

# Guidance on the Ethical Development and Use of Artificial Intelligence



PCPD



HK



PCPD.org.hk



香港個人資料私隱專員公署  
Office of the Privacy Commissioner  
for Personal Data, Hong Kong



守護 · 私隱 · 廿五載  
GUARDIAN · PRIVACY · 25 YEARS



# TABLE OF CONTENTS



<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data Stewardship Values</b>	<b>7</b>
2.1	Being Respectful	7
2.2	Being Beneficial	7
2.3	Being Fair	7
<b>3</b>	<b>Ethical Principles for AI</b>	<b>8</b>
3.1	Accountability	8
3.2	Human Oversight	8
3.3	Transparency and Interpretability	8
3.4	Data Privacy	9
3.5	Fairness	9
3.6	Beneficial AI	9
3.7	Reliability, Robustness and Security	9
<b>4</b>	<b>Practice Guide</b>	<b>11</b>
4.1	AI Strategy and Governance	12
4.1.1	AI Strategy	12
4.1.2	Governance Structure	13
4.1.3	Training and Awareness Raising	15
4.2	Risk Assessment and Human Oversight	16
4.2.1	Risk Factors to Consider	18
4.2.2	Determining the Level of Human Oversight	20
4.3	Development of AI Models and Management of AI Systems	22
4.3.1	Data Preparation for AI	23
4.3.2	Development of AI Models	26
4.3.3	Management and Monitoring of AI Systems	27
4.4	Communication and Engagement with Stakeholders	30
<b>5</b>	<b>AI Systems Provided by Third Parties</b>	<b>32</b>
<b>6</b>	<b>Closing Remarks</b>	<b>33</b>
	<b>Appendix A - Self-assessment Checklist</b>	<b>34</b>
	<b>Appendix B - Data Protection Principles under the Personal Data (Privacy) Ordinance</b>	<b>38</b>
	<b>Appendix C - References</b>	<b>40</b>

# 1 INTRODUCTION



## WHAT IS ARTIFICIAL INTELLIGENCE?

Artificial intelligence (“AI”) refers to a family of technologies that involve the use of computer programmes and machines to mimic the problem-solving and decision-making capabilities of human beings. Examples of AI applications include image recognition, speech recognition, chatbots, data analytics and automated decision-making or recommendation. AI technologies are still evolving, and more new applications may emerge.

## SCOPE AND OBJECTIVES OF THIS GUIDANCE

Personal data is commonly used in the development and use of AI. This Guidance on the Ethical Development and Use of AI (“Guidance”) applies to the development or use of AI systems that involve the use of personal data or the identification, assessment or monitoring of individuals, either of which would potentially impact the privacy of individuals in relation to personal data.

The objectives of this Guidance are to facilitate the healthy development and use of AI in Hong Kong and assist organisations in complying with the provisions of the Personal Data (Privacy) Ordinance (Cap. 486) (“PDPO”) in their development and use of AI.

In this Guidance, the term ‘AI’ refers generally to the technology; ‘AI models’ refers to the mathematical algorithms that were built and trained on datasets; and ‘AI system’ refers to the substantive programme used by organisations to assist their operations. While the terms vary in meaning, they could be understood interchangeably in this Guidance. The values, principles and practices suggested in this Guidance would not be substantially affected by the use of the terms.

Appendix A to this Guidance is a Self-assessment Checklist that assists organisations to determine whether the practices recommended in the Guidance have been adopted in their development and use of AI.



## BENEFITS OF AI

Organisations (including business entities, government departments and public bodies) increasingly use AI in their operations. Banks use AI to assess the creditworthiness of their customers and detect money laundering activities. Healthcare providers use AI to analyse medical records and assist doctors in diagnoses. Government departments use AI to monitor and optimise road traffic in order to reduce congestions. Other organisations also use AI to assess the resumes of job applications and respond to customers' enquiries, etc. AI presents huge opportunities and benefits to different sectors by saving manpower, improving operational efficiency, optimising resource allocation, personalising services and generating new insights. Research shows that global GDP could be up to 14% higher in 2030 as a result of using AI<sup>1</sup>.

## RISKS OF AI

The potential of AI is being realised by the increasing amount of big data generated in the digital age, resulting in personal data commonly involved in the development and use of AI, in particular for the new generation of AI, which acquires its "intelligence" by analysing a vast amount of training data with the use of complex machine learning algorithms. Therefore, AI poses challenges to privacy and the protection of personal data by stretching the limits of conventional data protection principles, such as transparency, data minimisation and limitation of use. Furthermore, data protection risks of AI intersect with the potential ethical impact of AI as individuals who have had their personal data analysed by an AI system may have their rights, freedom and interests impacted by automated decisions made by that very AI system. Therefore, if used improperly, AI may undermine human rights (including privacy right), human dignity, individual autonomy and fairness. Organisations that use AI may lose the trust of consumers as a result.

## COMPLIANCE WITH THE PERSONAL DATA (PRIVACY) ORDINANCE

Personal data belongs to individuals and its collection, holding, processing and use are regulated by the PDPO. It is important to note that organisations have to collect, hold, process and use personal data lawfully in accordance with the PDPO when they develop and use AI. Appendix B to this Guidance provides a brief introduction to the requirements under the six Data Protection Principles ("DPPs") in Schedule 1 to the PDPO. The six DPPs represent the core requirements of the PDPO and cover the entire life cycle of the handling of personal data from collection to destruction.

---

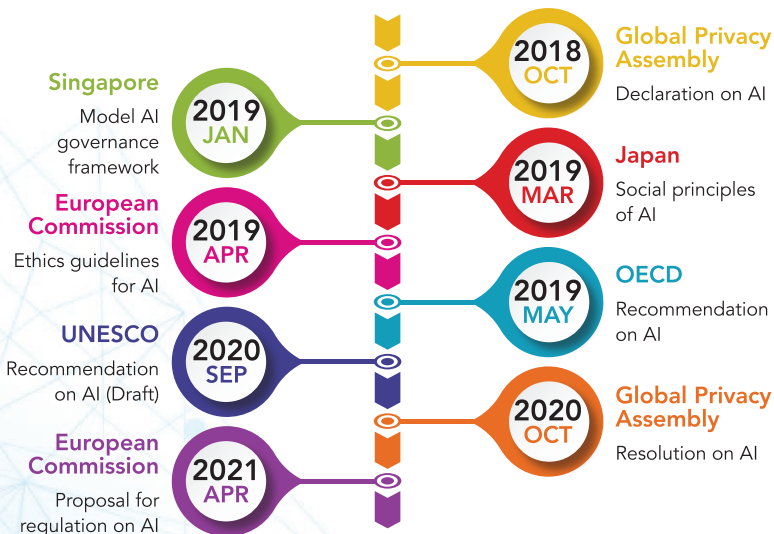
<sup>1</sup> PwC, *Sizing the Prize - What's the real value of AI for your business and how can you capitalise?* (2017)

## ETHICS IN THE DEVELOPMENT AND USE OF AI

Given the potential ethical risks of AI, organisations are also encouraged to embrace good data ethics in their operation and in the development and use of AI. In this regard, organisations should take into account the rights, freedom and interests of all stakeholders concerned (i.e. adopting a multi-stakeholder approach), ensuring that both privacy risks and wider ethical risks are mitigated.

Against this background, calls for accountable and ethical use of AI have mounted in recent years. Principles and guidance relating to the use of AI also spring up around the globe. For example, the Global Privacy Assembly<sup>2</sup>, the European Commission<sup>3</sup>, OECD<sup>4</sup>, UNESCO<sup>5</sup>, Japan<sup>6</sup> and Singapore<sup>7</sup> have published their respective guidance notes in recent years. Some common principles, such as accountability, transparency, fairness, data privacy and human oversight, can be found in the guidance notes, signalling a global consensus in the area. The European Commission made a proposal for regulating AI by legislation means<sup>8</sup> in April 2021. If passed, it may become the world's first regulation on AI.

**Figure 1** *Timeline: Recent Development of AI Governance around the Globe*



<sup>2</sup> The Global Privacy Assembly is a leading international forum for over 130 data protection regulators from around the globe to discuss and exchange views on privacy issues and the latest international developments. The Assembly adopted a *Declaration on Ethics and Data Protection in Artificial Intelligence* in 2018, endorsing six guiding principles to preserve human rights in the development of AI. In 2020, the Assembly adopted a *Resolution on Accountability in the Development and Use of AI*, recommending organisations that develop and use AI to adopt 12 accountability measures in order to build trust with stakeholders.

<sup>3</sup> European Commission's Independent High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (2019)

<sup>4</sup> OECD, *Recommendation of the Council on Artificial Intelligence* (2019)

<sup>5</sup> UNESCO, *First Draft of the Recommendation on the Ethics of Artificial Intelligence* (2020)

<sup>6</sup> Japan, *Social Principles of Human-Centric AI* (2019)

<sup>7</sup> Singapore, *Model Artificial Intelligence Governance Framework (First Edition)* (2019). The second edition of the framework was published in 2020.

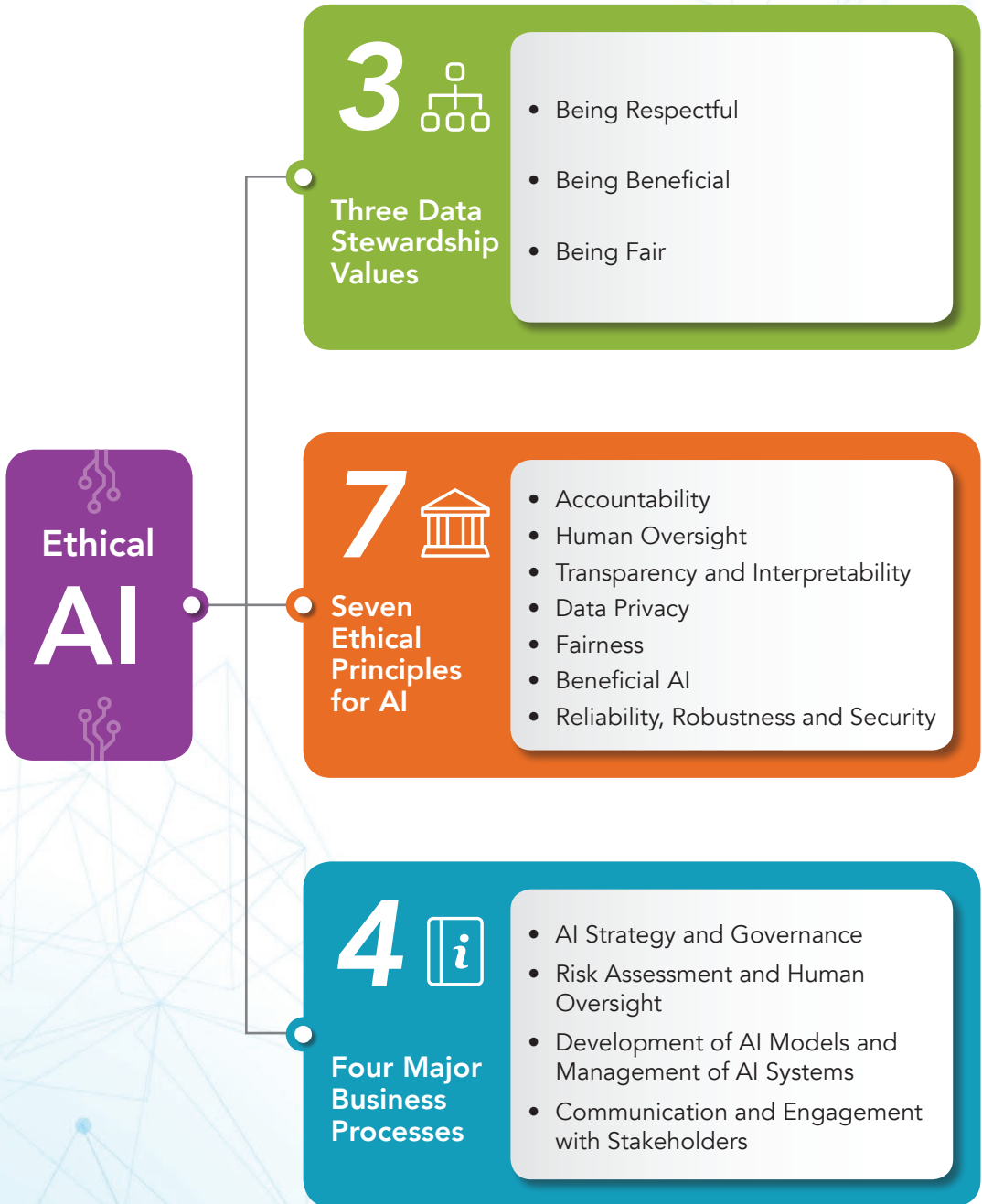
<sup>8</sup> European Commission, *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence* (2021)

This Guidance is based on the principles and various guidance notes relating to the use of AI mentioned above, and encompasses the experience of the Office of the Privacy Commissioner for Personal Data, Hong Kong (“PCPD”), in co-chairing the Global Privacy Assembly Working Group on Ethics and Data Protection in AI since 2019. It recommends a set of Data Stewardship Values and Ethical Principles for AI. The Guidance also provides a Practice Guide that follows the structure of a general business process to assist organisations in the development and use of AI in a lawful (insofar as the PDPO is concerned) and ethical manner, with the aim of enabling organisations to gain the necessary trust from their stakeholders, in particular individual consumers.

In the wake of the recent legislative proposal by the European Commission, consensus is yet to form as regards whether AI should be regulated through legislation or other means and the extent of regulation. In the United Kingdom (“UK”), there are even voices calling for scraping the provision under the country’s data protection law that allows individuals to opt out from being subject to fully automated decision-making for the sake of facilitating innovation and the development of AI. In Hong Kong, we believe that providing guidance on privacy-friendly and ethical practices in the development and use of AI would facilitate innovation and the wider use of AI in the community.

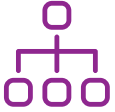
Hong Kong is striving to become a data hub and innovation centre for the Greater Bay Area and the Asia Pacific region as well as a world-class smart city. Given that data is the lifeblood of AI, Hong Kong may capitalise on its advantage of being a data hub to boost the development of AI. The healthy development and use of AI will also contribute greatly to making Hong Kong an innovation centre and a world-class smart city. We believe that this Guidance will help organisations in Hong Kong unlock the gate of success in their development and use of AI.

**Figure 2** Structure of this Guidance





# 2 DATA STEWARDSHIP VALUES



Values define how an organisation carries out its activities and achieves its mission and vision. To ensure that the development and use of AI are ethical, organisations should first and foremost define their core ethical values. The Ethical Accountability Framework for Hong Kong, China, published by the PCPD in October 2018<sup>9</sup>, recommends organisations to embrace three Data Stewardship Values, namely, being respectful, being beneficial and being fair. These values represent the starting point for formulating ethical principles and practices for the development and use of AI.

## 2.1 Being Respectful

It is crucial to respect the dignity, autonomy, rights, interests and reasonable expectations of individuals in processing their data. In this regard, every individual should be treated ethically, instead of as an object or a piece of data.

## 2.2 Being Beneficial

The value of being 'beneficial' emphasises the need to provide benefits to stakeholders, which include individuals affected by the use of AI and the wider community, where possible. Meanwhile, any harm to the stakeholders should be prevented or minimised.

## 2.3 Being Fair

The value of being 'fair' concerns both the processes and the results. In respect of the processes, being 'fair' entails that decisions are made reasonably without unjust bias or unlawful discrimination. Highly accessible and effective avenues should be established for individuals to seek redress for unfair treatments. In respect of the results, being 'fair' entails that like people should be treated alike. Any differential treatments between different individuals or different groups of people should be justifiable with sound reasons.

<sup>9</sup> Accessible at: [https://www.pcpd.org.hk/misc/files/Ethical\\_Accountability\\_Framework.pdf](https://www.pcpd.org.hk/misc/files/Ethical_Accountability_Framework.pdf)

# 3 ETHICAL PRINCIPLES FOR AI



Having regard to the three Data Stewardship Values and their corporate values, organisations should devise compatible principles and policies which enshrine those values. In this regard, organisations are encouraged to adopt the following Ethical Principles for AI.

## 3.1 Accountability

Organisations should be responsible for what they do and be able to provide sound justifications for their actions. Measures should be put in place to assess and address the risks of AI, with participation by senior management and interdisciplinary collaboration in the process.

## 3.2 Human Oversight

Users of AI systems should be able to take informed and autonomous actions regarding the recommendations or decisions of the AI systems. The level of human involvement in the process should be proportionate to the risks and impact of using the AI systems. The option of human intervention should always exist if the use of AI is assessed to be of high risk.

## 3.3 Transparency and Interpretability

Organisations should clearly and prominently disclose their use of AI and the relevant data privacy practices while striving to improve the interpretability<sup>10</sup> of automated and AI-assisted decisions. Transparency and interpretability are instrumental in demonstrating accountability as well as protecting individuals' rights, freedom and interests in the use of AI.

<sup>10</sup> Interpretability refers to the ability to determine the cause and effect from an AI system. In other words, it is the extent to which a person can predict what will happen when there is a change in the input to the AI system or its algorithmic parameters.

### 3.4 Data Privacy

Privacy is a fundamental human right. Effective data governance should be put in place to protect individuals' privacy in the development and use of AI. Personal data involved in the development and use of AI should be processed and protected in accordance with the PDPO, in particular the six DPPs in Schedule 1 to the PDPO. The six DPPs represent the core requirements of the PDPO and cover the entire life cycle of the handling of personal data from collection, retention, use to deletion. Details of the DPPs are set out in Appendix B.

### 3.5 Fairness

Individuals are entitled to be treated in a reasonably equal manner, without unjust bias or unlawful discrimination. Differential treatments between different individuals or different groups of people should be justifiable with sound reasons.

### 3.6 Beneficial AI

AI should provide benefits to human beings, businesses and the wider community. Provision of benefits encompasses prevention of harm. Where the use of AI may cause harm to stakeholders, measures should be taken to minimise the probability and severity of the harm.

### 3.7 Reliability, Robustness and Security

Organisations should ensure that AI systems operate reliably as intended over their expected lifetime. The AI systems should be resilient to errors during operations in order to prevent or minimise unintentional harm. The AI systems should also be protected against attacks, such as hacking and data poisoning<sup>11</sup>. Fallback plans should be put in place in the event that the AI systems cannot function properly.

<sup>11</sup> Data poisoning is a kind of attack against AI systems by polluting their training data, thereby impacting the systems' ability to produce correct predictions.

**Figure 3** Mapping of Ethical Principles for AI to Data Stewardship Values

	Data Stewardship Values	Ethical Principles for AI
1	Being Respectful	<ul style="list-style-type: none"><li>• Accountability</li><li>• Human Oversight</li><li>• Transparency and Interpretability</li><li>• Data Privacy</li></ul>
2	Being Beneficial	<ul style="list-style-type: none"><li>• Beneficial AI</li><li>• Reliability, Robustness and Security</li></ul>
3	Being Fair	<ul style="list-style-type: none"><li>• Fairness</li></ul>



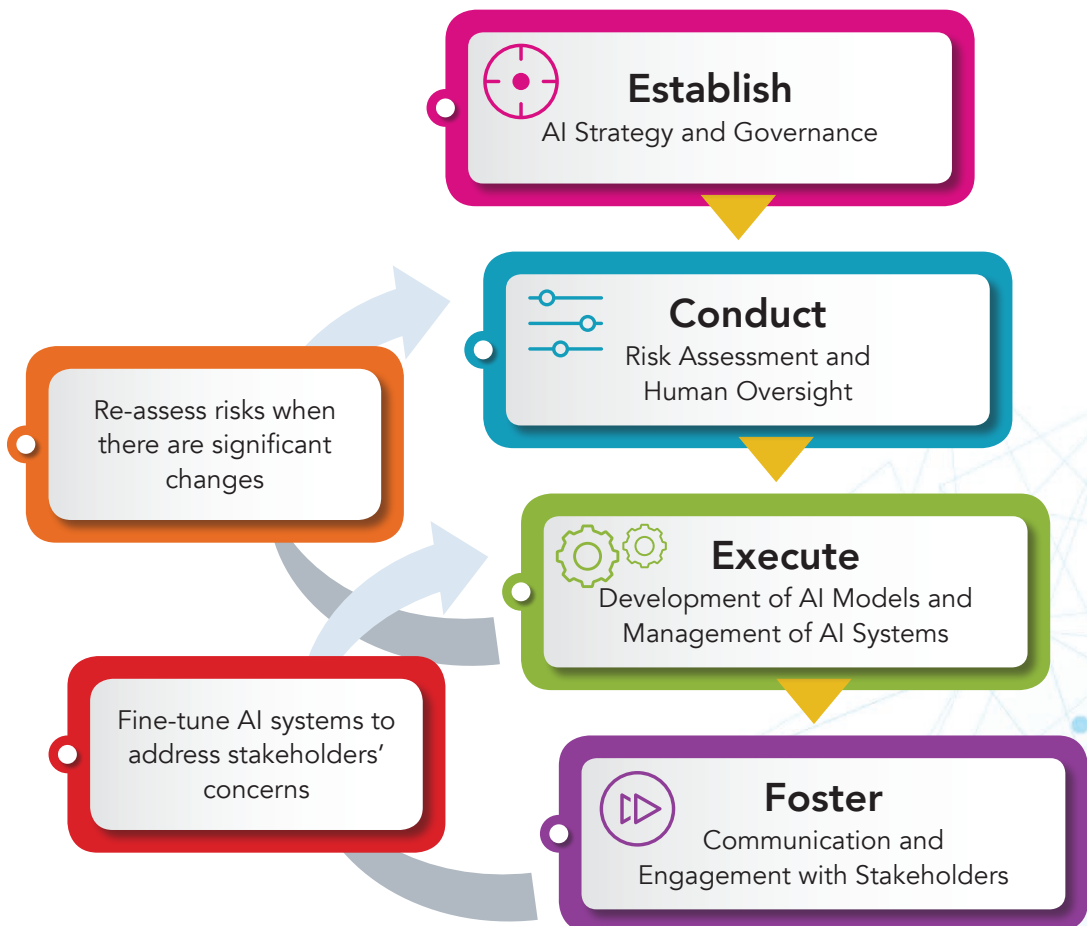
# 4 PRACTICE GUIDE



To ensure that the aforesaid values and ethical principles are practicable, organisations should formulate appropriate policies, practices and procedures. In this regard, organisations should take into consideration the recommended practices in the following areas when they develop and use AI throughout the life cycle of their business processes:

- AI Strategy and Governance;
- Risk Assessment and Human Oversight;
- Development of AI Models and Management of AI Systems; and
- Communication and Engagement with Stakeholders.

**Figure 4** *Workflow of Ethical Development and Use of AI*



In general, organisations should follow a risk-based approach to developing, managing and using AI systems. The following recommendations should therefore be considered and adopted in proportion to the risk that the AI systems may pose. The recommendations are by no means exhaustive and organisations should adopt any other measures as appropriate to implement the Data Stewardship Values and the Ethical Principles for AI in the development and use of AI.

## 4.1 AI Strategy and Governance

Buy-in from and active participation by the top management are essential ingredients of success in the implementation of AI systems. Organisations should therefore establish an internal governance structure to steer the development and use of AI. The internal governance structure should generally comprise both an organisational level AI strategy and an AI governance committee (or a similar body).

As the development and use of AI systems require a large amount of data, which often include personal data, organisations should devise policies on the application of privacy and data security by design in the AI life cycle. They may consider leveraging and adapting existing data governance or accountability frameworks in relation to the handling of personal data, such as the Privacy Management Programme advocated by the PCPD, and incorporate elements of this Guidance into the existing workflow so as to readily manage the development and use of AI systems.

Organisations should establish an AI strategy and an AI governance committee (or a similar body) to steer the development and use of AI.

### 4.1.1 AI Strategy

#### Key principle: Accountability

Organisations should formulate an AI strategy to demonstrate the commitment of the top management to the ethical development and use of AI. The AI strategy should also provide directions on the purposes for which AI may be used and how AI should be used.

The AI strategy of an organisation may include the following elements:

- (i) Determining the business objectives of using an AI system, such as the problems that the AI system would help to solve;
- (ii) Defining the functions that the AI system would serve in the technological ecosystem of the organisation;
- (iii) Setting out the ethical principles for the development and use of AI that are specific and applicable to the organisation by making reference to the Ethical Principles for AI introduced above;
- (iv) Determining acceptable uses of the AI system and specifying what uses are disallowed. The organisation may adopt a traffic light system for the use of AI<sup>12</sup>;
- (v) Ensuring that the use of the AI system conforms with the organisation's vision, mission and values;
- (vi) Setting up specific internal policies and procedures on how to design, develop and use AI ethically, including an institutionalised decision-making process with escalation criteria; and
- (vii) Communicating regularly the AI strategy, policies and procedures to all relevant personnel, including internal staff at all levels and, where appropriate, external stakeholders such as business partners.

#### 4.1.2 Governance Structure

##### Key principles: Accountability / Human Oversight

Expertise in different fields, such as computer engineering, data science, cybersecurity, user experience design, laws and compliance, public relations, etc. are required for the development and use of AI. An internal governance structure with sufficient resources, expertise and authority should be established to steer the implementation of the AI strategy while overseeing the development and use of AI. An AI governance structure may include the following elements:

- (i) An AI governance committee (or a similar body) which oversees the whole life cycle of AI from development, use to termination;

<sup>12</sup> In a traffic light system, AI use cases will be separated into green, amber and red categories. The red category comprises AI use cases of which the risks are too high that they should not be allowed. The green category comprises low-risk AI use cases and they may be adopted without a stringent risk assessment process. The amber category comprises all other use cases that are not in the red or green category. Stringent risk assessment should be conducted to decide whether the use of AI in the amber category should be allowed.

### **AI Governance Committee**

Participation by senior management and interdisciplinary collaboration should be the most significant attributes of the AI governance committee. A cross-functional team with a mix of skills and perspectives should be set up, and the team should include business and operational personnel, system analysts, system architects, data scientists, cybersecurity professionals, legal and compliance professionals, human resources personnel, customer service personnel, etc.

A C-level executive (such as Chief Executive Officer, Chief Information Officer, Chief Privacy Officer or any similar role) should be designated to lead the cross-functional team.

(Optional) Independent AI and ethics advice may be sought from external experts by the AI governance committee. An additional ethical AI committee may also be set up to conduct independent review when a project is sufficiently large with a great impact and a high profile and its ethical values may be challenged.

- (ii) Clear roles and responsibilities for different divisions or personnel regarding the development and use of AI;

#### **Examples of roles and responsibilities:**

- System analysts, system architects and data scientists should focus on the design, development, monitoring and maintenance of the AI system;
- Legal and compliance professionals should focus on ensuring compliance with laws and regulations (including data protection laws) as well as internal policies in the development and use of AI;
- Business and operational personnel should use AI in accordance with the policies and procedures of the organisations; and
- Customer service and public relations personnel should communicate with stakeholders, including customers, regulators and the general public, and address their concerns.



- (iii) Adequate resources in terms of both finance and manpower for the development and use of AI; and

**Examples of where adequate resources are required:**

- Hiring necessary internal and external experts with relevant technical skills, experience and expertise to develop and use the AI system;
- Conducting risk assessment when necessary to identify and mitigate risks, including privacy and security risks, arising from the use of AI;
- Establishing information systems that allow the monitoring, documentation and review of the AI system; and
- Providing adequate training to relevant personnel (see section 4.1.3 below).

- (iv) Effective internal reporting mechanisms in relation to the development and use of AI, such as reporting any system failure or raising any data protection or ethical concerns, to facilitate proper monitoring by the AI governance committee.

### 4.1.3 Training and Awareness Raising

**Key principle: Accountability**

A good strategy, plan or policy has to be executed by competent personnel in order to be successful. To ensure that AI-related strategies and policies are properly executed, relevant and adequate training should be provided to all relevant personnel to ensure that they have appropriate knowledge, skills and awareness to work in an environment using AI systems. Examples of training include:

- (i) Training on compliance with laws, regulations, internal policies and cybersecurity risks for system analysts, system architects and data scientists;
- (ii) Training on AI technology for legal and compliance professionals as well as users of AI (including business and operational personnel); and

- (iii) Training for human reviewers in charge of overseeing the decision-making of AI systems, with a view to enhancing their capabilities to detect and rectify any unjust bias, unlawful discrimination and error in the decisions made by AI systems.

To ensure that human reviewers perform their duties conscientiously and that human oversight is not just a gesture, the relevant personnel should have the ability to weigh up and interpret the recommendations made by AI. Reviewers should also be able to properly exercise their discretion and authority to veto the recommendations made by AI when necessary.

In addition to the core personnel identified above, other divisions of the organisation may be involved in the implementation of AI systems. The personnel awareness of the organisational strategy and policies on AI, as well as the risks of AI, should also be raised. This can be achieved, for example, by:

- (i) Providing general briefing or training to those personnel whose work relates to the AI system but they do not interact with the AI system directly (e.g. customer service and public relations personnel), so that they understand the benefits, risks, functions and limitations of the AI system used by the organisation; and
- (ii) Conveying to all relevant personnel the importance of ethical AI and applicable principles through staff meetings or other internal communications (such as circulars) in order to cultivate and promote a respectful and ethical culture in the development and use of AI.

## 4.2 Risk Assessment and Human Oversight

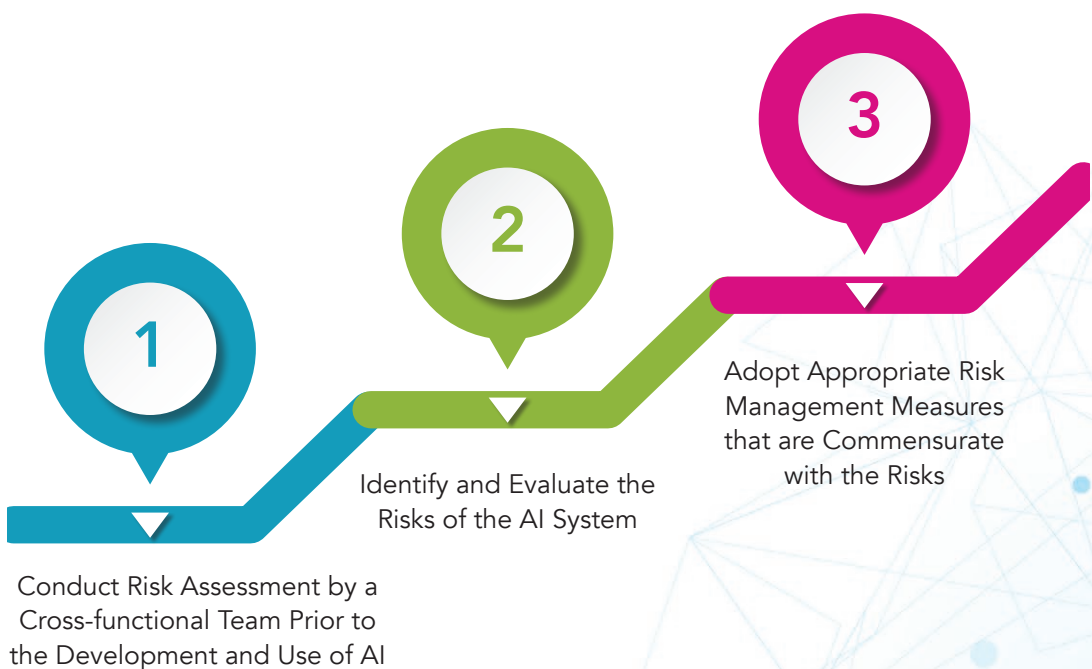
Risk levels of different AI systems vary, depending on, among others, the purposes for which the AI systems are used and how the AI systems are used. For example, an AI system used to assess the credit worthiness of individuals tends to carry higher risks than another one used to serve personalised advertisements because the former may deny individuals' right to access credit facilities while the latter may not have a significant impact on individuals. Furthermore, an AI system that is fully autonomous may be riskier than one that only provides recommendations to human actors. A risk-based approach should therefore be adopted in the management of AI systems. In this regard, comprehensive risk assessment

is necessary for organisations to systematically identify, analyse and evaluate the risks, including privacy risks, in relation to the development and use of AI. For high-risk AI systems, a risk management system should be put in place, implemented, documented and maintained throughout the entire AI life cycle.

Comprehensive risk assessment is necessary for organisations to systematically identify, analyse and evaluate the risks, including privacy risks, in relation to the development and use of AI.

Risk assessment should be conducted by a cross-functional team comprising personnel from different functional departments before the development and use of a new AI system or when there are significant updates to an existing AI system. Depending on the circumstances, people from different social, cultural and religious backgrounds as well as of different genders and races may need to be involved in the risk assessment in order to identify any unjust bias and unlawful discrimination in the process of development of AI. All risk assessments should be appropriately documented, and the risk assessment results should be reviewed and endorsed by the AI governance committee (or a similar body).

**Figure 5** *Process of Risk Assessment*



## 4.2.1 Risk Factors to Consider

### Key principles: Beneficial AI / Data Privacy / Fairness

As the development and use of AI usually involve the use of personal data, data privacy risk must be addressed. From the perspective of the protection of personal data privacy, factors to be considered in a risk assessment include:

- (i) The allowable uses of data that would be used to train AI models or fed into AI systems to make decisions, taking into account the requirements of the PDPO, in particular DPP 3<sup>13</sup>;
- (ii) The volume of data, in particular personal data, required for training AI models or the operation of AI systems<sup>14</sup>;
- (iii) The sensitivity of the data involved. Data that is generally considered to be more sensitive include biometric data, health data, and personal data of vulnerable groups, such as children;
- (iv) The quality of the data involved, taking into account its source, reliability, integrity, accuracy, consistency, completeness, relevance and usability<sup>15</sup>;
- (v) The security of personal data in the development or use of AI systems, taking into account how personal data may be transferred in and out of the AI systems across the organisation's technological ecosystem<sup>16</sup>; and
- (vi) The probability that the privacy risks (e.g. excessive collection, misuse or leakage of personal data) will materialise and the potential severity of the harm that may result.

From a wider ethical perspective, and insofar as the use of AI systems may have an impact on the rights, freedom or interests of stakeholders, in particular individuals, factors to be considered in a risk assessment should also include:

- (i) The potential impact (including benefits and harms) of the AI system on the affected individuals and the wider community;
- (ii) The probability that the impact of the AI system will occur as well as its severity and duration; and

13 DPP 3 stipulates that personal data must not be used for new purposes without the prescribed consent of the data subjects. For details of the requirements under DPP 3, please see Appendix B.

14 DPP 1 stipulates that the amount of personal data to be collected shall be adequate but not excessive in relation to the purpose of collection. For details of the requirements under DPP 1, please see Appendix B.

15 DPP 2 requires a data user to take all practicable steps to ensure that personal data is accurate having regard to the purpose for which the personal data is used. For details of the requirements under DPP 2, please see Appendix B.

16 DPP 4 requires a data user to take all practicable steps to safeguard the security of personal data held by the data user. For details of the requirements under DPP 4, please see Appendix B.



- (iii) The adequacy of mitigation measures (both technical and non-technical) to minimise the risks.

As for individuals, the impact may affect their legal rights, human rights (including privacy rights), employment or educational prospects as well as their access and eligibility to services, etc. An AI system that is likely to cause a significant impact on stakeholders, in particular individuals, is considered to be of high risk.

An AI system that is likely to cause a significant impact on stakeholders, in particular individuals, is considered to be of high risk.

**Figure 6** Factors to Consider in Risk Assessment (Non-exhaustive)



## 4.2.2 Determining the Level of Human Oversight

### Key principle: Human Oversight

The primary objective of risk assessment is to enable organisations to adopt appropriate risk management measures to mitigate identified risks. Organisations should adopt a risk-based approach in the development and use of AI. Therefore, the types and extent of risk mitigation measures (including human oversight) to be adopted should correspond with and be proportionate to the identified risks, as well as the levels of the risks. The overall residual risks which cannot be eliminated should be communicated to the users of the AI system. In any event, the residual risks of the AI system should be reduced to an acceptable level. The residual risks are considered acceptable if they are as low as reasonably practicable and proportionate to the benefits that the AI system will bring to stakeholders.

Human oversight is a key measure for mitigating the risks of using AI. The results of risk assessment for the AI system would indicate the appropriate level of human oversight required in the use of the AI system. In any event, human actors should ultimately be held accountable for the decisions made by AI.

In any event, human actors should ultimately be held accountable for the decisions made by AI.

In general, an AI system with a higher risk profile, i.e. likely to cause a significant impact on stakeholders, suggests that a higher level of human oversight is required:

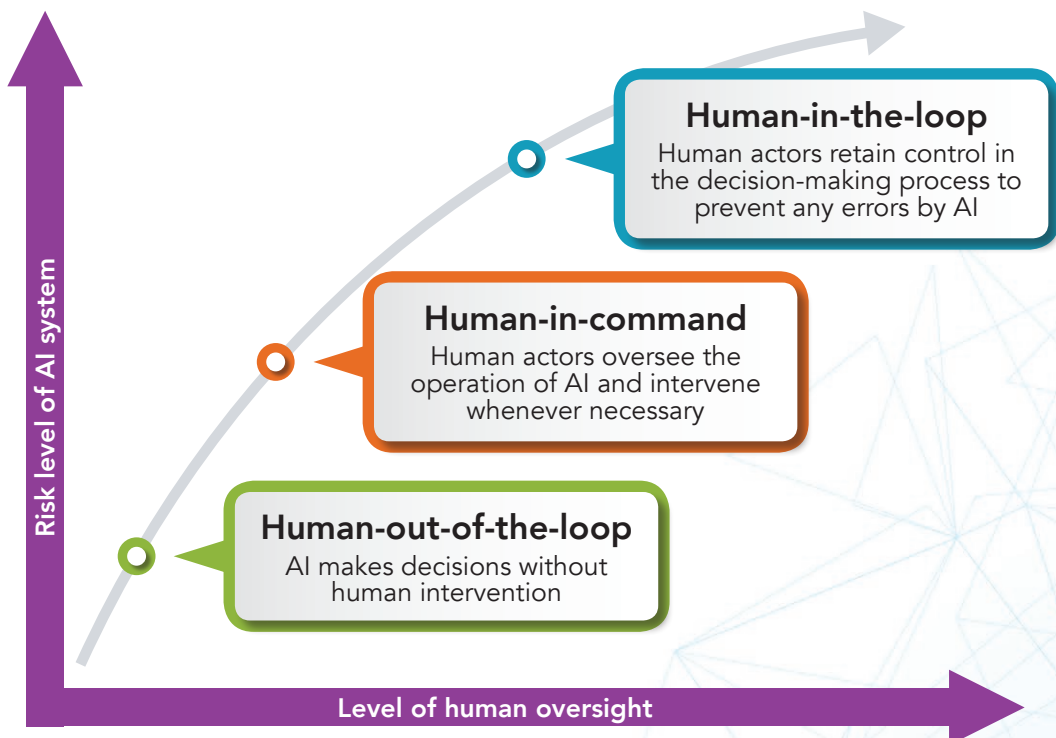
- (i) A high-risk AI system should take the 'human-in-the-loop' approach to human oversight, where human actors retain control in the decision-making process in order to prevent any errors or improper decisions made by the AI.
- (ii) An AI system with no or low real risks may take the 'human-out-of-the-loop' approach to human oversight, where the AI system is given the capability to make decisions without human intervention, so as to achieve fully automated decision-making.
- (iii) If both 'human-in-the-loop' and 'human-out-of-the-loop' approaches are not suitable, such as when the risks are not

negligible and the 'human-in-the-loop' approach is not cost-effective or practicable, organisations may consider the 'human-in-command' approach, where human actors will oversee the operation of the AI system and intervene whenever necessary.

Examples of AI use cases that may incur higher risks and may require a higher level of human oversight include:

- (i) Real-time identification of individuals by using biometric data, such as facial recognition, voiceprint recognition and gait recognition, which may result in taking adverse actions against the individuals;
- (ii) Recruitment, evaluation of job performance or termination of employment contracts;
- (iii) Evaluation of individuals' eligibility for social welfare or public services by public authorities; and
- (iv) Evaluation of the creditworthiness of individuals for making automated decisions in the offer of loans or other financial services.

**Figure 7** Risk-based Approach to Human Oversight

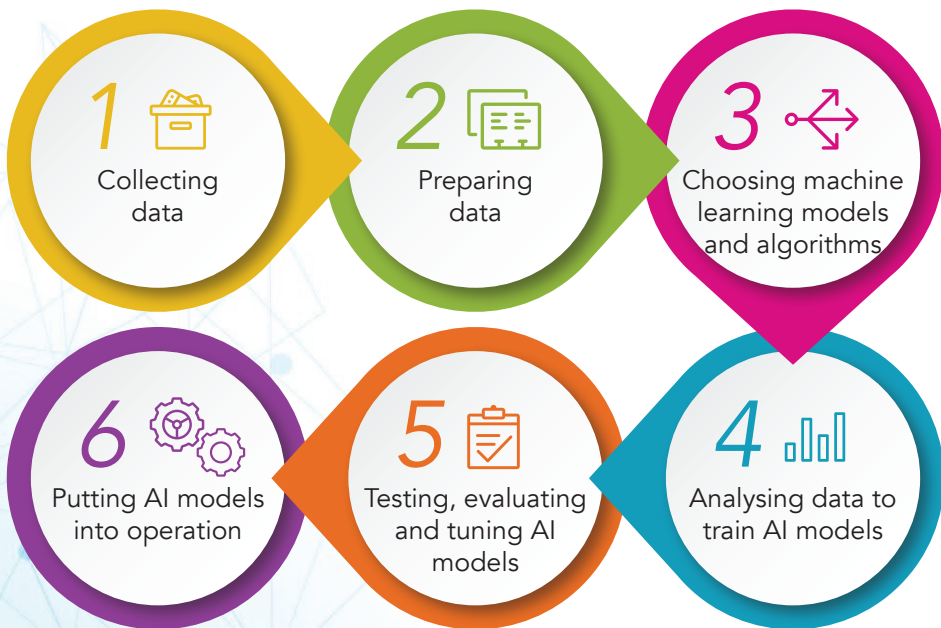


### 4.3 Development of AI Models and Management of AI Systems

The development of AI models by way of machine learning involves several steps, namely, (1) collecting data; (2) preparing data; (3) choosing the types of machine learning models (such as the supervised learning model<sup>17</sup> and the unsupervised learning model<sup>18</sup>) and the algorithms; (4) developing AI models by feeding training data to the machine learning algorithms, and (5) testing, evaluating and tuning the AI models. The quantity and quality of training data, as well as the types of machine learning models and algorithms used, will have a significant impact on the accuracy and reliability of the AI models.

The AI models may continue to learn and evolve after they are put into use. The environment in which the AI systems operate may also be changing. Continuous monitoring, review and user support are therefore required after the adoption of AI models in order to ensure that the AI systems remain effective, relevant and reliable. The following session will provide recommended practices in the development of AI models and the management of AI systems.

**Figure 8** Process of Development of AI Models



17 Supervised learning is a kind of machine learning in which labelled datasets are used to train AI models so that the trained AI models can classify data and make predictions. For example, pictures of cats are labelled as “cat” and fed to machine learning algorithms to train an AI model. The trained AI model will be able to identify cats in pictures. AI models developed by using supervised learning may provide more accurate predictions. However, they may not be able to generate new insights.

18 Unsupervised learning involves the use of machine learning algorithms to analyse unlabelled datasets and let the algorithms to discover patterns and draw insights from the datasets by themselves, without the need for human intervention. For example, an online retailer may use unsupervised learning to analyse the online behaviour of its customers in order to identify the preferences and needs of different customers. AI models developed by using unsupervised learning may provide more new useful insights. However, these AI models may also be less transparent and their predictions may be less interpretable.

### 4.3.1 Data Preparation for AI

#### Key principles: Data Privacy / Fairness

AI uses data in both the training and decision-making stages to discern patterns, draw inferences and make recommendations or decisions. Personal data is commonly involved. Effective data governance not only protects individuals' privacy in relation to personal data but also ensures the quality of data, which is critical to the fairness of AI systems. Poorly managed data would lead to 'garbage in, garbage out' and would have an adverse effect on the results that an AI system produces.

Data quality is critical to the fairness of AI systems.

Before data is used to train AI models, organisations should take the following steps in the preparation of datasets:

(i) **Measures must be taken to ensure compliance with the requirements under the PDPO**, including-

- Collecting an adequate but not excessive amount of personal data by lawful and fair means<sup>19</sup>;
- Refraining from using personal data for any purpose that is not compatible with the original purpose of collection, unless express and voluntary consents of the data subjects have been obtained, or the personal data has been anonymised<sup>20</sup>;
- Taking all practicable steps to ensure the accuracy of personal data before use<sup>21</sup>;
- Taking all practicable steps to ensure the security of personal data<sup>22</sup>; and
- Erasing or anonymising personal data when the original purpose of collection has been achieved<sup>23</sup>.

19 See DPP 1

20 See DPP 3

21 See DPP 2(1)

22 See DPP 4

23 See section 26 of the PDPO and DPP 2(2)



(ii) **Minimising the amount of personal data used in the development and use of AI would reduce privacy risks.** To minimise the collection and use of personal data, organisations should adopt the following practices and techniques, where appropriate:

- Collecting only the data that is relevant to the particular purpose of the AI in question and discard the data containing characteristics of individuals that are irrelevant to the purposes concerned;
- Using anonymised<sup>24</sup>, pseudonymised<sup>25</sup> or synthetic<sup>26</sup> data to train AI models;
- Applying 'differential privacy'<sup>27</sup> techniques to datasets before releasing the datasets for use in training AI models;
- Using federated learning<sup>28</sup> for training AI models so as to avoid unnecessary sharing of training data from different sources; and
- Erasing personal data from the AI system when the data is no longer required for the development and use of AI.

(iii) **The quality of the data used to train AI models should be managed,** especially when the decisions made by the AI system may have a significant impact on individuals. The data should be reliable, accurate, complete, relevant and without unjust bias or unlawful discrimination. In this regard, organisations should consider the following:

- Understanding the source, reliability, integrity, accuracy, consistency, completeness, relevance and usability of the data;
- Conducting relevant data preparation processes, such as annotation, labelling, cleaning, enrichment and aggregation;

24 Anonymised data refers to a dataset that has been processed in such a manner that no individual can be identified from it. As anonymised data cannot be used to identify individuals, it is not personal data.

25 Pseudonymised data refers to a dataset that has all personally identifiable information removed from it and replaced by other values, preventing direct identification of individuals without additional information. Pseudonymised data is personal data because individuals can still be identified from it indirectly, with the aid of additional information.

26 Synthetic data refers to a dataset that has been generated artificially and does not relate to real people. It therefore should have no privacy risks.

27 Differential privacy is an approach to privacy protection in the release of datasets, usually by adding noises (i.e. making minor alterations) to the datasets before release. Unlike de-identification, differential privacy is not a specific process, but a quality or condition of datasets that a process can achieve. A released dataset achieves differential privacy if it is uncertain whether a particular individual's data is included in it. Differential privacy is generally considered to have stronger protection of privacy than de-identification.

28 Federated learning refers to the collaborative development of AI models by separate computer systems. AI models will first be developed on the separate systems by using the data in the respective systems. This avoids the transmission of training data to a central database, reducing privacy and data security risks. Only the trained AI models will be transferred out of the respective systems to further develop a consolidated and shared AI model.

- Identifying outliers and anomalies in the datasets and removing or replacing the values as necessary;
- Testing the data for fairness before using it to train AI models; and

For example, unjust bias may inherently exist in training datasets if certain groups of individuals are under or over-represented. To address the issue, sampling techniques may be used to rebalance class distribution. Examples of sampling techniques include random over-sampling (i.e. duplicating samples from the minority class) and random under-sampling (i.e. deleting samples from the majority class).

- Designating personnel to regularly review and update the training datasets to ensure data quality.
- (iv) **Proper documentation of the handling of data should be in place** to ensure that the quality and security of data are maintained over time, as well as ensuring compliance with the requirements of the PDPO. The kinds of documentation include:
- The sources of the data;
  - The allowable uses of the data;
  - How the data used was selected from the pool of available data;
  - How data was collected, curated and transferred within the organisation;
  - Where the data is stored; and
  - How the data quality is maintained over time.

### 4.3.2 Development of AI Models

Key principles: Transparency and Interpretability / Reliability, Robustness and Security

After the preparation of data, organisations may apply machine learning algorithms to analyse the training data in order to develop AI models. Organisations should understand the characteristics of different types of machine learning algorithms and select the ones that meet their needs, taking into account, for example, the desired level of accuracy and interpretability of the output to be generated by the AI system.

Organisations should understand the characteristics of different types of machine learning algorithms and select the ones that meet their needs.

In addition to selecting appropriate machine learning algorithms, organisations should consider adopting the following measures to improve the AI system:

- (i) Performing rigorous testing of the AI models to ensure their reliability, robustness and fairness by, for example-
  - Comparing the AI decisions with decisions made by human beings or traditional non-AI models;
  - Using edge cases, unseen data<sup>29</sup> or potential malicious input to test the AI models; and
  - Conducting repeatability and reproducibility<sup>30</sup> tests for the AI system;
- (ii) Implementing measures to minimise the risk of malicious input or training data being fed into the AI system;
- (iii) Establishing multiple layers of mitigation to stop system errors or failures at different levels or modules of the AI system;
- (iv) Putting in place controls that allow human oversight and intervention of the operations of the relevant AI system;

<sup>29</sup> Unseen data refers to the datasets that have not been used for training an AI model. Instead it is used to test the performance of an AI model and therefore also known as test data.

<sup>30</sup> Reproducibility refers to whether an AI system produces the same results when the same datasets or methods of prediction are used. Reproducibility is important in assessing the reliability of an AI system.

- (v) Putting in place security measures to protect the AI system and the data against attacks and leakages;
- (vi) Establishing contingency plans of promptly suspending the AI system when needed and triggering fallback solutions if necessary;
- (vii) Establishing mechanisms to ensure that operations of the AI system are sufficiently transparent to enable users to interpret their output; and
- (viii) Establishing mechanisms to enable the traceability<sup>31</sup> and auditability of the AI system by, for example, automatically recording events (i.e. logs) while the AI system is operating.

### 4.3.3 Management and Monitoring of AI Systems

**Key principles: Reliability, Robustness and Security / Human Oversight**

AI systems should be monitored and reviewed continuously because the risk factors regarding the application of AI systems, including the relevance of the training data and the reliability of the AI models, may change over time. This will affect the reliability, robustness and security of the AI systems. The approach to continuous monitoring and reviewing of AI systems would vary depending on the risk levels. High-risk AI systems would necessitate more frequent and stringent monitoring and reviewing.

In this regard, organisations should consider incorporating the following review mechanisms:

- (i) Keeping proper documentation of the risk assessments, design, development, testing and use of the AI system;
- (ii) Conducting re-assessment of the risks of the AI system to identify and address new risks when there has been a significant change to the functionality or operation of the AI system, or a significant change to the regulatory or technological environment<sup>32</sup>;
- (iii) Conducting a periodic review of the AI models to ensure that they are operating and performing as intended;
- (iv) Regularly tuning and re-training the AI models with new data;

<sup>31</sup> Traceability refers to the capability to keep track of the development and use of an AI system, including the training and decision-making processes, as well as the data used, typically by means of documentation. Ensuring traceability can help enable auditability.

<sup>32</sup> Simple security patches and bug-fixing usually do not trigger the need for re-assessing the risks of the AI system.

- (v) Ensuring that an appropriate level of human oversight for the AI system is in place, taking into account the risk profile of the AI system;

Human oversight should aim at preventing and minimising the risks posed by AI to individuals. Personnel who exercise human oversight should be able to:

- Fully understand the capacities and limitations of the AI system;
- Remain aware of the possible tendency of over-reliance on the output produced by AI (i.e. 'automation bias');
- Correctly interpret the output produced by AI;
- Disregard, override or reverse the output produced by AI if the output is abnormal; and
- Intervene and interrupt the operation of the AI system where appropriate.

- (vi) Maintaining robust security measures throughout the AI system life cycle, from development, use, monitoring to termination;
- (vii) Establishing ongoing operational support and feedback channels for users of the AI system; and
- (viii) Evaluating regularly the wider technological landscape to identify gaps in the existing technological ecosystem of the organisation in order to make adjustments to the AI strategy and governance structure as necessary.

Internal audit should be conducted periodically to ensure that the development and use of AI comply with relevant policies of the organisation, and are in line with the AI strategy. The results of internal audit should be reported to both top management and governance bodies, such as the audit committee, of the organisation.



**Figure 9** *Development and Management of AI Systems*



## 4.4 Communication and Engagement with Stakeholders

### Key principle: Transparency and Interpretability

The use of AI should be transparent to stakeholders in order to demonstrate the organisation's adherence to the three Data Stewardship Values and avoid or reduce the harm possibly caused by the use of AI. Organisations that develop and use AI systems should therefore communicate and engage efficiently with stakeholders, in particular individual consumers and regulators. Effective communications are also essential to trust building.

In this regard, organisations should consider incorporating the following steps in their communications with stakeholders:

- (i) Clearly and prominently disclosing the use of the AI system to individuals unless it is obvious under the circumstances and the context of use;
- (ii) Providing adequate information on the purposes, benefits, limitations and effects of using AI systems in their products or services, unless the disclosure will compromise commercially sensitive information; and
- (iii) Disclosing the results of risk assessment of the AI systems, unless the disclosure will compromise commercially sensitive information.

For an AI system which may have a significant impact on individuals, organisations should also provide channels for the individuals to correct any inaccuracies, provide feedback, seek explanation, request human intervention and/or opt out from using AI, where possible.

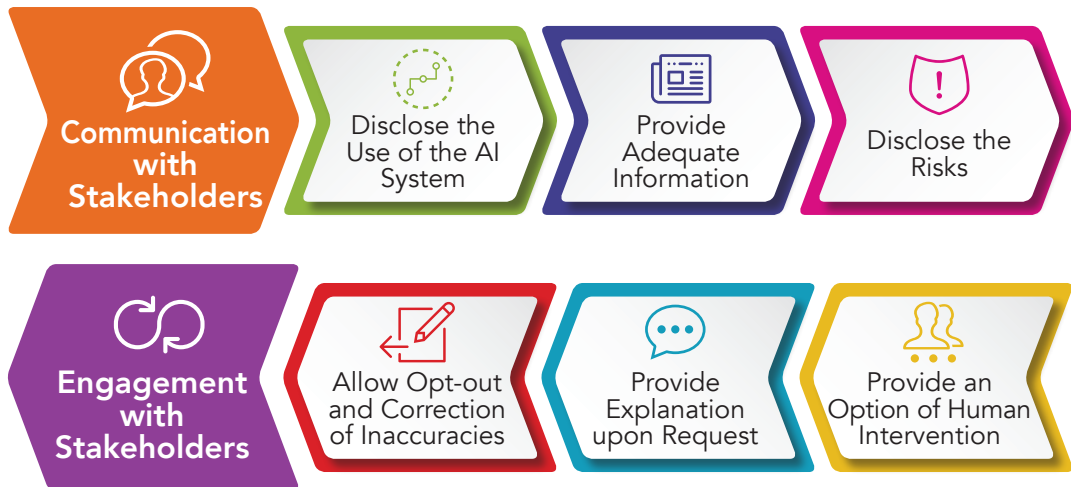
The explanation of decisions made or assisted by AI may include the following, where possible<sup>33</sup>:

- (i) How and to what extent AI has been involved in the decision-making process, including an overview of the key tasks of the AI system deployed as well as the involvement of human actors;
- (ii) The types of data that have been used in the automated or AI-assisted decision-making process and why these types of data are considered relevant and necessary;

<sup>33</sup> Organisations may consider referencing the guidance on *Explaining Decisions Made with AI* published by the Information Commissioner's Office, UK and The Alan Turing Institute in 2020 for more advice on how automated decisions made by AI may be meaningfully explained.

- (iii) How individuals' profiles used in the automated decision-making process are built, including any statistics used in the analysis and why the profile is relevant to the automated decision-making process; and
- (iv) The major factors leading to the automated decisions and final decisions, if different.

**Figure 10** *Communication and Engagement with Stakeholders*



In deciding the types of information to be disclosed and the level of details, organisations should take into account the stakeholders' understanding of the information, their needs and whether the disclosure will adversely impact the security and legitimate purposes of the AI system, among others. For example, the information disclosed to ordinary consumers should not be too technical, otherwise they may not understand it. For an AI system used to detect frauds committed by customers, the organisation concerned may not need to disclose the fraud indicators that the AI system is looking for lest the customers will 'game' the system.

Communications with stakeholders, in particular consumers, should be in plain language, clear and understandable by laymen, and drawn to the attention of stakeholders. They may also be included in the privacy policies of the organisations.

Communications with stakeholders should be in plain language, clear and understandable by laymen, and drawn to the attention of stakeholders.

# 5 AI SYSTEMS PROVIDED BY THIRD PARTIES



This Guidance primarily provides recommendations for organisations that develop their own AI systems. For organisations engaging third-party service providers to develop AI systems or purchasing off-the-shelf AI systems, they should take appropriate steps to ensure that the principles and practices recommended in this Guidance are adhered to. For example, organisations may request third-party contractors to follow the recommendations in this Guidance in the development of AI. Organisations may also test the reliability, robustness and fairness of the off-the-shelf AI systems before putting them into use.

Even if the AI systems are developed by third parties, organisations using the systems would still be held accountable for decisions made by the systems and compliance with the requirements under the PDPO as well as the wider ethical principles.



# 6 CLOSING REMARKS



*"AI holds the potential to deliver enormous benefits to society, but only if it is used responsibly.<sup>34</sup>"* While authorities worldwide are considering if, and how, the development and use of AI should be explicitly and directly regulated by laws and regulations, it is imperative that those who develop and use AI must comply with applicable laws on the protection of personal data and should uphold good data ethics meanwhile. To this end, we urge organisations to abide by the values, principles and practices recommended in this Guidance.

All in all, trust is pivotal in a data-driven economy. The values, principles and practices recommended in this Guidance are the very tools which might be used by organisations to gain the necessary trust from customers, other stakeholders and the community at large.



<sup>34</sup> Professor Klaus Schwab, Founder and Executive Chairman of the World Economic Forum. Quote from the press release of the World Economic Forum, *World Economic Forum Launches New Global Initiative to Advance the Promise of Responsible Artificial Intelligence* (28 January 2021)



# APPENDIX A - Self-assessment Checklist

## AI STRATEGY AND GOVERNANCE

	Question	Answer (Yes/No)	Further actions required
1	Has your organisation formulated an AI strategy before the development and use of AI?		
2	Did your organisation set up internal policies and procedures specific to the ethical design, development and use of AI?		
3	Did your organisation establish an AI governance committee (or a similar body) that would oversee the life cycle of the AI system, from its development, use to termination?		
4	<p>Does the AI governance committee (or a similar body) have:</p> <ul style="list-style-type: none"> <li>• Members from different disciplines and departments to collaborate in AI development and use?</li> <li>• A C-level executive (or management in a similar role) to oversee its operation?</li> </ul>		
5	Did your organisation set out clear roles and responsibilities for the personnel involved in the development and use of AI?		
6	Has your organisation set aside adequate resources in terms of finance and manpower for the development and use of AI?		
7	Has your organisation provided training to the personnel involved in the development and use of AI that is relevant to their respective roles?		
8	Has your organisation arranged regular awareness-raising exercises to the use of AI with all relevant personnel?		

## RISK ASSESSMENT AND HUMAN OVERSIGHT

	Question	Answer (Yes/No)	Further actions required
1	Did your organisation conduct a risk assessment before the development and use of AI?		
2	Did the risk assessment of your organisation take into account personal data privacy risks and other ethical impact of the AI system?		
3	Were the risk assessment results reviewed and endorsed by the AI governance committee (or a similar body)?		
4	Has your organisation put in place an appropriate level of human oversight and other mitigation measures for the AI system, taking into account the risk profile of the AI system?		

## DEVELOPMENT OF AI MODELS AND MANAGEMENT OF AI SYSTEMS

	Question	Answer (Yes/No)	Further actions required
<b>Preparation of Data</b>			
1	Has your organisation taken steps to minimise the use of personal data and ensure compliance with the requirements under the PDPO (e.g. using anonymised or synthetic data; understanding the sources and allowable uses of personal data; checking the accuracy of personal data, etc.)?		
2	Did your organisation take steps to ensure the reliability, integrity, accuracy, consistency, completeness, relevance, fairness and usability of data before putting it to use?		

	Question	Answer (Yes/No)	Further actions required
<b>Development of AI Models</b>			
3	Did your organisation evaluate the characteristics of the machine learning algorithms before putting them to use?		
4	Did your organisation perform rigorous testing of AI models to check their reliability, robustness and fairness?		
5	Has your organisation put in place adequate risk mitigation measures, including human oversight, to deal with errors or failures that may arise in the use of the AI system?		
6	Did your organisation put in place adequate security measures to protect the AI system against attacks?		
7	Did your organisation establish contingency plans of suspending the AI system and triggering fallback solutions when it is necessary?		
<b>Management and Monitoring</b>			
8	Does your organisation keep appropriate documentation of the handling of data, risk assessments and the design, development, testing and use of the AI system?		
9	Does your organisation have any plans in place to re-assess the risks of AI when there is a significant change to the functionality or operation of the AI system, or a significant change to the regulatory or technological environment?		
10	Has your organisation reviewed, tuned and re-trained AI models periodically?		
11	Did your organisation put in place an appropriate level of human oversight for the AI system based on the assessed level of risk?		
12	Did your organisation establish operational support and feedback channels for users of the AI system?		

	Question	Answer (Yes/No)	Further actions required
13	Did your organisation implement appropriate security measures throughout the AI system life cycle, from development, use, monitoring to termination?		
14	Does your organisation have any plans to conduct regular evaluation of the wider technological landscape to identify gaps in its existing technological ecosystem?		
15	Does your organisation conduct internal audit periodically to ensure compliance with internal policies in the development and use of AI?		

## COMMUNICATION AND ENGAGEMENT WITH STAKEHOLDERS

	Question	Answer (Yes/No)	Further actions required
1	Did your organisation clearly and prominently disclose the use of AI to individual consumers?		
2	Did your organisation inform individual consumers of the purposes, benefits and effects of using the AI system in its products or services?		
3	Did your organisation disclose the results of risk assessment of the AI system where appropriate?		
4	Did your organisation provide channels for individuals to opt-out from using AI where possible?		
5	Were channels provided for individuals to correct any inaccuracies, provide feedback, seek explanation and request human intervention where possible?		
6	Are the communications with stakeholders made in a plain, clear and layman-understandable language?		

# APPENDIX B - Data Protection Principles under the Personal Data (Privacy) Ordinance

The Personal Data (Privacy) Ordinance (Cap. 486) (“PDPO”) governs the collection, holding, processing and use of personal data by both private and public sectors. The PDPO is technology-neutral and principle-based. The Data Protection Principles (“DPP”) in Schedule 1 to the PDPO represent the core requirements of the PDPO and cover the entire life cycle of the handling of personal data from collection to destruction.

## DPP 1 - PURPOSE AND MANNER OF COLLECTION

DPP 1 provides that personal data shall only be collected for a lawful purpose directly related to a function or activity of the data user. The means of collection shall be lawful and fair. The data collected shall be necessary and adequate but not excessive for such purpose.

Data users shall also be transparent as regards the purpose of collection and the potential classes of persons to whom the personal data may be transferred, and the data subjects’ right and means to request access to and correction of their personal data. Usually, the information is presented in a Personal Information Collection Statement.

## DPP 2 - ACCURACY AND DURATION OF RETENTION

DPP 2 requires data users to take all practicable steps to ensure that personal data is accurate and is not kept longer than is necessary for the fulfillment of the purpose for which the data is used. Section 26 of the PDPO contains similar requirements for the erasure of personal data that is no longer required.

If a data user engages a data processor for handling personal data, the data user must then adopt contractual or other means to prevent the personal data from being kept longer than is necessary by the data processor.

## DPP 3 - USE OF DATA

DPP 3 prohibits the use of personal data for any new purpose which is different from and unrelated to the original purpose of collection, unless express and voluntary consent has been obtained from the data subjects.



## DPP 4 - DATA SECURITY

DPP 4 requires data users to take all practicable steps to protect the personal data they hold against unauthorised or accidental access, processing, erasure, loss or use.

If a data user engages a data processor in processing the personal data held, the data user must adopt contractual or other means to ensure that the data processor complies with the aforesaid data security requirement.

## DPP 5 - OPENNESS AND TRANSPARENCY

DPP 5 obliges data users to take all practicable steps to ensure certain information, including their policies and practices in relation to personal data, the kind of personal data held and the main purposes for which the personal data is held, is generally available to the public.

## DPP 6 - ACCESS AND CORRECTION

DPP 6 provides data subjects with the right to request access to and correction of their own personal data.

DPP 6 is supplemented by the detailed provisions in Part 5 of the PDPO which covers the manner and timeframe for compliance with data access requests and data correction requests, the circumstances in which a data user may refuse such requests, etc.

## APPENDIX C - References

- Global Privacy Assembly, *Declaration on Ethics and Data Protection in Artificial Intelligence* (2018)
- European Commission – Independent High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (2019)
- Hong Kong Monetary Authority, *High-level Principles on Artificial Intelligence* (2019)
- Japan, Cabinet Office, *Social Principles of Human-Centric AI* (2019)
- OECD, *Recommendation of the Council on Artificial Intelligence* (2019)
- Global Privacy Assembly, *Resolution on Accountability in the Development and Use of Artificial Intelligence* (2020)
- Infocomm Media Development Authority and Personal Data Protection Commission, Singapore, *Model Artificial Intelligence Governance Framework (Second Edition)* (2020)
- Infocomm Media Development Authority and Personal Data Protection Commission, Singapore, World Economic Forum, *The Companion to the Model AI Governance Framework – Implementation and Self-Assessment Guide for Organizations* (2020)
- Information Commissioner's Office, UK, *Guidance on AI and Data Protection* (2020)
- Information Commissioner's Office, UK, and The Alan Turing Institute, *Explaining Decisions Made with AI* (2020)
- UNESCO, *First Draft of the Recommendation on the Ethics of Artificial Intelligence* (2020)
- European Commission, *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence* (2021)
- Office of the Government Chief Information Officer, Hong Kong SAR, *Ethical Artificial Intelligence Framework* (2021)





香港個人資料私隱專員公署  
Office of the Privacy Commissioner  
for Personal Data, Hong Kong



查詢熱線 Enquiry Hotline : (852) 2827 2827  
傳真 Fax : (852) 2877 7026  
地址 Address : 香港灣仔皇后大道東248號大新金融中心13樓1303室  
Room 1303, 13/F, Dah Sing Financial Centre,  
248 Queen's Road East, Wanchai, Hong Kong  
電郵 Email : [communications@pcpd.org.hk](mailto:communications@pcpd.org.hk)



私隱公署網頁  
PCPD website  
[pcpd.org.hk](http://pcpd.org.hk)



下載本刊物  
Download  
this publication



本刊物使用署名4.0國際(CC BY 4.0)的授權條款，只要你註明原創者為香港個人資料私隱專員，便可自由分享或修改本刊物。詳情請瀏覽[creativecommons.org/licenses/by/4.0/deed.zh](http://creativecommons.org/licenses/by/4.0/deed.zh)。

This publication is licensed under Attribution 4.0 International (CC By 4.0) licence. In essence, you are free to share and adapt this publication, as long as you attribute the work to the Privacy Commissioner for Personal Data, Hong Kong. For details, please visit [creativecommons.org/licenses/by/4.0](http://creativecommons.org/licenses/by/4.0).

#### 免責聲明 Disclaimer

本刊物所載的資訊和建議只作一般參考用途，並非為法例的應用提供詳盡指引，亦不構成法律或其他專業意見。私隱專員並沒有就本刊物內所載的資訊和建議的準確性或個別目的或使用的適用性作出明示或隱含保證。相關資訊和建議不會影響私隱專員在《個人資料(私隱)條例》下獲賦予的職能及權力。

The information and suggestions provided in this publication are for general reference only. They do not serve as an exhaustive guide to the application of the law and do not constitute legal or other professional advice. The Privacy Commissioner makes no express or implied warranties of accuracy or fitness for a particular purpose or use with respect to the information and suggestions set out in this publication. The information and suggestions provided will not affect the functions and powers conferred upon the Privacy Commissioner under the Personal Data (Privacy) Ordinance.