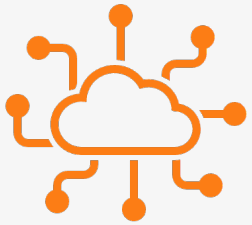
of major companies in the same sectors from 4 of the largest developing countries including **India, Brazil, Mexico and Indonesia** are deploying AI.

This expansion of AI will **eliminate 85 million “traditional” jobs** and create **97 million new job opportunities** or categories of jobs by 2025

- Deloitte Report, 2021 -

This confirms how AI is being aggressively used by all parties in carrying out their daily tasks. It is used to stay connected with new and returning customers through auto-reply emails, appointment reminders, and feedback surveys. In this regard, AI is capable of improving business practices.

Transition Towards Adoption of AI Technology And Its Economic Value



Global AI market Value is to reach **USD $267** billion by **2027**

(Fortune Business Insights, 2021)

Global economy contribution currently **USD $400** million expected to increase to **USD $15.7** trillion by the year **2030**

PwC Global,

Global economic activity of around **$13 trillion** by **2030**

McKinsey

ASEAN

AI contribution to GDP

USD 950 billion by **2030**

Malaysia

USD 115 billion (14%) by **2030**

(Kearney Analysis, 2020)

1.2 KEY CONSIDERATIONS FOR THE NATIONAL GUIDELINES DEVELOPMENT

The following are taken into consideration in the development of the National Guidelines on AI Governance & Ethics for Responsible and Inclusive AI (National Guidelines on AIGE):

AI for All	The National Guidelines need to be inclusive and cater for a range of stakeholders, and must not create an exclusive space. This is to ensure that no one is left behind (AI for members of the public).
Building Trust	The National Guidelines constitute a policy document for stakeholders, and showcase government intervention and commitment to governing AI as part of Responsible Government, and to build trust, and obtain the trust of the nation.
Comprehensible	The National Guidelines need to be simple so that the stakeholders can understand them.
Alignment with National, Regional and Global AI Practices	The National Guidelines refer to documents at national, regional and global levels, ensuring their coherency worldwide, and also draw on local and indigenous knowledge.
Synergy	The National Guidelines reflect synergy among- government agencies and the stakeholders operationalize and institutionalize AI in implementing national policies and action plans.
Call to Action	The National Guidelines serve as a reference and call to action for different sectors (healthcare, finance, etc) to develop their implementation guidelines to serve their own sectors.
Living Document	Since AI is a dynamic technology, these National Guidelines will be reviewed regularly and updated to meet the requirements and expectations of stakeholders. The National Guidelines will be made available in both Malay and English.

1.3 OBJECTIVES OF THE NATIONAL GUIDELINES AIGE

As mentioned in the Preamble section, these national guidelines are on a voluntary for the stakeholders. Nevertheless, these National Guidelines are intended to disseminate the importance of AI and Responsible AI to the entire nation with the theme of:

AI for Malaysia, AI for All

While Sections 1 and 2 are for general knowledge, Section 3 is divided into three separate parts according to the different stakeholders:

National Guidelines for AI End Users	**National Guidelines for Policy Makers in Government, Agencies, Organizations and Institutions**	**National Guidelines for Developers, Designers, Technology Providers and Suppliers**

Further explanation concerning the stakeholders are included in the Introduction to Section 3

The national guidelines are prepared with the stakeholders in mind. The objectives of these National Guidelines are:

I. To support the implementation of the Malaysia National AI Roadmap 2021 - 2025

II. To facilitate the implementation of Responsible AI according to the 7 AI Principles

III. To build trustworthiness in AI, which is emphasized by Responsible AI.

IV. To manage risks caused by the development and deployment of AI technology.

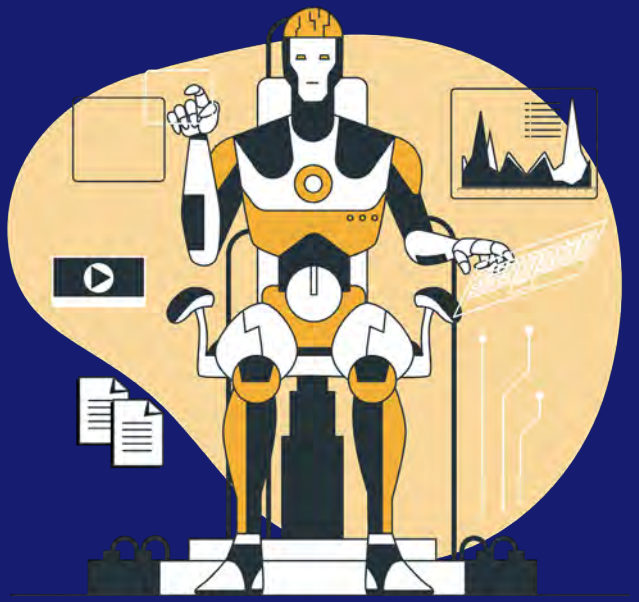
V. To maximize the benefits of AI to enhance the national productivity, economic growth and competitiveness.

The most urgent problem today is not to limit the development of AI or even to ensure it’s aligned with human values. It’s to ensure we have a plan to ensure it does what we want.

- Stuart Russell-

Section 2

UNDERSTANDING AI



2.1	Understanding AI	8
2.2	A Case for Change - AI in Modern Transformational World	11
2.3	Ethical Issues in AI	13
2.4	Building Trust	14
2.5	Global Calls on the Ethics of AI	16
2.6	National Considerations for Responsible AI	18
2.7	Responsible AI	21
2.8	Human-Friendly AI (HF-AI) Platform	28

2.1 UNDERSTANDING AI

This introduction is intends to introduces topics and explanations on the terminologies of AI in the hope that readers will find them useful and enlightening. AI is a rapidly evolving field of research and innovative technologies. Thus, in order to give readers, the idea of how AI has arrived at the current state of the art, this section traces its nature and evolutions over the decades.

Artificial Intelligence (AI)
The field of computer science that seeks to create intelligent machines that can replicate or exceed human intelligence.

Machine learning
A subset of AI that incorporates aspects of computer science, mathematics, and coding. Focuses on developing algorithm and models that help machines learn from data and predict trends and behaviours, without human assistance.

Deep learning
A function of AI that imitates the human brain by learning from how it structures and processes information to make decisions. Instead of relying on an algorithm that can only perform one specific task, this subset of machine learning can learn from unstructured data without supervision.

Generative AI
Models of algorithms that create new output such as text, video, photo, code, data, statistics from huge amounts of data that they are training of.

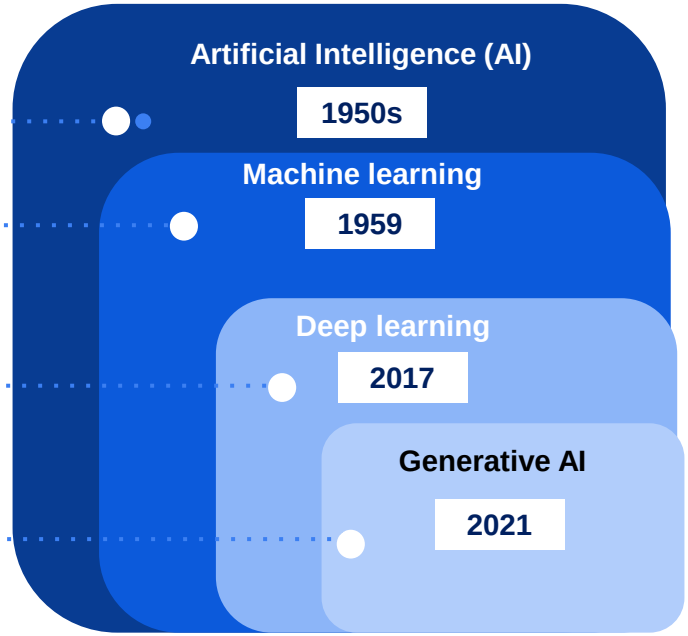


Figure 1: The evolution of AI technology

2.1 UNDERSTANDING AI

a. Traditional AI

AI, a technology developed since the mid-20th century, has revolutionized various fields, including Humanities and Natural Science. Its development has been influenced by the need to replace physical work, such as washing clothes, with AI, which has evolved to encompass various interpretations and applications.

The has grown significantly from its roots in basic computation, possesses the remarkable capability to undertake cognitive tasks. This empowers machines to handle a wide range of activities, from mundane chores such as garment cleaning to complex mathematical computations. However, unlike humans, machines do not attempt to replicate the intricate workings of the human brain. This fundamentally different approach results in their "artificial" intelligence. Paradoxically, this non-human method can create a convincing illusion of sentience when computer programs utilize language processing.

In the 1960s, the ELIZA program emerged as a response to the Turing test, a seminal concept designed to assess a machine's ability to convincingly simulate human conversation. ELIZA's objective was to comprehend and respond to user input, thereby prompting inquiry into the very nature of human intelligence. As advancements facilitated machines in mimicking human intelligence with increasing efficacy, CAPTCHAs were subsequently developed to distinguish between human and machine interaction.

Advancements in artificial intelligence (AI) have yielded significant progress in solving complex problems through text analysis. This approach involves identifying fundamental linguistic components such as nouns, adjectives, and verbs. The Hidden Markov Model (HMM) played a pivotal role in enabling linguists to comprehend how computers solve problems. This model facilitates the discovery of novel solutions rather than relying solely on pre-programmed ones. Furthermore, AI aspires to replicate human intuition, akin to the steering actions employed by human drivers, a development proving crucial for the advancement of self-driving car technology.

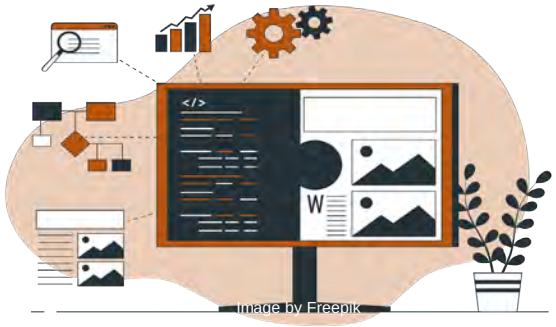
Artificial intelligence (AI) is revolutionizing facial recognition technology by drawing inspiration from the brain's communication and problem-solving mechanisms. Through the integration of machine learning algorithms, AI strives to replicate the natural processes involved in human face recognition. This approach holds the potential to not only improve facial recognition accuracy but also to deepen our scientific understanding of how neural networks within the brain communicate and facilitate intelligent tasks.

b. Generative AI

Artificial intelligence (AI) has significantly revolutionized the field of machine learning by leveraging extensive datasets. The vast amount of data readily available on the internet fuels this progress, enabling the creation of novel machine learning models capable of generating new forms of data, including text, images, and audio files. This advancement has demonstrably enhanced large-scale computing capabilities.

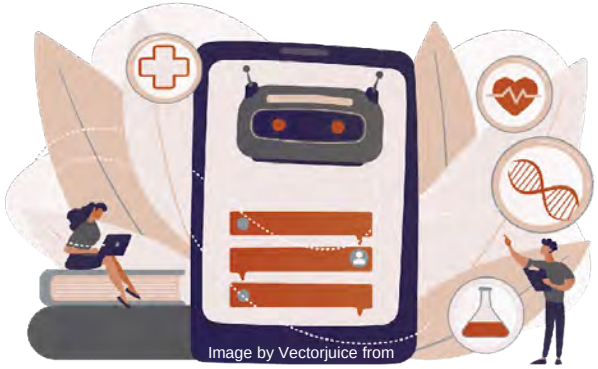
Generative AI, a subfield of artificial intelligence, has garnered significant popularity in recent times. A prominent example is ChatGPT, which allows users to interact with seemingly intelligent machines. However, it is paramount to distinguish between machine intelligence and human intelligence. While machines are meticulously trained to emulate human behavior, they lack the fundamental characteristics of human-like consciousness. Attributing sentience or pseudo-consciousness to machines can lead to misconceptions.

Since generative AI approach introduced in 2023, the government is exploring the policy, framework, opportunities and the risk including the intellectual property matters. All these will be incorporated at the next version of the AIGE.



c. AI and General Intelligence

Recent research indicates that human intelligence is not a single ability, but a combination of multiple abilities. While humans excel in specific tasks like kicking a ball, machines are designed for general intelligence (AGI), which is expected to be achieved in the near future. AGI aims to build machines capable of performing multiple tasks like humans.



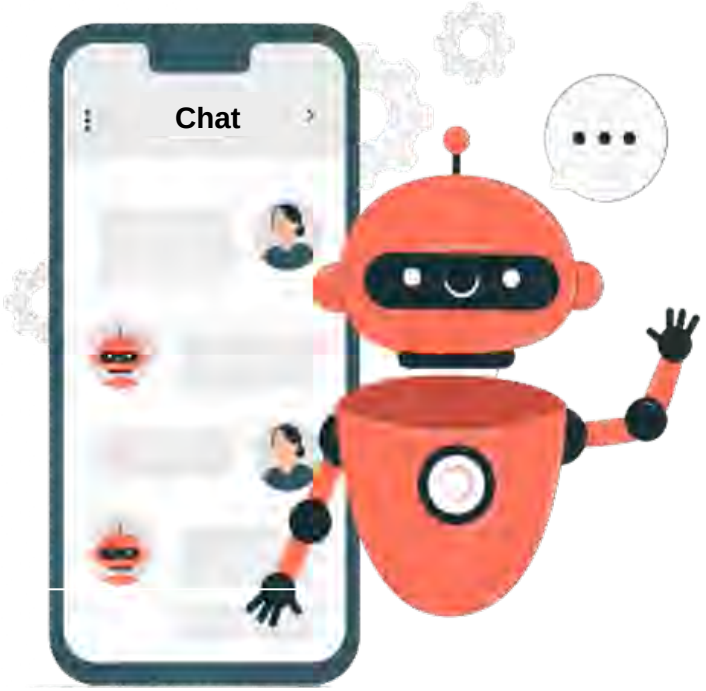
OpenAI, a leading AI development organization, aims to advance AGI, or AI systems that are smarter than humans, benefiting everyone. However, the exact definition and definition remain unclear. Reliable sources like the Alan Turing Institute provide valuable insights into AI and AGI, introducing a moral aspect to AI research and technology.

AI's current capabilities are being assessed, but potential dangers are also being recognized. The generative AI system recommending inappropriate ingredients underscores the need for a balance between recognizing potential dangers and assessing AI's capabilities, as a machine's comprehension of its output is crucial.

2.2 A CASE FOR CHANGE - AI IN A MODERN TRANSFORMATIONAL WORLD

The following are some examples of AI and Generative AI in different sectors in Malaysia:

Examples of Real-world AI and GenAI in Malaysia



Business, Banking and Finance Sectors

- AI-driven chatbots are used in customer service, provide instant responses to enquiries, and improve the overall banking experience.
- **CIMB Bank** was the first bank in Malaysia to launch an AI-driven chatbot, namely Enhanced Virtual Assistant (EVA).
- **HSBC Bank** launched a similar Chatbot Amy for its corporate banking clients. Amy understands both English and simplified Chinese, and can provide prompt responses to banking queries 24/7.
- **RHB Bank** and **RinggitPlus,** jointly launched My RHB Easy, a chatbot which gives users flexibility to apply for a personal loan, without the manual completion and submission of application forms.
- **Hong Leong Bank Berhad** has introduced a virtual assistant named HALI to improve the efficiency of bank operations. The chatbot developed for HLB is designed to assist two main core support functions, namely Human Resources and Branch Operations. It will take over staff queries on policies and procedures, previously undertaken by staff members from the two divisions

Examples of Real-world AI and GenAI in Malaysia

The transportation sector

- AirAsia has been at the forefront of using chatbots and artificial intelligence to enhance customer service and streamline their operations. Virtual personal assistant, **AVA** has been the first onboarding platform created in 2019 and designed to assist customers with tasks such as flight bookings, seat selections, check-in procedures, and answering frequently asked questions in eleven languages. In 2023, ‘Ask Bo’, an AI-powered concierge, was launched in connection with AirAsia’s commitment to transparent communications. Guests have more autonomy by being given the opportunity to talk live to human agents and reveal live information during the Ask Bo interaction.

The healthcare sector

- M3DICINE Technology Sdn Bhd, together with its strategic partner CREST (Collaborative Research in Engineering, Science and Technology) has launched Stethee Pro, the world’s first AI-enabled stethoscope system.

- The Stethee employs machine learning that can identify patterns and detailed analysis while monitoring the progress in the health of individual patients and uncover new trends in the fight against heart and lung disease globally (The Legal 500, 2021).

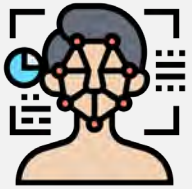
2.3 ETHICAL ISSUES IN AI

The internet has created amazing opportunities, but it has also given both good and malicious people access to them. However, the advancement of AI (e.g Generative AI) has increased the risk of spreading lies and enabling harmful actions, as machines are not trained for truth or ethical values. Some examples of ethical issues in AI are:

Limiting access/discriminatory pricing of products/services due to consumers' race/gender

Collecting and processing personal data in AI algorithms **without individual consent**

Processing data in AI algorithms for purposes other than for which it was collected

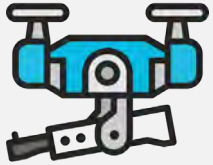


Citizens object to using **facial recognition technology** by police for mass surveillance

Citizens **object to the collection and use** of their data such as biometrics, by an AI system.

Over-reliance on machine-led decisions without disclosure in the banking sector

Customers demand reasoning/ clarification for decision taken by an **AI algorithm to deny credit/ transaction**



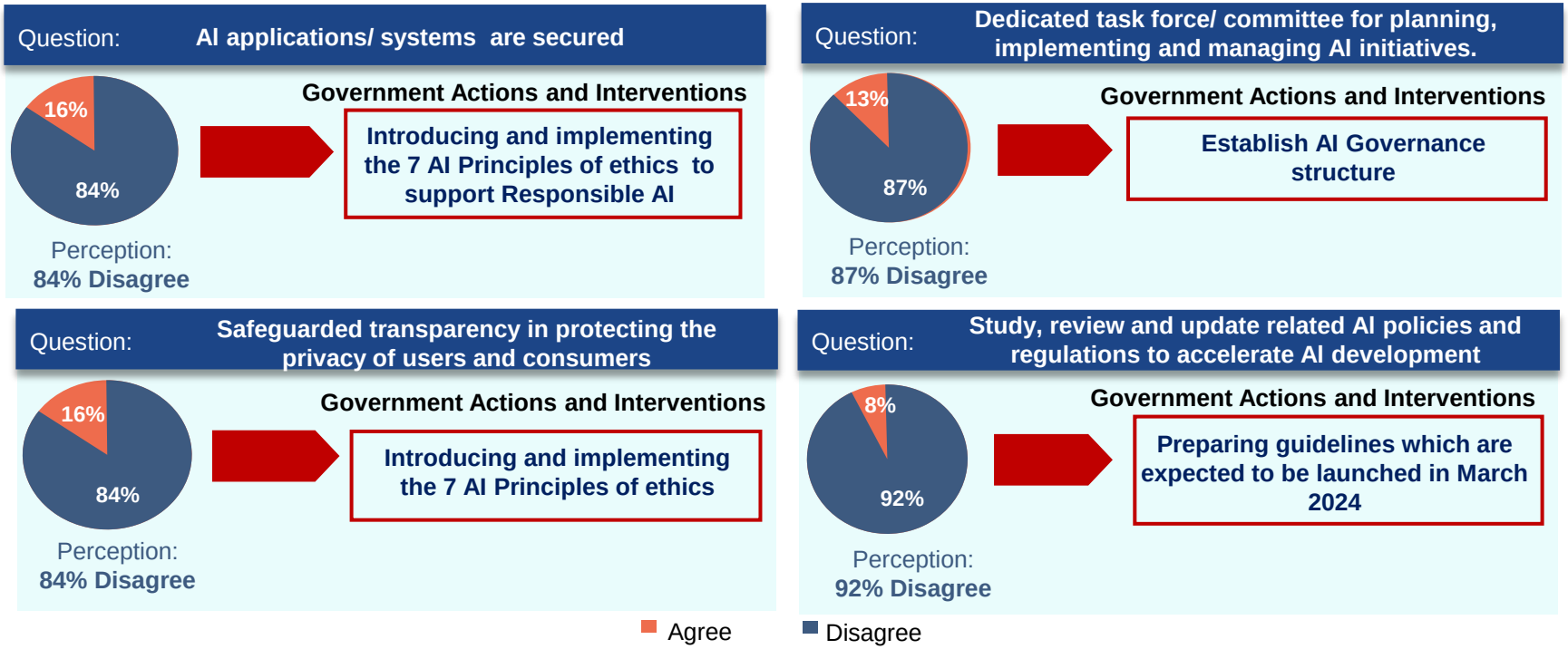
Crewless Ships and self directed drone swarms; using off-the-shelf tech for the **next stage of human conflict**.

Ethical consideration of colliding with wild animals or other vehicles. Autonomous Vehicle can not make **a decision to avoid a collision**.

2.4 BUILDING TRUST

It is essential to build trust in AI in order to overcome resistance and pave the way for its widespread adoption. This Explains why a study was conducted from 2021 to 2022 to ascertain stakeholder perceptions of AI. Its findings are illustrated in Figure 2 below. In 2023, relevant government agencies introduced several actions and develop interventions to address the concerns of stakeholders identified in the study.

Figure 2: Stakeholder perception of AI



Respondent: 250 government staff, industry and associations totalling of 800 respondents in the Klang Valley

2.4.1 Trust and Benefits

As illustrated in the study, many people were worried about the safety of AI technology. This document serves as a bridge to gain stakeholders' trust through the implementation of responsible AI. Trust is essential for protecting stakeholders in their use of emerging technologies. Digital Trust, as defined by various expert organizations, is the expectations of individuals that "digital technologies and services – and the people, processes, and organizations providing them – will protect all stakeholders' interests and ensure a secure digital world" (ISACA, WEF Digital Trust Initiative, 2022). The National Guidelines address the need of having AI systems governed, developed and deployed while adhering to the 7 principles. The following are some potential benefits of AI when used responsibly:

1 Enhancing AI Technology Development

Leadership in AI technology development can create the ability to set global norms and standards for the ethical, safe, and equitable use of AI.

2 Increasing Economic Growth and Innovation

AI drives innovation in industry, creating new products and services, improving efficiency and productivity, and fostering economic growth. Many companies invest in AI to gain a competitive edge on the global stage. .

3 Ensuring Security and Safety

AI plays a critical role in national security and safety, from cybersecurity to the analysis of intelligence data and autonomous weaponry.

4 Improving the Healthcare and Societal Well Being

In the healthcare sector, AI technologies are providing more accurate diagnostics, personalized medicine, and predictive analytics for patient care, leading to better health outcomes and efficiency in healthcare systems. At the same time, AI has the potential to address major societal challenges, managing large-scale public health crises by analyzing data and creating predictive models, for example during the last COVID pandemic.

5 Improving Talent For Job Opportunities

As AI changes the job landscape, countries see the importance of investing in AI to create new job opportunities and to upskill and reskill the workforce in line with the future needs of the industry and building the trust in AI.

6 Enhancing The Quality of Life

AI may also improve the quality of life by enhancing services such as transportation, communication, and entertainment, making them more personalized, efficient, and accessible.

Given all these benefits, countries tend to regard AI as a strategic asset that is central to their future development, security, and competitiveness, prompting significant investment and policy focus in this area. As AI technology continues to evolve, its importance in national agendas is likely to grow even further.

2.5 GLOBAL CALLS ON THE ETHICS OF AI

The first ethical code for AI systems was introduced by the famed science fiction writer Isaac Asimov, who presented his Three Laws of Robotics in Runaround (Asimov 1942). These three were later supplemented by a fourth law, called the Zeroth Law of Robotics, in Robots and Empire (Asimov 1986).

The four laws are as follows:

- A robot may not injure a human being or, through inaction, allow a human being to be harmed;
- A robot must obey the orders given it by human beings except where such orders would conflict with the first law;
- A robot must protect its own existence as long as such protection does not conflict with the first or second law;
- A robot may not harm humanity or, by inaction, allow humanity to suffer harm.

With the evolution of the technology and advancement of the modern world, the discussion of AI ethics started at an international platform on 2019, when the organization for Economic Co-operation and Development (OECD) introduced its Principles on AI.

However, the first global wide discussion on the Ethics of AI was formally started by UNESCO with the participation of its 193 member states in November 2019 at the 40th General Conference of UNESCO, when they adopted 40 C/Resolution 37, which mandated the Director-General "to prepare an international standard-setting instrument on the ethics of artificial intelligence (AI) in the form of a recommendation". This was later presented in 2021 as a Recommendation on the Ethics of AI.

UNESCO Recommendation on the Ethics of AI - Nov 2021

- The Recommendation on the Ethics of AI (UNESCO, 2021) which is the first-ever global standard on for AI Ethics was adopted by all 193 member states on 23 November 2021. The recommendation emphasises 10 AI Principles
- The protection of **human rights and dignity** is the cornerstone of the Recommendation, based on the advancement of fundamental principles such as transparency and fairness.
- **This recommendation** enables policymakers to translate the core values and principles into action with respect to data governance, environment and ecosystems, gender, education and research, and health and social well-being etc.
- UNESCO encouraged each of the country to develop AI governance for their country.

Malaysia Alignment with Global Calls : The AI Roadmap - August 2022

- To **support** the **UNESCO recommendation on AI ethics**, the Malaysian government introduced 7 AI Principles in August 2022 as explained in section 2.7.1
- The principles are included in the Malaysian National Artificial Intelligence Roadmap 2021-2025 (AI-RMAP)
- The Malaysian government has also taken the initiative to develop these National Guidelines with the intention to **enhance AI technology development and adoption.**

ASEAN AI Governance and Ethics Guidelines - February 2024

2022 → **2024**

- At the regional level, ASEAN Member States have approved the ASEAN Guide on AI Governance and Ethics in February 2024 to empower organizations and governments in the region to use AI responsibly and increase users' confidence in AI.
- This document serves as a practical guide for organizations in the region that wish to develop and deploy AI technologies.

These National Guidelines for AI Governance & Ethics for Responsible and Inclusive AI (National Guideline AIGE) are aligned to international best practices and standards in principles.

There are currently varying approaches to AI governance and in line with the emerging global consensus, Malaysia supports a risk-based approach that will allow for alignment with international approaches, which will support national competitiveness.

2.6 NATIONAL CONSIDERATIONS FOR RESPONSIBLE AI

2.6.1 Rukun Negara (National Principles)

Generally, any ethical guidelines or framework that supports values must be consistent with the constitution - a constitutional ethic' comprising principles and values, particularly inalienable human rights and civil liberties that line up with the aspirations of the framers of the Federal Constitution (Sadhu Singh, 2023). In addition, other national values are laid down in the Constitution Establishing Islam as the official religion of the Federation while allowing freedom to practise other faiths and the Constitutional Monarch, Yang di-Pertuan Agong ('YDPA'), as the supreme head of the Federation, and the supremacy of the Constitution.

Conceived by the National Consultative Council, a diverse body tasked with rebuilding harmony, the Rukun Negara was officially proclaimed on 31 August 1970. Rukun Negara is not just a set of principles, but constitutes a declaration of shared values, a common ground upon which all Malaysians can unite, regardless of ethnicity, religion, or background. Its five core principles -Belief in God, Loyalty to King and Country, Supremacy of the Constitution, Rule of Law, and Courtesy and Morality -aimed to bridge the divides and foster a sense of shared identity.

The Rukun Negara and the spirit of the Federal Constitution of Malaysia are intimately linked. Both documents enshrine core values that guide the nation's development. Both emphasize the importance of a pluralistic society based on mutual respect and understanding. The Rukun Negara can be seen as complementary text that strengthens and clarifies the spirit of the Federal Constitution.

The fundamental or core values of a nation-state are found in its constitution, comprising political, religious, moral, cultural and economic values (Faruqi, 2019). These values must naturally form an ethical framework adopted in Malaysia, namely the Responsible AI Principles. With the supremacy of the Constitution Preserved in Article 4(1) of the Federal Constitution, which declares that the 'Constitution is the supreme law of the Federation and that any law passed after Merdeka Day which is inconsistent with this Constitution shall, to the extent of the inconsistency be void', the Constitution provides a bedrock of values which when distilled operate alongside the Responsible AI Principles.

2.6.2 Rukun Negara and Responsible AI

Building a responsible AI ecosystem in Malaysia requires necessitates interweaving the nation's core values with the 7 AI Principles of responsible AI

1 Belief in God and Fairness

Adhering to this core value involves:

- Aligning the development of AI with the 7 AI principles that encompassing diverse religious perspectives, and preventing algorithmic bias based on religious beliefs , and
- Ensuring that the development is consistent with the Federal Constitution's spirit of equality and non-discrimination avoiding any form of discrimination on account religion and any other protected characteristics.

2 Loyalty to King and Country and Transparency

This includes

- Accountability and Public Trust which involves ensuring that there is transparency in AI decision-making processes, upholding the King's role as a symbol of unity and the Constitution's emphasis on accountability, and
- The National Interest and Security which involves the development of AI applications that safeguard national security and the public interest, aligning with the principle of loyalty while adhering to legal and ethical frameworks.

3 & 4 Supremacy of the Constitution and Rule of Law

These two core values include matters related to:

- Safety and Security to ensure that AI systems comply with legal and ethical frameworks, respecting the fundamental rights and liberties enshrined in the Federal Constitution.
- Inclusivity and Diversity which incorporate diverse perspectives in AI development and governance, reflecting the Malaysian multicultural nature of Malaysia and the Federal Constitution's emphasis on equality.
- Human Control and Oversight in order to maintain human control over AI systems, preventing autonomous actions that could violate constitutional rights or lead to unintended consequences.

5 Courtesy & Morality and Human Dignity

This core value requires that the AI systems are designed to:

- Respect individual privacy and dignity, aligning with the principle of courtesy and the Federal Constitution's protection of privacy.
- Emphasize Social Responsibility: Utilizing AI to address societal challenges and promote social good, upholding the spirit of ethical conduct and the Federal Constitution's commitment to a just and equitable society.

The thoughtful integration of these values and principles makes it possible for Malaysia to harness the power of AI, and at the same upholds its national identity and commitment to justice, inclusivity, and human dignity.

2.6.3 Responsible AI from an Islamic Perspective

Islamic principles guide the ethical development of AI, as Islam emphasizes social good for public interest and human welfare. The 7 AI principles align with Islamic practices and can be used as a foundation to analyze the implications of AI. Muslim participants believe that understanding and applying Islamic principles can contribute to a knowledgeable and ethical approach to adopting AI, as they align with the teachings of Allah and the Islamic principles of responsible AI.

Islam, a way of life, guides all aspects of human existence, including the ethical adoption of AI technology. As viceregents (khalifah) in the world, humans are responsible for the ethical consumption of resources provided by Allah. AI technologies have shown positive impacts on various aspects, such as education, business, and healthcare management. Islam accepts all technological progress, including AI, but requires ethical guidance to ensure it does not transgress Islamic boundaries and result in negative impacts.

AI in decision-making processes must adhere to ethical principles, promoting social justice and avoiding unfair treatment. Islam emphasizes the importance of human authority over AI systems to prevent misapplication and violate ethical principles. AI can optimize resource utilization, reduce waste, and promote sustainable practices, contributing to environmental preservation. Muslims seek guidance from Islamic teachings to address moral challenges posed by AI, such as self-governing weapons and stewardship of resources and the environment.

AI technology offers immense possibilities for Muslims, but it's crucial to respect human dignity, rights, and privacy. AI systems should promote inclusivity and diversity, avoiding disregard for human connection, compromise, and ethical judgment. Muslims believe AI should enhance human capabilities, not replace them, and upskilling and reskilling are essential for relevance in rapid technological changes.

The Islamic perspective on AI emphasizes the importance of ethical regulation and innovation, aligning with the 7 AI Principles. AI can address societal challenges and benefit humanity, but it also requires careful consideration of ethical dilemmas and human values preservation. By actively engaging in AI development and regulation, Muslims can contribute to a future where technology and ethics align with Responsible AI, fostering collaboration and open dialogue.

AI technology, while providing hardware and software, also plays a crucial role in spiritual values, providing the necessary heartware to drive activities forward.

2.7 RESPONSIBLE AI

2.7.1 AI Principles

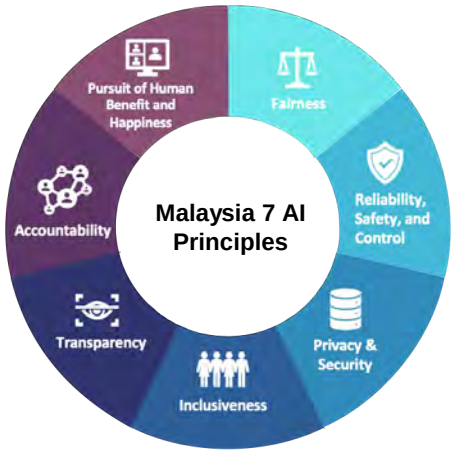
The aspiration for these National Guidelines is for the voluntary adoption of the seven AI Principles by developers and deployers of AI alongside existing laws to ensure that the adoption of AI technology is undertaken in consideration of both the potential benefits to the nation as a whole as well as minimising any harms and risks posed to the users. These national guidelines should be operationalised as part of a self-governance mechanism alongside the use of AI assurance techniques when assessing and auditing AI systems.

To comply with all of the 7 principles themselves, stakeholders need to understand the main component in each principle and how it is applied to them, which to a certain extent, needs consideration on its own. While the national guidelines (as specified in Section 3) do provide general guidance, it is important to note that each stakeholder group may have its own interpretation of the principles, in according to its sector and industry to which the technology applies. It is noted that the complexity of the way the AI system works and adhering to the national guidelines maybe far more intricate. However, the principles remain for the benefit of all stakeholders.

There will be both convergence and divergence in the definition of these principles as these National Guidelines continue to iterate and evolve alongside other standards being developed for regional and global adoption.

To ensure comprehensive and consistent application, of the operational definitions for the 7 AI Principles for Responsible AI in Malaysia will be developed as the nation progresses on the trajectory of AI adoption and evolution. This will ensure that the application of these principles is in line with converging values found in existing ethical frameworks worldwide. As an illustration, the first principle, Fairness is often associated with values such as justice, which includes respect for diversity, inclusion and equality, as well as avenues for redress and remedy to appeal, review or challenge decisions. Similarly, the seventh principle, Transparency, can be interpreted to include efforts to increase explainability, interpretability or other acts of communication and disclosure.

This is an approach to develop and deploy AI from both an ethical and a legal point of view. The goal of responsible AI is to employ AI in a safe, trustworthy and ethical manner. There should be a clear understanding of who is responsible for the outcomes of AI systems, and that these systems are transparent and understandable to the end users.



2.7.2 Morals, Principles and Ethics

When you read about AI, or listen to podcasts on AI you will frequently come across references to ethics and principles, which have become increasingly important as AI has developed. Less frequently, you will see references to morals, especially in the phrase moral principles. The three words ethics, principles and morals are very close in meaning, and in many contexts they can replace each other without much changing the meaning. However, they come from different sources, and there are slight differences in the way they are appropriately used. Explaining the use of these terms is taken as the point of departure for this section. If 7 AI principles are considered important, it is essential to ask why this is the case, which is the topic of the section entitled “The need for ethical principles”. This completes the essential background for the discussion of the seven AI principles adopted by the MOSTI, which forms the focus of this whole section on ethics, principles and morals

Morals and Moral Principles

Morals are rules based on socially acceptable behaviour, and they come from the Latin word for customs. Principles are general rules of conduct, and moral principles often focus on prohibiting socially unacceptable actions like stealing or assaulting other people. These principles help us determine what is right and wrong, and they are connected to the Golden Rule which involves treating others as we would like to be treated.

Moral values are shaped by social customs and can change as society evolves. For instance, a hundred years ago, concerns focused more on sexual behaviour, but racism became a significant issue by the end of the century. Racism is considered immoral because it goes against the Golden Rule. Our understanding of principles can also shift over time. Stealing used to refer mainly to taking physical possessions like money or jewellery, but in recent times, it includes the unauthorized use of ideas or creative work, known as intellectual property. This poses challenges for AI companies, especially with generative AI, as the training data may contain unauthorized intellectual property, impacting the output.

If we develop AI without any understanding of ethics, it will magnify our biases. If we develop AI without concern for its alignment with the human values, it will lack principles. It is up to us to ensure the technology is designed and used responsibly.

- Max Tegmark-

The Need For Ethical Principles

In an ideal world, people would behave following moral and ethical principles. That is alas not the kind of world in which AI has been developed. In the real world, AI is widely perceived to have the potential to bring with it catastrophic dangers, including in the extreme case the destruction of the entire human race. AI machines make use of new technologies, and many real dangers associated with AI belong to the long history of machines and new technologies. Although these dangers are real, they are not caused specifically by AI, and so belong to a different group than dangers specific to AI. The following subsections deal in turn with alleged catastrophic dangers, machines and technologies, and dangers associated with AI.

Alleged Catastrophic Dangers

There are on the internet many warnings claiming that AI is leading to catastrophe and presents an existential threat to humanity ('a danger that threatens the very existence of human beings). While there are real dangers that have to be taken very seriously, it is important not to invent imaginary dangers. Much confusion is caused by referring to machines as though they were human, and using expressions such as machine "learning", "training" and "intelligence". These are metaphors and they have to be understood as metaphors, which capture some resemblance between two entities without implying that they are the same thing. There is of course no foot at the foot of a mountain, and no head at the head of a valley. The development of machines is like learning and training, and the output is like intelligence; but these resemblances do not make them identical. If the distinction between resemblance and identity is disregarded, and if we assume without evidence that machines will one day turn hostile and compete with humans for mastery of the world, the development of "superintelligent" machines is indeed frightening. The outlook is even more bleak if predictions that machines will become billions of times more intelligent than humans are to be believed. However, there are real dangers that we know will have to be faced.

AI should be used responsibly, respecting individual rights and liberty, and adhering to legal obligations and policy requirements. It should identify and mitigate potential bias, be robustly tested for reliability and safety, benefit everyone, use explainable methods, and be accountable for its effects. AI systems should be safe and secure, benefit everyone, and comply with federal constitutions. Accountability should be established at each stage and across its lifecycle, including maintaining records. Human-centered values should be respected. These AI elements should be periodically reviewed to ensure continued purpose and resolution of issues. To enhance public trust in AI, AI systems must comply with the Principles for Responsible AI outlined in the Malaysian National AI Roadmap 2021-2025.

The section on Responsible AI Principles presents variations in their meaning across sections, potentially causing confusion about their fundamental nature. A suggested revision is to focus on a single, detailed explanation of each principle, not repeated elsewhere. Other sections can cross-refer to this section, and consistency in language and substance should be maintained. Modeling the language and descriptions from the UNESCO Recommendation on the Ethics of Artificial Intelligence could be a possible approach.

The AI principle of Accountability requires clarity on the definition of AI model owners and actors, as well as the definition of success or failure. In a B2B context, if an AI model is developed with responsible practices but biased data is submitted, the model may work successfully but produce undesirable output.

AI governance is crucial for responsible development and use of technology, considering the complexity of AI ecosystems. Each party in the value chain should be assigned risk-proportionate obligations based on their role and control over data and context. AI developers often create customizable tools with low-risk purposes, but customers often control data submission, configuration, and output usage. B2B companies like Salesforce, as data processors, may not own or control customers' data, making it difficult for them to know or control how customers use their data in AI systems. Therefore, a reasonable division of responsibilities should be established between entities at different points along the AI supply chain, including B2B companies like Salesforce.

2.7.3 Seven (7) AI Principles

We would appreciate the clarification on how these additional values will be integrated into and/or how these values will interact with the 7 AI Principles.

These sections all set out some version of the Responsible AI Principles but there are variations of what each principle means, from section to section. This may lead toA suggested revision is that a single, detailed and clear explanation of each of the Responsible AI Principles be set out as the main focus of this document and that it is not repeated elsewhere. Other sections can instead cross-refer to this section. In addition, there should be consistency in language and substance of the detailed description. Modelling on the language and descriptions from the UNESCO Recommendation on the Ethics of Artificial Intelligence (https://unesdoc.unesco.org/ark:/48223/pf0000381137) is a proposed approach.

lack of clarity on the fundamental nature of the principles, that should be the same across all use of AI.

1

FAIRNESS

The development of AI must be designed to avoid bias or discrimination against the target users. All AI systems should not have one-size-fits-all approaches and should address the widest possible range of factors (e.g. age, gender, religion, and ethnicity). This enables equitable asses and active participation of all stakeholders

2

RELIABILITY, SAFETY AND CONTROL

AI systems should be safe and secure, and perform as intended, and resist being compromised by unauthorised parties.

Developers should where necessary obtain consent from individuals before using or disclosing personal data for AI development and deployment. The systems must guarantee privacy and data protection throughout the entire life cycle of the AI system. Information and data collected from users must not be used in an unlawful or discriminatory way against them. Developers also should incorporate security-by-design and privacy-by-design principles and refer to international standards related to information security and privacy when implementing the AI system.

3

PRIVACY AND SECURITY

AI systems or solutions must be robustly tested to be reliable, safe and fail-safe by default so that users can trust and depend on the AI solution. To ensure proper access, control and protection of critical situations, AI systems should work reliably, consistently and operate in a real world under normal circumstances and unexpected conditions. In order to prevent or mitigate negative outcomes, AI systems should be able to response quickly according to the intended purposes.

Developers and end-users are encouraged to carry out relevant testing and certification as well as risk and impact assessments to prevent any potential harm and mitigate risks. Autonomous systems must have safeguards to ensure ultimate controllability by humans, particularly when it comes to the applications of AI in high-risk areas such as autonomous vehicles, military applications, and when human life is at stake.

4

INCLUSIVENESS

AI must be inclusive for all stakeholders in order to avoid unequal access to AI as in the case of social clefts.

AI systems should benefit everyone and address national needs and experience inclusively in compliance with the Federal Constitution and National Principles. This principle should be adhered in 3 ways: using inclusive development techniques; strategically developing tools which are sensitive to the specific needs of vulnerable groups in society; and proactively ensuring diversity amongst AI developers and decision makers.

5

TRANSPARENCY

AI algorithms should be transparent to ensure that any capabilities can be explained, both technical processes of an AI systems and related human decisions. This will allow stakeholders to evaluate the risks of AI and address issues that may arise.

Transparency principles are mainly applicable to contexts where AI is being used as part of decision-making process. There are 5 element need to be adhered: full disclosure about the facts that an AI system is being used in decision making; the intended purpose of the AI systems; the training data (describing the data that was used in training, what historical and social biases exist in the data; what procedures were used to verify data quality); maintenance and assessment of AI systems; and the ability to challenge the decisions made by the AI systems.

6

ACCOUNTABILITY

The developers, owners of AI models, and AI actors should be accountable for the success or failure of the AI solutions. They should take responsibility for ensuring the proper functioning of AI systems in compliance with AI Acts, governances, and ethical principles.

While designing AI systems, 4 elements (system purpose, technology capability, quality and reliability, and sensitive users) need to be considered to avoid consequential harm.

7

THE PURSUIT OF HUMAN BENEFIT AND HAPPINESS

AI systems should respect human-centered values, pursue human benefit for human society, enhance the quality of life, and increase human happiness.

AI systems should not be used for malicious purposes in making decisions. Understanding the way users interact with AI systems and how it is perceived if there are any negative outcomes from the adoption of AI can improve happiness and human well-being.

Human oversight helps to ensure that AI systems do not undermine human autonomy through governance mechanism such as human-in-the-loop (HITL), human-on-the-loop (HOTL) and human-in-command (HIC). HITL refers to the capability for human intervention in every decision cycle of the AI systems, while HOTL refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. HIC refers to the capability to oversee the overall activity of the AI systems and the ability to decide when and how to use the AI system.

2.8 HUMAN FRIENDLY AI (HF-AI) PLATFORM

Establishing an effective AI communication platform is crucial to ensure total engagement from all stakeholders in enhancing AI development and adoption in Malaysia. The Human-Friendly AI (HF-AI) Platform was introduced to provide immediate 24/7 national support, answering queries, solving problems, and providing information without delay or zero downtime. This HF-AI Platform can offer the following:

Operational Efficiency

Streamline and automate routine tasks and inquiries, freeing up human employees to focus on more complex and strategic tasks. This leads to higher productivity and often reduced operational costs..

Data-Driven Insights

This platform can collect and analyse interaction data, providing valuable insights into customer behaviour, preferences, and trends.

Personalization at Scale

AI can analyse users' data and tailor interactions accordingly, providing personalized recommendations and support for each user.

Enhanced Accessibility

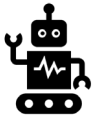
The platform can overcome language barriers and provide services in many languages, making information and services accessible to all.

Scalability

The AI platform can easily scale up to handle increased load without the need for a proportional increase in human resources.

Risk Management

AI can identify and mitigate risks by monitoring communications for fraud, misinformation, and other anomalies.



Innovation and Competitive Edge

Implementing cutting-edge AI communication technology can set a company apart from competitors, making it more attractive to users.

The existing platforms can be used to support the implementation of HF-AI Platform:

Knowledge Resource for Science and Technology Excellence, Malaysia (KRSTE.my)

KRSTE.my utilizes SPAF to offer advanced knowledge management and technology services, facilitating context-based search, informed decision-making, and visual content display.

Malaysian Government Central Data Exchange (MyGDX)

MyGDX facilitates the provision of online services through data integration, enhancing coordination and efficiency across agencies. This service reduces infrastructure provision costs and system integration, thereby enhancing the overall efficiency of online services.

Outreach Programmes-Malaysia Open Science Platform (MOSP)

The MOSP, a strategic initiative in Malaysia, aims to enhance the STI Collaborative Ecosystems, ensuring research data accessibility and sharing aligned with national priorities. It is connected to Jabatan Digital Negara's Open Government Data portal and MASTIC's Raw Database for Research and Science.

AI untuk Rakyat is an online program aimed at raising public awareness about Artificial Intelligence (AI) among senior citizens, students, professionals, and those interested in developing a digital first mindset, promoting a clear explanation of AI. The programme is divided into two sections:

<u>Section 1 AI Aware:</u>
The aim is to raise awareness about AI, and includes defining the terms Artificial Intelligence and Human Intelligence and explaining the differences between AI and non-AI machines, and addressing popular misconceptions about AI technology.

<u>Section 2 AI Appreciation:</u>
This section includes describing the different domains of AI and their impact on various industries, and explaining the principles of responsible AI and AI ethics.

Section 3

NATIONAL GUIDELINES FOR STAKEHOLDERS

3.1 Introduction: Summary of Part A, Part B and Part C 32

INTRODUCTION: SUMMARY OF PART A, B, AND C

AI System as Structure for National Guidelines development

The OECD defines AI systems as machine-based systems that, for explicit or implicit objectives, generate outputs like predictions, content, recommendations, or decisions, varying in autonomy and adaptiveness post-deployment.

These national guidelines cover stakeholders within the systems and include policy makers to govern within the system components. Stakeholder scope and national guidelines for stakeholders (End Users, Policy Markers, Developers) are illustrated in figure 4. While these three stakeholder groups may not be exhaustive, they are necessary and fairly sufficient to establish the fundamental building blocks of an AI system. With this, a clear shared responsibility framework is developed to detail different roles and responsibilities among the key stakeholders. The shared responsibility framework needs to ensure that there is a clear delineation of roles and responsibilities between group of stakeholders. It is important to note that, in reality, a single entity or person can certainly play more than one role.

For example a specific stakeholder can hold multiple roles at the same time, depending on the particular use of AI. These categories of stakeholders can be mutually exclusive and collectively exhaustive. Each stakeholder can refer to parts A, B or/and C according to their interest and requirements.

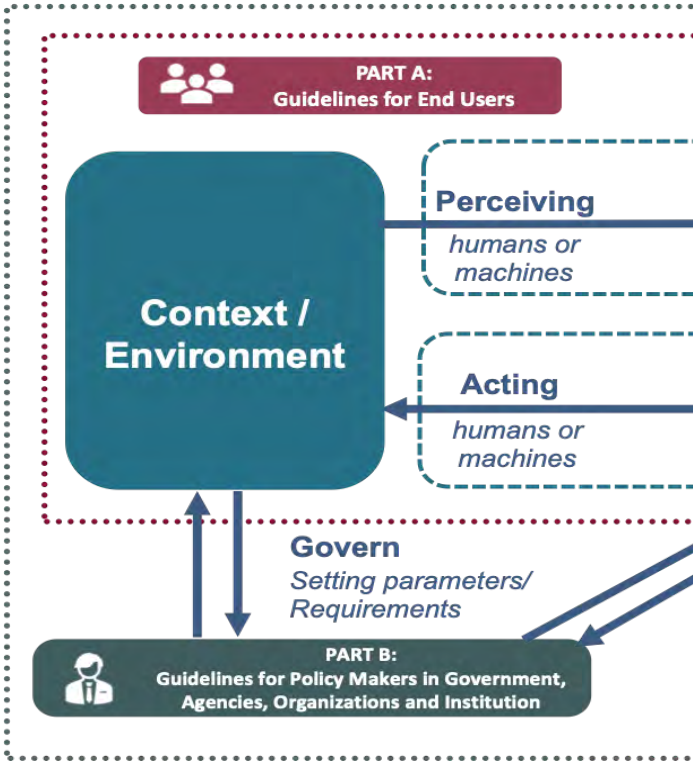


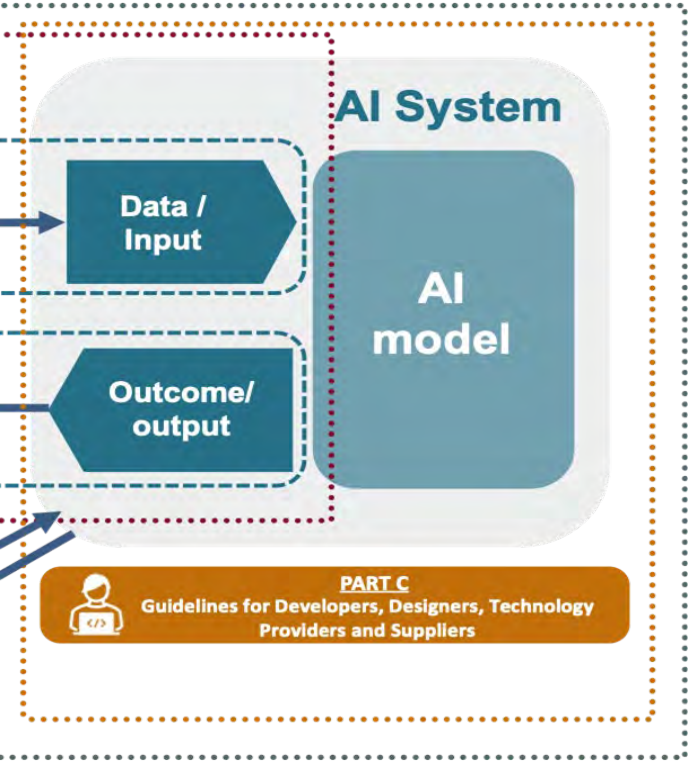
Figure 2: The scope of stakeholder and national guidelines covering

PART

A **End Users**

End users are individuals or organizations who use AI products. Different groups of end users use AI products and services in different ways. Some use AI-powered virtual assistants, smart home devices, or AI applications on their smartphones.

Others working in different agencies and industries use AI-generated applications: government agencies use AI for security, law enforcement, or public services; manufacturers use it for quality control and to optimize their supply chain; banking institutions rely on AI for fraud detection; and content creators use AI for video editing or recommendation systems. Since these consumers may or may not be aware that they are adopting AI technologies, these guidelines may raise their awareness of responsible AI use, inform people about their rights and responsibilities, and empower them to make informed decisions regarding AI technologies and applications.



PART B

Policy Makers of Government, Agencies, Organizations and Institutions

The primary target audiences for the National Guidelines are the policy makers, planners and managers responsible for AI workforce policy and planning at national and local levels. Government officials, policy makers and regulatory agencies are individuals or agencies that are responsible for formulating and implementing policies and regulations concerning the use of AI, and for ensuring ethical and Responsible AI practices across different sectors.

The National Guidelines provide a framework and practical recommendations to ensure the responsible and ethical use of AI within the jurisdictions of those responsible, and guide them in ensuring compliance, consumer protection, and fair competition in AI-related activities.

PART C

Developers, Designers, Technology Providers and Suppliers

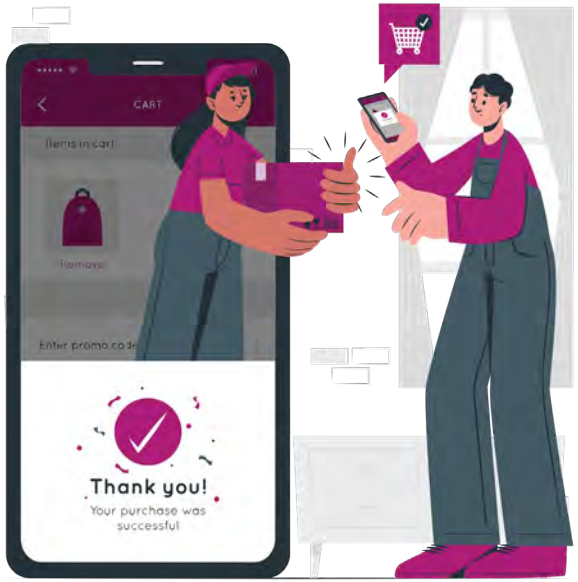
Developers and designers are those involved in developing and designing AI products. They come from different fields of expertise, such as computer science, data science, engineering and AI research, and they work for different AI users and industries. The National Guidelines outline best practices, ethical considerations, and technical standards for Responsible AI development and deployment. The Guidelines also enable these professionals to align their work with recognized benchmarks for better outcomes, and to reduce or mitigate any AI ethical risks.

Despite helping to improve productivity, efficiency, the quality of life and creativity of human, the modern and dynamics Generative AI technology also has a dark side that society needs to be aware of. Human values, data protection and the responsible use of technology are all affected and given a new meaning by this technology. Therefore, all developers, researchers, data scientists and decision-makers to work together to develop and adopt AI-related technology Responsibly and Ethically.

Dato' Seri Anuar Ibrahim - The Prime Minister of Malaysia

PART A

AI FOR END USERS



A1.0	Introduction	36
A2.0	Responsible AI for The End Users	38
A3.0	What You Can and Cannot do as AI Users	42
A4.0	Aligning Users Contributions with Responsible AI and Sustainability	43

A1.0 INTRODUCTION

1.1 INTRODUCTION TO THE NATIONAL GUIDELINES FOR AI END USERS

The National Guidelines for end users are essential because they mitigate the risk or harm, protect rights and values, build trust and confidence, enhance accountability, transparency, as well as align and empower end users' rights and societal values. As an illustration, the first ethical principle Fairness is often associated with values such as justice, which includes respect for diversity, inclusion and equality; as well as avenues for redress and remedy to appeal, review or challenge decisions. The principle of transparency can be interpreted to include efforts to increase explainability,interpretability or other acts of communication and disclosure. There will be both convergence and divergence in the definition of these principles as these National Guidelines continue to evolve alongside other standards being developed for regional and global adoption.

This National Guidelines will be voluntarily adopted by all users. To ensure comprehensive and consistent application, the definition of each seven AI Principles in Malaysia as explained above (section 2.7.3) is part of the AI-RMAP action plan towards enhancing AI technology adoption. By adhering to these AI principles and National Guidelines, individuals can ensure that AI technologies are developed and adopted in a manner that benefits communities, while at the same time respecting the rights and dignity of all end users. To ensure that the adoption of AI technology will not cause any harm or risk, it is important to approach AI with an understanding of the dos and don'ts that need to be adhered to by all users as explained under item A3.0. On a positive side, the country can exploit the full potential of AI technologies for the benefit of its citizens, while addressing ethical concerns and mitigating environmental impacts. This alignment will foster an inclusive and sustainable AI ecosystem that is expected to contribute to the well-being and progress of Malaysia as a whole.

A1.0 INTRODUCTION

1.2 BENEFITS OF ADOPTING AI

Education

- Reference Management
- E-Learning
- Assessment and evaluation

Transportation

- Grab service
- Wave apps
- Smart traffic light
- Smart cities (LRT)

Agriculture

- Drone
- Smart farming

Tourism

- E-booking

Healthcare

- Heartrate monitoring device
- Online consultation
- MySejahtera

Business and Financial

- E-banking
- Online Shopping

Entertainment

- Youtube
- Netflix

A2.0 RESPONSIBLE AI FOR THE END USERS

2.1 AI ETHICS

Ethics can be defined as a set of moral principles and standards that guide an individual's conduct and differentiate between right and wrong, good and bad. In the context of Malaysia, a diverse country with various ethnicities, cultures, languages, and religions, Malaysia's Rukun Negara/ National Principles) was introduced on 31 August 1970 with the main purpose of forming a strong unity. Malaysia's the Rukun Negara/ National Principles directly states the belief in God as a doctrine to be followed by all Malaysians irrespective of religious and cultural differences, implying the importance of religion in Malaysian society despite cultural diversities. Malaysians embrace religious teachings that promote fundamental principles and virtues to lead fulfilling lives. Religion provides a moral compass that guides individuals towards ethical behavior, emphasizing positive values such as compassion, justice, and fairness.

However, as societal norms change and new challenges emerge, ethical considerations must be adapted and address contemporary issues.

AI ethics is a field that deals with the moral issues related to the development, use, and impact of AI. It focuses on how developers, manufacturers, operators and users should behave to prevent ethical problems caused by AI, such as unethical design, misuse, or wrong applications. The goal is to create a fair digital society where AI is overseen by trustworthy humans, used responsibly, and respects our choices and values.'

"The goal is to create a fair digital society where AI is overseen by trustworthy humans, used responsibly, and respects our choices and values"

2.2 CASCADING THE 7 AI PRINCIPLES TO END USERS

The ubiquity of AI has made it necessary to consider how AI improves people's lives while avoiding harm. For example, according to the non-profit OpenAI (https://openai.com): "Our mission is to ensure that Artificial General Intelligence – AI systems that are generally smarter than humans – benefits all of humanity". Ethical developments in the AI industry are complemented by government initiatives, including MOSTI's National AI Roadmap 2021-2025 (https://my-ai.my/) which includes (p. 30) a table containing a set of seven AI principles. These principles are explained cascading for end users in detail below:

1. FAIRNESS

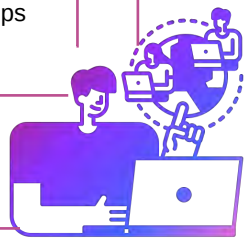
Fairness is a crucial principle in AI to prevent discrimination and ensure equal treatment. AI systems should not discriminate based on race, gender, or religion, and algorithm developers must be cautious of unintentional biases in the data. It is also important to consider the equitable distribution of AI benefits to avoid leaving certain groups without access to its advantages.

2. RELIABILITY, SAFETY AND CONTROL

When interacting with AI, it is crucial for the algorithms to be reliable and safe. This applies to robots in restaurants delivering food and face recognition systems accurately identifying individuals. Safety is especially important in areas like driverless cars, healthcare, and financial services. In addition, measures need to be taken to prevent intentional misuse and ensure control over AI systems.

3. PRIVACY AND SECURITY

Privacy and security are crucial when dealing with personal data like financial and health information. Proper procedures, informed consent, and secure storage are necessary to handle collected data responsibly. Complying with data protection laws is often required to ensure privacy. In addition, security measures are essential to prevent hacking and other malicious attacks by bad actors.



4. INCLUSIVENESS

Inclusiveness in AI means ensuring that no groups are unfairly excluded. Kedah rice farmers are unlikely to be thought of as a group who have much to gain from AI, and they themselves are unlikely to take much interest in AI as a discipline. Nevertheless, they have an interest in the potential of AI in developing and choosing the best seeds for their soil and climate, and in maximising the yield from their plots. Considering diverse stakeholders and promoting inclusiveness is the goal of these guidelines.

5. TRANSPARENCY

Transparency in AI is about being able to understand what an AI company does with personal data and having the right to access relevant information. It means that the company openly shows how it operates, especially regarding personal data, and provides the information that end user are entitled to know. Lack of transparency occurs when a company refuses to share important information upon request, which can be unfair.

6. ACCOUNTABILITY

Accountability in AI involves taking responsibility for the actions and outcomes of the technology. In cases where something goes wrong, it can be challenging to pinpoint a single person or entity. In this respect, accountability needs to be identified and addressed during the design and deployment of AI systems to ensure that responsibility is properly assigned and fulfilled.

7. PURSUIT OF HUMAN BENEFITS AND HAPPINESS

The pursuit of human benefit and happiness is given prominence in AI because it serves as a central principle that other principles are connected to. It challenges the conventional view that companies should prioritize delivering dividends to shareholders and emphasizes the moral responsibility of organizations. AI practitioners, including OpenAI, prioritize this moral duty to promote human well-being and explicitly express it in their mission statement

2.3 CONSUMER PROTECTION PRINCIPLES IN AI

There is a need for strengthening consumer protection through Consumer Rights Principles for a forward-looking regulatory framework to make AI technology safe, reliable and fair so that consumers are not used as laboratory experiments for new technologies, in this case AI.

The principles set below provide a foundation for how government and enforcement agencies should approach the opportunities and pitfalls of AI.

Strongly advocates the following consumer rights to be adhered to in AI:

1. Consumer rights must be respected at all times concerning AI products and services.
2. Consumers should always have the right to information. They should be made aware when an algorithm is using their personal information to provide offers for goods and services, uses this data to make decisions or reports their data to third parties.
3. The consumer must have the right to object and to be given explanation.
4. Consumers must have the “right to be forgotten” to have personal data deleted.
5. Consumers must have the right to interact with a human instead of AI.
6. Consumers must have the right to redress and compensation for any damages.
7. Consumers must have the right to collective redress. This would enable consumers to go to court as a group if a company has not respected their rights.
8. Consumers must have the right to complain to a supervisory authority or take legal actions.
9. Developers and deployers of AI must establish systems to ensure that these rights are available.

Collaboration between developers and end users is valuable in ensuring that fundamental rights and consumer rights are upheld. Since technology is evolving rapidly, legislation should be adaptable to address new challenges. In areas without established norms, codes of conduct can help with self-regulation.

FOMCA

“Every end user or consumer of AI has the right to be well-informed and empowered. It is important to know when algorithmsare using individual personal information to make an offer or a decision. Consumers should also be awareif their data is being shared with third parties. If the outcomes generated by an AI-based compliant handling system isinaccurate, it should always have the option to ask for human intervention. Access to resources like documentation, tests,datasets, and data prediction tools is essential for understanding and assessing the algorithms being used byan AI system. Awareness campaigns targeting consumers, companies, and policy makers can promote understanding and responsible use of AI technology”

Consumer Protection

Regulation of AI is still at an early stage of development. However, some suggestions below can perhap be considered:

1 Define Generative AI and its Scope

The amendment to the current laws should define generative AI and its various applications to ensure clarity in the law. This definition should encompass AI systems that can create new content, such as text, images, or videos, based on patterns and data inputs.

2 Mandatory Disclosure

Companies deploying generative AI should be required to disclose immediately when AI is used in generating content or recommendations. This transparency allows consumers to make informed decisions and understand the source of the information they are consuming.

3 Quality and Accuracy Standards

Establish quality and accuracy standards for the content generated by AI systems. Companies should be held responsible for ensuring that the content generated by their AI systems meets these standards. This could include measures to minimize misinformation, bias, or harmful content.

4 Liability for Harmful Content

Define liability for companies whose generative AI systems produce harmful or misleading content. This could include provisions holding companies accountable for damages caused by content generated by their AI systems, particularly in cases of defamation, infringement of intellectual property rights, or dissemination of false information.

5 Data Protection and Privacy

Strengthen data protection and privacy provisions to safeguard consumer data used to train generative AI systems. Companies should be required to obtain explicit consent from users before using their data for training AI models, and ensure that data privacy rights are respected throughout the process.

6 Monitoring and Enforcement

Allocate resources to monitoring the use of generative AI systems by companies and enforce compliance with the regulations. This could involve establishing specialized units within regulatory agencies overseeing AI-related issues and conducting regular audits of companies' AI systems.

7 Consumer Education and Awareness

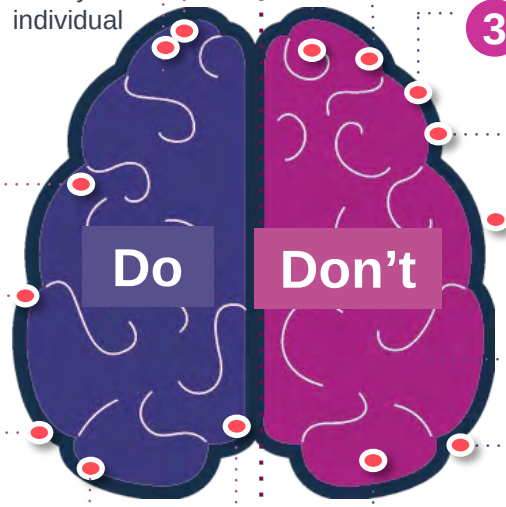
Launch public awareness campaigns to educate consumers about the capabilities and risks of generative AI technology. Empowering consumers with knowledge about how AI works and its potential impacts can help them make more informed choices and protect themselves from potential harm.

8 Consultation with Experts

Consult with AI experts, consumer rights advocates, industry stakeholders, and other relevant parties during the amendment process to ensure that the regulations are effective, balanced, and practical.

A3.0 WHAT YOU CAN AND CANNOT DO AS END USERS?

Adopting AI technology is a transformative move for any organization or individual, but it's important to approach it with an understanding of both the potential benefits and the challenges or risks involved. Below is a list of Dos and Don'ts when adopting AI technology.



Do

1. **Understand Your Needs**
   Clearly identify problems you want to solve with AI and understand how AI can specifically address those needs.
2. **Explore**
   Begin exploring the AI technology to know and to be competitive
3. **Protect Personal Data**
   AI's performance is heavily depending on the availability of data from various sources. Therefore protecting individual personal data is crucial
4. **Adopting Ethical Guidelines**
   Adopting ethical guidelines and checks to ensure that the AI system is fair and unbiased
5. **Upgrade AI Skills**
   Educate and train ourselves with AI to remain competitive
6. **Stay Updated**
   AI is a rapidly evolving field. Regularly update your knowledge and skills to keep up with new developments.
7. **Accepting the benefits of AI**
   To recognize and accept AI outstanding
8. **Invest in AI**
   Start invest in AI and carry out assessment for future benefits and impact assessment for future investment

Don't

1. **Underestimate the Complexity**
   Implementing AI is not just about technology; it involves change lifestyle and cultural shift.
2. **Ignore the Human Element**
   AI should augment human abilities, not replace them. Focus on how AI can assist and empower your daily life.
3. **Neglect Data Privacy**
   Be mindful of privacy laws and ethical considerations related to data collection and usage.
4. **Overlook Safety and Security**
   Comply to relevant regulation to safeguard our personal information and data
5. **Chase Hype**
   Don't adopt AI just because it's a buzzword. Ensure it genuinely adds value to your wellbeing and happiness
6. **Overdependent on AI**
   Don't implement AI without understanding of how it fits into your lifestyle and avoid rely solely on AI
7. **Ignore on Reliable Source**
   Don't adhered and trust unreliable data and source
8. **Disregard Regulations**
   Stay compliant with all relevant laws and regulations affecting AI in your industry and region.

A4.0 ALIGNING END USERS CONTRIBUTION WITH RESPONSIBLE AI AND SUSTAINABILITY

AI is revolutionizing Malaysia by promoting responsible use and sustainability, focusing on economic, social, and environmental sustainability. Advances in AI can aid in understanding climate change, transitioning to sustainable transport systems, and accelerating agrotech to help farmers end food poverty and malnutrition, contributing to the achievement of other Sustainable Development Goals.

Economic Sustainability:

- AI technologies are revolutionizing economic growth and sustainability by driving innovation, productivity, and efficiency. AI-powered systems streamline processes, reduce costs, and create new job opportunities. However, fair and inclusive AI systems are crucial to avoid biases and discrimination, ensuring economic sustainability in various industries.

Social Sustainability:

- AI systems are revolutionizing healthcare, particularly in remote areas where access to medical professionals is limited. By enhancing services, AI can provide equitable access to healthcare for all Malaysians. Transparency in AI systems is crucial for building trust and ensuring responsible adoption, as it allows users to understand and question outcomes, thereby enhancing the overall healthcare experience.

Environment Sustainability:

- AI plays a crucial role in optimizing resource allocation, reducing waste, and promoting environmental sustainability. It aids in monitoring and managing natural resources like water, thereby reducing reliance on fossil fuels.

Malaysia is embracing AI as a tool for innovation, aiming to create an inclusive and sustainable ecosystem. A collaborative approach is being adopted to ensure AI is developed ethically and transparently, benefiting citizens and governments. This approach will ensure the full potential of AI technologies is harnessed, while addressing ethical concerns and mitigating environmental impacts, contributing to Malaysia's overall progress.

The development of full artificial intelligence could spell the end of the human race… It would take off on its own, and re-design itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete and would be superseded."

-Stephen Hawking-

PART B

POLICY MAKERS OF GOVERNMENT, AGENCIES, ORGANIZATIONS AND INSTITUTIONS

B1.0	Introduction	46
B2.0	Responsible AI for Policy Makers	51
B3.0	What You Can Do as a Policy Maker?	61
B4.0	The Independent Advisory Body	69
B5.0	Aligning Government Initiative with Responsible AI and SDG & ESG Principles	70

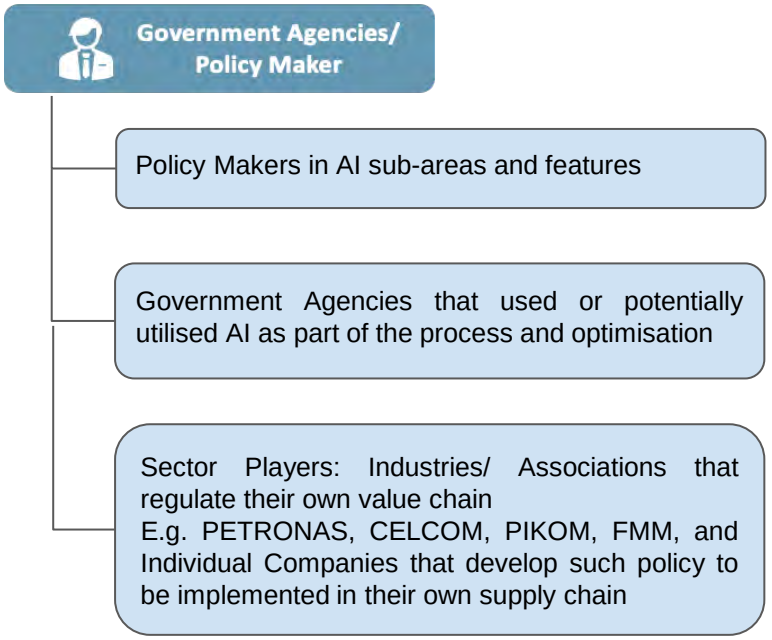
B1.0 INTRODUCTION

1.1 Introduction to the National Guidelines AIGE for Policy Makers of Government, Agencies, Organizations, and Institutions

AI is increasingly being integrated into governmental processes to improve efficiency, transparency, and service delivery. However, it is crucial to establish a framework for responsible and ethical AI practices. The National Guidelines for Responsible AI for Government, Agencies, Organizations, and Institutions aim to balance technological innovation with fundamental values, fostering a future where AI can advance societal progress while minimizing risks and ensuring inclusivity. This demonstrates the commitment to navigate the complex terrain of AI governance, promoting accountability, fairness, and responsible use of AI in public and self-governing sectors.

AI is a crucial tool in various sectors, and its use is governed by national responsibilities. These National Guidelines aim to guide decision-makers, regulators, planners, and managers in ensuring ethical and responsible AI use. They emphasize the need for international collaboration and the development of technical standards to ensure fair competition in AI-related activities. The national guidelines aim to ensure the interoperability of secure and trusted AI systems across borders.

For ease of understanding, the **Government and Policy Makers are not restricted to public sectors alone**. The image below can summarize on the differentiation of type for policy makers



1.2 AI INNOVATION ECOSYSTEM- SYNERGISING MULTIPLE STAKEHOLDER ENGAGEMENT

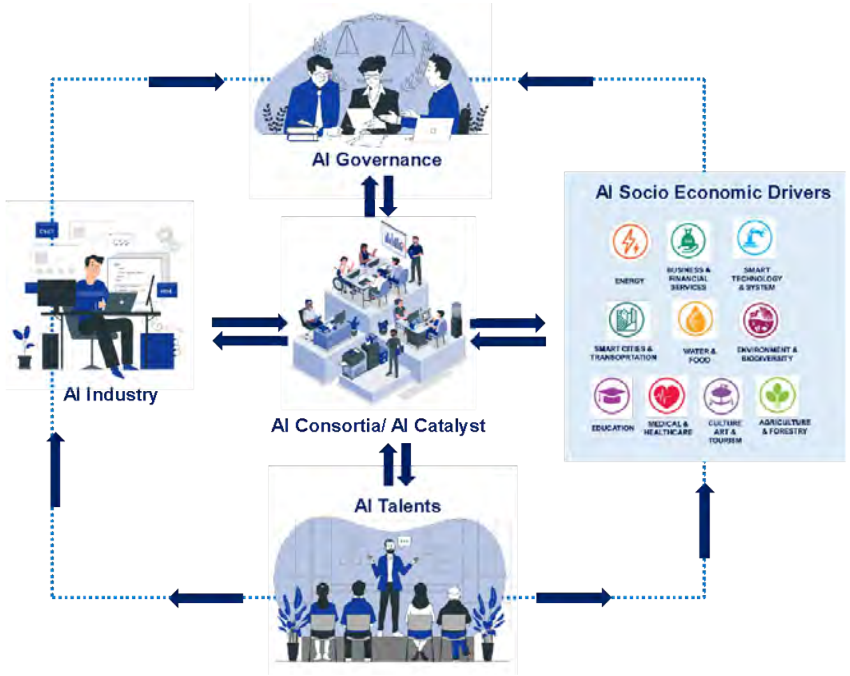


Figure 3: AI Innovation ecosystem

AI policy development requires synergy among multiple stakeholders to ensure comprehensive, inclusive, and effective frameworks. Identifying key stakeholders within the value chain is crucial for creating well-informed, balanced, and considering interests and concerns. This includes policymakers, data scientists, legal experts, IT professionals, and representatives from affected communities, ensuring a balanced approach to AI systems.

AI innovation ecosystem involves stakeholders, technologies, and organizations working together to advance AI. Trust is crucial for proper functioning, supported by policies and technical standards. Responsible AI is shared across government and stakeholders, ensuring a shared responsibility in the AI field.

AI-Catalyst: The AI Innovation Ecosystem, centered around the AI-Catalyst, comprises consortia tasked with tackling specific industrial or public sector AI challenges.

AI Governance: The AI ecosystem is undergoing significant transformation, fostering a dynamic and self-sustaining environment through strategic investments, supportive interventions, and good governance.

AI Industry: AI, a rapidly evolving technology, is being utilized by various sectors including MNCs, Malaysian companies, startups, SMEs, and National Research Institutes, enhancing their domain expertise.

AI Talent: AI and data science professionals require continuous training to enhance their skills. Tertiary institutions, schools, and training centers are crucial in sourcing expertise and providing training for applied and fundamental research and development activities in AI.

AI Socio Economic Drivers: For socio economic drivers under the 10-10 Malaysian Science, Technology, Innovation and Economy Framework see here: https://www.akademisains.gov.my/10-10-mystie

MOSTI empowering the synergy for Responsible AI Implementation and AI Governance Challenges

Synergy for Responsible AI Implementation

MINISTRY OF SCIENCE, TECHNOLOGY AND INNOVATION

National Guidance on Responsible AI:

- AI Roadmap: national strategy
- Overarching guidance: Responsible AI
- Alignment and coordinate with regional and international approaches

Vertical
E.gs. BNM, SC, MOHE, MoH, Ministry of Communication, Ministry of Digital

Ministerial Activities:

- Alignment with national strategy
- Incorporation of overarching guidance into sectoral/activity policy and regulation
- Alignment and coordination with regional and international sectoral approaches

Horizontal
E.gs:AG, JPDP, NACSA

Adherence to Sector/Activity: Specific Policy/Regulation

- Monitoring and enforcing compliance with sectoral/activity policy and regulation
- Where relevant, issuing approvals/ certifications related to activities within purview

Synergy to overcome AI Governance Challenges and Barriers

Awareness on AI & Ethics

Sustainability AI ethics Centre

Institutional support & regulations

Ownership & Data Sharing

Buy-in Stakeholder

Funding & infrastructure

Champion certified professional trained & skilled workforce

All findings were derived from total engagements through Town Hall, interview and survey.

PART B

1.3 RESPONSIBLE GOVERNANCE

As policymakers undertake the integration of Artificial Intelligence (AI) into governance structures, the role of responsible governance takes centre stage. Responsible governance, in the context of AI policy making, underscores the significance of ethical considerations, transparency, and accountability to ensure that the adoption of AI technologies aligns with societal values and does not inadvertently perpetuate biases or exacerbate inequalities.

Being a responsible governing body in AI policy making requires a commitment to fairness, transparency, and inclusivity. It involves actively mitigating the potential risks associated with AI, ensuring that its benefits are equitably distributed, and prioritizing the well-being of citizens. The National Guidelines of Responsible AI for the Government and Policy Makers serves as a compass for navigating these responsibilities, providing a framework to uphold ethical standards and foster a positive and inclusive AI-driven future.

AI governance helps to create a comprehensive framework to guides the development, deployment, and use of AI in various socio economic sectors. AI systems need to be safe and reliable, especially when they are used in critical areas like healthcare, transportation, or finance. Governance frameworks can set standards and national guidelines to ensure the robustness and reliability of AI systems, preventing accidents or malfunctions that could lead to significant harm.

As AI systems often handle vast amounts of data, including personal information, a responsible government will do its best to protect privacy and ensure data security. A responsible government will also establish regulations and standards for data handling, storage, and processing, preventing misuse and breaches. At the same time, it prevents monopolies, manages job displacement due to automation, and ensures that the economic benefits of AI are shared widely across sectors and society.

The integration of AI into policy making represents a transformative leap, revolutionizing the way policy makers analyze data, make informed decisions, and address complex societal challenges. As a powerful tool, AI offers unprecedented capabilities to enhance the efficiency, effectiveness, and responsiveness of policy formulation and implementation. The use of AI in policy making is multifaceted, touching upon various aspects of the decision-making process. While the integration of AI in policy making holds tremendous potential, it also raises ethical considerations to be addressed by a responsible governing. Striking a balance between leveraging AI's capabilities and addressing these challenges is crucial to ensure that the benefits of AI in policy making are equitably distributed and contribute to the overall well-being of society or the value chain it impacted.

1.4 THE PRACTICES OF AI GOVERNANCE

AI governance refers to a set of national guidelines, rules, and best practices that work together to limit bias risks and optimise intended advantages when it comes to the development and application of AI technologies that aim to comprehend, manage, and sculpt the responsible application of AI. The Global AI Initiative has suggested several approaches to promote responsible AI governance, and these include:

- Prioritizing a people-centered approach
- Leveraging AI for sustainable development and global challenges
- Respecting national sovereignty and laws
- Aligning AI development with values of peace, development, equity, justice, democracy, and freedom
- Combating the misuse of AI
- Adopting responsible practices in military applications
- Promoting mutual respect and benefit
- Implementing risk-based governance
- Enhancing explainability and data accuracy
- Ensuring privacy and security
- Upholding principles of fairness and non-discrimination.

AI Governance is a critical aspect of integrating AI into societal frameworks, with far-reaching implications for the implementation of the National Artificial Intelligence Roadmap (AI-RMAP) implementation. AI has the potential to affect many aspects of life, and therefore, without proper governance, there is a risk of its being used in ways that are unethical or harmful. AI Governance helps ensure that AI technology is developed and used in a manner consistent with societal values, human rights, and ethical principles. Proper governance helps in building public trust in AI technologies. When nations understand that there are regulations, standards, and ethical guidelines governing AI, they are more likely to trust and accept these technologies.

AI governance can guide the equitable distribution of economic and effective governance, and can also align the AI RMAP with international norms and standards, facilitating global cooperation and ensuring that our AI technologies are compatible and competitive internationally. In addition, clear and effective governance can stimulate innovation and investment in AI by providing a stable and predictable environment for researchers, developers, and businesses. AI governance may also help in maintaining accountability and transparency in the development and deployment of AI and in identifying and mitigating risks associated with AI, including biases, inequalities, and other societal impacts, thereby ensuring that the technology benefits community and society.

Fundamental Questions about AI Governance

- **What or who are you governing? (An organization , System of interest, type of AI, or Use Case for AI? Contextually and culturally bound)**
- **How are you governing it now? (current state)**
- **How do you want to be governing it (ideal / preferred state)**
- **Who is governing? (Roles & Responsibilities, Horizontal or Vertical)**
- **What does good governance look like? (Budget, Competence, Capability, Capacity, Expertise, Empowered)**
- **Purpose driven: What are you governing it for (safeguarding)?**
- **What incentivises the culture of governance?**
- **How can you tell if governance has been done well?**
- **What are the consequences for governance failures?**

B2.0 RESPONSIBLE AI FOR POLICY MAKERS

Translating and Implementing the 7 AI Principles of Responsible AI

Translating and implementing the principles of Responsible AI for policymakers involves developing a regulatory and ethical framework that guides the development, deployment, and use of AI technologies. The following are the 7 Principles of Responsible AI:

1. FAIRNESS

Develop policies that explicitly address fairness concerns in AI systems. This may involve creating guidelines for unbiased data collection, algorithmic decision-making, and regular audits to detect and rectify biases.

2. RELIABILITY, SAFETY AND CONTROL

Develop certification standards for AI systems to ensure their reliability. Policymakers can mandate adherence to these standards and establish a certification process to assess the resilience of AI technologies in various conditions.

3. PRIVACY AND SECURITY

Strengthen data protection laws that govern the ethical use of personal data in AI applications. This involves setting standards for obtaining informed consent, ensuring data security, and defining the permissible uses of personal information. AI system is a subset of computing systems. Therefore, it is subjected to all security and privacy standards and regulatory framework. However, the AI Principle of Privacy and Security should not be enacted under Responsible AI.

Translating and Implementing AI

4. INCLUSIVENESS

Integrate diversity and inclusivity requirements into AI policies. This could involve promoting diverse representation in AI development teams, considering the needs of underrepresented communities, and assessing the impact of AI on different demographic groups.

5. TRANSPARENCY

Enforce regulations that mandate transparency in AI systems. This includes requirements for organizations to provide clear explanations of how their AI algorithms work, disclose data sources, and communicate the impact of AI decisions on individuals or groups.

6. ACCOUNTABILITY

Establish a legal framework that assigns responsibility for AI systems. This could involve defining clear lines of accountability, specifying the roles of stakeholders, and implementing mechanisms for oversight and enforcement.

7. PURSUIT OF HUMAN BENEFITS AND HAPPINESS

Incorporate ethical guidelines into policies that prioritize societal benefit and sustainability. Policymakers can encourage the development of AI technologies that contribute positively to the nation and minimize negative impact.

2.1 DATA GOVERNANCE

In today's rapidly evolving digital landscape, data governance has emerged as a critical discipline for organizations worldwide. As data volumes grow exponentially, becomes paramount to ensure the availability, integrity, usability, and security of valuable data assets. The development of comprehensive and robust data governance framework, rooted on the globally recognized DAMA Data Management Body of Knowledge (DMBoK), has become essential reference to safeguarding and unlocking the full potential of valuable data assets.

Data governance encompasses the holistic management of an organization's data. It involves establishing policies, processes, and standards to ensure responsible data handling and in compliance with relevant acts and regulations. Effective data governance aims to maximize the value of data assets, minimize risks, and foster a data-driven culture.

Figure 4: DAMA DMBoK Wheel

DMBoK is a comprehensive framework that outlines best practices and guidelines for effectively managing data within organizations. It consists of 11 interconnected knowledge areas, each addressing a specific aspect of data management. Data governance frameworks typically have a wide reach and could vary based on organizational needs.

Data Liberalisation is a strategic imperative which ensures transparency and accessibility of data across the entire organization. It ensures data flow in the organization and not unencumbered by silos. Transparency becomes bedrock for stakeholders' access to relevant data seamlessly. Organization must confront the organization drag of data accessibility to improve timeliness and the accuracy of data for operational efficiency. Meanwhile, data liberalisation must not compromise data sensitivity but to harmonize it with policies and guidelines, ensuring compliance.

Figure 5: DIKW Pyramid

The relationship of data, information, knowledge, and wisdom (D-I-K-W) highlights its incremental value over its utilization. As data should be managed according to data governance, Knowledge Management governance focuses on ensuring that knowledge assets are effectively replenished, replicated, and utilized to achieve desired results within an organization. This ensures a holistic approach in managing knowledge assets to achieve desired results and drive organizational success. Proper knowledge management enhances decision-making by providing access to relevant information and expertise, fostering innovation and creativity through collaboration and idea-sharing. It provides a competitive advantage and essential for organizations to thrive in a dynamic and competitive landscape.

2.1.1 Data Ethics

As organizations increasingly adopt advanced technologies like Artificial Intelligence (AI), prioritizing ethical considerations in Data Governance practices becomes imperative. Data Ethics assumes a pivotal role in this context, emphasizing openness and accountability in data handling. Ethical data governance necessitates transparency regarding data collection, usage, and sharing within AI systems. Clear explanations of AI algorithms and decision-making processes bolster trust and empower individuals to assess the implications of AI applications on their lives.

Ethical data collection, processing, and decision-making should actively mitigate biases and uphold fairness in outcomes. Accountability in Data Ethics entails taking responsibility for the ethical implications of AI systems, facilitated by mechanisms such as establishing clear roles, conducting regular audits, and addressing ethical concerns promptly. By fostering accountability among individuals and organizations involved in AI-related decisions, Data Ethics promotes responsible and ethical AI development and deployment

Emphasizing reliability, safety, and control underscores the importance of responsible data handling practices to safeguard the integrity and safety of AI systems. Ethical data governance encompasses measures aimed at improving data accuracy, preventing data breaches, and empowering users with control over their data. Privacy and security efforts entail respecting individuals' privacy rights, obtaining informed consent for data collection, and implementing robust security measures to safeguard sensitive information, adhering to data protection laws and regulations to uphold privacy and security standards in AI applications.

.Integrating Data Ethics into data governance frameworks equips organizations with the tools and protocols necessary to identify and address potential risks stemming from unethical data practices. By incorporating ethical considerations into their data management strategies, organizations can proactively mitigate the chances of data misuse, breaches, or biases, thus safeguarding the integrity of their operations.

Moreover, this integration of Data Ethics serves as a cornerstone to fostering trust among stakeholders. By adhering to ethical principles and transparent data practices, organizations can instill confidence among stakeholders, including customers, employees, partners, and regulatory bodies. This trust is pivotal for maintaining positive relationships and credibility in the eyes of the broader community.

Furthermore, prioritizing Data Ethics underscores an organization's commitment to responsible AI development and deployment. By ensuring that AI technologies are developed, implemented, and utilized in alignment with ethical standards and societal values, organizations demonstrate their dedication to ethical conduct and social responsibility. This commitment not only mitigates potential harm from AI applications but also maximizes the positive impact these technologies can have on society.

Ultimately, organizations that prioritize Data Ethics are positioned as leaders in driving positive change and innovation. By leveraging AI technologies responsibly and ethically, these organizations contribute to sustainable development and societal progress. Their actions not only benefit their own operations but also set a precedent for ethical behaviour in the broader AI ecosystem, fostering a culture of responsible innovation and positive societal impact.

2.2 OPEN-SOURCE DATA SHARING

Open source data sharing refers to the practice of making datasets freely available to the public, allowing anyone to access, use, modify, and distribute the data without restrictions. This approach encourages collaboration, transparency, and the collective advancement of knowledge across various domains.

Open source datasets are often accompanied by documentation that provides information about the data's source, structure, and any relevant details. This transparency helps users understand the context and limitations of the data. While promoting openness, open source data sharing must also adhere to ethical principles. Sensitive information should be handled responsibly, and measures should be in place to protect individuals' privacy.

Open source data sharing is instrumental in addressing challenges, fostering innovation, and advancing research across diverse fields. It promotes a culture of openness and collaboration, empowering individuals and organizations to leverage shared resources for the greater benefit of society.

The practice of open-source sharing has a range of benefits for institutions and communities:

- Collaboration in collecting AI datasets leads to continuous improvement, forming an “AI commons” as individuals worldwide contribute to a shared pool of open-source training data and improve it over time
- Open-source AI allows for easier identification of biases and malfunction (algorithmic auditing), allowing for rectification before and after deployment. Hence, open AI systems promote transparency, accountability and therefore help to discover and mitigate biases and other risks.
- Open data minimizes legal liabilities associated with opaque data collection approaches such as web-scraping or unlawful discrimination.
- Regarding cybersecurity, open AI training data and AI models might enable more systematic and crowd-based detection of malicious action and prevention of tampering, and therefore, have the potential to be more secure.

Source: https://medium.com/@openforgood/open-source-ai-data-sharing-yes-data-colonialism-no-3062a922de03

Showcase:

OpenDOSM : The backbone of effective governance.

OpenDOSM is a platform for open data access on official statistics which also a showcase for catalogues, visuals, and analyses of DOSM's wealth of data. This platform provides access to machine-readable data format on a wide range of datasets, including social, economic & financial and environmental. OpenDOSM is an initiative taken by the department to expand the open data audience and promote open data usage among the public.

OpenDOSM Nextgen supports the government’s aspirations in the formation of data infrastructure which is one of the main elements of being data driven. Access to open data through this platform enables analytics and insights into new information produced by various parties. The platform also uses a centralized data pipeline system which ensures the data accessed is reliable, accurate and authentic.

2.3 AI REGULATIONS

In the era where fake news and propaganda has become a threats to the society, the introduction of such AI principles may provies means to overcome the misused of AI technology. Incidents of ethical misuse of AI has skyrocket due to the lack of regulations to govern its development and deployment. Concerning the misuse of AI technology, countries like the European Union and the United Kingdom have begun initiating regulatory frameworks for AI. The EU AI Act was voted in the Parliament in March 2024 will have a significant impact on the development of other AI-related laws outside the EU.

Malaysia has not enacted specific legislations on AI governance. However, all the adoption of AI must be ethically carried out that promotes accountability and transparency, protects data privacy and security, mitigates risks, and builds public trust parallel to Malaysia's existing laws.

Regulations are necessary to protect the stakeholders' rights, including matters pertaining to privacy, non-discrimination, and freedom of expression. AI has the potential to infringe on these rights if not properly regulated, especially in areas like surveillance, data handling, and biased decision-making. AI systems applications are increasingly used in critical areas such as medical and healthcare, smart manufacturing, transportation, smart cities and criminal justice.

Regulations are crucial to ensure these systems introduced and developed are safe, reliable, and accountable for decision making. Regulations can mandate a level of transparency and explainability, ensuring users understand how and why decisions are made, which is essential for trust and accountability.

Without regulation, there are risk of AI being used for harmful purposes, including creating deepfakes, automating cyber attacks, or enabling mass surveillance. Regulation helps to prevent such misuse. AI systems can unconsciously perpetuate or even intensify bias. Regulations can enforce fairness standards and require regular auditing to ensure AI does not discriminate against individuals or groups.

AI has the potential to significantly disrupt job markets through automation. Without regulations to guide this transition,it could lead to economic inequality and social unrest. If the government does not regulate AI while other countries do, it is difficult to compete at the international stage. Regulation may promote innovation by setting clear rules that encourage the development of ethical, safe, and beneficial AI. Without it, the market might prefer cheapest solutions, not the most ethical or beneficial ones. AI regulation is therefore important to ensure that the development and deployment of AI technologies is conducted ethically, safely, and beneficially. The absence of regulation may lead to a wide array of negative outcomes, including loss of public trust, unfair practices, threats to privacy, unsafe applications, economic disruption, and international isolation. Proper regulation may promote a healthy ecosystem in which AI benefits all sectors while minimizing or mitigating its risks.

Initiatives on the formulation of AI Act to support Responsible AI

USA

In 2024 will move closer to the establishment of comprehensive AI regulation and drafting potential acts namely the Algorithmic Accountability Act, National AI Commission Act, Digital Platform Commission Act, and Transparent Automated Governance Act.

European Union (EU)

EU have successfully endorsed a comprehensive regulatory framework for AI called EU AI Act. June 2024 Emphasis on High-Risk Systems to support Ethical AI. Introduced Data Governance Act, Digital Service Act and Digital Market Act.

China

In 2024, anticipates the formalization and enactment of AI regulation which provide a consolidated and overarching framework for governing various aspects of AI development and deployment within the country. Measure for the management of Generative AI Services

There are few countries that are in the process of drafting and finalizing AI Act and Framework and will introduce and implement them soon. Details of initiatives on the formulation of AI Acts with the intention to support Responsible AI are provided in Appendix 2.

Malaysia

Responding to the Global Calls

- National Roadmap including “AI for All” and a framework o the responsible use of AI.
- The Cyber Security Act 2024 has been officially gazetted on 26 June 2024. This legislation is a major milestone in strengthening Malaysia's cyber defence and enhancing our resilience against emerging threats.
- Currently review the Personal Data Protection Act 2010 to algin and support digitalisation strategic initiatives including AI technology.
- Mapping relevant acts to support the implementation of 7 AI Principles .
- Omnibus Bill, which deals with data sharing are being finalised. This bill would regulate data sharing among civil service agencies.

At the time of writing, there is no specific law that regulates AI. However, the Federal Constitution of Malaysia and all laws currently in force can be applied to govern and regulate AI. All aspects of the seven principles of Responsible AI in the National Guidelines must be aligned with the national principles contained in the spirit and provisions of the Federal Constitution and the five principles of Rukun Negara. Should any conflict arise between these principles and the Constitution, the latter shall always prevail and take precedence.

The following list identifies the existing Malaysian laws relevant to the adoption of AI principles in Malaysia.Although there is no specific law in AI responsible governance, the interoperability of existing provides general support in principle.

- Personal Data Protection Act 2010 (PDPA)
- Consumer Protection Act 1999 – unfair trade practices;
- Malaysian Code on Corporate Governance (MCCG)
- Industrial Relations Act 1967
- Malaysia Employment Act 1955 – Amend 2023
- Corruption Act 1997
- Digital Signature Act 1997- for electronic transactions.
- Consumer Protection Law 1999 [Act 599].
- Electronic Commerce Act 2006
- Computer Crimes Act 1997
- Communications and Multimedia Act 1998 (CMA)
- Woman and Girls Protection Act 1973
- Employment Information Act 1953
- Social Inclusion Act 2014
- Malaysian Anti-Corruption Commission Act 2009
- Electronic Government Activities Act 2007
- Adoption Act 1952
- Intellectual Property Corporation of Malaysia Act 2002,
- Telemedicine Act 1997
- Human Rights Commission of Malaysia Act 1999 (Act 97)
- Environmental Quality Act 1974

Cyber Security Malaysia (CSM)

CSM is aware of the potential for AI technologies such as deepfakes to spread misinformation and the effectiveness of AI phishing, or malicious Generative Pre-trained Transformers (GPT), in gathering dangerous information. Generative AI could potentially be used to disseminate false information, create deception and modify malicious code to evade cyber defences. CSM is proactively monitoring and investigating fraudulent activities facilitated by AI technologies. Programmes such as Jelajah Anti-Scam Kebangsaan, CyberSAFE Program (Cyber Security Awareness For Everyone) and the Cyber Security Awareness Talk (CSAT) have been launched to combat various AI-focused crimes.

It is to also to be noted that the list is not exhaustive and not exclusive specifically to AI. Please refer to Appendix 3 for more information on how the existing legal framework supports the Responsible AI principles.

Disclaimer: The Act list and the clauses referring to the Act quoted in this document shall not be treated as exclusive legal reference for interpretation.

While AI is often categorized as "technology," "innovation," or an "intellectual property process or product," its regulatory landscape in Malaysia remains under development.

However, under the Federal Constitution, being the supreme law of the land, AI as subject matter, falls under the purview and jurisdiction of the Federal Government as expounded in Section 8 of the Federal List. In the meantime, while waiting the Federal Government develops a specific AI law or amendment to existing statutory laws and regulations to address its unique characteristics and potential cross-border implications, all current statutory laws of Malaysia are deemed to be applicable in governing and regulating AI.

All aspects of the 7 principles of responsible AI, including definitions, interpretations, approaches, enactment, amendments and implementation must be aligned with the spirit and provisions of the Federal Constitution and the 5 principles of Rukun Negara. As explained under Section 2.6. should any conflict arise between these principles and the Constitution, the latter shall always prevail and take precedence.

To ensure comprehensive and consistent application, the operational definitions for the 7 responsible AI principles in Malaysia such as fair, transparent, in pursuit of happiness and others must be established, meticulously reviewed and compared the with existing definitions in the relevant Acts for a more seamless alignment with the objectives, goals, and legal framework.

While existing laws aim to protect human dignity from unwanted intrusion and infringement, treating individuals as moral subjects rather than mere objects, their application often falls short in the face of emerging artificial intelligence (AI). In the efforts to protect people and their rights in the age of AI, some adjustments on legal frameworks, such as data privacy and non-discrimination laws, are required to address the unique nature and applications of these technologies. This adjustment should be mindful of both the potential benefits AI can offer and the risks it poses, striving for a responsible balance that safeguards human dignity while fostering responsible technological advancement and socio-economic growth.

When statutory laws lay the cornerstone for AI governance, wielding the authoritative power of enforceability and sanctions, existing roadmaps, guidelines, or code of ethics shall continue to act as a complementary and vital layer, guiding and shaping the evolution of AI in a more flexible and adaptable manner. Their application is mainly voluntarily and based on the willing participation of stakeholders and interested parties.

It is best to have different and separate guidelines on AI for government, private sector and industry and the general public due to the different level of expectations, responsibilities and duties towards responsible AI.

Showcase

Personal Data : How are the Public Being Protected?

The Department of Personal Data Protection (JPDP) is an agency under the Ministry of Digital. This Department was established in May 16, 2011 after the Parliament passed the Personal Data Protection Bill 2009. The main responsibility of this Department is to enforce and regulate PDPA in Malaysia. PDPA focuses on the processing of personal data in commercial transactions and the misuse of personal data. In enforcing in the PDPA, this Department has been mandated to register all classes of Data Users under the Order of the Minister.

The Principles of Data Protection includes:

1. General : Personal data shall be adequate, relevant and not excessive. Processed with consent and for a lawful purpose
2. Notice and Choice:Information must be given concerning purposes for which the personal data is being processed, collected or disclosed
3. Disclosure: Disclosure without consent is not permissible
4. Security: Data must protected from loss, misuse, unauthorized access, etc.
5. Retention : Personal data shall not be kept longer than necessary
6. Data Integrity : Personal data shall be accurate, up-to-date, and verifiable
7. Access: Persons have the right to access their own personal data

Personal Data Protection Standard 2015

The Personal Data Protection Standard is a minimum requirement issued by the Commissioner, that provides for common and repeated use, rules, guidelines or characteristics for activities or their results. This standard applies to:

- Any person who processes; and
- Any person who has control over or authorizes the processing of, any personal data in respect of commercial transactions.

This is a minimum standard which includes three personal data protection principles, namely security, retention and data integrity.

2.4 AI PRINCIPLES AND ETHICAL STANDARDS

AI standards play a major role in design, manufacturing and business practices, where it can function as a quality management system for AI users, consumers, organizations, research institutions and governments by offering the ability to recognise and encourage ethical conduct through their decisions. The objectives of AI standards are to facilitate responsible AI development, widespread global adoption of AI technologies, and inform emerging regulatory and legal considerations on the use of AI.

Beyond conventional standards, co-regulatory approaches exist: accreditation systems, professional codes of ethics or standards for fundamental rights compliant design. Many standards exist and the examples can be used to ensure the responsible development, deployment, and use of AI.

Some work has currently been initiated at the global level for AI standards - ISO/IEC and IEEE. Although Market requirements commonly demand both IEEE and ISO/IEC standards, the ISO/IEC has had global success with many more standards than IEEE. The Malaysian Department of Standards (DSM) as a government agency is also a member of ISO, and so it has access to the ISO standards development process. Participation through DSM at ISO level means bringing and putting forward national stance in the ISO standards making work. The committee is developing the national ISO/IEC standards in AI based on the ISO/IEC JTC 1/SC 42. A list of standards can be found in Appendix 4.

As standards are required to ensure safety, quality and interoperability for AI applications, Malaysia is leveraging this platform to position Malaysia as one of the leaders in this region in promoting the adoption of AI. Participation in ISO/IEC JTC 1/SC 42 is imperative to safeguard national interests as well as to influence the standards developed. For an effective participation, DSM has established a National Mirror Committee (NMC), consisting of the relevant stakeholders including the Government, academia, industries and civil societies, was established in 2023 to mirror the work at the international level.

IEEE has established several standards that examine into Responsible Development, Deployment, and Use of AI such as:

- IEEE 7000-2021 Model Process for Addressing Ethical Concerns During System Design
- IEEE 7002 - 2022 Standard for Data Privacy Process

ISO also has established standards related to AI such as:

- ISO/IEC 42001:2023 AI Management system
- ISO/IEC 23053 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
- ISO/IEC 23894:2023 AI Guidance on risk management
- ISO/IEC 27001 Information security management systems
- ISO/IEC 27002 for privacy information management … and others

- In the early 2024, the government of Malaysia through DSM has established the AI Technical Committee comprised of 20 members from government agencies and industry.
- Current status in March 2024, of ISO/IEC JTC 1/SC 42:
  - 25 published ISO Standards
  - 32 ISO Standards under development
  - 40 Participating Members (including Malaysia)
  - 24 Observing Members

B3.0 WHAT YOU COULD DO AS POLICY MAKER

3.1 THE AI GOVERNANCE AND ETHICS (AIGE) FRAMEWORK FOR AI SYSTEMS

An ethical framework for AI serves this precise function by applying principles and values, whether based on ethical considerations or human rights, to assess and measure the way AI is to minimise the risks, harmful practices, and negative consequences for individuals and society. Identifying and mitigating these risks and harms in the AI's design and use in accordance with ethical considerations makes it possible to hold developers and deployers are held responsible and accountable for their unethical actions.

Ethical debates often predate legal regulatory initiatives, and the lifecycle of AI is no exception. Where the risks and challenges arising from the design and use of AI require managing, in that case, ethical frameworks can be introduced in place of or prior to legal regulation. Ethical frameworks promote trustworthy and responsible AI that can be developed and deployed to promote innovation and serve altruistic societal and economic development benefits. The accelerated use of AI, whilst yielding benefits, must be compatible with value-based principles within these governance frameworks.

The values in an ethical framework for Malaysia must ideally place national interests and policies as informing and directing the values to be included, adopting a model where the nation-state takes the lead drawing up a national framework with special consideration of any additional principle(s) unique to its national values and aspirations (Sadhu Singh and Segaran, 2021).

The AI Governance and Ethics (AIGE) Framework was developed to provide a structured approach to responsible AI development and use. It serves as a guide for stakeholders to navigate the complex ethical landscape of AI, promoting all 7 principles of Responsible AI in the development and deployment of AI technologies.

When developing a policy that utilizes AI, the AIGE Framework can be adopted as part of the thought process. Policy maker adopting the AIGE framework have to understand the AI Governance System and assess the landscape within their jurisdictions:

I. Engage a diverse set of stakeholders for valuable insights and define a set of ethical principles that align with the 7 AI principles
II. Develop clear and comprehensive performance measures and evaluation mechanisms
III. Foster an environment that supports the technical pillars while maintaining ethical standards
IV. Regularly review and adapt the framework to address evolving challenges.

AI Governance and Ethics (AIGE) Framework and Process Flow

For further explanation, can refer to Section 3, Part C 3.2.

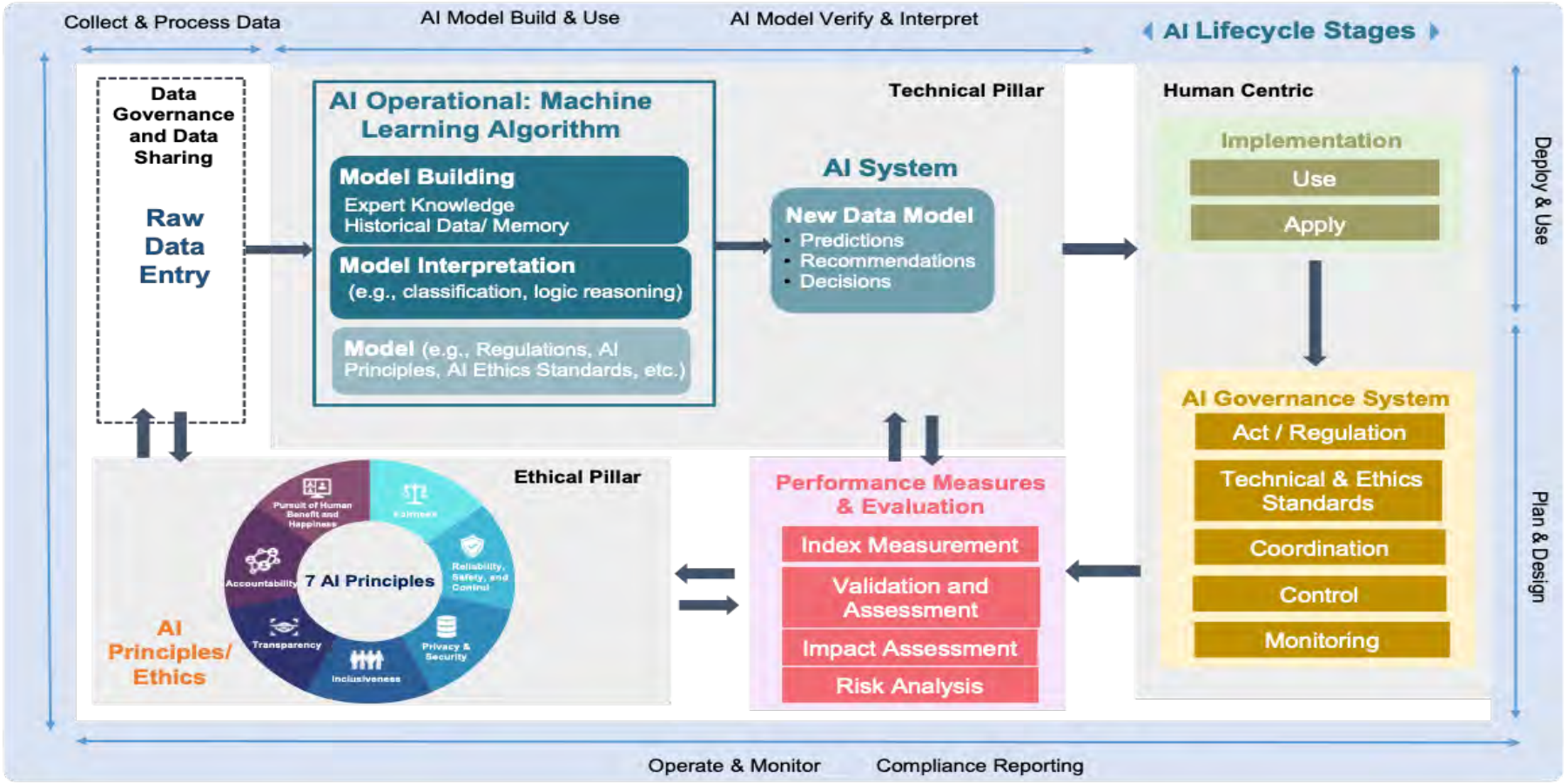


Figure 6: AIGE framework and process flow

PART B

3.2 EMBEDDING THE LIFE CYCLE IN HUMAN-CENTRED AI

Human-centered AI learns from human input and collaboration, focusing on algorithms that exist among a larger, human-based system. Human-centered AI is defined by systems that are continuously improving because of human input while providing an effective experience between human and robot. By developing machine intelligence with the goal of understanding human language, emotion and behavior, human-centered AI pushes the boundaries of previously limited artificial intelligence solutions to bridge the gap between machines and human beings.

To overcome prejudice, humans must be involved, and algorithms can contribute to prejudice. Without data and human empathy for the needs of others, approaches based entirely on machine learning will fall short of realising the full potential of artificial intelligence. The development of an artificial intelligence perspective on the optimal way to meet human and societal needs should be accelerated as much as possible to enable us to see the promised benefits.

Human-centered design is a method of problem-solving that places a premium on empathy with the human being at every stage of the process. Human-centered design is the ideal bridge between technological possibilities and the actual needs of users. Human-centered artificial intelligence entails prioritising the human and emphasising empathy as a core value. The fundamental tenet of human-centered AI design is not to create visually appealing products, but begins with the people's desirable human perspective and considers what human want and need.

Ethical AI design focuses on enhancing rather than replacing human capabilities. Ethical AI design also necessitates HCI design in order to ensure that human operators can take control of intelligent systems in the event of an emergency, and to avoid fatal accidents such as those mentioned above involving autonomous cars. Human-centred AI design encompasses more than just user interface design; and also considers the broader implications of AI, such as responsibility for error, ethics, bias, and governance.

As a policy maker, you may want to embed and emphasize the human-centric approach and thinking in the development process; embedding a human-centric approach throughout the life cycle of AI, which involves integrating ethical considerations, transparency, and inclusivity into the development, deployment, and use of AI technologies. You could advocate the integration of human-centric design principles in AI development. Encourage user-centred design processes that prioritize users' experience, inclusivity, and accessibility. In addition to this, you could promote empowerment for users by allowing individuals greater control over their data and the AI systems they interact with.Provide options for users to customize AI settings and make informed choices.

Artificial intelligence applications will be valuable only if they are designed to address the needs of human end users. When an AI solution works for its users, rather than forcing them to work with it, it becomes productive, successful, and responsible.

Source:
https://www.cognizant.com/us/en/glossary/human-centered-ai
https://www.linkedin.com/pulse/ai-requires-life-centred-design-sandeep-ozarde/

3.3 ESTABLISHMENT OF AI GOVERNANCE SYSTEM

Establishing an AI governance system involves developing a structured framework that guides the ethical, legal, and responsible use of artificial intelligence technologies within your value chain.

By establishing your own AI Governance System related to your policy, you will gain a thorough understanding of AI technologies, their potential applications, and the ethical considerations associated with their use. This will also help you to develop a more comprehensive policy that details how AI technologies should be developed, deployed, and used within your value chain and include specific guidelines on data handling, algorithmic decision-making, transparency, bias mitigation, and accountability.

However, keep in mind that the governance system needs to implement a feedback loop and mechanisms for continuous improvement. Regularly reassess and adapt your AI governance system based on the evolving landscape of AI technologies, emerging ethical considerations, and lessons learned from previous projects.

Act / Regulation

Ensure that your AI governance system aligns with existing legal frameworks and regulations. Address issues related to the 7 principles that may apply to AI applications in your jurisdiction.

Technical & Ethics Standards

Refine a set of 7 AI principles that will serve as the foundation for your policy. The standard should align with the 7 principles and the policy expectations.

Coordination

Clearly define roles and responsibilities for all stakeholders involved in the AI life cycle. Specify who is accountable for ethical considerations, compliance with policies, and oversight of AI projects that relate to the policy.

Control

Implement oversight mechanisms to ensure that AI systems adhere to established policies and ethical principles. This may involve regular audits, reviews, and assessments of AI projects to identify and address any ethical or legal concerns.

Monitoring

Integrate monitoring and evaluation processes into the AI governance system. Establish key performance indicators (KPIs) and metrics to assess the performance, impact, and ethical implications of AI systems throughout their life cycle.

3.4 MEASURE AI PERFORMANCE AND EVALUATION

Measuring AI performance and evaluation as a policy maker is crucial for policy makers to ensure that AI systems align with ethical standards, legal requirements, and the intended goals of their deployment.

By systematically measuring and evaluating AI performance using these principles, policy makers can ensure that AI technologies align with ethical considerations, provide value to users, and contribute positively to society. Regular assessments and adjustments based on these evaluations are essential for responsible AI governance.

In order to comply with Responsible AI requirements, all AI technologies should be subject to performance measures and evaluation. Relevant index measurements for each AI system need to be developed for validation and assessment purposes, and subsequently to conduct impact assessment and risk analysis to determine the risk level for each AI system.

Showcase: Microsoft: Responsible AI Toolbox

Responsible AI is an approach to assessing, developing, and deploying AI systems in a safe, trustworthy and ethical manner. The Responsible AI toolbox is a collection of integrated tools and functionalities to help operationalize Responsible AI in practice. The components of the Responsible AI Toolbox include

Model Statistics - A high-level view of model prediction distribution for AI performance. Analysis along a variety of metrics can help characterize errors more specifically.

Data Explorer - Determine whether your findings related to fairness, error analysis and causality are a result of your dataset's distribution. This will help to mitigate errors arising from representation issues, label noise, feature noise, label bias, etc.

Interpretability - The interpretability component generates human-understandable descriptions of the predictions of a Machine Learning model. This will help regulatory audit of an AI system to validate models and monitor the impact of model decisions on humans.

Causal Inference - estimates the effect of a feature on an outcome of interest on average, across a population or a cohort and on an individual level which allow decision makers to apply new policies and effect real-world change.

https://responsibleaitoolbox.ai/introducing-responsible-ai-dashboard/#dashboard-components

3.5 MANAGE RISK AND IMPACT

Managing AI risk and impact from a policy maker's perspective involves developing and implementing policies and regulations that mitigate potential risks, ensure ethical use, and maximize positive outcomes.

AI systems are different from legacy technologies and may pose unfamiliar risks. A robust risk management approach should be taken at every stage of the model lifecycle. This helps build trust and towards the acceptance and greater use of AI technologies in the region, gauge the effectiveness of the proposed AI governance system and its implementation as well as the impact of responsible AI on stakeholders.

AI poses key potential risks that can impact organizations, consumers, and create broader detrimental effects for society. Such potential risks may arise in whole or in part from sources including the data used to train the AI system; potential risks arising from the AI system itself; potential risks arising from the usage of the AI system; and potential risks arising from poor overall governance of the AI system

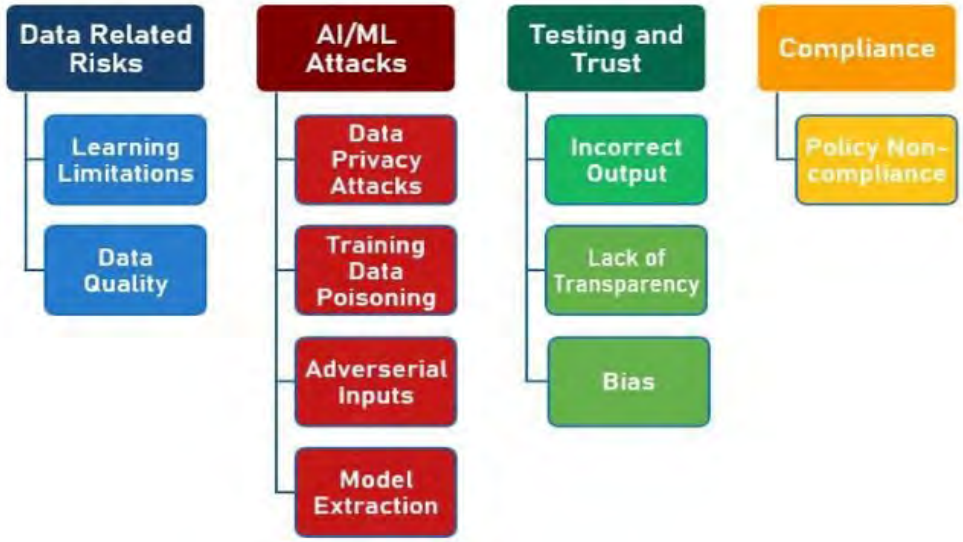


Figure 7: AI risk categorization defined by Artificial Intelligence/Machine Learning Risk and Security Working Group (AIRS, 2023)

While much of the risk is associated on the technological side of AI, it is also important to bear in mind that the risk and impact of AI on Regional and Global security, safety and peace also play part in the Responsible use of AI by every Nation State like Malaysia.

The risk of like miscalculation of AI forecasting, escalation of conflict from disinformation by AI and proliferation of new kind of weapon (be it conventional or cyber) by the adoption of AI technology.

The EU Artificial Intelligence Act (AIA) however was classified AI risk based on its potential harm, namely unacceptable, high, limited, and minimal (Figure 4). The definition of AI risk (European Parliament, 2023) is as follows:

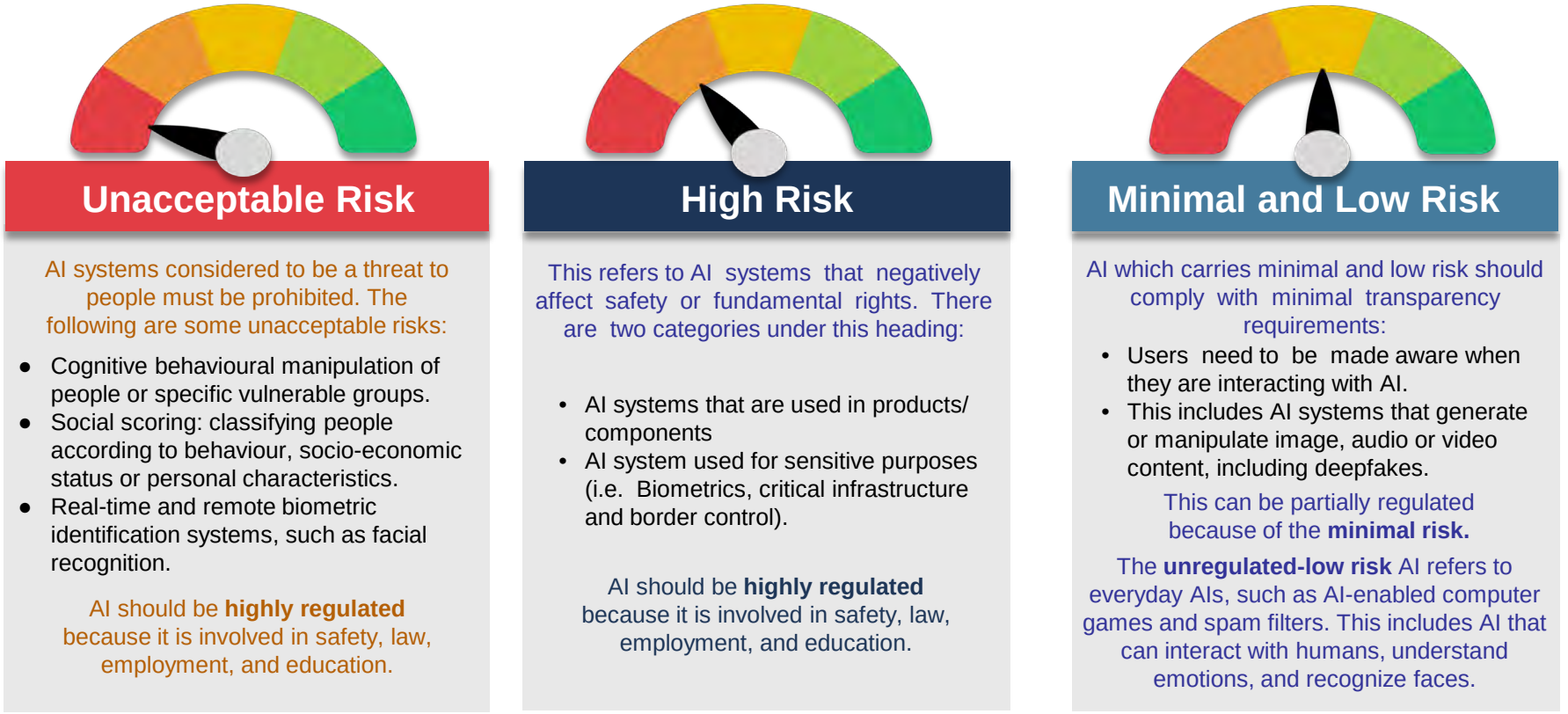


Figure 8: AI risk categorization based on EU Artificial Intelligence Act (AIA)

Source: The EU Artificial Intelligence Act definition on AI risk (European Parliament, 2023)

3.6 SELF-DEVELOPMENT OF RESPONSIBLE AI CHECKLIST: EXAMPLE

It is understandable that each policy maker's space and scope is unique. There is no specific formula that can cater for the stakeholders that you engage with. However, with the understanding of the AI Governance and Ethics Framework,including the life cycle and each component with its risk management, you could develop your own checklist to cater for your own specific needs.

Table 1: Example of a Policy Development Checklist related to AI in view with Cyber Security measures

Principles	Description	✓
Fairness	• Require audits of AI training datasets to ensure they are fair, unbiased, and representative. • Implement algorithmic auditing procedures to regularly check AI systems for signs of unfair bias or discrimination. • Set up whistleblower protections and reporting channels for those who identify problems.	
Reliability, Safety & Control	• Set cybersecurity requirements like encryption, access controls, and vulnerability testing for high-risk AI systems. • Establish regulatory bodies to actively monitor the cybersecurity of critical infrastructure AI. • Require contingency plans and failsafes in case AI systems are compromised.	
Privacy & Security	• Set cybersecurity requirements like encryption, access controls, and vulnerability testing for high-risk AI systems. • Establish regulatory bodies to actively monitor the cybersecurity of critical infrastructure AI. • Require contingency plans and failsafes in case AI systems are compromised.	
Inclusiveness	• Fund education programmes on AI and cybersecurity skills for underrepresented communities. • Subsidise access to AI technologies for disadvantaged demographic groups and regions.	
Transparency	• Mandate documentation about an AI system's architecture, data sources, security protections etc. • Set up frameworks for external cybersecurity researchers to conduct audits of an AI system's codebase as needed.	
Accountability	• Require detailed cybersecurity audit logs showing who have accessed AI systems and what changes have been made. • Implement chain-of-custody cybersecurity tracking regarding AI data usage and algorithmic development.	
The Pursuit of Human benefit and Happiness	• Require the consideration of the appropriate level of human control for the particular AI system to ensure human benefits and happiness • Require the formation of a self-audit, response mechanism and an autonomous system to determine human rights and well being	

B4.0 THE INDEPENDENT ADVISORY BODY

An Independent Advisory Agency is essential to ensure that AI technology is developed and implemented in a way that maximises benefits while minimising risks, ensuring ethical compliance, and fostering public trust and confidence inAI systems. A feasibility study needs to be carried out in order to determine the requirements for the establishment of an independent national Advisory Agency.

This agency can play many roles and needs to work closely together with the Hi-Tech National Council and the Fourth Industrial Revolution (4IR) Council and relevant ministries to implement effective Responsible AI. The entity would also advise on the use of AI technologies to increase Malaysia's competitiveness. The agency may provide expert advice to government policymakers on AI-related issues, helping to shape informed and effective AI policies and regulations.

As an independent body, the agency could provide unbiased recommendations, free from commercial or political interests. This ensures that AI policies and practices would be aligned with the public interest rather than being driven by specific corporate agendas or political biases. The agency coul monitor AI developments, ensuring compliance with ethical standards and legal requirements, and could intervene when necessary to protect public interests.

By ensuring that AI is developed and used responsibly, the agency may help build public trust in AI technologies, which is essential for their widespread acceptance and adoption. The agency could establish ethical guidelines for AI development and use, ensuring that AI systems are designed and deployed in a manner that respects human rights, privacy, and democratic values.

This independent agency may play a vital role in ensuring that AI technologies adhere to AI principles to be adopted by relevant stakeholders. By providing clear guidelines and standards, the agency could help to balance the need for innovation in AI with the need for responsible development and use. It can also act as a mediator between the public, the government, and AI developers, addressing concerns and queries about AI and its impact on society.

With rapid advancements in AI, the agency could help ensure that policies remain relevant and effective in the face of evolving technologies. It can set standards for transparency, accountability, and fairness in AI systems, promoting trust and confidence in AI technologies among the public and stakeholders. At the same time the agency could collaborate with international bodies to ensure that the country's AI policies and practices are in line with global standards and best practices. The agency could represent the country in global forums, ensuring that its interests and values are reflected in international AI governance frameworks.

The agency also could support research in ethical AI, bias mitigation, and other areas crucial for the responsible advancement of AI technology. In case of AI-related incidents or ethical breaches, the agency could provide guidance on managing and resolving such situations. Subsequently, playing a key role in identifying potential risks and unintended consequences of AI, and in developing strategies to mitigate these risks

B5.0 ALIGNING GOVERNMENT INITIATIVE WITH RESPONSIBLE AI AND SDG & ESG PRINCIPLES

5.1 SUSTAINABILITY PILLARS IN POLICY MAKING

Sustainability in AI policy making involves balancing the three pillars of sustainability - economic, social, and environmental—while embracing responsible principles for AI development and deployment. This will ensure that AI technologies contribute to the long-term, maximise the well-being and prosperity of society while minimising adverse impacts and risks.

Sustainability focuses on meeting the needs of the present without compromising the ability of future generations to meet their needs. The concept of sustainability is composed of three pillars: economic, environmental, and social.

For policymakers developing policies related to AI, sustainability encompasses several key dimensions which include:

Environmental Impact: Taking account of the environmental sustainability of AI technologies involves assessing the energy consumption and carbon footprint associated with AI hardware infrastructure, data centres, and computing resources. Sustainable AI policies may encourage the development and adoption of energy-efficient algorithms, hardware, and practices to mitigate environmental impact.

Social and Economic Impact: Focusing on the social and economic sustainability of AI involves addressing potential societal system disruptions as well as its improvement. Sustainable AI policies aim to promote inclusive growth, create new job opportunities, and support workers in transitioning to new roles through education, training, and re-skilling programmes.

PART C

DESIGNERS, DEVELOPERS, TECHNOLOGY PROVIDERS, & SUPPLIERS

C1.0	Introduction	72
C2.0	Responsible Principle for Sector Player	74
C3.0	What You Can Do as Sector Player	78
C4.0	Demonstration and Sectors Cases	93
C5.0	Code of Ethics	102
C6.0	Aligning Industrial Economic Contributions with Responsible AI and ESG Principle	103

C1.0 INTRODUCTION

1.1 INTRODUCTION TO THE NATIONAL GUIDELINES AIGE FOR DEVELOPERS, DESIGNERS, TECHNOLOGY PROVIDERS & SUPPLIERS

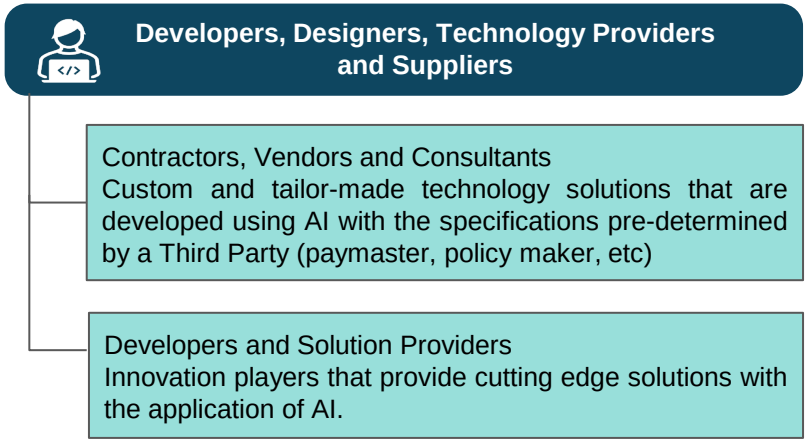
As artificial intelligence (AI) continues to advance and spread in various sectors, it is crucial to ensure that its adoption is guided by responsible principles and ethical considerations. However, the adoption of AI needs a robust governance framework to ensure responsible and ethical deployment. The AI governance frameworks rely on three key components: regulations, standards, and ethics. These components are important for responsible and ethical AI development, deployment, and use, helping to ensure that AI technologies are aligned with societal values and contribute positively to the well-being of individuals and communities.

The National AIGE Guidelines for Designers, Developers, Technology Providers and Suppliers are developed specifically for sector players, including organizations and businesses that play a vital role in the adoption of responsible AI. These National Guidelines include putting policies in place to remove bias from AI systems, protecting data privacy, and considering the social and environmental effects of AI applications. Sector players also should actively contribute to the development and maintenance of open datasets, fostering collaboration, innovation, and ensuring privacy and security safeguards. Open data sharing facilitates transparency, reproducibility, and accountability in AI research and development.

In order to evaluate the ethical and responsible use of AI, establishing a performance index (an example can be refer at Appendix 9) is of utmost importance. This index can measure adherence to all seven principles for decision-making algorithms in the AI systems. By continuously monitoring and improving performance, industry players can promote responsible AI practices.

In addition, effective risk management is a crucial aspect of responsible AI adoption. It involves identifying, testing, and assessing potential risks associated with AI systems. Proactive identification and assessment are necessary to address unintended consequences, cybersecurity threats, and bias. Thorough testing and evaluation procedures will ensure the reliability, safety, and ethical use of AI technologies, while certification is required to provide additional assurance and accountability.

These national guidelines are intended to cater for Designer, Developers, Technology Providers and Suppliers, which can be categorized as follows:



1.2 CHALLENGES AND BARRIERS

1 Lack of infrastructure

To fully leverage AI, a strong infrastructure is required, including high-speed internet and advanced computing resources. However, in some parts of Malaysia, there is inadequate infrastructure, which hampers the efficient implementation of AI technologies. In addition, there may be a scarcity of available data, making it difficult to train AI models effectively and limiting the potential benefits of AI.

2 Financial constraints

Developing and implementing AI technologies can be costly, particularly for small and medium-sized enterprises (SMEs). The costs associated with acquiring hardware, software, and skilled personnel can pose a financial barrier to AI adoption. Furthermore, the reliance on imported AI technologies due to limited local capabilities can further strain financial resources and hinder the wider adoption of AI in Malaysia.

3 Human capital and AI skills gaps

The successful integration of AI requires a skilled workforce with expertise in AI development, deployment, and maintenance. However, there is a shortage of professionals with AI skills in Malaysia. Bridging the human capital gap through training programmes and educational initiatives is essential to equip individuals with the necessary knowledge and skills to work with AI technologies effectively.

4 No AI regulatory framework

Malaysia currently lacks a comprehensive regulatory framework specific to AI. The absence of clear guidelines and regulations raises concerns about accountability, transparency, and the ethical use of AI. Establishing an AI regulatory framework is vital to address potential risks and ensure responsible AI adoption across sectors in Malaysia.

5 Limited trust and transparency

Building trust in AI systems is crucial for their widespread acceptance and adoption. However, limited transparency in AI decision-making processes can undermine trust. It is essential to promote transparency and explainability in AI systems, enabling users and stakeholders to understand how AI arrives at its conclusions and ensuring fairness and reliability.

6 Data sharing concerns

AI relies on access to data for training and improvement. However, concerns regarding privacy and data quality can hinder data sharing. In Malaysia, there is a need to strike a balance between protecting individual privacy and enabling responsible data sharing for AI development. Implementing robust data protection measures and frameworks can help alleviate concerns and promote responsible data practices

AI UNTUK RAKYAT

Challenges and barriers also have been highlighted in the AI for Rakyat (https://ai.gov.my/#/home)

C2.0 RESPONSIBLE PRINCIPLES FOR SECTOR PLAYERS

2.1 RESPONSIBLE PRINCIPLES AND ETHICS

As artificial intelligence (AI) continues to revolutionize various sectors, it is essential for sector players to embrace responsible AI practices in alignment with the seven AI principles and ethics in order to demonstrate their commitment to ethical AI practices and build public trust in their products and services. This, in turn, can lead to positive outcomes, enhanced consumer loyalty, and contribute to the responsible development and deployment of AI technologies. This can also be implemented even when a sector player is only adhering to contract and specification by the hiring party.

1. FAIRNESS

Ensure that AI systems are unbiased and treat all individuals fairly, regardless of their race, gender, ethnicity, or other characteristics. This involves considering diverse perspectives during the design phase and implementing measures to mitigate bias in data and algorithms.

2. RELIABILITY, SAFETY AND CONTROL

Prioritize the reliability and safety of AI systems, ensuring they perform as intended and avoid harmful outcomes, This includes rigorous testing, monitoring, and implementing fail-safe mechanisms to prevent unintended consequences. Also providing users with control over AI systems enables them to intervene when necessary and mitigate potential risks.

3. PRIVACY AND SECURITY

Uphold the privacy rights of individuals and safeguard sensitive data collected or processed by AI systems; implementing robust security measures and adhering to relevant privacy regulations to protect user data from unauthorized access or misuse.

Designers, Developers, Technology Providers & Suppliers

4. INCLUSIVENESS

Prioritize accessibility features and consider the diverse needs and perspectives of all users throughout the design and development process.

5. TRANSPARENCY

Provide clear and understandable explanations of how AI systems work, including their capabilities, limitations, and potential impacts on users and society. This involves transparent communication about data usage, algorithmic decision-making processes, and potential biases.

6. ACCOUNTABILITY

Take responsibility for the outcomes of AI systems and be accountable for any unintended consequences or harms they may cause. This includes establishing mechanisms for redress and ensuring that users have avenues for addressing grievances or seeking recourse in case of adverse outcomes.

7. PURSUIT OF HUMAN BENEFIT AND HAPPINESS

Aligning AI objectives with societal values and prioritizing applications that have positive impacts on individuals and communities.

2.2 RESPONSIBLE AI ALGORITHM DEVELOPMENT

The AI algorithm is the core component of an AI system, and responsible AI algorithm development is essential to uphold ethical standards, ensure fairness, and mitigate potential biases. There are two main components of AI algorithm development: model building and model interpretation. model building often uses historical data/ memory/ and/or expert knowledge by humans and/or by automated tools to aggregate data automatically into the algorithm.

Model building involves the process of designing and training the AI model. It often relies on historical data, human expertise, and automated tools to aggregate and process data. During this process, it is important to consider the objectives of the AI system, such as the desired output variables, and the performance measures, such as accuracy and representativeness of the dataset.

Model interpretation is the process by which humans and/or automated tools derive outcomes from the AI model, such as recommendations, predictions, or decisions. In some cases, a model may offer a single recommendation based on deterministic rules, while in other cases, such as probabilistic models, a range of recommendations may be provided along with associated performance measures like confidence levels or risk assessments.

Responsible AI during algorithm development is crucial for sector players to uphold ethical standards, ensure fairness, and mitigate potential biases. By prioritizing data collection, transparency, ethical considerations, validation, users feedback, and ongoing monitoring, sector players can contribute to the development of AI algorithms that are trustworthy, unbiased, and aligned with societal values.

Responsible AI algorithm development involves several key practices:

Data Collection: Ensuring that the data used to train AI models is diverse, representative, and free from biases. This includes considering potential sources of bias in the data and taking steps to mitigate them.

Transparency: Providing explanations and justifications for the decisions made by AI algorithms. This includes making the decision-making process transparent and understandable to users.

Ethical Considerations: Incorporating ethical considerations into the design and development of AI algorithms. This involves identifying potential ethical challenges and addressing them proactively.

Validation: Conducting rigorous testing and validation to ensure that the AI algorithm performs as intended and is reliable and accurate.

Users Feedback: Incorporating users feedback into the algorithm development process. This includes actively seeking input from users and incorporating their perspectives and needs into the design of the AI system.

Ongoing Monitoring: Continuously monitoring the performance and impact of AI algorithms to identify and address any biases, risks, or unintended consequences that may arise.

2.3 THE ESTABLISHMENT OF AN AI GOVERNANCE SYSTEM

All sectoral players are encouraged to practice the highest ethical standards when designing, developing and implementing AI. This may reduce the risk of negative impact on those affected by AI adoption. Establishing an AI governance system involves implementing structures, processes, and guidelines to ensure that AI technology is developed, deployed, and used responsibly and ethically. This governance system helps designers, developers, providers, and suppliers navigate the complex ethical, legal, and social implications of AI technology.

An effective AI governance system encompasses a combination of policy, technical and ethical standards, coordination, control, and monitoring mechanisms. By integrating these component in AI governance system can promote responsible, trustworthy, and beneficial AI deployment while mitigating potential risks and safeguarding societal interests.

AI governance system enables the sector players to navigate the complex landscape of AI technology responsibly and ethically, while promoting trust, transparency, and accountability in AI systems and applications.

Technical and Ethical Standards: Technical standards play a crucial role in ensuring interoperability, reliability, and safety of AI systems. These standards define technical specifications, protocols, and best practices for AI development, deployment, and operation. It helps in minimizing risks. Ethical standards are another vital component, guiding AI actors to adhere to 7 AI principle and norms during the entire AI lifecycle.

Internal Policy and Guidelines: Establishing policies and adhering to relevant regulations governing the development, deployment, and use of AI technology includes compliance with data protection laws, industry standards, and guidelines issued by regulatory bodies.
Developing and adhering to ethical guidelines that outline principles for responsible AI development and use may cover issues such as fairness, transparency, accountability, privacy, and human rights. These national guidelines provide for data privacy, algorithmic transparency, fairness, and other ethical considerations.

Coordination, Control and Monitoring: Coordination mechanisms bring together various stakeholders, especially the potential users (or paymasters), facilitating their active engagement in shaping AI development. Control mechanisms will ensure compliance with regulations and standards, monitoring the implementation and use of AI technologies as well as assess the adherence to the standards. Monitoring mechanisms, such as reporting frameworks and evaluation processes, continuously assess the societal impact and performance of AI systems. It provides feedback loops to identify risks, biases, or unintended consequences, enabling corrective actions and continuous improvement.

2.4 DATA SHARING

Data sharing plays a crucial role in the responsible development and deployment of AI systems. It enables the availability of diverse and representative datasets, fosters collaboration and innovation, and ensures transparency and accountability.

Data sharing can provide access to larger and more diverse datasets, which are essential for training AI models. By combining data from different sources, AI systems can learn from a broader range of examples and improve their accuracy and performance. Data sharing also helps mitigate biases that may be present in individual datasets, as diverse data sources can provide a more balanced representation of the population.

Collaboration and innovation are fostered through data sharing. When organizations share data, it allows researchers, developers, and practitioners to work together, exchange knowledge, and build upon each other's work. This collaborative environment promotes the development of more advanced and effective AI algorithms and applications.

Include information related to the lineage of datasets used, model training and selection process and expected behaviour of the AI solution to help organizations that do not develop AI models in-house to be able to deploy AI models, and the AI systems should still be set up and comply with AI Governance guidelines

Transparency and accountability are enhanced through data sharing. When organizations share data, they contribute to the transparency of AI systems by enabling independent validation and scrutiny of the algorithms and models used. This transparency helps build trust and confidence among users and stakeholders. Additionally, data sharing enables accountability by allowing for external audits and assessments of the data used and the decision-making processes employed.

Data sharing must be done responsibly and in compliance with privacy and security regulations. Appropriate measures must be in place to protect sensitive and personal information. Anonymization and data de-identification techniques can be employed to preserve privacy while enabling data sharing.

"Establish data-driven policy development and improve the data sharing environment to ensure data quality"

Outcome

- Establishment of open data guidelines to produce data with transparency, integrity, and accountability
- Improvement of Malaysia's position in open data global rankings

https://amcham.com.my/wp-content/uploads/DEB_ENG_FINAL.pdf

C3.0 WHAT YOU CAN DO AS SECTOR PLAYERS

3.1 ADVOCATE FOR RESPONSIBLE AI ACROSS THE VALUE CHAIN

I am just a contractor or vendor to develop this system that includes the use of AI. Everything has to follow the contract, which does not require me to adopt Responsible AI principles. How can I do this?

Advocating for responsible AI practices when designers, developers, technology providers, and suppliers are primarily focused on fulfilling specifications, contracts, and terms set by the paymaster (either Business-to-Business (B2B) or Business-to-Government (B2G) relationship) can be challenging but not impossible.

While it may require persistence and concerted effort, advocating for responsible AI practices within downstream (B2B or B2G) organizations that obligated via contract and financial specifications is crucial for promoting ethical behavior and mitigating potential risks associated with AI technologies. By building awareness, fostering dialogue, and leading by example, stakeholders can drive positive change and contribute to the responsible development and deployment of AI solutions.

Educate stakeholders about the importance of responsible AI practices and the potential risks of disregarding ethical considerations. Raise awareness about the broader societal impacts of AI technologies and the value of incorporating responsible AI principles into the development process. You may also foster dialogue and engagement with stakeholders across the organization, including executives, project managers, and procurement teams. Encourage open discussions about ethical considerations, risk management, and opportunities for promoting responsible AI within the organization.

Incorporating Responsible AI into Contracts: Advocate for the inclusion of clauses related to responsible AI practices in contracts and agreements with paymasters. Negotiate terms that require adherence to ethical guidelines, transparency requirements, and accountability mechanisms throughout the project lifecycle.

You can also practice internally within your organization on your own to demonstrate the benefits of responsible AI practices through concrete examples and case studies. Highlight successful implementations of responsible AI that have resulted in positive outcomes for both businesses and society, including those shared in demonstration and sector cases at section C4.0. Embrace a culture of continuous improvement and adaptation within the organization. Encourage teams to regularly review and refine processes, policies, and practices to ensure they align with evolving ethical standards and regulatory requirements.

3.2 ADOPT NATIONAL GUIDELINES AIGE FRAMEWORK AND PROCESS FLOW

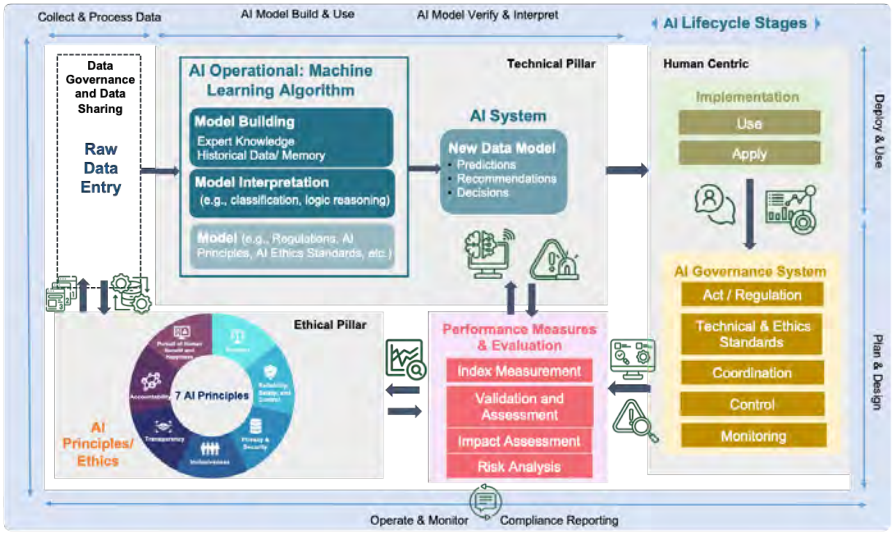


Figure 9: Responsible AI – AI Governance and Ethics Framework

Designers, developers, technology providers, and suppliers can adopt an AI governance and ethics framework to promote ethical and responsible AI development, deployment, and ongoing management. This process helps mitigate risks, enhance transparency, and ensure that AI systems contribute positively to society.

Figure 9 illustrates the formulation process of a responsible AI governance and ethics framework, specifically tailored to facilitate the implementation of responsible AI in Malaysia. The development of AI system, basically started with raw data sources that can included structured and unstructured types.

It is crucial for you to adopt AI principles at this stage [Raw Data Entry] to ensure the quality of the data input and output of the AI product.

The information generated from the data governance and data sharing step is fed into the AI operational machine learning algorithm. It is crucial to ensure that the algorithms and models used are transparent, explainable, and free from bias. Designers, developers, technology providers, and suppliers should prioritize the use of ethical AI techniques and adhere to established guidelines for algorithmic fairness and transparency.

3.2 ADOPT NATIONAL GUIDELINES AIGE FRAMEWORK

The governance process for an AI system involves several key steps to ensure ethical and responsible AI development and deployment:.

Establish Data Governance System

Gather relevant data from various sources, ensuring that it is representative and unbiased.
Data Annotation: Annotate and label data accurately, clearly defining what each data point represents. Implement data privacy and security measures to protect sensitive information and ensure compliance with privacy regulations.

Choose an appropriate machine learning model based on the nature of the problem and available data. Train the model using the collected and annotated data. Validate the model's performance on a separate dataset to assess its generalization capabilities and mitigate overfitting. Deploy the trained model in the intended application or system.

Users Interaction where to ensure that users understand the system's capabilities, limitations, and potential biases. Encourage users to provide feedback and report issues related to system behavior. Also need to adhere to 7 AI principle and guidelines, such as fairness, transparency, accountability, and non-discrimination. Also to comply with applicable laws and regulations governing AI use in the specific domain or region.

To **continuously monitor the AI system's performance and behavior** in real-time.
Detect and address data drift or changes in data distribution. Monitor for biases in model predictions and correct them as needed. Subsequently implement a controlled process for updating the AI model, ensuring thorough testing and validation before deployment.

Emergency Shut-off

Design an emergency shut-off mechanism to disable the system in case of unexpected behavior or ethical concerns.

Develop a **performance measurement index** that includes metrics related to ethics and responsible AI, such as fairness, bias, explainability, and accountability. Regularly evaluate the AI system's performance against these ethical metrics and track improvements or issues over time. Conduct regular assessments to validate the AI system's behaviour against the 7 AI principle and objectives. Assess the system's impact on stakeholders, including any unintended consequences or ethical concerns.

Perform a **comprehensive risk analysis** to identify potential risks associated with the AI system's use. Mitigate identified risks through appropriate measures, such as bias reduction techniques, transparency enhancements, or policy adjustments.

Establish a **feedback loop** that integrate users feedback, monitoring results, and risk analysis findings into the AI governance process. Continuously refine and improve the AI system based on feedback and evolving ethical considerations.

Lastly **maintain comprehensive documentation** of the AI governance process, including data sources, model architecture, ethical considerations, and actions taken.

Report on AI system behavior, performance, and ethical standards adherence to relevant stakeholders, including users, regulators, and the public. Adopt the AI governance process as needed to address emerging ethical challenges, technological advancements, and changes in regulations.

3.3 EMBEDDING THE LIFE CYCLE IN HUMAN-CENTRED AI

In embedding AI lifecycle on human centric, the data generated by the AI system will be managed by human for decision making. Users with valuable insights and relevant expertise leverage the knowledge provided by the AI system to make informed decisions. This collaboration between humans and AI systems is crucial to ensure responsible and ethical decision-making. Designers, developers, technology providers, and suppliers should design user interfaces and interaction mechanisms that facilitate effective communication, transparency, and accountability between humans and AI systems.

AI life cycle requires involvement of AI actors in the following four phases, typically: i) Design, data collection, and algorithm development and training ii) verification and validation; iii) deployment, and iv) operation and monitoring.

AI life-cycle stages	i) Design, plan and model	ii) Verification and validation	iii) Deployment & use	iv) Operation and monitoring
AI actors	1. Planning and design: Data scientists, domain experts, and governance experts. 2. Data collection and processing: Data scientists, domain experts, data engineers, data providers 3. Algorithm development and training: modellers, model engineers, data scientists, domain experts.	data scientists, data/model/systems engineers, governance experts	system integrators, developers, systems/software engineers, testers, operators and practitioners, evaluators, and domain experts with expertise in human factors, socio-cultural analysis, and governance.	governance experts, domain experts, and systems/software engineers, system operators, product developers, evaluators and auditors, compliance experts, organizational management, and members of the research community.

Figure 10: AI life cycle stages and AI actors

3.3 EMBEDDING THE LIFE CYCLE IN HUMAN-CENTRED AI

Stage 1- Design, data collection, and algorithm development

The first stage of the AI life cycle is planning and design. The AI actors responsible for creating the concept and objectives of AI systems and responsible AI practices at this early stage contribute to build trust, mitigate risks, and ensure the ethical use of AI technologies. All actor players should establish clear ethical guidelines that align with societal values and legal requirements. A multidisciplinary team that includes experts from various fields such as AI, ethics, law, and social sciences at this stage can provide different perspectives and insights, helping to identify and address potential ethical and social implications of the AI system.

During data collection and processing, all actor players must ensure the ethical sourcing of data, respecting privacy and data protection regulations. AI actors should also be transparent about the data they collect, ensuring that it is relevant, representative, and free from bias. Additionally, they should implement data quality checks to maintain the integrity of the data.

Tools: Web scraping tools such as BeautifulSoup, Scrapy, or Selenium.
Metrics: Data completeness, accuracy, consistency, uniqueness, and validity.

During the algorithm development phase, AI actors have a responsibility to ensure fairness, accuracy, and transparency. All actors should strive to minimize bias in algorithms and models by carefully selecting training data and considering the potential impact on different groups. AI actors also should document the development process, including the choice of algorithms, parameters, and any ethical considerations for external scrutiny and help in identifying and addressing potential biases or unintended consequences.

Tool: Python libraries such as Scikit-learn, TensorFlow, and PyTorch.
Metrics: Accuracy, precision, recall, F1 score, AUC-ROC, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Confusion Matrix

3.3 EMBEDDING THE LIFE CYCLE IN HUMAN-CENTRED AI

Stage 2- Testing and validation

The testing and validation stage of the AI life cycle is a critical step to ensure that the developed machine learning models are accurate, reliable, and effective. The AI system or its components should be tested against new data to ensure that it can handle new situations and scenarios effectively. This process also involves identifying any errors in the model, which are then corrected before moving on to the next stage. The testing process should involve diverse datasets and scenarios to identify potential biases or discriminatory behavior.

Tools:

1. Python libraries such as Scikit-learn, Keras, and TensorFlow for building and testing machine learning models.
2. Testing frameworks such as PyTest, Nose, and Robot Framework for automated testing.
3. Model explainability tools such as LIME and SHAPE for interpreting and explaining model predictions.

Metrics:

Accuracy: Measures how well the model predicts the outcome of a given input.
Precision: Measures the proportion of true positives compared to all positive predictions made by the model.
Recall: Measures the proportion of true positives compared to all actual positive outcomes in the data set.
F1 Score: A combination of precision and recall that measures the overall accuracy of the model.
Confusion Matrix: A table that shows the number of true positives, true negatives, false positives, and false negatives generated by the model.
Receiver Operating Characteristic (ROC) Curve: A curve that summarizes the performance of a binary classification model over all possible decision thresholds.
Lift Curve: Evaluates the effectiveness of a predictive model at identifying the top candidates in a group, relative to a random selection.
Mean Squared Error (MSE): A measure of the average squared difference between the predicted and actual target values.

3.3 EMBEDDING THE LIFE CYCLE IN HUMAN-CENTRED AI

Stage 3- Deployment and use.

At this stage, the AI system is integrated into the platform or system where it will be used. This involves integrating the AI model with other systems, APIs, and databases to ensure seamless operation with existing infrastructure.
AI Deployment and Use actors are responsible for contextual decisions relating to how the AI system is used to assure deployment of the system into production. The tasks include piloting the system, checking compatibility with legacy systems, ensuring regulatory compliance, managing organizational change, and evaluating users experience.

Tool: IBM Watson Machine Learning, TorchServe, TensorFlow Serving, Kubeflow.
Metrics: Real-time processing speed, model uptime, and response time.

Stage 4 - Operation and monitoring.

The AI model that has either been deployed independently or integrated as a hyperautomation process will be monitored and evaluated. It involves closely monitoring the AI system's performance, detecting anomalies, and identifying potential issues that could lead to errors or malfunctions. These tasks are carried out by AI actors who are responsible for operating the AI system and working with others to regularly assess system output and impacts. The main evaluation criteria are representative of the technology itself, diverse individuals in diverse settings utilizing the technology and value generated by the technology.

Tool: MLflow, TensorBoard, Kibana, and Prometheus.
Metrics: Model accuracy, model efficiency, model robustness, and model fairness.

3.4 INDEX MEASUREMENT, IMPACT ASSESSMENT AND RISK

Responsible AI governance and ethics are ongoing processes. It is essential to continuously monitor the performance and impact of AI systems and make improvements as necessary. This includes monitoring for bias, unintended consequences, and evolving ethical considerations. Designers, developers, technology providers, and suppliers should engage in regular audits, feedback loops, and updates to ensure that AI systems remain aligned with 7 AI principle and societal values.

Tools for risk and impact assessment

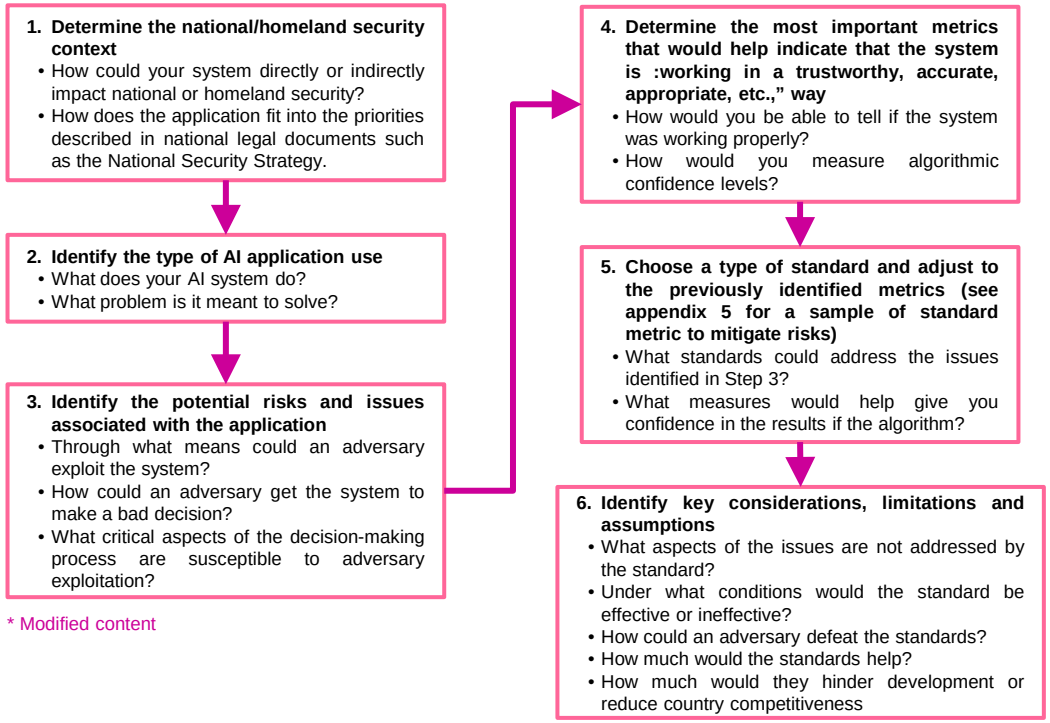
Tools available that can help in the governance of AI models includes:

1. The Responsible AI Impact Assessment (RAIIA) Tool is an example of an assessment tool to help the industry toevaluate risks and to develop risk mitigation strategies to responsibly implement AI systems. This is a free access tool and isavailable at the following address: https://www.itechlaw.org/sites/default/files/RAIIA%20Template%20Final_0.xlsx. For an overview of this tool see Appendix 6.

2. In January 2022, the Japanese Expert Group on How AI Principles Should be Implemented has published the Governance Guidelines for the Implementation of AI Principles. The guidelines introduce an evaluation tool/ checklist of conformity to the AI governance goals in the development and operation of AI systems. An overview of the evaluation tools is given in Appendix 6.

3. Tools concerning how to assess the responsible AI when developing, deploying or using AI systems are listed in Appendix 6, which is adopted from the High-Level Expert Group on AI (AI HLEG)

In order to ensure the suitability of AI technology to improve existing services as well as to mitigate future risk, the following steps can be taken (NIST)



* Modified content

PART C

3.4 INDEX MEASUREMENT, IMPACT ASSESSMENT AND RISK

The risk and impact assessment of AI products and services needs to be carried out to identify, evaluate, and address potential issues associated with the use of AI. It involves analysing the potential impact on individuals, society, and the environment, considering factors such as privacy,security, fairness, and accountability. Designers,developers, technology providers, and suppliers should proactively assess and address these risks to ensure responsible and ethical AI deployment.

Using index measurement in conjunction with risk and impact assessment provides a more comprehensive framework to evaluate the performance, ethical alignment, and societal impact of AI products and services. Leveraging tools, designers, developers,technology providers, and suppliers can lead to informed decisions, address potential risks, and promote responsible and ethical AI deployment.

Examples of components for each of the responsible AI principles can be found in Table 1 on the next page,and examples of indexes can be found in Attachment 9.

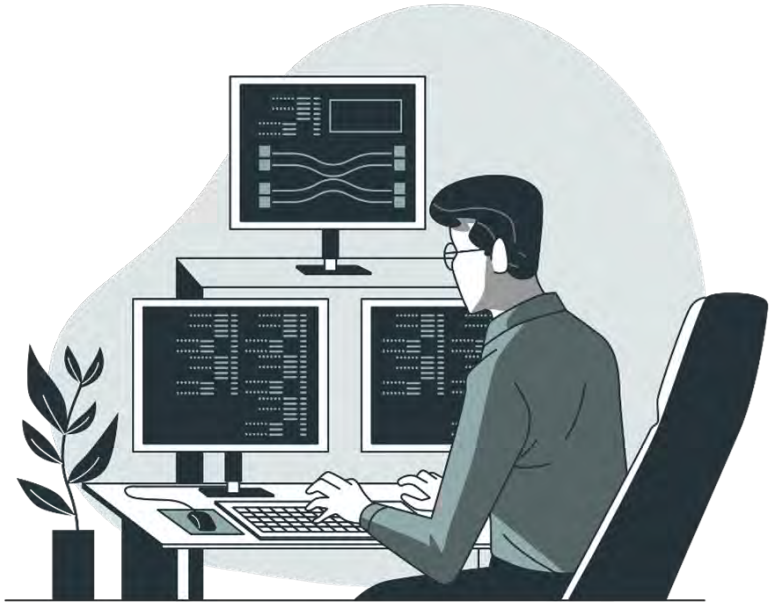
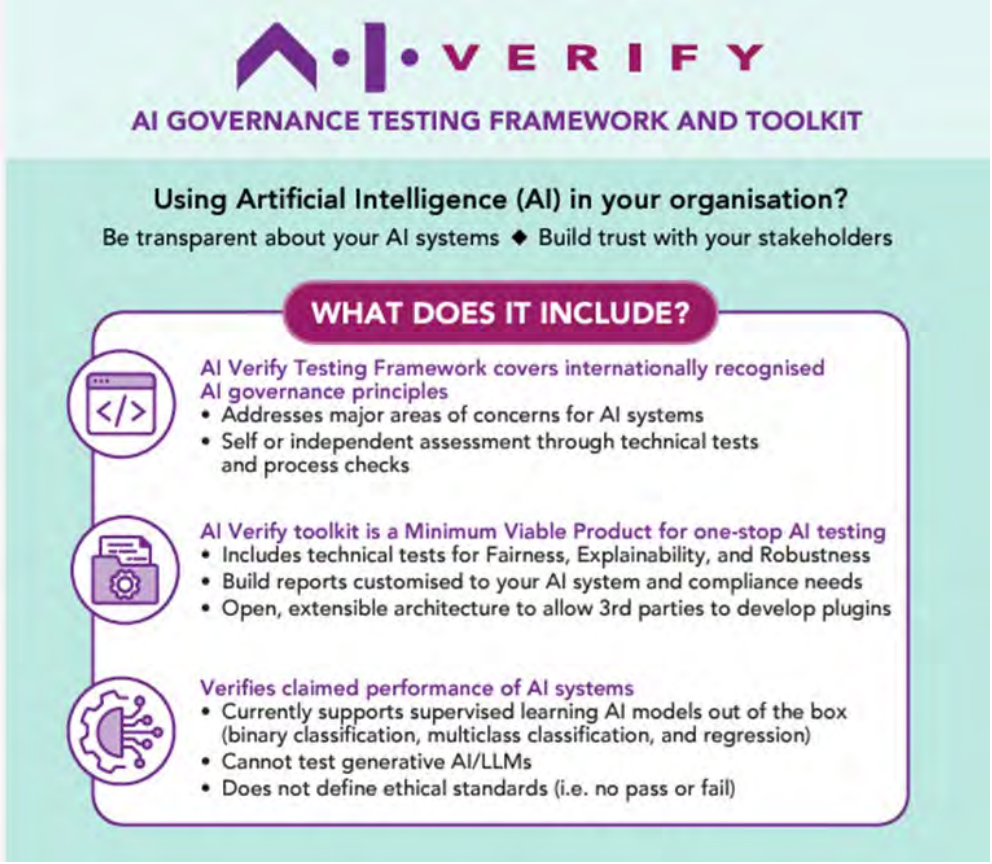


Table 1: Examples of components for each responsible AI principles

AI Principles	Index Measurement	Impact Assessment	Risk Analysis
Fairness	• Equal Access • Gender • Income Group • Marital Status • Age • Calibration metrics • Demographic Parity • Bias Metrics	• Context • Data metrics • Algorithms • Trade-offs involved job opportunity • Fairness-aware pre-processing and post-processing • Generalization gap	• FEAT principles • Unemployment issues at national level
Reliability, Safety, Control	• Service Availability • Human intervention • Standard • Limitation / threshold • Lost Time Injury Frequency Rate (LTIFR) • Total Recordable Incident Rate (TRIR) • Fatality Rate • Safety Performance Index (SPI) • Safety Climate Survey • Vehicle Safety Ratings • Severity Rate • Safety Integrity Level (SIL) • Aviation Safety Index • Crime Rate • Environmental Safety Index	• Testing • Zero Tolerance • Loss control	• Risk management • Cease Operation / business
Privacy & Security	• Personal data • PDPA • GDPR • Cyber security • Fair Information Practice Principles (FIPPs) • Privacy Impact Assessments (PIAs) Privacy by Design (PbD): • Privacy Scorecards • Data Protection Impact Assessments (DPIAs) • Privacy Enhancing Technologies (PETs) • GDPR Compliance • Privacy Impact Metrics • User-Centric Assessments • effective de-identification and anonymization techniques	• Personal privacy / security • Privacy invasion • Identity Theft • Scam • Ransom Attack	• Human Safety • Privacy • Loss Reputations • Financial Loss
Inclusiveness	•Minority •Diversity	• Large knowledge gap • Income gap	• Lack of talent
Pursuit of human benefits and happiness	• Happiness Index • Custom tailored Metrics • Hybrid Metrics • AI Impact Assessments • Gross National Happiness (GNH)	• Depression • Stress	• Mental Health • Society Problem

> AI systems must be understandable, transparent, and accountable. We cannot simply hand over decision-making to algorithms without oversight and responsibility.
>
> - Kate Crawford-

3.5 TESTING, ASSESSMENT AND CERTIFICATION



An example of an AI governance testing framework and a software toolkit to helps organizations validate the performance of their AI systems against these principles through standardized tests.(https://aiverifyfoundation.sg/what-is-ai-verify/_

3.5 TESTING, ASSESSMENT AND CERTIFICATION

These are some of the example of Toolkits for the 7 Principles:

Fairness

AI Fairness 360 (AIF360): This open-source toolkit developed by IBM Research provides a suite of algorithms and metrics to assess and mitigate bias in AI systems, promoting fairness and equal treatment across different demographic groups.

Reliability

ModelOp Center: This enterprise AI operationalization platform includes testing and validation capabilities to ensure the reliability and robustness of AI models. It provides tools for monitoring model performance and detecting anomalies.

Safety and Control

OpenAI Safety Gym: OpenAI's Safety Gym is a toolkit designed to assess and develop safe and controlled AI systems. It provides a set of simulated environments that allow researchers to evaluate the safety and robustness of AI algorithms.

Privacy and Security

PySyft is a privacy-preserving deep learning framework that allows for secure and private computation on distributed data. It enables organizations to assess and implement privacy and security measures in AI systems.

Inclusiveness

Microsoft Fairlearn: Fairlearn is a Microsoft toolkit that focuses on promoting inclusivity and reducing bias in AI systems. It provides algorithms and techniques to achieve fairness and mitigate disparate impact.

Transparency

Captum is an open-source interpretability library developed by PyTorch. It offers techniques and tools for interpreting and explaining the decisions made by AI models, enhancing transparency.

Accountability and Pursuit of Human Benefit and Happiness

Ethically Aligned Design (EAD) Certification: The IEEE's EAD certification program aims to promote ethical considerations and accountability in AI systems. It provides a framework for assessing and certifying the ethical qualities of AI technologies.

These are for example purposes only and are not officially mandated or endorsed by MOSTI

3.6 CRITICAL THINKING - CONSIDERING AND AVOIDING

- **Understand your needs and responsibilities:** Ask the question whether the AI system you are developing or implementing will be used to make decisions or otherwise has a significant impact (positive or negative) on people (including marginalized groups), the environment or society
- **Start small:** Don't try to do too much too soon. Start with small, manageable projects to test the waters before fully committing to AI.
- **Gather quality data:** AI is only as good as the data it is trained on. Make sure you are using high-quality data that is relevant to your needs.
- **Consider ethics and bias:** Be aware of the potential for bias in AI systems and take steps to mitigate it. Choose AI products and services that are committed to ethical development.
- **Invest in talent:** Make sure you have the right people in place to implement and manage AI technology effectively.
- **Stay updated:** AI is a rapidly evolving field. Keep up with the latest developments and trends to ensure you are using AI effectively.
- **Focus on integration:** Make sure your AI systems integrate smoothly with your existing processes and infrastructure.
- **Measure ROI:** Continually assess the return on investment and the impact of AI on your operations to make informed decisions about future projects.
- **Stay compliant:** Be aware of all relevant laws and regulations affecting AI in your industry and region
- **Be explainable and transparent:** Describe of what kind of AI is incorporated descriptions and the rationale behind why certain features or models have been selected during the model development process.

- **Underestimate the complexity:** Implementing AI is not just about technology; it involves change management, process redesign, and often a cultural shift.
- **Ignore the human element:** AI should augment human abilities, not replace them. Focus on how AI can assist and empower your workforce.
- **Neglect data privacy:** Be mindful of privacy laws and ethical considerations related to data collection and usage.
- **Overlook maintenance:** AI systems require ongoing maintenance, updates, and retraining to stay effective.
- **Chase hype:** Don't adopt AI just because it is a buzzword. Ensure that genuinely adds value to your operations.
- **Skip the strategy:** Don't implement AI without a clear strategy and understanding of how it fits into your overall business objectives.
- **Forget scalability:** Plan for the future by considering how your AI solution can scale up as with your business grows.
- **Disregard regulations:** Stay compliant with all relevant laws and regulations affecting AI in your industry and region.

C4.0 DEMONSTRATION AND SECTOR CASES

The government has identified 10 Socio-economic drivers through via its document of 10-10 MySTIE Framework.

In this Section 4, practical examples are described to show how AI ethics/ principles are introduced in sectoral industry which are:

1. **DEMONSTRATION 1**: Smart Technology and Systems Sector – WISE AI using eKYC-Customer digital onboarding and identity verification solution to onboard customers in just a few minutes.

2. **DEMONSTRATION 2:** Energy sector – PETRONAS using Enterprise Data Hub (EDH) platform to liberalise trusted data across organization , equipped with advanced analytics to enhance efficiency and improve decision-making.

3. **DEMONSTRATION 3**: Education sector – ANH-LEYZE using AI ChatBot onboarding Validation offers parents the convenience of communicating with teachers and getting updates on their child's education.

4. **DEMONSTRATION 4**: Smart Cities and Transportation Sector - TIME DOTCOM using AI technology to enhance efficiency, flexibility, and speed of service delivery, catering to customers' increasing demand for ultra-low latency network performance.

4.1 SMART TECHNOLOGY AND SYSTEMS SECTOR - WISE AI

About
An international award-winning Malaysian AI company

Vision
To become the AI powerhouse of Southeast Asia

AI Solutions & Services
•eKYC*
•Enterprise AutoML (automated machine learning)
•AI Training

Client Industries
Banking, finance, insurance, fintech, telecommunication, public sector, ICT and hospitality

**The ONLY Malaysian company with PROPRIETARY full-fledged eKYC AI tech)*

Solution: eKYC
Customer digital onboarding and identity verification.

Align with AI Ethics (refer to next page) 7	Process				
5 4 2	AI Technology	Identity Fraud Detection	Identity Card OCR	Liveness Detection	Face Recognition
5 6	AI Technology Ownership (Owner/Reseller) *if reseller, state the technology owner	Owner	Owner	Owner	Owner
6	AI Technology Certification & Certification Ownership	Pending MOSTI NTIS Sandbox validation	N/A	1. ISO30107 2. MOSTI NTIS Sandox Owned by WISE AI	1. NIST FRTE 2. MOSTI NTIS Sandbox Owned by WISE AI
3	Personal Data Captured	ID card image	NRIC information	Face biometric	Face biometric
5 3	Personal Data Processing Party (ies) &U Cross Border Data Flows	WISE AI only, data stored locally	WISE AI only, data stored locally	WISE AI only, data stored locally	WISE AI only, data stored locally
1 4	Platform	Web, IOS, Android, HarmonyOS			

PART C

Alignment with AI Ethics

	7 AI Principles/ Technology	AI Technology	Software Solution
1	Fairness	No bias in the data collection and development of AI modelling	Wide access to the solution – making things affordable, catering to diverse smartphones
2	Reliability, Safety and Control	System robustness, complete end-to-end EKYC an AI tech component	A system ready any time with 99.5% uptime
3	Privacy & Security	Data remains anonymous in the model training process	Comply with PDPA and protect the country's sovereignty by ensuring the data remains local.
4	Inclusiveness	Face of ASEAN, the solution is tailored made for wide variety of ASEAN faces	Wide access to the solution – catering to diverse mobile operating systems.
5	Transparency	Disclosure of AI technology ownership	Disclosure of AI software solution ownership
6	Accountability	Adopt best practice for AI model development and deployment (eg FAR<5%)	Adopt best practice for system deployment through regular vulnerability checks (eg: pentest)
7	Pursuit of Human Benefit and Happiness	Continuous evaluation and refinement to ensure AI relevancy through time	AI for everyone, instant onboarding anywhere, anytime

Certifications

a) ISO 30107*
WISE AI's proprietary technology - Liveness Detection, is certified for reliability and effectiveness against malicious identity attacks. Technology ownership also promotes accountability.

b) MOSTI NTIS Sandbox
Graduated from MOSTI's NTIS Sandbox program. It speaks volumes about the robustness and reliability of our technology in real-world applications.

c) NIST FRTE of USA
WISE AI's proprietary Facial Recognition technology (Face of ASEAN) is certified for high accuracy and promotes trust and inclusivity (faces of different ethics).

4.2 ENERGY SECTOR - PETRONAS

About

A global Malaysian energy group with presence in over 100 countries, producing & delivering energy and solutions that power society's progress

Statement of Purpose

A progressive energy and solution partner enriching lives for a sustainable future

AI Solutions & Services

- Enterprise Contextualise Search & Chat
- Cognitive Services Capabilities

Client Industries

- Oil and Gas (Exploration & Production, Gas & Power, Chemical & Products, Retail & Marketing), Clean Energy Solutions

Solution:

Enterprise Data Hub (EDH) platform equipped with Advanced Analytics capabilities, whereby trusted data from various sources are liberalised for valuable insights across PETRONAS

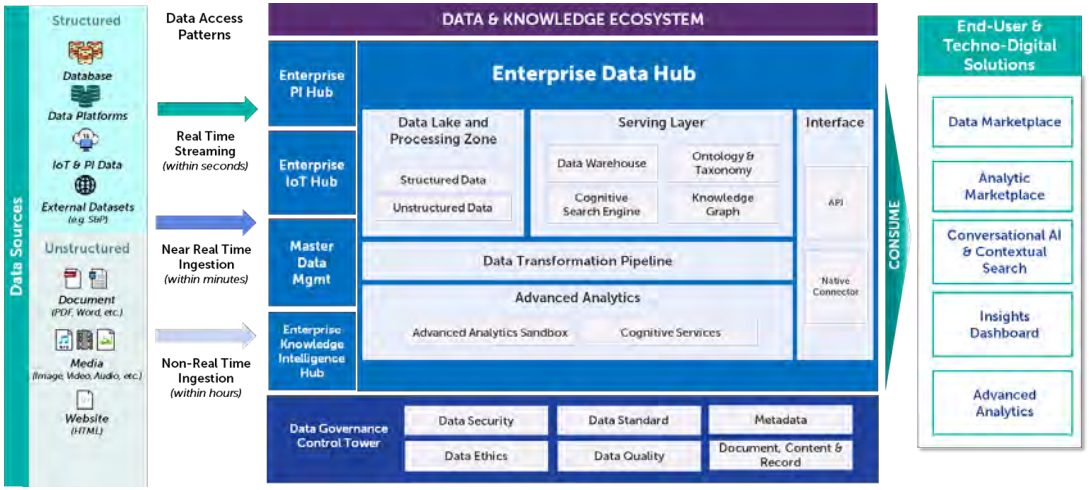


Figure 11: PETRONAS Enterprise data hub platform and knowledge ecosystem

PART C

Alignment with AI Ethics

	7 AI Principles/ Technology	AI Technology	Software Solution
1	Fairness	Diverse training data used in development of AI models, incorporating human-in-loop for review & correction	Multi-modal A.I cognitive services deployed to extract critical data from documents to improve operational efficiency
2	Reliability, Safety and Control	Robust data quality checks across data & business domains, with model testing & validation embedded in lifecycle of AI models	Data quality health is made visible across the group with confidence score to measure accuracy of AI models, ensuring trustworthy data & reliable models for decision making
3	Privacy & Security	Compliance to data acts and regulatories, further guided by the 7 unique categories of data prior to sharing and protection done through masking and tokenisation	Secret and confidential data are secured with entitlement request available across the group
4	Inclusiveness	Use of natural language processing to accept human-like query with multi-lingual support, and inclusion of users feedback to improve search and chat result	Facilitate easy access to information, made accessible to everyone in the organization with users feedback loop embedded in application
5	Transparency	Full disclosure on the use of contextualise search and chat, powered by knowledge graph, cognitive search & generative AI, highlighting its capabilities, limitations and list of systems connected	Facilitate insights discovery by accelerating search & chat results, equipped with traceability of result within the parameters of connected systems
6	Accountability	Clear ownership of data & AI models, with lineage tracking on origin and transformation used in analytics	Centralised governance allows for uniformity of standard practices across group while decentralised assurance ensures adoption and compliance on the practices, empowering accountabilities across the group
7	Pursuit of Human Benefit and Happiness	Contextualise search able to connect and crawl across other systems to surface relevant results, powered by cognitive search and knowledge graph.	Improve work experience in finding relevant data and knowledge from diverse systems to generate insights and enable decision making in respective business across the group

4.3 EDUCATION SECTOR - ANHSIN TECHNOLOGY

About

A dynamic and forward-thinking A.I Technology company, specialize in the development of cutting-edge AI-powered solutions designed to transform educations and revolutionize the way it operates in the digital era.

Vision

To Empower AI technology in order to elevate better study experiences and to simplifying the complexities of digitalization making it accessible and effective for students across various environments.

AI Solutions

- eANHSIN* (Anh-Nyn, Anh-Nyn(ATAS), Anh-Leyze
- Enterprise NLP (automated and customized Natural Language Processing)
- Enterprise ML (automated and customized Machine Learning)
- Enterprise NLA (automated Natural Language Action)

Client Industries

Educations (Public sector & Private Sector), Insurance, ICT, Healthcare (Private Sector)

ANHSIN

Solution: AI ChatBot onboarding Validation

Alignment of AI Ethics with the 7 Core Principles	Processes	Chatbot (Headmasters and Teachers)	Chatbot (Parents & Guardians)	Chatbot (School Children)
7	Processes			
1 2 5	AI Technology	School Identity	Parent/Guardian Identity	Student Identity (by Age Group)
1 3 6	AI Technology Ownership (Developer/Owner)	AI Developer and School (PTA) Ownership	AI Developer and School (PTA) Ownership	AI Developer and School (PTA) Ownership
3	AI Technology Certification and Certification Ownership	Pending MOSTI approval Llama2 (License) OpenAI (API Key) GPT4All Embedding	Pending MOSTI approval Llama2 (License) OpenAI (API Key) GPT4All Embedding	Pending MOSTI approval Llama2 (License) OpenAI (API Key) GPT4All Embedding
2	Personal Data Captured	School Database Live Quiz Profiling Test Co-curriculum	MyKad Information	MyKad Information
1 2	Personal Data Processing Party (es) and Criss Border Data Flows	eANHSIN only, data stored locally	eANHSIN only, data stored locally	eANHSIN only, data stored locally
4 5 6	Platform	Web, iOS, Android		

108

PART C

Alignment with AI Ethics

	7 AI Principles/ Technology	AI Technology	Software Solution
1	**Fairness**	No biasedness in data collections, datasets, and algorithms are used to develop AI models and Use-Cases	Ensuring that the software processes and solutions accessible by all users equitably
2	**Reliability, Safety and Control**	Robustness AI's pillars for AI Models (e.g Volume, Velocity, Variety, Variability, Veracity, Value & Visualization)	Ensuring that the solution performs its intended function under the stated conditions for a specified period of time without problems
3	**Privacy & Security**	Data Collections remain encrypted and confidential and in compliance with the PDPA Act (if personal records) or BNM regulations for others	Protecting the confidentiality, integrity, and availability of information in compliance with PDPA act including complete with appropriate Disclaimer regulated by law (BNM)
4	**Inclusiveness**	Customized and automated-ML, NLA, NLP, with solutions tailor made for user needs.	Creating software that is accessible from different platform such as web and mobile
5	**Transparency**	Disclosing limitations (e.g. ML, NLA, NLP) and potential problems with the AI models	Developers should be open about the capabilities and limitations of the software solution
6	**Accountability**	Adopting to best practice on designing, developing and deploying AI models with 95% accuracy	Adopting to best practice and perform regular source code vulnerability scanning
7	**Pursuit of Human Benefit and Happiness**	Regularly review, evaluate and enhance to ensure AI models align with human values and ethical considerations.	Ensuring the solution is accessible, user-friendly, and makes a positive contribution

4.4 TELECOMMUNICATION SECTOR : TIME DOTCOM BERHAD

About

Telco provider with businesses covering fixed line, global networks, cloud and data centres

Vision

To become the Trusted Sovereign Cloud Advisor in the region

AI Solutions & Services

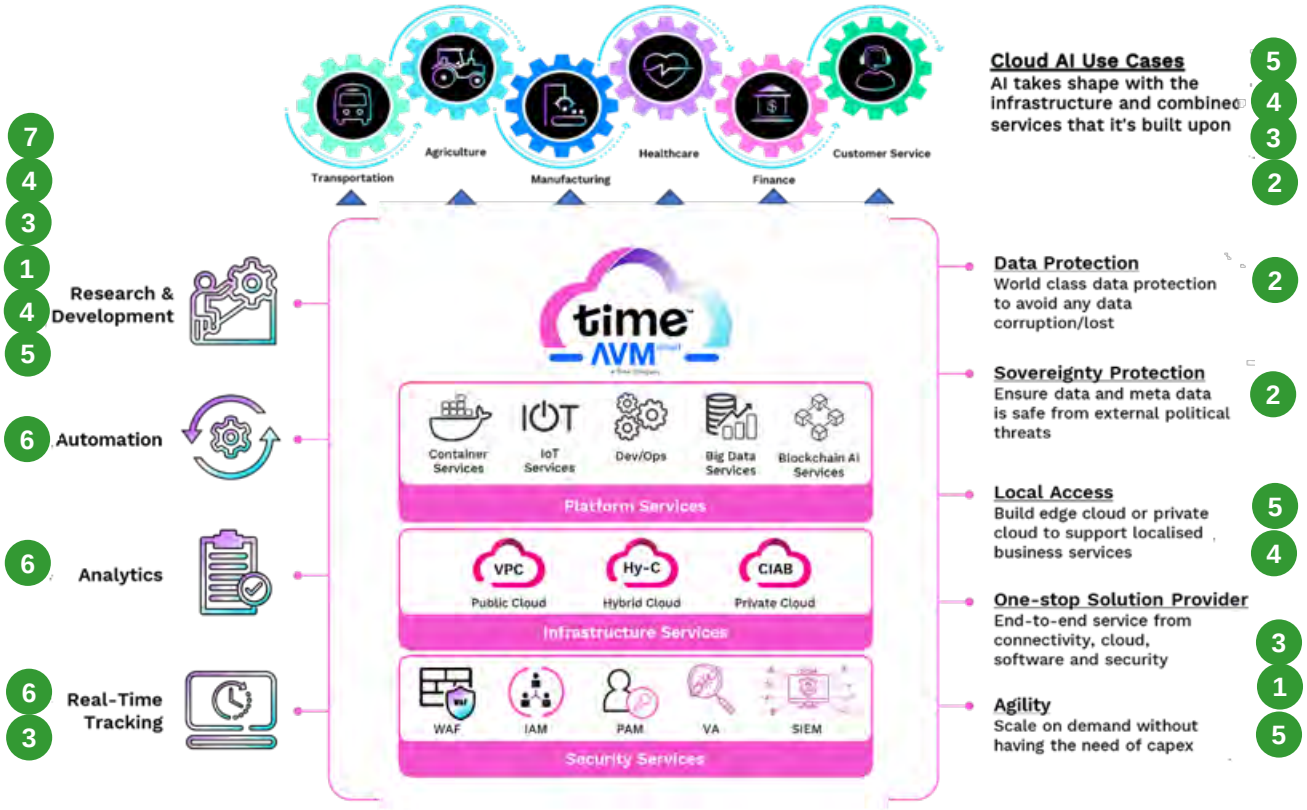
•Sovereign Cloud*
•Zero-Trust Architecture**
•AI-enabled network automation**

Client Industries

Banking, FSI, fintech, Oil and Gas, healthcare, hospitality, media, ICT, public sector, education

* The ONLY Malaysian Cloud Provider with Certified Sovereign Cloud architecture)
** AI enabled 3C infrastructure

Solution: Sovereign Cloud Infra
Protecting and unlocking your Critical Data Values



PART C

Alignment with AI Ethics

1	2	3	4	5	6	7
Fairness	**Reliability, Safety And Control**	**Privacy And Security**	**Inclusiveness**	**Transparency**	**Accountability**	**Pursuit of Human Benefit And Happiness**
Digital Infra at zero data transfer cost across all access - AI use case of all scenarios, regardless enterprises and individuals	Reliable infra with higher SLA, Lower acceptable margins for packet loss and latency and AI-enabled 3C operation	Protect national data sovereignty with adherence to Local Laws, Standard Quality and securitv compliance and without external jurisdiction control	Digital infra of all platform & deployment - Public, private and hybrid, catering to different operating systems to support different data and AI innovation	TIME fully owns the infrastructure, spanning across cloud platforms, DC and fiber connectivity to the Internet Exchange and international cable, available for customers	Cloud infra design based on sovereign cloud framework in accordance with industrial best practice and guidelines and agile build to users' critical needs	The trusted Sovereign cloud provider, empowering everyone with cloud & AI innovation that accelerate country digital economy and sustain people's income

C5.0 CODE OF ETHICS

AI Principles and Code of Ethics are voluntary guidelines promoting ethical behaviors in AI development and adoption. The principles focus on addressing ethical challenges and mitigating unethical risks in AI algorithms and systems. The code of ethics, tailored to specific industries, organizations, or professions, provides detailed guidelines for ethical conduct, ensuring responsible AI development and adoption at the national level.

AI Principles	Item	Code of Ethics
AI technologies, despite their potential, have significant ethical considerations and implications, necessitating a deep understanding of their ethical implications.	Scope and Application	The AI Code of Ethics applies ethical principles and standards to specific organizations, professions, or industries, ensuring adherence to ethical guidelines within their respective domains.
International organizations, academic institutions, and other international organizations are developing non-compulsory guidelines, which are not legally binding unless incorporated into legislation.	Regulatory Status	Organizations and professionals may adopt the Code of Ethics, which may eventually be incorporated into regulatory laws or regulations.
The National Guidelines provide high-level principles to guide ethical decision-making in AI development and deployment, providing a comprehensive guide for ethical AI practices.	Specificity	The Code of Ethics offers detailed guidance on the practice of ethical conduct in the profession, ensuring adherence to ethical standards and disciplinary measures.
Guidelines evolve as AI technologies advance, incorporating ethical considerations as technologies evolve, necessitating continuous reflection and updating of AI Principles to address emerging ethical challenges.	Evolution & Adaptation	The Code of Ethics is subject to regular reviews and revisions to align with societal norms and professional practices, ensuring their continued relevance and effectiveness.
Technologists are working towards a broader range of stakeholders, ensuring recognition of diverse perspectives in the field of AI.	Stakeholder Engagement	The Code of Ethics is a set of guidelines that guide the conduct of professional activities, ensuring the ethical conduct of the profession.

C6.0 ALIGNING INDUSTRIAL ECONOMIC CONTRIBUTIONS WITH RESPONSIBLE AI AND SDG & ESG PRINCIPLES

AI becomes a powerful tool that can drive sustainable development and contribute to the achievement of the Sustainable Development Goals (SDGs) while aligning with Environmental, Social, and Governance (ESG) principles. ESG compliance has become increasingly important in the face of increasing climate risks as it deals with a company's impact on the environment and society. ESG data includes any indicators that shed light into the sustainability context of an asset, facility, company or region, whether historic, current or expected. ESG data is collected under three primary umbrellas:

- Environmental data should capture environmental information such as annual carbon emissions and energy consumption, water usage, waste and pollution output, and more.
- Social data focuses on statistics related to workforce diversity, gender equity, human rights, and more.
- Governance data tracks company input regarding corruption, labor practices, the gender composition of the board of directors, etc..

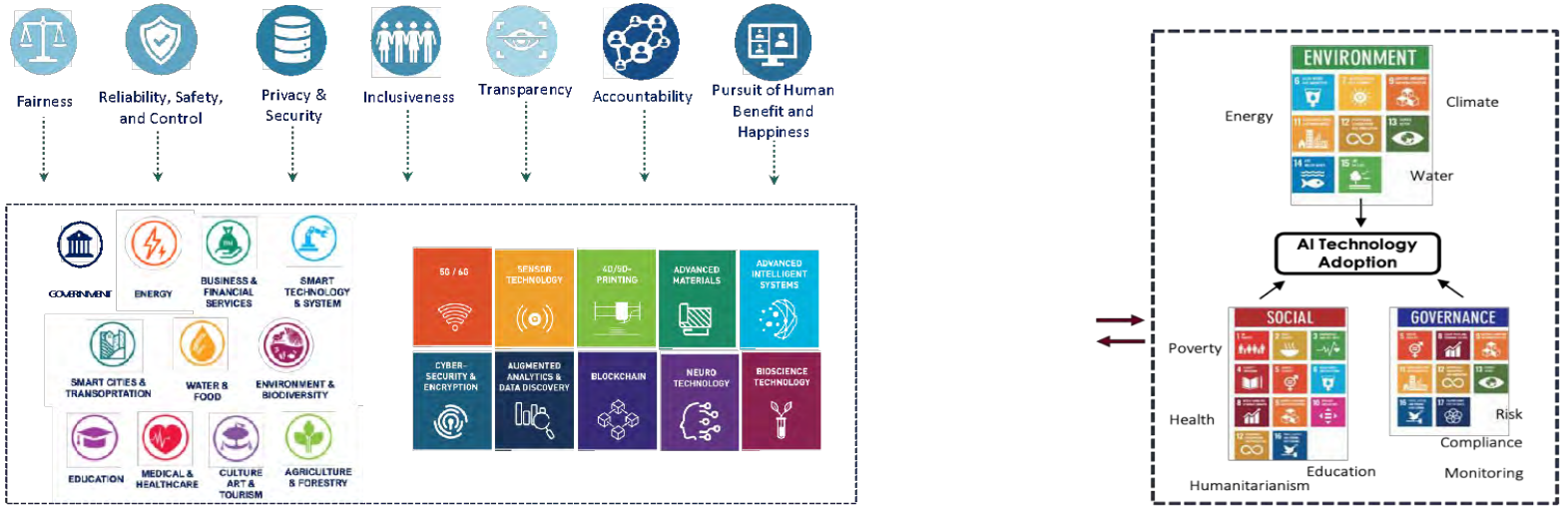


Figure 12: Guidelines AI Governance and Ethics for each AI Socio Economic Drivers (2024-2025)

Section 4

Benchmarking and International Smart Partnership

4.1 BENCHMARKING AND ENHANCING INTERNATIONAL SMART PARTNERSHIP

Benchmarking analysis helps in identifying best practices in AI governance and ethics, identifying commonalities and differences, enabling organizations and policymakers to align their approaches and ensure consistency in ethical standards across different AI systems and applications. The benchmarking of AI governance and ethics framework and guidelines with other countries or organizations is shown in the following table.

Table 2: Benchmarking analysis AI guidelines

Guidelines, Countries/ Organization (Year)	**Principles on Artificial Intelligence, OECD (2019)**	**Recommendations of AI Ethics, UNESCO (2021)**
Description	To promote use of AI that is innovative and trustworthy and that respects human rights and democratic values. The principles also focus on how governments and other actors can shape a human-centred approach to trustworthy AI.	To provide a basis to make AI systems work for the good of humanity, individuals, societies and the environment and ecosystems and to prevent harm, and stimulating the peaceful use of AI systems.
Ai Principles	**AI principles divided into 2 categories:** • **Values-based principles:** (1) Inclusive growth, sustainable development and well-being; (2) Human-centred values and fairness; (3) Transparency and explainability; (4) Robustness, security and safety; and (5) Accountability • **Recommendations to governments**: (1) Investing in AI R&D; (2) Fostering a digital ecosystem for AI; (3) Shaping an enabling policy environment for AI; (4) Building human capacity and preparing for labour market transformation; (5) International co-operation for trustworthy AI	**10 AI principles:** (1) Proportionality and Do No Harm, (2) Fairness and non-discrimination, (3) Safety and security, (4) Right to Privacy, and Data Protection, (5) Sustainability, (6) Human oversight and determination, (7) Transparency and explainability, (8) Responsibility and accountability, (9) Awareness and literacy, (10) Multi-stakeholder and adaptive governance and collaboration.
Who it is for?	For policymakers working on data, AI and international relations.	All Member States, both as AI actors and as authorities responsible for developing legal and regulatory frameworks throughout the entire AI system life cycle.
Special features	The principles encourage collaboration among countries, organizations, and stakeholders to promote the responsible development and use of AI technologies for inclusive growth and sustainable development.	Providing a guiding framework for the development and implementation of 11 policies areas which allow policymakers to translate the core values and principles into action.
Similarity	Emphasis on fairness, transparency, accountability, human rights, and the responsible development and use of AI, while recognizing the imperative to address societal concerns, promote diversity, and safeguard the well-being of individuals and communities.	

4.1 BENCHMARKING AND ENHANCING INTERNATIONAL SMART PARTNERSHIP

Table 2: Benchmarking analysis AI guidelines (continued)

Guidelines, Countries/ Organization (Year)	AI procurement guidelines, Office for AI and World Economic Forum, UK (2020)	BLUEPRINT FOR AN AI BILL OF RIGHTS: Making Automated Systems Work for the American People, White House Office of Science and Technology Policy, US (2022).	Ethics Guidelines for Trustworthy AI, European Commission (2019)
Description	A guide that addresses the ethical uncertainty around AI procurement so that officials feel equipped to use innovative technology while being able to mitigate risks.	The document serves as a guide to incorporating protections and ethical considerations into the design and deployment of AI-powered automated systems, with the aim of ensuring that these systems work for the benefit of the American people while upholding their rights and values.	To promote 3 components of trustworthy AI which is legality, ethics, and robustness. The development, deployment and use of the AI system must respect for human autonomy, prevention of harm, fairness, and explicability.
Ai Principles	**The 5 AI principles are:** (1) safety, security and robustness; (2) Appropriate transparency and explainability; (3) Fairness; (4) Accountability and governance\; (5) Contestability and redress	**5 Ai principles includes:** (1) Safe and effective system; (2) Algorithmic Discrimination Protections; (3) Data Privacy; (4) Notice and Explanation; (5) Human Alternatives, Consideration, and Fallback	**7 key requirements for Trustworthy AI**: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability.
Who it is for?	For public servants and commercial specialists procuring AI.	Local-level policymakers across sectors and technical teams preparing to deploy or already deploying AI systems.	All AI actors and as authorities responsible for AI systems made in Europe as well as those developed elsewhere and deployed or used in Europe.
Special features	Providing best practices addressing specific challenge of acquiring AI in public sectors	Prioritizing the protection of civil rights, democratic values, and public engagement in the context of AI systems.	Providing Assessment lists and methodologies for implementing ethical AI.
Similarity	Emphasis on fairness, transparency, accountability, human rights, and the responsible development and use of AI, while recognizing the imperative to address societal concerns, promote diversity, and safeguard the well-being of individuals and communities.		

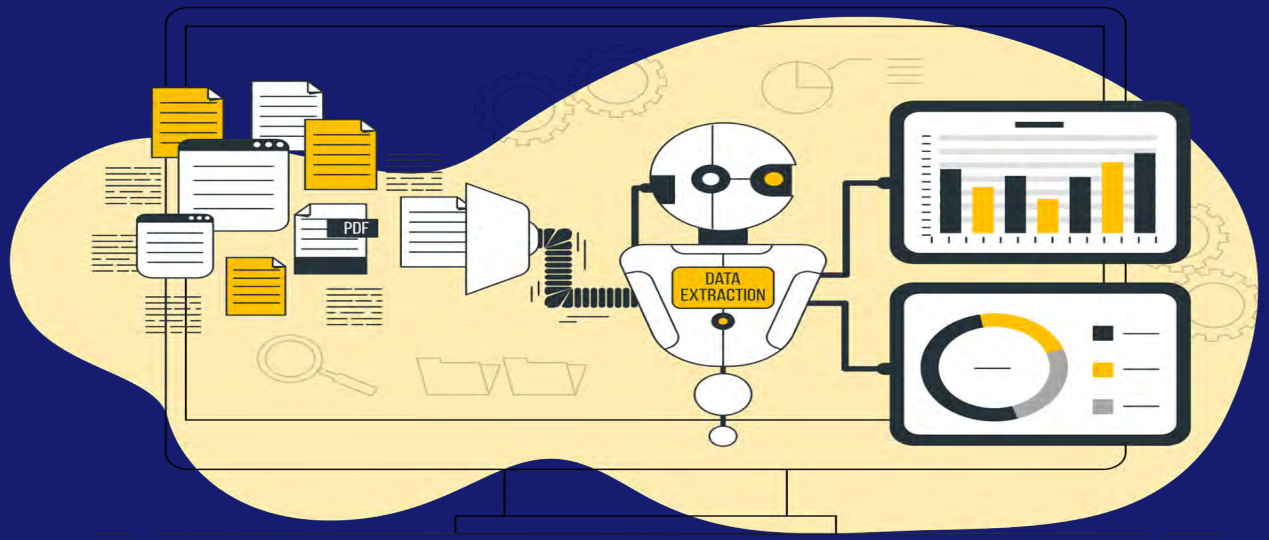
4.1 BENCHMARKING AND ENHANCING INTERNATIONAL SMART PARTNERSHIP

Table 2: Benchmarking analysis AI guidelines (continued)

Guidelines, Countries/ Organization (Year)	AI Ethics and Governance in Practice, The Alan Turing Institute (2019)	Governance Guidelines for Implementation of AI Principles ver. 1/1, Japan (2022)
Description	The workbook describe a multitiered governance model namely Process Based Governance (PBG) Framework, that enables project teams to integrate ethical values and practical principles are successfully operationalized and documented across the AI project life-cycles.	The guideline present action targets to be implemented by an AI company, with the aim of supporting the implementation of the AI principles that are required to facilitate AI deployment.
Ai Principles	**SSAFE-D principles:** (1) Sustainability, (2) Safety, (3) Accountability, (4) Fairness, (5) Explainability, (6) Data stewardship	**7 social principles for AI:** (1) Human-centric, (2) Education/Literacy, (3) Privacy Protection, (4) Ensuring Security, (5) Fair Competition, (6) Fairness, Accountability, and Transparency, and (7) Innovation.
Who it is for?	Civil servants engaging in the AI Ethics and Governance, facilitators	AI companie (AI system developers (i.e., companies that develop AI systems), AI system operators (i.e., companies that operate AI systems), and data providers).
Special features	Providing tools and activities (for participants) to applying of the AI and Ethics and Governance. The activities will help demystify AI and ML by exploring the key technical components that make up AI systems, different types of ML, the AI lifecycle model, and an introduction to Responsible Research and Innovation.	Including action targets to be implemented by AI companies and examples of how to evaluate for gap analysis and the implementation of agile governance
Similarity	Emphasis on fairness, transparency, accountability, human rights, and the responsible development and use of AI, while recognizing the imperative to address societal concerns, promote diversity, and safeguard the well-being of individuals and communities.	

Section 5

Conclusion



5.1 Conclusion 110

5.2 Moving Forward 111

5.1 CONCLUSION

Referring to the comprehensive elaboration we can agreed that AI governance and ethics guidelines are crucial for all stakeholders, spanning from end-users and consumers to policymakers, developers, and technology providers. These guidelines serve as the foundation for responsible AI development, deployment, and utilization, ensuring that AI systems are developed and used in a manner that aligns with societal values, legal frameworks, and ethical principles. Guidelines provide assurances to end-users and consumers that AI systems they interact with are designed with their well-being and safety in mind. With the introduction of Responsible AI, stakeholders can trust AI systems more when they know they adhere to ethical guidelines, fostering confidence in their usage. In addition, with clear guidelines in place, stakeholders have recourse if they encounter AI systems that behave unethically or harmfully.

In order to foster trust in AI, consumer should always have the right to information. They should be made aware when an algorithm is using their personal information to provide offers for goods and services, uses this data to make decisions or reports their data to third parties. Developers and end users should collaborate in conducting a fundamental rights and consumer rights assessment.

Governance guidelines inform policymakers about the potential risks and benefits of AI technologies, helping them craft effective governance and regulations to ensure public safety and welfare. Guidelines aid in the creation of regulatory frameworks that require AI developers and users to comply with ethical standards, fostering responsible AI innovation. At the same time these Guidelines provide developers with clear principles and best practices for designing and implementing AI systems ethically, helping them avoid unintentional biases, discrimination, or harm. By adhering to governance and ethics guidelines enhances the reputation of developers and their organizations, showcasing their commitment to responsible AI practices. Adherence to governance guidelines helps developers set industry standards for ethical AI development and adoption, promoting consistency and interoperability across different AI systems. Emphasizing and adopting AI governance and ethics can gain a competitive edge by appealing to consumers and businesses who prioritize ethical considerations in their AI investments.

AI has becoming a powerful tool that also can drive sustainable development and contribute to the achievement of the SDGs while aligning with ESG principles. ESG compliance has become increasingly important in mitigate climate risks as it deals with a company’s impact on the environment and society. We can conclude, AI governance and ethics guidelines are essential for the entire AI innovation ecosystem of AI development and adoption. By adopting and adhering to these guidelines, together we can collectively contribute to the advancement of AI in a manner that is beneficial for the country economic growth and societal well-being. In embracing the 7 AI principles as outlined in these guidelines, we call upon all stakeholders to join hands in advancing the responsible and ethical use of artificial intelligence. Together, let us propel AI innovation forward while steadfastly upholding the values of fairness; transparency; accountability; inclusiveness; reliability,safety and control; privacy and security; as well as the pursuit of human benefit and happiness.

5.2 MOVING FORWARD

This document constitutes a framework for Continuous Improvement Process (CIP) in the domain of Responsible Artificial Intelligence (AI). It underscores the imperative of perpetual advancement, emphasizing the necessity to consistently evaluate and enhance AI systems to align with evolving ethical standards and societal needs. Forward progress is not merely encouraged but deemed essential in our commitment to fostering AI technologies that uphold principles of responsibility and ethics.

With CIP in mind, it is imperative to emphasize in the importance of forward momentum and ongoing enhancement. National Guidelines on AI Governance & Ethics for Responsible and Inclusive AI (National Guidelines on AIGE) serve as a blueprint for ensuring that AI systems operate in a manner that aligns with ethical principles and societal values. Central to these guidelines is the this mindset of CIP, as mentioned in the Section 1.2 as key consideration that are infused in the development process of the guidelines, recognizing that the landscape of AI is dynamic and constantly evolving. As such, the document advocates for an iterative approach to AI development and adoption, wherein stakeholders are encouraged to continually assess, adapt, and refine their AI systems.

Moving forwards is imperative within the realm of AI ethics. As technology advances and our understanding of ethical considerations evolves, it is essential that we proactively address emerging challenges and opportunities. This forward-thinking approach ensures that AI systems not only meet current ethical standards but also anticipate and adapt to future ethical dilemmas. By prioritizing forward momentum, we demonstrate our unwavering commitment to ethical progression in AI. This involves not only rectifying existing ethical concerns but also anticipating and preemptively addressing potential ethical pitfalls. Through ongoing evaluation, learning, and improvement, we strive to cultivate AI systems that not only function effectively but also uphold the highest ethical standards.

In embracing the 7 AI principles as outlined in these national guidelines, we call upon all stakeholders within the AI ecosystem to join hands in advancing the responsible and ethical use of artificial intelligence. Together, let us propel AI innovation forward while steadfastly upholding the values of fairness, transparency, accountability, and societal well-being.

Continuous Improvement via friendly platform to address an unethical AI risk

Sectoral National AIGE guidelines and code of ethics development

Integration and synergy of Responsible AI and ESG

International harmonization for standard regulation, certification and accreditation of AI product and service

Continuous Improvement via friendly platform to address an unethical AI risk

As mentioned in section 2.8 (Human Friendly AI (HF-AI) Platform), the platform is crucial to ensure total engagement from all stakeholders in enhancing this national guidelines which is treated as a living document. Incorporating HF-AI Platform into our continuous improvement process entails establishing an accessible and intuitive communication system that actively solicits stakeholders feedback. This platform features user-friendly interfaces and multiple channels for feedback submission, ensuring that stakeholders can easily engage and provide input. Transparency is upheld through clear communication about the purpose of feedback collection and the process for addressing the input, fostering trust and engagement among stakeholders. Utilizing data analytics tools, feedback is analyzed to identify patterns and trends, guiding iterative improvements to these national guidelines.

Responsive feedback handling and a commitment to closing the feedback loop demonstrate the government dedication to listening to stakeholders' needs and driving continuous enhancement. Accessibility considerations and stringent privacy measures ensure that the platform is inclusive and prioritizes user privacy and security. By empowering stakeholders to actively participate in the improvement process, the human-friendly platform serves as a catalyst for innovation and fosters a collaborative relationship between stakeholders, ultimately leading to more responsive and user-centric outcomes.

Sectoral National AIGE guidelines and code of ethics development

As part of advancing the national guidelines, an essential initiative involves urging sectoral industry players (as promoted in Section 3, Part C) to craft their own industry-specific guidelines of responsible AI, including a code of ethics and capacity building efforts. Recognizing the diverse applications of AI across sectors, these tailored guidelines address sector-specific challenges, ethical considerations, and regulatory requirements, fostering a more contextually relevant approach to AI adoption. By establishing clear ethical standards, industry players ensure that AI technologies align with sector-specific values and stakeholder expectations, enhancing trust and accountability within each sector.

Furthermore, the initiative emphasizes capacity building efforts, encompassing upskilling and reskilling initiatives related to AI. By investing in workforce development programs and educational resources, industry sectoral players empower their workforce to navigate the complexities of AI technologies effectively and responsibly. These capacity building efforts not only drive innovation and competitiveness within sectors but also contribute to the broader goal of promoting responsible AI deployment for societal benefit. Ensuring equitable adoption of AI requires more upskilling on low-code/no-code skills and digital trust, new mindsets such as computational thinking, and an understanding of how to use open-source AI tools.

Integration and synergy of Responsible AI and ESG

An imperative facet of advancing the 7 AI Principles is the integration and synergy of AI governance frameworks within organizations and across industry sectors. Given the pervasive nature of AI applications that transcend traditional sector boundaries, siloed approaches to governance are inadequate. Instead, there's a pressing need to foster collaboration and coherence within internal ecosystems, ensuring that AI governance mechanisms are seamlessly integrated into organizational structures, processes, and decision-making frameworks. Moreover, extending this integration beyond organizational borders to encompass collaboration across industry sectors is essential. By establishing platforms for cross-sector dialogue and cooperation, stakeholders can share insights, best practices, and resources, fostering a collective approach to addressing the ethical, legal, and societal implications of AI deployment.

This acknowledges that AI governance is not solely the responsibility of a single sector or organization but requires collective action and coordination. By promoting integration and synergy, stakeholders can leverage shared expertise and resources to develop robust and adaptable AI governance frameworks. This entails aligning internal policies and practices with external regulatory requirements, industry standards, and emerging ethical norms. Furthermore, it involves fostering a culture of collaboration, where stakeholders collaboratively monitor, evaluate, and evolve AI governance mechanisms in response to evolving challenges and opportunities; so it can navigate the complexities of AI adoption more effectively while upholding the 7 AI Principles.

International harmonization for standard regulation, certification and accreditation of AI product and service

The initiative in advancing responsible AI involves prioritizing international harmonization efforts for cross-regulation. Recognizing the borderless nature of AI technologies and their impact on global society, it is imperative to establish cohesive regulatory frameworks that transcend national boundaries. This initiative emphasizes the need for collaboration among international organizations, governments, and stakeholders to develop harmonized standards and guidelines for responsible AI deployment. By aligning regulatory approaches and principles across jurisdictions, we can address inconsistencies and gaps in current regulations, ensuring that Malaysia's AI governance remains coherent, effective, and equitable on a regional and global scale.

This initiative underscores the importance of leveraging collective expertise and resources to navigate the complex landscape of AI governance. By fostering regional and international dialogue and cooperation, stakeholders can exchange insights, share best practices, and collectively address emerging challenges and opportunities in AI regulation. Moreover, harmonizing global calls for responsible AI regulation enhances interoperability and facilitates the seamless exchange of AI technologies and data across borders, promoting innovation while safeguarding our 7 AI Principles and establish a unified and inclusive framework for AI governance that fosters trust in AI technologies.

List of Abbreviations

MOSTI	-	Malaysian Ministry of Science, Technology and Innovation
AI-RMAP	-	Malaysian National Artificial Intelligence Roadmap
GenAI	-	Generative AI
AGI	-	Artificial General Intelligence
MED4IR	-	National Council of Digital Economy & the Fourth Industrial Revolution
HF-AI	-	The Human Friendly AI
KRSTE.my	-	Knowledge Resource for Science and Technology Excellence, Malaysia
SPAF	-	Single Point Access Facility
MyGDX	-	Malaysian Government Central Data Exchange
E2E	-	End to End
JDN	-	Jabatan Digital Negara
RADARS	-	Raw Database for Research and Science
MASTIC	-	Malaysian Science and Technology Information Center
JDN	-	Jabatan Digital Negara
CGSO	-	Chief Government Security Officer
NACSA	-	National Cyber Security Agency
JPDP	-	Jabatan Perlindungan Data Peribadi
MARA	-	Majlis Amanah Rakyat
MATRADE	-	Malaysia External Trade Development Corporation
MDEC	-	Malaysia Digital Economy Corporation
FMM	-	Federation of Malaysia Manufacturers
MNCs	-	Multinational Corporations
SMEs	-	Small and Medium Enterprises
BNM	-	Bank Negara Malaysia
SC	-	Securities Commission Malaysia
MOHE	-	Malaysian Ministry of Higher Education
MoH	-	Ministry of Health of Malaysia
AG	-	The Attorney General of Malaysia
PDSQ	-	PETRONAS Data Standard and Quality
PDF	-	PETRONAS Data Framework
DSM	-	Department of Standards Malaysia
4IR	-	Fourth Industrial Revolution

List of Figures

Figure 1 : The evolution of AI technology
Figure 2 : The scope of stakeholder and guidelines covering
Figure 3 : AI Innovation ecosystem
Figure 4 : DAMA DMBoK Wheel
Figure 5 : DIKW Pyramid
Figure 6 : AIGE framework and process flow
Figure 7 : AI risk categorization defined by Artificial Intelligence/Machine Learning Risk and Security Working Group (AIRS, 2023)
Figure 8 : AI risk categorization based on EU Artificial Intelligence Act (AIA)
Figure 9 : Responsible AI – AI Governance and Ethics Framework
Figure 10 : AI life cycle stages and AI actors
Figure 11 : PETRONAS Enterprise data hub platform and knowledge ecosystem
Figure 12 : Guidelines AI Governance and Ethics for each AI Socio Economic Drivers (2024-2025)

List of Tables

Table 1 : Example of a Policy Development Checklist related to AI in view with Cyber Security measures
Table 2 : Benchmarking analysis AI guidelines

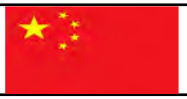
List of Appendices

Appendix 1	Descriptions of Seven AI Principles
Appendix 2	Initiatives on the Formulation of AI Act to Support Responsible AI
Appendix 3	Example of Related Law to Support Responsible AI
Appendix 4	Standard for Responsible Development, Deployment, and Use of AI
Appendix 5	Sample Standard Matrix to Mitigate Risks
Appendix 6	Overview of the RAIIA Tools Used for Responsible AI Impact Assessment
Appendix 7	Example of Performance Index for Each AI Principle
Appendix 8	Example of Code of Ethics

Appendix 1: Descriptions of Seven AI Principles

AI Principle	Description	Examples
Fairness	Ensuring AI fairness is vital. AI should provide equitable guidance in domains like healthcare, loans, and employment, making unbiased recommendations for individuals with similar attributes to uphold dignity and respect	1. Have you formulated a strategy or a set of procedures to prevent the creation or reinforcement of unfair bias within the AI system, both about input data usage and algorithm design? 2. Have you conducted assessments and acknowledged potential limitations stemming from the composition of the utilised datasets? 3. Did you consider the diversity and representativeness of users within the data, including testing for specific populations or problematic use cases?
Reliability, Safety & Control	AI systems must prioritise reliability and safety. Their complexity has raised concerns about potential risks and misuse. Trust in AI depends on its consistent and secure performance, even in unforeseen situations. This is crucial in healthcare and finance, where decisions have significant consequences.	1. Have you ensured that your system has a robust fallback plan to address adversarial attacks or other unforeseen situations, including technical switching procedures or requesting human intervention before proceeding? 2. Did you carefully evaluate the level of risk posed by the AI system in this specific use case, taking into account potential consequences and mitigations? 3. Have you conducted an assessment to determine whether there is a reasonable probability that the AI system might cause harm or damage to users or third parties?
Privacy & Security	For individuals to willingly share their data, they must trust that it will be securely stored, used responsibly, and for ethical purposes. AI systems must adhere to privacy laws governing data collection, usage, and storage, with robust safeguards in place to protect personal data from malicious intent, including data theft and harm	1. Did you consider different types and natures of vulnerabilities, such as data pollution, physical infrastructure, and cyber-attacks? 2. Did you put measures or systems in place to ensure the integrity and resilience of the AI system against potential attacks?
Inclusiveness	AI systems should be inclusive, benefiting all and meeting diverse human needs. They can empower individuals with disabilities, opening up opportunities in education, employment, and public services. This enhances their well-being, socioeconomic status, and societal participation	1. How can we ensure that AI and technology are designed to be inclusive and accessible to individuals of all backgrounds, abilities, and demographics? 2. What steps can be taken to bridge the digital divide and ensure that marginalized or underserved communities have equal access to the benefits of AI and technology? 3. How can AI and technology contribute to greater inclusion in education, employment, and other societal domains for individuals with disabilities?
Transparency	Transparency is vital, as it fosters trust and reduces suspicion. In Malaysia, people highly value organizations openly sharing how they use personal data. Malaysians are generally more open to data usage by both private and government entities and seek a clear understanding of associated risks, surpassing the global average.	1. Have you implemented measures to guarantee traceability, including documenting the following procedures? a. Have you ensured an explanation for why the AI system made specific choices leading to particular outcomes, making it understandable for all users? b. Have you informed (end-)users, whether through disclaimers or other means, that they are interacting with an AI system and not a human? Have you properly labelled your AI system? 2. Have you established mechanisms to keep (end-)users informed about the rationale and criteria behind the AI system's outcomes?
Accountability	Those who create and implement AI systems must take responsibility for their functioning. Drawing from other sectors like healthcare, we can establish standards and best practices. Internal review boards can offer guidance and oversight for AI system development and deployment.	1. Have you implemented mechanisms to enable the auditability of the AI system, including ensuring traceability and logging of the system's processes and outcomes? 2. Did you conduct a comprehensive risk or impact assessment of the AI system, considering various stakeholders who may be directly or indirectly affected? 3. Have you established processes that allow third parties, such as suppliers, consumers, distributors, vendors, or workers, to report potential vulnerabilities, risks, or biases within the AI system?
Pursuit of Human Benefits & Happiness	AI is fundamentally a tool designed to enhance human well-being. By incorporating the aim of improving human happiness and quality of life into our national AI Ethics charter, w can systematically address each AI principle to resolve societal issues and enhance overall well-being in the country	1. How can we use technology and innovation to enhance the overall well-being and happiness of individuals and communities? 2. What role can AI and automation play in addressing societal challenges and improving the quality of life for people? 3. In what ways can AI and technology contribute to the pursuit of human benefits, particularly in areas like healthcare, education, and sustainability?

Appendix 2: Initiatives on the Formulation of AI Act to Support Responsible AI

EUROPEAN UNION (EU)	USA	 CHINA
Implementation of the EU AI Act EU have successfully endorsed a comprehensive regulatory framework for AI. This legislative milestone significantly shapes the regulatory landscape, provide clear guidelines and standards for the development and deployment of AI systems across member states. **Emphasis on High-Risk Systems to support Ethical AI** In 2024, the classification of high-risk AI systems, along with the prescribed requirements for compliance and ensuring responsible AI development and usage within the European Union will be fully adopted. **Governance Framework and Innovation Measures** The establishment of a governance framework divided between the EU AI Board and national authorities, coupled with innovation measures, indicates a holistic approach to fostering a robust AI ecosystem. The formation of codes of conduct for non-high-risk AI systems and confidentiality requirements for authorities dealing with proprietary data may contribute to a balanced and responsible AI landscape in the European Union. In 2024, it anticipates the EU AI Act contributing to the growth of innovation in AI technologies while maintaining a well-regulated environment.	**Towards Comprehensive AI Regulation** In 2024 will move closer to the establishment of comprehensive AI regulation. In the middle of drafting potential acts namely the Algorithmic Accountability Act, National AI Commission Act, Digital Platform Commission Act, and Transparent Automated Governance Act. **Strengthening International Collaboration in AI Research** The emphasis on a principled and coordinated approach to international collaboration in AI research where in 2024, the country will play a pivotal role in fostering global cooperation. Engagement with international partners and the development of frameworks for ethical and responsible AI research on a global scale. **Enhanced Public Input and Trust in AI Systems.** The Biden administration's initiatives, such as the request for information on AI's impact and the National Telecommunications and Information Administration's AI Accountability Policy Request for Comment, indicate a commitment to public input and trust-building in AI systems. These initiatives are expected to have had a tangible impact on shaping policies, foster transparency, and build public trust in AI technologies.	**Maturation of China's Comprehensive AI Regulation** China, as one of the pioneers in implementing AI regulations and will witness the maturation of its comprehensive AI regulation process. In 2024, anticipates the formalization and enactment of this regulation which provide a consolidated and overarching framework for governing various aspects of AI development and deployment within the country. **Evolution of Specific AI Regulations** The existing regulations such as the Algorithmic Recommendation Management Provisions and the Interim Measures for the Management of Generative AI Services are already in force. These regulations will have evolved and matured, with possible amendments and additional guidelines to address emerging challenges in their respective domains. **Deep Synthesis Management Provisions: Integration into AI Regulatory Framework** The draft for Deep Synthesis Management Provisions implies a forthcoming regulation specific to deep synthesis in AI. This draft to have progressed to a more mature stage, potentially being officially implemented in 2024. This development will signify China's focus on staying ahead in regulating advanced AI technologies, particularly in areas like deep synthesis, to ensure responsible and ethical use.

Appendix 2: Initiatives on the Formulation of AI Act to Support Responsible AI

AUSTRALIA	UK	JAPAN
Anticipated Regulatory Developments in Australian AI Governance Given the absence of specific AI-related laws in Australia, it is foreseeable In 2024, the government may enact new regulations tailored specifically to AI governance. This could stem from the growing importance and impact of AI technologies, prompting a need for a dedicated legal framework to address emerging challenges and ensure responsible AI use. **Evolving Landscape: The Impact of The State of AI Governance Report** In 2024 the Human Technology Institute to release "The State of AI Governance in Australia", a notable shift in the approach to AI governance. The findings from this report could influence policymakers, leading to the development of targeted strategies and fostering a more robust AI governance ecosystem. **Strategic Emphasis on AI in Key Sectors** Australia's AI Roadmap, with its focus on high-potential areas like natural resources and infrastructure, indicates a strategic vision for AI development. In 2024, this roadmap will be translated into tangible advancements in natural resource management and infrastructure development, potentially transforming these industries and contributing to overall economic growth.	**Context-Based AI Regulation in the U.K. by 2024** This year, the U.K. may have adopted and implemented innovative regulatory model. This approach indicates a nuanced and flexible framework that considers the specific context in which AI systems are deployed, providing a more adaptable and effective means of regulating emerging technologies. **Guardrails within Existing Sectoral Laws: Integration and Impact** In 2024, the U.K. will have successfully integrated AI-related considerations into various industry-specific regulations which include the tangible impact of these guardrails, ensuring responsible and ethical AI practices across sectors while avoiding the imposition of burdensome and redundant regulations. **Pro-Innovation Regulatory Environment in Fostering Growth and Ethical Practices** the U.K. will maintain a regulatory environment that encourages innovation in AI technologies. The availability of resources such as "A pro-innovation approach to AI regulation" where a balance between fostering growth in the AI sector and maintaining ethical practices, may reflect the government's commitment to supporting technological advancement while safeguarding against potential risks.	**Evolution of Agile Governance in Japan's AI Landscape by 2024** Japan's promotion of "agile governance" suggests that by this year, the country may witness the continued evolution of its approach to AI regulation. The government's role in providing nonbinding guidance and deferring to the private sector's voluntary efforts indicates a flexible regulatory environment, possibly adapting to emerging technologies and industry practices while maintaining ethical standards. **Continued Emphasis on Self-Regulation: Voluntary Efforts in the Private Sector** The preference for the private sector's voluntary efforts for self-regulation, as seen in Japan's current approach, is likely to persist in 2024. This prediction anticipates ongoing collaboration between the government and industry stakeholders, with the private sector taking a proactive role in shaping and adhering to AI principles and governance guidelines. **Guiding Policies: Updates and Iterations for AI Governance** The issuance of white papers such as "AI Governance in Japan Ver. 1.1" and "Governance Guidelines for Implementation of AI Principles" suggests that by this year, Japan will continue to refine its guiding policies for AI governance. Updates and iterations to these documents may reflect the evolving landscape of AI technologies, incorporating new insights and lessons learned to ensure effective and ethical governance in the Japanese context

Appendix 3: Example of Related Law to Support Responsible AI

AI Principle	Related Law / Regulations In Malaysia
Fair	Consumer Protection Act 1999 – unfair trade practices, Malaysian Code on Corporate Governance (MCCG), Industrial Relations Act 1967 [Act 177], Employment Act 1955 [Act 265]
Reliability Safety & Control	Digital Signature Act 1997 [Act 562]. Consumer Protection Act 1999 [Act 599]. Electronic Commerce Act 2006 [Act 658]
Privacy & Security	Personal Data Protection Act 2010 [Act 709] , Computer Crimes Act 1997 [Act 563], Communications and Multimedia Act 1998 [Act 588]
Inclusiveness	Social Inclusion Act 2014
Transparency	Malaysian Anti-Corruption Commission Act 2009 [Act 694]; Electronic Government Activities Act 2007 [Act 680]
Accountability	Intellectual Property Corporation Of Malaysia Act 2002 [Act 617], Telemedicine Act 1997 [Act 564]
Human Rights & Happiness	Human Rights Commission of Malaysia Act 1999 [Act 597]

Appendix 4: Standard for Responsible Development, Deployment

4.1 ISO Standard for Responsible Development, Deployment

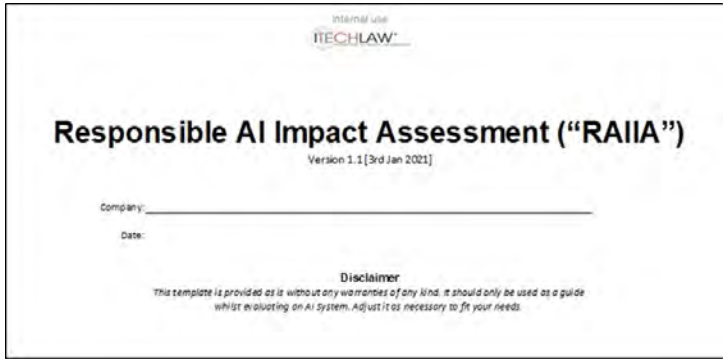
1. **ISO/IEC TS 4213:2022** Information technology Artificial intelligence -Assessment of machine learning classification performance
2. **ISO/IEC 5338:2023** Information technology Artificial intelligence-AI system life cycle processes
3. **ISO/IEC 5339:2024** Information technology Artificial intelligence-Guidance for AI applications
4. **ISO/IEC TR 5469:2024** Artificial intelligence- Functional safety and AI systems
5. **ISO/IEC 8183:2023** Information technology- Artificial intelligence -Data life cycle framework
6. **ISO/IEC 22989:2022** Information technology Artificial intelligence concepts and terminology
7. **ISO/IEC 23053:2022** Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
8. **ISO/IEC 23894:2023** Information technology Artificial intelligence Guidance on risk management
9. **ISO/IEC TR 24027:2021** Information technology Artificial intelligence (AI) Bias in AI systems and AI aided decision making
10. **ISO/IEC TR 24028:2020** Information technology Artificial intelligence Overview of trustworthiness in artificial intelligence
11. **ISO/IEC TR 24029-1:2021** Artificial Intelligence (AI) Assessment of the robustness of neural networks Part 1: Overview
12. **ISO/IEC 24029-2:2023** Artificial intelligence (AI) Assessment of the robustness of neural networks Part 2: Methodology for the use of formal methods
13. **ISO/IEC TR 24030:2021** Information technology Artificial intelligence (AI) Use cases
14. **ISO/IEC TR 24368:2022** Information technology Artificial intelligence Overview of ethical and societal concerns
15. **ISO/IEC TR 24372:2021** Information technology Artificial intelligence (AI) Overview of computational approaches for AI systems
16. **ISO/IEC 24668:2022** Information technology Artificial intelligence Process management framework for big data analytics
17. **ISO/IEC 25059:2023** Software engineering Systems and software Quality Requirements and Evaluation (SQuaRE) Quality model for AI systems
18. **ISO/IEC 38507:2022** Information technology Governance of IT Governance implications of the use of artificial intelligence by organizations
19. **ISO/IEC 42001:2023** Information technology Artificial intelligence Management system
20. **ISO/IEC 20546:2019** Information technology Big data Overview and vocabulary
21. **ISO/IEC TR 20547-1:2020** Information technology Big data reference architecture Part 1: Framework and application process
22. **ISO/IEC TR 20547-2:2018** Information technology Big data reference architecture Part 2: Use cases and derived requirements
23. **ISO/IEC 20547-3:2020** Information technology Big data reference architecture Part 3: Reference architecture
24. **ISO/IEC TR 20547-5:2018** Information technology Big data reference architecture Part 5: Standards roadmap

Appendix 5: Sample Standard Matrix to Mitigate Risks

Type of Standard	How applicable to AI?	Where are the standards applied?	How can it reduce AI risk from an adversary?
Analytic & Research	Standards that evaluate the quality of analysis and scrutability of algorithms	Back end explainability and transparency	• Identify faulty logic or reasoning, increase the difficulty of deceiving and/or manipulating analysis from AI • Determine how much to trust system inputs and outputs
Legal & Regulatory	Standards based on governance and regulatory oversight into preserving privacy and consent	Front end usability and personalization. Back ends standardized architecture	• Change understanding of liability for mistakes and enhance attribution • Transform notion of jury of peers and evolve crime and punishment
Moral & Ethical	Standards that prevent AI from performing actions that are contrary to a moral or ethical norm	Back end failsafes	• Reduce likelihood that AI will do the 'wrong thongs' (i.e.: immoral or unethical behavior) of exploited or infiltrated by and adversary
Technical & Industry	Standards to measure the performance of an algorithm on relevant tasks	Front end performance	• Meet appropriate technical specifications (e.g.: low number of false positives) to be robust against adversary denial and deception activities
Data & Information Security	Standards for the protection sharing or use of data relevant to a task	Front end training Back ends integrity and availability	• Limiting access to and information about how an AI system works to appropriate people could help prevent exploitation by an adversary • Preventing manipulation of training data

Application of a standard or combination of standards such as analytic, research, legal, regulatory, moral, ethical, technical, industry, data, and information security that can help to reduce the risk of adversary exploitation. It also include a description of how it is applicable to AI and how it can reduce AI risk.

Appendix 6: Overview of the RAIIA Tools Used for Responsible AI Impact Assessment



Internal use

ITECHLAW

Responsible AI Impact Assessment ("RAIIA")

Version 1.1 [3rd Jan 2021]

Company: _____

Date:

Disclaimer

This template is provided as is without any warranties of any kind. It should only be used as a guide whilst evaluating an AI System. Adjust it as necessary to fit your needs.

Example of Components

1. Project Summary
2. Key factors for Conducting RAIIA
3. Main Assessment
   a. Principle 1: Ethical Purpose and Societal Benefits
   b. Accountability
   c. Transparency and Explainability
   d. Fairness and Non Discrimination
   e. Safety and reliability
   f. Open Data , Fair Competition and Intellectual Property
   g. Intellectual Property
   h. Privacy
4. Risk Assessment Summary

Content of the RAIIA tools

	Factors to Evaluate Need for RAIIA		
Context	1. Describe the context in which the AI System will be used or deployed.	Human Oversight	14. Will the AI System make or participate in making decisions with material impacts on individuals or society?
Context	2. Will the use of the AI System be citizen-facing?	Human Oversight	15. What is the expected degree of autonomy of the AI System? Will, for instance, human operators or decision-makers have oversight on individual AI decisions, if any?
Context	3. What is the market, industry or sector targeted?	Human Oversight	16. How frequently will there be human oversight over the operation of the AI System?
Laws and regulations	4. Do the jurisdiction(s) in which the AI Solution will be deployed have data protection laws or regulation that are applicable to its use?	Human Oversight	17. What measures would be taken to avoid automation bias or anchoring to the AI System?
Laws and regulations	5. Does the jurisdiction(s) in which the Project will take place abide by rule of law principles?	Human Oversight	18. What will be the Organisation's degree of control and responsibility over the finalized AI System?
Laws and regulations	6. Does this jurisdiction have antidiscrimination laws?	Data and Privacy	19. What is the type and origin of the data that will be used to train the AI System?
Laws and regulations	7. What are the main regulatory requirements relevant to the use and deployment of the AI System within the targeted market, industry or sector?	Data and Privacy	20. Will the training data include personal information?
Laws and regulations	8. Will the AI System be used across legal jurisdiction borders (whether they be across federal states or national borders)?	Data and Privacy	21. If personal information are used in the context of the AI System, who are the data subjects?
Laws and regulations	9. What are the main ethical concerns relevant to the use and deployment of the AI System for the targeted market, industry or sector?	Data and Privacy	22. What is the level of sensitivity of the data in term of privacy?
Stakeholders	10. Who will be the main stakeholders affected by the AI System?	Human-understandable AI	23. What are the technical characteristics of the AI System that could influence the explainability and auditability of the algorithm?
Stakeholders	11. Who are the expected contributing third parties?	Human-understandable AI	24. Can the results of the AI System be explained in humanly understandable terms?
Stakeholders	12. What individual rights and interests will be at stake as a consequence of the use of the AI System?		
Stakeholders	13. Are those rights fundamental or human rights?		

Appendix 7: Example of Performance Index for Each AI Principle

FAIRNESS INDEX

E.g Demographic Parity:

This metric assesses whether the distribution of outcomes or predictions is the same across different demographic groups. It can be measured using the difference in acceptance rates, impact, or other relevant factors for various groups.

SAFETY INDEX

E.g Environmental Safety Index:

These indices assess the safety of the environment, considering factors like air and water quality, pollution levels, and ecosystem health.

PRIVACY INDEX

E.g User-Centric Assessments:

Gathering feedback from users or affected individuals about their perceived privacy and consent experiences can be crucial in evaluating an AI system's privacy performance.

E.g Accountability Assessments:

Examine the proper functioning throughout AI lifecycle. AI system that designed, developed, operated or deployed, accordance with their roles and applicable regulatory frameworks.

ACCOUNTABILITY INDEX

INCLUSIVENESS INDEX

E.g Diversity and Representation Index:

This index evaluates how diverse and representative the training data and the development team are. It looks at whether different groups, including underrepresented minorities, are adequately represented and whether the team designing AI systems is diverse.

minorities

TRANSPARENCY INDEX

E.g Transparency Documentation:

Examines the availability of comprehensive documentation, transparency reports, and disclosure of AI system capabilities and limitations.

HUMAN CENTERED & HAPPINESS INDEX

E.g Happy City Index:

This index focuses on the urban environment and measures factors that contribute to the well-being and happiness of city residents, such as transportation, education, housing, and the environment.

Some potential measurements to be considered

Appendix 8: Example of Code of Ethics

Scope of AI Usage

AI is generally the use of technology to carry out intelligent behaviour based on input into a computer system. AI is sometimes developed by training the AI to act in certain ways based on data, and AI often involves data processing, and automatic or semi-automatic decisioning. AI may be used internally in business as well as externally as part of any products or services and can range from physical robots, including human-like robots, industrial robots and autonomous vehicles, to more digital AI, such as chatbots, image recognition and analytic tools. Because certain AI may operate automatically or semi-automatically, it can cause physical or reputational harms or infringe on a person's rights. Given the potential impact of the adoption of AI and potential misuse of AI, we need to develop and use AI in a human-centered manner, and in a responsible, secure, accountable, transparent, fair and ethical manner.

AI in the Workplace and Fair Working Conditions

Developers or Adopters of AI should promote fair working conditions, free from any form of discrimination. When AI is used in the workplace, it should be fair. AI in the workplace could include automated or semi-automated employment decisions, such as hiring, promotion, or employee monitoring. Before using AI in the workplace, AI Developers/ Owners should:

- Ensure that the use of AI in the workplace is consistent with human relations policies.
- Ensure that the use of the AI is transparent and that employees are notified that is AI is being used.
- Respect employees' privacy and confidence as required by law.
- Ensure that AI actions are free from bias

Health, Safety, Environment and Product Integrity

Developers/Adopters of AI should promote **health, safety, and environmental protection**. AI Owners should:

• Ensure that the AI has safeguards against uncontrolled behavior (especially for fully or semi-automated physical robots) that may cause harm.
• Ensure that the AI meets legal safety requirements.
• Provide the necessary instructions, warnings and safety equipment for products that may use AI.

The **product integrity** of AI should be ensured by design. AI Owners should:

• Develop and apply strong safety and secure practices to avoid unintended results that can create risks and cause harm
• Ensure that the AI is well designed, developed, and tested with appropriate training data.
• Understand the scope of AI including the business case and geography under which the AI is intended to be used.
• Evaluate the potential harm that could occur from the use of the AI and also the processes to mitigate such harm.
• Test AI technologies in constrained environments and monitor their operation after deployment.
• Involve relevant stakeholders in the development of the AI.

Usage of AI that Complies with Laws, Regulations, Standards and Instructions

To comply with the laws and regulations applicable in the countries of operation, it is necessary to understand how the use of AI may be regulated. As with any technology, there may be a number of legal concerns (in the area of data protection, data security, intellectual property, safety, employment, product liability and more), and the relevant policies related to those areas should be reviewed to ensure that use of the AI complies with those policies, including the Code of Conduct.

Diversity, Non-Discrimination and Fairness in the Use of AI

The development, deployment and use of AI systems must be fair to avoid the potential for unfair bias or discrimination. For this reason, AI developers/ owners should:

- Be aware of the potential for AI to unintentionally promote unfair bias, disparate impacts, or discrimination due to improper programming or training.
- Be vigilant of the use of data which may introduce unfair bias, including limited or skewed data sets.
- When using an AI tool to filter potential candidates for employment interviews , determine whether the developer is trained with data that may cause discrimination based on gender, age, race, skin color, social or ethnic origin, nationality, sexual orientation, ability, religious or political beliefs, or any other characteristics protected by applicable law.
- Identify potential harms to certain segments of the populations and implement controls to mitigate the risk of harm.

Data Protection and Cybersecurity in the Usage of AI

Data is an integral part of any AI system, and every developer is committed to respecting people's rights to privacy and personal data.

- Ensure that AI involving personal data is compliant with data protection laws, including where applicable, the general Data Protection Regulation (GDPR) and any other applicable data protection legislation in force at the time.
- For example, under the GDPR, data subjects shall have the right not to be subjected to a decision based solely on automated processing, including profiling, which produces legal effects concerning them under the GDPR.
- To employ security practices including encryption and access control methodologies to prevent users' data from unauthorized theft, misuse or destruction.
- Developers of AI should be transparent, including striving to explain to users, where appropriate, why an AI made its decisions. Transparency shall also mean disclosing to users that they are interacting with an AI.

Responsibility and Accountability

AI developers/owners of AI should be accountable and controlled. Those who are responsible for developing or adopting AI need to understand these guidelines and be responsible for the training of their staff on these guidelines. In general, AI Developers /AI Owners should:

- Be vigilant in inquiring where the data are acquired from and what are the terms of use that apply and their implications, if any third-party data are used for AI training.
- Ensure that use of the AI is aligned with these ethical guidelines.
- Require vendors to apply these guidelines before acquiring and deploying AI solutions and services.
- Ensure that is human accountability for the training, development and use of the AI.
- Ensure that are controls in place for people to change the AI's behavior to prevent or reduce harm, particularly where human empathy and judgement may be needed.
- Ensure the AI is subject to monitoring and periodic reviews.
- Have a general understanding of how the AI makes its decisions. When an AI's decision cannot be (easily) understood or explained, increased care should be used in the deployment of the AI

Confidential Information, Intellectual Property and Our AI

Everybody is responsible for protecting confidential business information and trade secrets ("Confidential Information") and intellectual property embodied and processed in AI.

With regard to all information and intellectual property including trade secrets, copyrights, databases, patents, software, algorithms, data, enhancements to the AI algorithm and the outputs of the AI, AI Developers/Owners should:

- Ensure that Confidential Information and intellectual property embodied in AI are held in strict confidence.
- Ensure that the Confidential Information that may be processed by the AI, including any outputs, are protected appropriately.
- Respect the Confidential Information embodied in clients and or suppliers' AI and only use them in accordance with applicable laws and/or contractual obligations.

Acknowledgement

Ministry of Science, Technology and Innovation	
Ministry of Digital	
AP. Dr. Aini Suzana Hj. Ariffin	Universiti Teknologi Malaysia (UTM)
AP. Nor Ashikin Mohamed Yusof	
Dr. Shazwin Mat Taib	
Prof. Dr. Zuraidah Mohd Don	
Dr. Norasikin Saman	
Nurul Nazleatul Najiha Mohd Nazif	
Mohd Rizalman Mohd Ali	KSK Group Berhad
Dr. Hon Hock Woon	MIMOS Berhad
Calvin Woo	MyDIGITAL Corporation
Mohamad Roshidi Bin Rozali	National Cyber Security Agency (NACSA)
Ts. Mohamed Kheirulnaim Mohamed Danial	
Nurul Aisyah Sim Abdullah	Jabatan Digital Negara (JDN)
Ts. Dr. Siti Istianah Mahdzur	
Malini Ramalingam	Malaysian Communications & Multimedia Commission (MCMC)
Jacinta Felix	
Athari Ismail	
Dr. Muhamad Sufyan Basni	Department Of Personal Data Protection (JPDP)
Ts. Tengku Azrul Tengku Azhar	Malaysia Digital Economy Corporation (MDEC)
Chan Chuey Hwye	
Ts. Nurul Haszeli Ahmad	SIRIM Berhad
Dr. Muhammad Azwan Ibrahim	National Metrology Institute of Malaysia,SIRIM Berhad
Ts. Nurul Huda Abdul Rahman	Akademi Sains Malaysia
Sabariah Binti Ahmad	CyberSecurity Malaysia (CSM)
Saiful Azri Mohamad	Malaysian Anti-Corruption Commission (MACC)

Noor Akmal Mustafa Kamal	Jabatan Peguam Negara (AGC)
Norhaliza Arifin	Jabatan Standard Malaysia (JSM)
Ilyana Norsaidah Ab Rahman	Perbadanan Produktiviti Malaysia (MPC)
Prof. Ts. Dr. Salwani Mohd Daud	University Malaysia of Computer Science & Engineering (UNIMY)
AP. Dr. Jaspal Kaur Sadhu Singh	Canterbury Christ Church University ,UK
Lim Tao Cheng (David)	WISE AI Sdn Bhd
Dr. Lim Chern Loon	
Abdul Fattah Yatim	Blockchain Consultant
Azree Nazri	Artificial Intelligence Society (ARTIS) Malaysia
Datin Anna Amalina Imam Baweh	Intel Corporation
Tan Kuan Thye	TIME dotCom
Camy Low	Anhsin Technology
Mohamad Syahir Mohamad Zaini	ALSPACE
Gan Wei Yuan	Asolute
Adilah Junid	Microsoft
Datin Habsah Nordin	PETRONAS Berhad
Zamri Abdul Halim	Innovation Associates Consulting (IAC)
Dr. Jun-E Tan	Khazanah Research Institute (KRI)
Datuk Indrani Thuraisingham	Federation of Malaysian Consumers Association (FOMCA)
Ong Kian Yew	PIKOM
Nurul Asyiqin Mohd Nasir	
Yan Xi Chan	Salesforce
Jeff Paine	Asia Internet Coalition
Amb. Brian D.McFeeters	US-ASEAN Business Council, Inc

References

AECOM (2024) Effective data governance: a key enabler for AI adoption. https://aecom.com/without-limits/article/effective-data-governance-a-key-enabler-for-artificial-intelligence-adoption/

Artificial Intelligence/Machine Learning Risk & Security Working Group, AIRS (2023) Artificial Intelligence Risk & Governance. https://ai.wharton.upenn.edu/white-paper/artificial-intelligence-risk-governance/

Deloitte's (2021) Global Human Capital Trends: Special report. https://www.deloitte.com/content/dam/assets-shared/legacy/docs/perspectives/2022/2021-deloitte-global-millennial-survey-report.pdf.html?icid=learn_more_content_click

European Commission (2019), Ethics Guidelines for Trustworthy AI, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

European Parliament (2023) EU AI Act: first regulation on artificial intelligence. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

Forbes (2020) Roundup Of Machine Learning Forecasts And Market Estimates, 2020. https://www.forbes.com/sites/louiscolumbus/2020/01/19/roundup-of-machine-learning-forecasts-and-market-estimates-2020/?sh=2b7c3d6d5c02

Fortune Business Insights (2020) Artificial Intelligence Market to Reach USD 266.92 Billion by 2027; Increasing AI Technology Users to Spur Market Growth, https://www.globenewswire.com/en/news-release/2021/05/07/2225220/0/en/Artificial-Intelligence-Market-to-Reach-USD-266-92-Billion-by-2027-Increasing-AI-Technology-Users-to-Spur-Market-Growth-Fortune-Business-Insights.html

Francis Heylighen (2017) Towards an intelligent network for matching offer and demand: From the sharing economy to the global brain. *Technological Forecasting and Social Change*, 114, 74–85.

Jacques Bughin, Jeongmin Seong, James Manyika, Michael Chui, and Raoul Joshi (2018) Notes from the AI frontier: Modeling the Impact of AI on the World Economy, *McKinsey Global Institute*, Discussion Paper. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy

Japan (2022). Governance Guidelines for Implementation of AI Principles (ver. 1.1) Expert Group on How AI Principles Should be Implemented AI Governance Guidelines WG. https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_2.pdf

Jaspal Kaur Sadhu Singh (2023), 'The values of an AI ethical framework for a developing nation: considerations for Malaysia' in Mark Findlay, Li Min Ong, and Wenxi Zhang (eds), *Elgar Companion to Regulating AI and Big Data in Emerging Economies*, Edward Elgar, 115-134. https://doi.org/10.4337/9781785362408.00013 Jaspal Kaur Sadhu Singh and Darmain Segaran (2021), An AI Ethical Framework for Malaysia: A Much-Needed Governance Tool for Reducing Risks and Liabilities Arising from AI Systems, *Malayan Law Journal*, 3MJL, cdxiv

Medium (2023). Open-Source AI Data Sharing: yes! Data Colonialism: no!. https://medium.com/@openforgood/open-source-ai-data-sharing-yes-data-colonialism-no-3062a922de03

Ministry of Science, Technology & Innovation, MOSTI (2021) Malaysia National Artificial Intelligence Roadmap 2021-2025 (AI-RMAP). https://airmap.my/apps.

National Institute of Standards and Technology, NIST. Using Standards to Mitigate Risks. https://www.nist.gov/document/nist-ai-rfi-dhs-001pdf

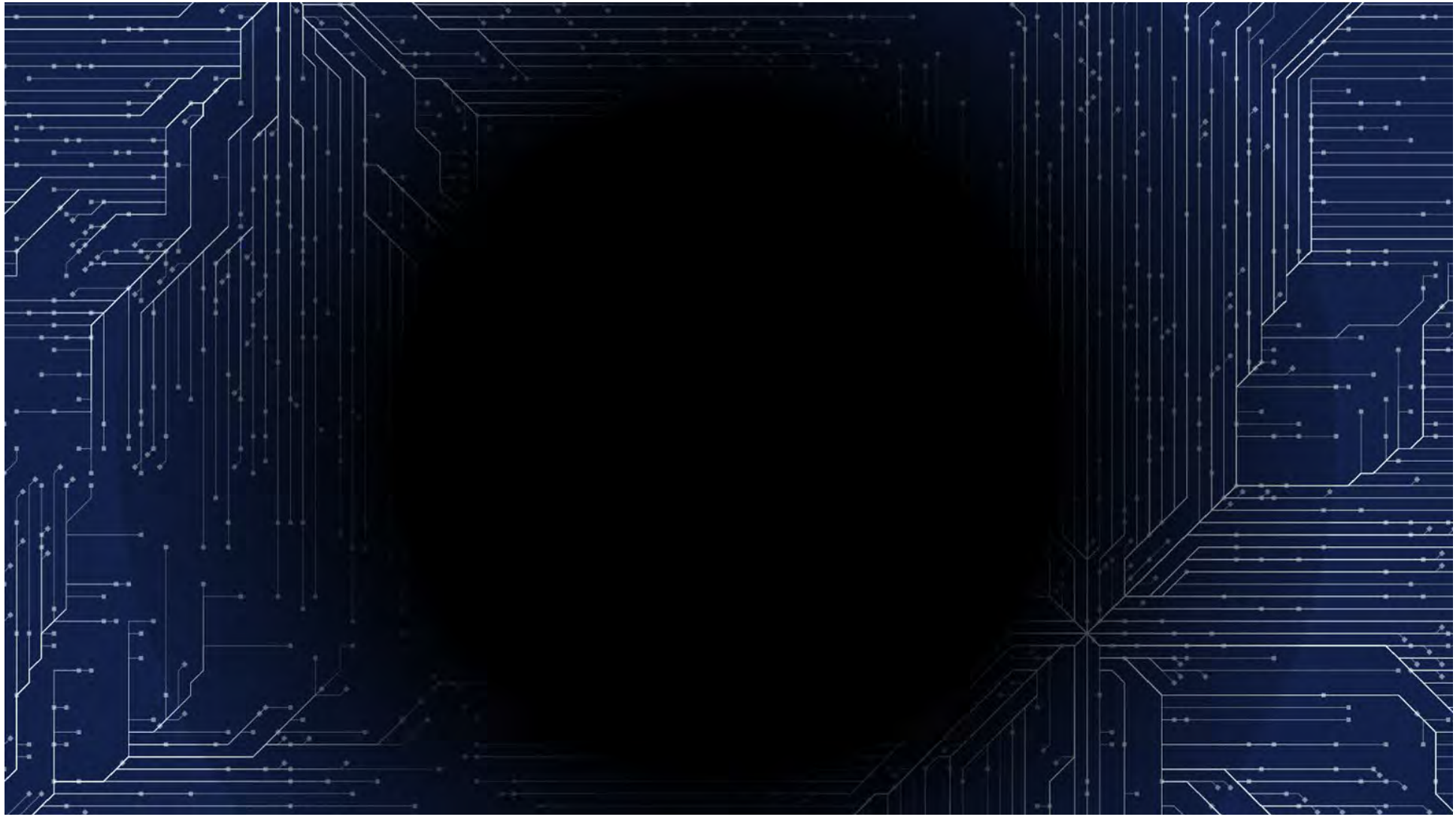
OECD (2019). Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#relatedInstruments

The Alan Turing Institute (2019). AI Ethics and Governance in Practice.https://www.turing.ac.uk/research/research-projects/ai-ethics-and-governance-practice

The Legal 500 (2021) Overview Of Artificial Intelligence in Malaysia. https://www.legal500.com/developments/thought-leadership/overview-of-artificial-intelligence-in-malaysia/

UK (2020) AI procurement guidelines, Office for AI and World Economic Forum. https://www.gov.uk/government/publications/guidelines-for-ai-procurement

UNESCO (2021). Recommendations of AI Ethics, https://unesdoc.unesco.org/ark:/48223/pf0000381137

White House Office of Science and Technology Policy (2022). Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People, https://www.whitehouse.gov/ostp/ai-bill-of-rights

Yong Qin, Zeshui Xu, Xinxin Wang, and Marinko Skare (2023) Artificial Intelligence and Economic Development: An Evolutionary Investigation and Systematic Review. *Journal of the Knowledge Economy*,1–35. doi: 10.1007/s13132-023-01183-2