

# **Chatbot ethical and legal considerations (2025)**

# Overview

- AI ethics and regulation materials published by :
  - Office Of the Privacy Commissioner for Personal Data (PCPD)
    - AI Security
  - Digital Policy Office
    - Data privacy and generative AI guidelines

# Overview

- AI ethics and regulation materials published by:
  - Cyberspace Administration of China (CAC), the People's Republic of China
    - Global AI Governance Initiative (2023)
    - [https://www.mfa.gov.cn/eng/zy/gb/202405/t20240531\\_11367503.html](https://www.mfa.gov.cn/eng/zy/gb/202405/t20240531_11367503.html)
  - European Parliament and Council
    - Regulation (EU) 2024/1689 – AI Act (European Union)
    - <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
  - US
    - State-level AI legislation also shows significant differences ([Hong Kong Generative Artificial Intelligence Technical and Application Guideline](#))

# Office Of the Privacy Commissioner for Personal Data

- [https://www.pcpd.org.hk/english/about\\_pcpd/commissioner/commissioner.html](https://www.pcpd.org.hk/english/about_pcpd/commissioner/commissioner.html)

About PCPD | Data Privacy Law | News & Events | Enforcement Reports | Compliance & Enforcement | Doxxing Offences | Data Security | AI Security **NEW!** | Anti-fraud Tips | Complaints | Education & Training | Resources Centre | Frequently Asked Questions | Contact Us

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

Keyword Search

Home > About PCPD > Welcome Message

RSS A A 繁體

## About PCPD

### Welcome Message

- Our Role
- Our Organisation
- Committees
- External Connection
- Job Vacancies
- Privacy Policy Statement
- Personal Information Collection Statement
- Performance Pledge
- An Inclusive Society

## Welcome Message from the Privacy Commissioner for Personal Data

Welcome to the website of the Office of the Privacy Commissioner for Personal Data (PCPD).

The PCPD is an independent body set up to oversee the implementation of and compliance with the provisions of the Personal Data (Privacy) Ordinance (Chapter 486 of the Laws of Hong Kong) (PDPO). The PCPD strives to ensure the protection of the privacy of individuals in relation to personal data through monitoring and supervising compliance with the PDPO, enforcing its provisions and promoting the culture of protecting and respecting personal data.



Established in August 1996, the PCPD is headed by the Privacy Commissioner for Personal Data and comprises different functional units, including the Complaints Division, Criminal Investigation Division, Compliance & Enquiries Division, Legal Division, Global Affairs & Research Division, Corporate Communications Division and Corporate Support Division.

Being a one-stop online portal which contains information on our work and the latest updates on personal data privacy issues, this website serves as an effective communication channel between us and our stakeholders. Other than providing useful information on the PDPO and compliance, the website offers comprehensive updates on the latest developments on local and international privacy issues, events, education materials and case summaries. It also informs you of our work such as answering enquiries and the handling of complaints and data breach notifications, etc., which aims to facilitate compliance with the legal requirements and adherence to the practice of good data ethics when one handles personal data.

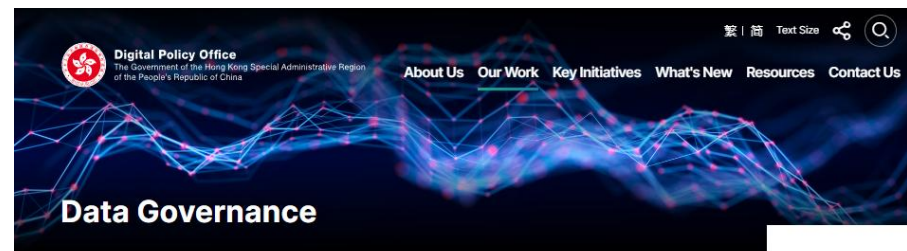
Besides this website, you may also visit our pages and channels on social media platforms, namely Facebook, Instagram, LinkedIn, X, Weibo and YouTube for related information.

We hope you find this website informative and useful. If you have any comments or suggestions on our work, you may send your views to us at [communications@pcpd.org.hk](mailto:communications@pcpd.org.hk).

Ms Ada CHUNG Lai-ling, SBS  
Privacy Commissioner for Personal Data

# Digital Policy Office

- [https://www.digitalpolicy.gov.hk/en/our\\_work/data\\_governance/policies\\_standards/ethical\\_ai\\_framework/](https://www.digitalpolicy.gov.hk/en/our_work/data_governance/policies_standards/ethical_ai_framework/)



[Home](#) > [Our Work](#) > [Data Governance](#) > [Enhancing Data Governance](#) > [AI and Data Ethics](#)

## Ethical Artificial Intelligence Framework

Artificial intelligence (AI) and big data analytics technologies present enormous opportunities for digital innovation and creativity to help drive smart city development, but at the same time, such technological advancement will also create various challenges. To realise the benefits and avoid adverse outcomes, the Government is well aware that it is important to pay due regard to AI and data ethics in implementing IT projects and services. With the growing adoption of data-driven technology, the Government is seeking a balanced approach which can safeguard public interest while facilitating innovation.

By drawing reference to the latest development in other countries and regions, the Ethical AI Framework has been developed for internal adoption within the Government regarding the applications of AI and big data analytics. The Framework is developed to assist B/Ds in adopting AI and big data analytics and incorporating ethical elements in the planning, design and implementation of IT projects or services and it consists of ethical principles, practices and assessment of AI. Nonetheless this framework, including guiding principles, practices and assessment template, is also applicable to other organisations in general and this customised version of framework is suitably revised (e.g. removal or adjustment of government specific terms) for general reference by organisations when adopting AI and big data analytics in their IT projects.

[Click here to download PDF file of the Ethical AI Framework \(customised version\)](#)

[Click here to download PDF file of the Ethical AI Framework Quick Reference Guide \(customised version\)](#)

# PCPD

- [https://www.pcpd.org.hk/english/about\\_pcpd/commissioner/commissioner.html](https://www.pcpd.org.hk/english/about_pcpd/commissioner/commissioner.html)

About PCPD | Data Privacy Law | News & Events | Enforcement Reports | Compliance & Enforcement | Doxxing Offences | Data Security | AI Security **NEW!** |  
Anti-fraud Tips | Complaints | Education & Training | Resources Centre | Frequently Asked Questions | Contact Us

Home > About PCPD > Welcome Message

## About PCPD

### Welcome Message

- Our Role
- Our Organisation
- Committees
- External Connection
- Job Vacancies
- Privacy Policy Statement
- Personal Information Collection Statement
- Performance Pledge
- An Inclusive Society

## Welcome Message from the Privacy Commissioner for Personal Data

Welcome to the website of the Office of the Privacy Commissioner for Personal Data (PCPD).

The PCPD is an independent body set up to oversee the implementation of and compliance with the provisions of the Personal Data (Privacy) Ordinance (Chapter 486 of the Laws of Hong Kong) (PDPO). The PCPD strives to ensure the protection of the privacy of individuals in relation to personal data through monitoring and supervising compliance with the PDPO, enforcing its provisions and promoting the culture of protecting and respecting personal data.



Established in August 1996, the PCPD is headed by the Privacy Commissioner for Personal Data and comprises different functional units, including the Complaints Division, Criminal Investigation Division, Compliance & Enquiries Division, Legal Division, Global Affairs & Research Division, Corporate Communications Division and Corporate Support Division.

Being a one-stop online portal which contains information on our work and the latest updates on personal data privacy issues, this website serves as an effective communication channel between us and our stakeholders. Other than providing useful information on the PDPO and compliance, the website offers comprehensive updates on the latest developments on local and international privacy issues, events, education materials and case summaries. It also informs you of our work such as answering enquiries and the handling of complaints and data breach notifications, etc., which aims to facilitate compliance with the legal requirements and adherence to the practice of good data ethics when one handles personal data.

Besides this website, you may also visit our pages and channels on social media platforms, namely Facebook, Instagram, LinkedIn, X, Weibo and YouTube for related information.

We hope you find this website informative and useful. If you have any comments or suggestions on our work, you may send your views to us at [communications@pcpd.org.hk](mailto:communications@pcpd.org.hk).

Ms Ada CHUNG Lai-ling, SBS  
Privacy Commissioner for Personal Data

# PCPD: AI Security

ing Offences | Data Security | AI Security **NEW!** |

tact Us



RSS



A

A

A

繁

簡

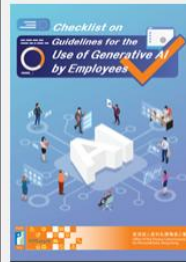
# PCPD : AI Security

## — A. PCPD's Guidance Materials on AI

Print

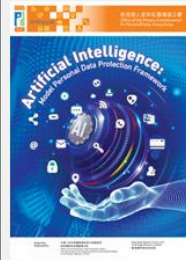
### Guidance Notes

#### [Checklist on Guidelines for the Use of Generative AI by Employees \(2025\)](#)



The "Checklist on Guidelines for the Use of Generative AI by Employees" aims to assist organisations in developing internal policies or guidelines on the use of Gen AI by employees at work while complying with the requirements of the Personal Data Protection Ordinance (PDPO).

#### [Artificial Intelligence: Model Personal Data Protection Framework \(2024\)](#)



The "Artificial Intelligence: Model Personal Data Protection Framework" (Model Framework) provides internationally well-recognised and practical recommendations and best practices to assist organisations to procure, implement and use AI, including generative AI, in compliance with the relevant requirements of the PDPO, so that organisations can harness the benefits of AI while safeguarding personal data privacy.

#### [Guidance on the Ethical Development and Use of Artificial Intelligence \(2021\)](#)



The "Guidance on the Ethical Development and Use of Artificial Intelligence" (AI Guidance) aims to help organisations understand and comply with the relevant requirements of the PDPO when they develop or use AI.



# PCPD: Guidelines for the user of Generative AI by Employees



# PCPD : Guidelines for the user of Generative AI by Employees

## Protection of Personal Data Privacy

- ✓ **Permissible types and amounts of input information:** Provide clear instructions on the types and amounts of information that can be inputted into Gen AI tools and the types of information that cannot be inputted (e.g., personal<sup>1</sup>, confidential, proprietary or copyrighted data)<sup>2</sup>.

**Data Protection Tip** 💡: If personal data is permitted for input into Gen AI tools, it is recommended that organisations instruct employees to anonymise the personal data (where possible and appropriate) and provide clear instructions on how personal data should be anonymised or cleansed before input.

- ✓ **Permissible use of output information:** Provide clear instructions on the permissible purposes<sup>3</sup> for using the information (including personal data) generated by Gen AI tools, and whether, when and how such personal data should be anonymised before further use.
- ✓ **Permissible storage of output information:** Require that the information generated by Gen AI tools, including any information used by employees, be stored according to the organisation's information management policy and deleted according to its data retention policy.
- ✓ **Compliance with other relevant internal policies:** Ensure that the policy on the use of Gen AI is aligned with the organisation's other relevant internal policies, including those on personal data handling and information security.

A photograph of a server rack with multiple units. The units have blue indicator lights. The background is blurred with yellow and blue bokeh lights.

Data collection

# PCPD : Data collection

- Collecting an adequate but not excessive amount of personal data by lawful and fair means;
- 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 anonymized
- Please refer to  
[https://www.pcpd.org.hk/english/resources\\_centre/publications/files/guidance\\_ethical\\_e.pdf](https://www.pcpd.org.hk/english/resources_centre/publications/files/guidance_ethical_e.pdf)



# PCPD : Data collection

- Taking all practicable steps to ensure the accuracy of personal data before use
- Taking all practicable steps to ensure the security of personal data
- Erasing or anonymising personal data when the original purpose of collection has been achieved
- Please refer to  
[https://www.pcpd.org.hk/english/resources\\_centre/publications/files/guidance\\_ethical\\_e.pdf](https://www.pcpd.org.hk/english/resources_centre/publications/files/guidance_ethical_e.pdf)

The image features a close-up, low-angle shot of a metallic combination lock on the left, with its dial showing numbers 4, 5, and 6. The lock is positioned over a dark, textured surface that appears to be a circuit board, with various electronic components and traces visible. The lighting is dramatic, highlighting the metallic surfaces of the lock and the intricate details of the circuitry. The overall composition suggests a theme of security and technology.

Data protection

# PCPD – AI: Model Personal Data Protection Framework



# PCPD – AI: Model Personal Data Protection Framework

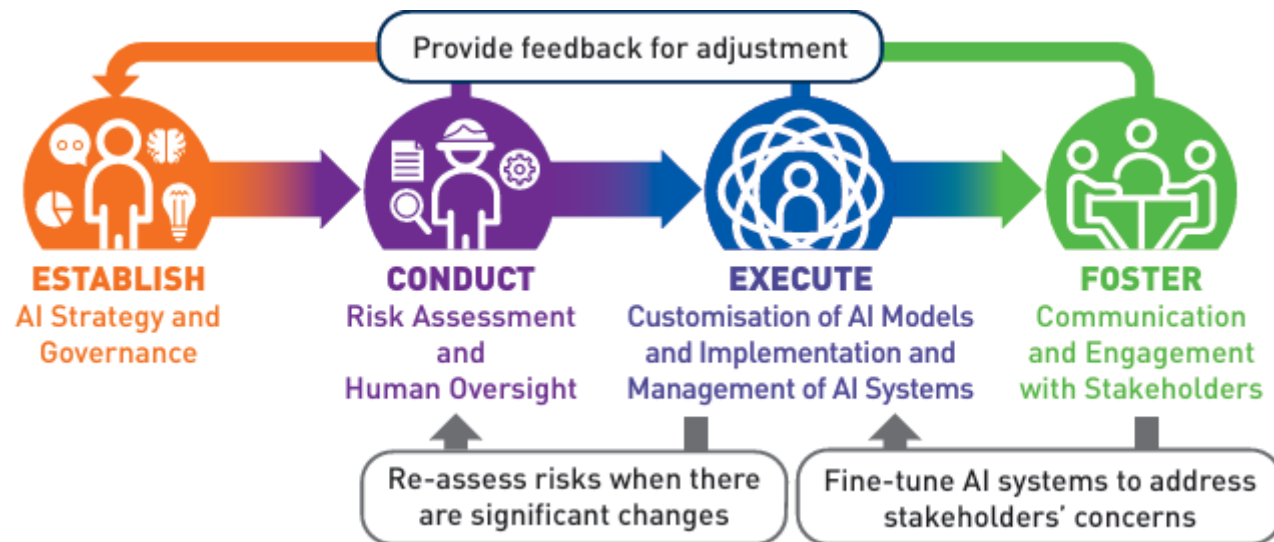
**Figure 1:** *Data Stewardship Values and Ethical Principles for AI*

	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>



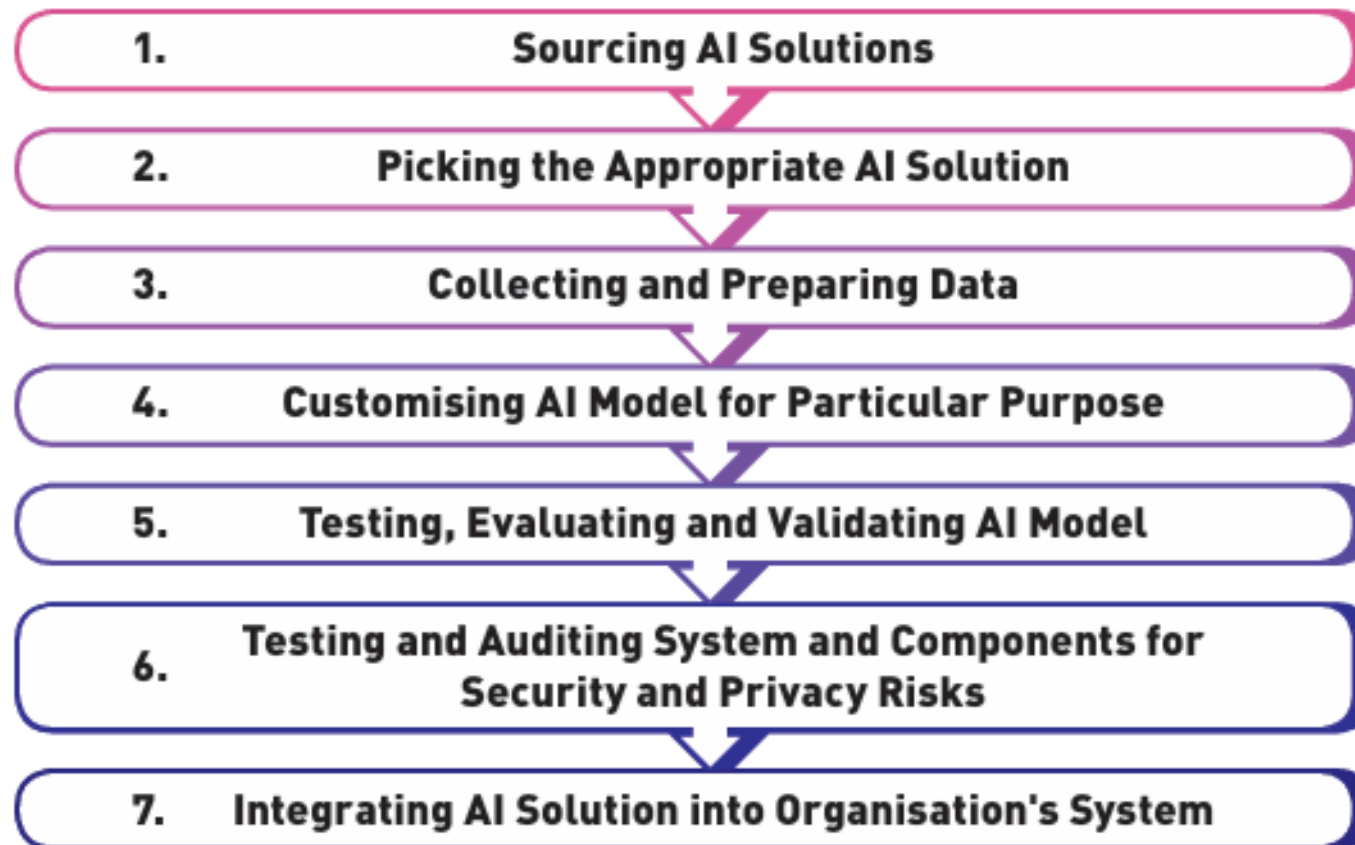
# PCPD – AI: Model Personal Data Protection Framework

- In general, organisations sourcing third-party AI solutions should adopt a risk-based approach to procuring, implementing and using AI systems, as part of a broader, holistic approach to AI governance in their organisations



# PCPD – AI: Model Personal Data Protection Framework

- Process of Procurement and Implementation of AI Models



# PCPD – AI: Model Personal Data Protection Framework

- Governance Considerations for Procuring AI Solutions

	Purpose(s) of Using AI
	Privacy and Security Obligations and Ethical Requirements
	International Technical and Governance Standards
	Criteria and Procedures for Reviewing AI Solutions
	Data Processor Agreements
	Policy on Handling Output Generated by the AI System
	Plan for Continuously Scrutinising Changing Landscape
	Plan for Monitoring, Managing and Maintaining AI Solution
	Evaluation of AI Suppliers

# PCPD – AI: Model Personal Data Protection Framework

- Key Data (Including Personal Data) Protection Compliance Considerations (Non-exhaustive)

## Who the data user is

- The party who has control of the collection, holding, processing or use of the personal data is the data user (section 2 of the PDPO).
- For example, an organisation that determines the types of personal data to be used for customising, testing, validating and / or operating an AI system is likely to be considered a data user.

# PCPD – AI: Model Personal Data Protection Framework

- Key Data (Including Personal Data) Protection Compliance Considerations (Non-exhaustive)

## Who the data processor is

- The party who processes personal data on behalf of another person and does not process the data for its own purposes is a data processor (section 2 of the PDPO).
- For example, an AI supplier that does not decide on the input data and the output of an AI model in the processing of personal data for customisation and only provides a platform for the customisation of AI is likely to be a data processor.

# PCPD – AI: Model Personal Data Protection Framework

- Key Data (Including Personal Data) Protection Compliance Considerations (Non-exhaustive)

## Legality of cross-border transfer

- If the customisation and use of AI involve processing personal data on cloud platforms with data centres distributed across multiple jurisdictions, and organisations (as data users) transfer personal data to places outside Hong Kong, the data user:
  - Must comply with the relevant requirements of the PDPO, including the 6 DPPs; and
  - Should ascertain if there are any restrictions or regulations pertaining to cross-border or cross-boundary transfers of data back to the data user from the jurisdiction where the data are processed.

# PCPD – AI: Model Personal Data Protection Framework

- Key Data (Including Personal Data) Protection Compliance Considerations (Non-exhaustive)







## Data security considerations

- If an organisation as the data user transfers personal data to the data processor for processing in the customisation and / or use of AI, it must adopt contractual or other means to prevent unauthorised or accidental access, processing, erasure, loss or use of the personal data, in compliance with the requirements of DPP 4(2) of the PDPO.

# PCPD – AI: Model Personal Data Protection Framework

- Training and Awareness Raising
- To ensure that AI-related policies are properly applied, adequate training should be provided to all relevant personnel to ensure that they have the appropriate knowledge, skills and awareness to work in an environment using AI systems.

## Examples of Training

Recommended Personnel	Training Topics
 <b>System analysts / architects / data scientists</b>	<ul style="list-style-type: none"><li>• Compliance with data protection laws, regulations and internal policies; cybersecurity risks</li></ul>
 <b>AI system users</b> (including business and operational personnel)	<ul style="list-style-type: none"><li>• Compliance with data protection laws, regulations and internal policies; cybersecurity risks; general AI technology</li></ul>
 <b>Legal and compliance professionals</b>	<ul style="list-style-type: none"><li>• General AI technology and governance</li></ul>
 <b>Procurement staff</b>	<ul style="list-style-type: none"><li>• General AI technology and governance</li></ul>
 <b>Human reviewers</b>	<ul style="list-style-type: none"><li>• Detection and rectification of any unjust bias, unlawful discrimination and errors / inaccuracies in the decisions made by AI systems or presented in the content</li></ul>
 <b>All staff performing work relating to AI system</b>	<ul style="list-style-type: none"><li>• Benefits, risks, functions and limitations of the AI system(s) used by the organisation</li></ul>



# PCPD: Data Protection Principles

## **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.

# PCPD: Data Protection Principles 1

## **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.

# PCPD: Data Protection Principles 2

## **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.

# PCPD: Data Protection Principles 3

## **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.

# PCPD: Data Protection Principles 4

## **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.

# PCPD: Data Protection Principles 5

## **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.

# PCPD: Data Protection Principles 6

## **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.

# Personal Data (Privacy) Ordinance (Cap. 486)

- <https://www.elegislation.gov.hk/hk/cap486>

The screenshot displays the official website for the Personal Data (Privacy) Ordinance (Cap. 486). The interface includes a search bar at the top with options for 'Match case' and 'Enable word stemming'. Below the search bar, there is a 'Point in Time' dropdown menu set to '01/10/2022\*' and a 'Go' button. The main content area is divided into two columns. The left column contains a table of contents with checkboxes for various sections, including 'Long Title', 'Part 1 Preliminary', '1. Short title and commencement', '2. Interpretation', '3. Application', '4. Data protection principles', 'Part 2 Administration', '5. Establishment, etc. of Privacy Commissioner for Personal Data', '6. Commissioner to hold no other office', '7. Filling of temporary vacancy', '8. Functions and powers of Commissioner', '9. Staff of Commissioner, etc.', '10. Delegations by Commissioner', '11. Establishment of Personal Data (Privacy) Advisory Committee', and '11A. Immunity'. The right column displays the text of the ordinance, starting with 'An Ordinance to protect the privacy of individuals in relation to personal data, and to provide for matters incidental thereto or connected therewith.' followed by the date '[1 August 1996] L.N. 343 of 1996' and a note '(Enacting provision omitted—E.R. 1 of 2013)'. The main heading is 'Part 1 Preliminary', and the first section is '1. Short title and commencement', which includes two subsections: '(1) This Ordinance may be cited as the Personal Data (Privacy) Ordinance.' and '(2) This Ordinance shall come into operation on a day to be appointed by the Secretary for Constitutional and Mainland Affairs by notice in the Gazette. (Amended L.N. 130 of 2007)'. The second section is '2. Interpretation', which includes a definition for 'act' and 'adverse action'.

Search:  ☐ Match case ☐ Enable word stemming

Point in Time: 01/10/2022\*  Monolingual Mode: Eng 繁 簡 Bilingual Mode: Eng / 繁 Eng / 簡

Show Whole Document ☐ Show highlight for: ☐ Matched Keywords ☐ Cross Reference(s) ☐ Source Note(s)

☐ Long Title  
☐ Part 1 Preliminary  
☐ 1. Short title and commencement  
☐ 2. Interpretation  
☐ 3. Application  
☐ 4. Data protection principles  
☐ Part 2 Administration  
☐ 5. Establishment, etc. of Privacy Commissioner for Personal Data  
☐ 6. Commissioner to hold no other office  
☐ 7. Filling of temporary vacancy  
☐ 8. Functions and powers of Commissioner  
☐ 9. Staff of Commissioner, etc.  
☐ 10. Delegations by Commissioner  
☐ 11. Establishment of Personal Data (Privacy) Advisory Committee  
☐ 11A. Immunity

An Ordinance to protect the privacy of individuals in relation to personal data, and to provide for matters incidental thereto or connected therewith.

[1 August 1996] L.N. 343 of 1996

(Enacting provision omitted—E.R. 1 of 2013)

(Format changes—E.R. 1 of 2013)

**Part 1**

**Preliminary**

1. **Short title and commencement**

(1) This Ordinance may be cited as the Personal Data (Privacy) Ordinance.

(2) This Ordinance shall come into operation on a day to be appointed by the Secretary for Constitutional and Mainland Affairs by notice in the Gazette. (Amended L.N. 130 of 2007)

2. **Interpretation**

(1) In this Ordinance, unless the context otherwise requires—

*act* (作為) includes a deliberate omission;

*adverse action* (不利行動), in relation to an individual, means any action that may adversely affect the individual's rights, benefits, privileges, obligations or interests (including legitimate expectations);

*appointed day* (指定日) means the day appointed under section





Risk level

# PCPD – AI: Process of Risk Assessment

- Comprehensive risk assessment is necessary for organisations to systematically identify, analyse and evaluate the risks, including privacy risks, involved in the procurement, use and management of AI systems



# PCPD – AI: Factors to Consider in Risk Assessment of AI Systems (Non-exhaustive)



Requirements under the PDPO



Volume, sensitivity and quality of data



Security of data



Potential impact on individuals, the organisation and community



Probability, severity and duration of impact



Mitigation measures



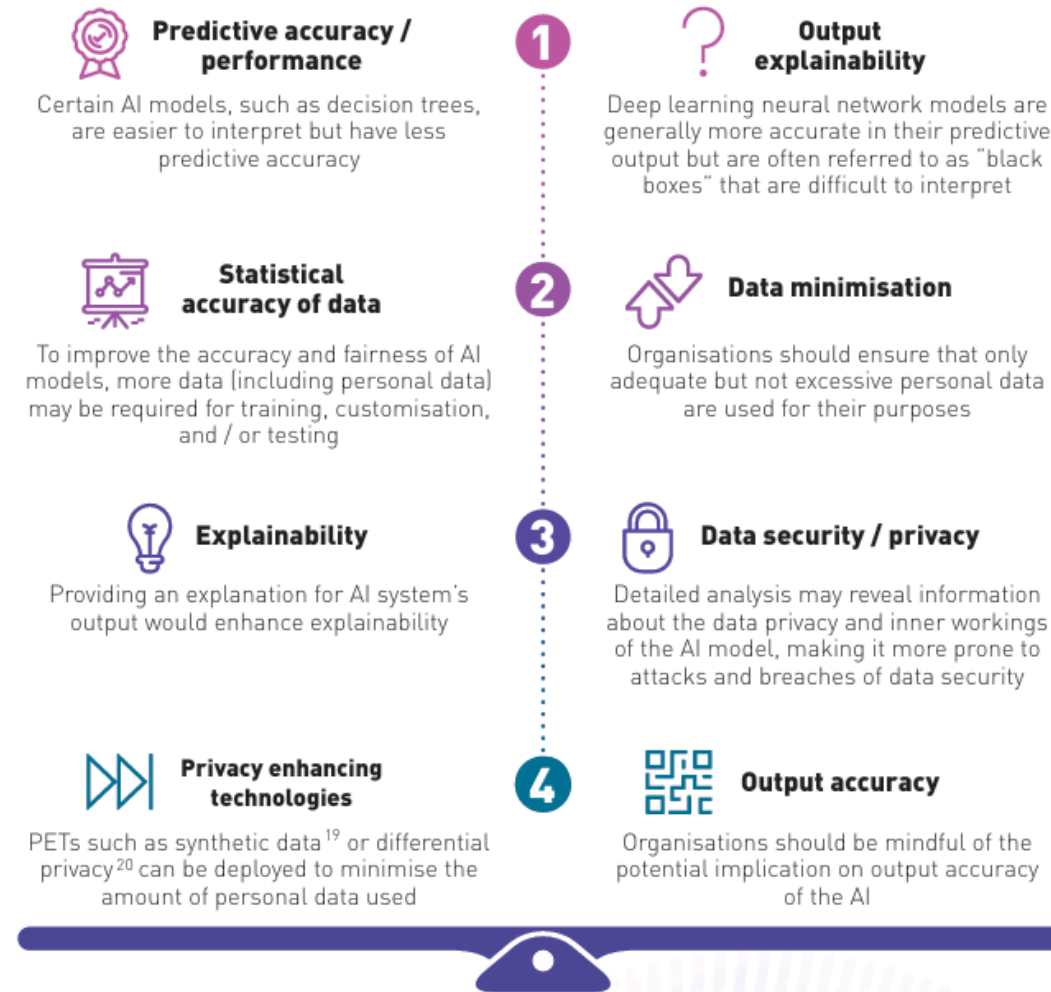
# PCPD – AI: Determining the Level of Human Oversight

In adopting a risk-based approach, the types and extent of risk mitigation measures should correspond with and be proportionate to the levels of the identified risks.

Risk-based Approach to Human Oversight:

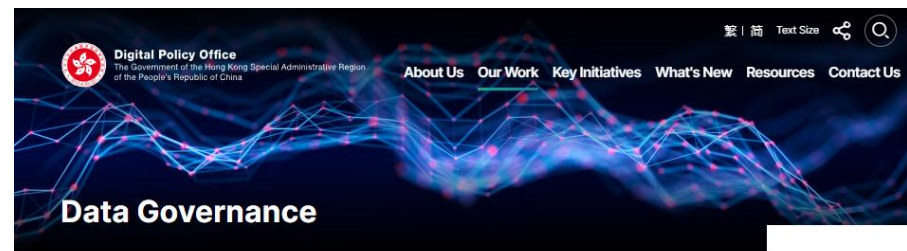


# PCPD – AI: Examples of Risk Mitigation Trade-offs



# Digital Policy Office

- [https://www.digitalpolicy.gov.hk/en/our\\_work/data\\_governance/policies\\_standards/ethical\\_ai\\_framework/](https://www.digitalpolicy.gov.hk/en/our_work/data_governance/policies_standards/ethical_ai_framework/)



[Home](#) > [Our Work](#) > [Data Governance](#) > [Enhancing Data Governance](#) > [AI and Data Ethics](#)

## Ethical Artificial Intelligence Framework

Artificial intelligence (AI) and big data analytics technologies present enormous opportunities for digital innovation and creativity to help drive smart city development, but at the same time, such technological advancement will also create various challenges. To realise the benefits and avoid adverse outcomes, the Government is well aware that it is important to pay due regard to AI and data ethics in implementing IT projects and services. With the growing adoption of data-driven technology, the Government is seeking a balanced approach which can safeguard public interest while facilitating innovation.

By drawing reference to the latest development in other countries and regions, the Ethical AI Framework has been developed for internal adoption within the Government regarding the applications of AI and big data analytics. The Framework is developed to assist B/Ds in adopting AI and big data analytics and incorporating ethical elements in the planning, design and implementation of IT projects or services and it consists of ethical principles, practices and assessment of AI. Nonetheless this framework, including guiding principles, practices and assessment template, is also applicable to other organisations in general and this customised version of framework is suitably revised (e.g. removal or adjustment of government specific terms) for general reference by organisations when adopting AI and big data analytics in their IT projects.

[Click here to download PDF file of the Ethical AI Framework \(customised version\)](#)

[Click here to download PDF file of the Ethical AI Framework Quick Reference Guide \(customised version\)](#)

# Digital Policy Office: Risks

Risk Tier	Definition	Regulatory Strategy
Unacceptable Risk	Systems posing existential threats (e.g., uses causing harm or affecting human safety, subliminal manipulation)	<ul style="list-style-type: none"><li>-Full prohibition</li><li>-Legal liability for development/deployment</li></ul>
High Risk	Critical infrastructure systems (e.g., healthcare diagnostics, autonomous vehicles)	<ul style="list-style-type: none"><li>-Conformity assessments</li><li>-Human-in-the-loop requirements</li><li>-Real-time monitoring</li></ul>
Limited Risk	Systems with moderate societal impact (e.g., recruitment tools, educational AI)	<ul style="list-style-type: none"><li>-Transparency obligations</li><li>-User opt-out mechanisms</li><li>-Annual compliance audits</li></ul>
Low Risk	Minimal-risk applications (e.g., spam filters, creative tools)	<ul style="list-style-type: none"><li>-Self-certification</li></ul>

Table 1: Risk Classification System

# Digital Policy Office: Safety Risks

- Even at the service stage, generative AI can still introduce new safety risks.
  - **Content Safety** is a critical issue for generative AI services. Such services pose risks of enabling users to create or disseminate harmful content, as well as exposing audiences to such content. For instance, users might exploit generative AI services to produce dangerous content involving pornography, violence, gore, terror, or child abuse. During interaction, generative AI services might also propagate harmful values, disseminating hate speech, discriminatory remarks, or inflammatory statements, thereby making users passive recipients of harmful content. When such harmful content is created, processed further, and widely distributed, it can exert a subtle negative influence on audiences, particularly teenagers who lack discernment, potentially inducing them to engage in illegal activities, criminal behaviour, or self-harm.



# Digital Policy Office: Safety Risks

- **Fabrication of Rumours** refers to the challenge wherein generative AI services can produce highly convincing text, images, audio, video, and other multimedia content. Due to their low cost, ease of use, and speed, these services can be exploited by users to deliberately create and disseminate rumours at scale, thereby confusing the public and misleading audiences. The general public often struggles to distinguish such fabricated rumours, which can significantly impact personal decision-making. As technology continues to advance, rumours generated using generative AI are expected to become even more realistic, posing severe challenges to the integrity of the societal information environment.

# Digital Policy Office: Safety Risks

- **Model Jailbreaking** refers to the practice of bypassing the safety mechanisms that developers have put in place to prevent generative AI services from being used for hazardous purposes and abuse. Under normal conditions, these models are designed to identify and reject unsafe requests that fall outside the established safety parameters. However, on the user side, attack methods targeting the safety perimeter continue to emerge, and these are referred to as model jailbreaks. For example, a user might input a carefully crafted sequence of commands as part of an attack, followed by an illicit request. If the model is successfully deceived by these commands, it may process requests that it should have otherwise refused, potentially leading to severe security risks.

# Digital Policy Office: Safety Risks

- **Data Leaks** refer to the challenge wherein generative AI services, particularly chat-based services, may collect user information in various forms. This includes information voluntarily provided by users, uploaded documents, and personal data accessed through devices. During the transmission and processing of this data, there is a risk of exposing private information belonging to individuals or enterprise users.

# Digital Policy Office: Lifecycle of Generative AI Model

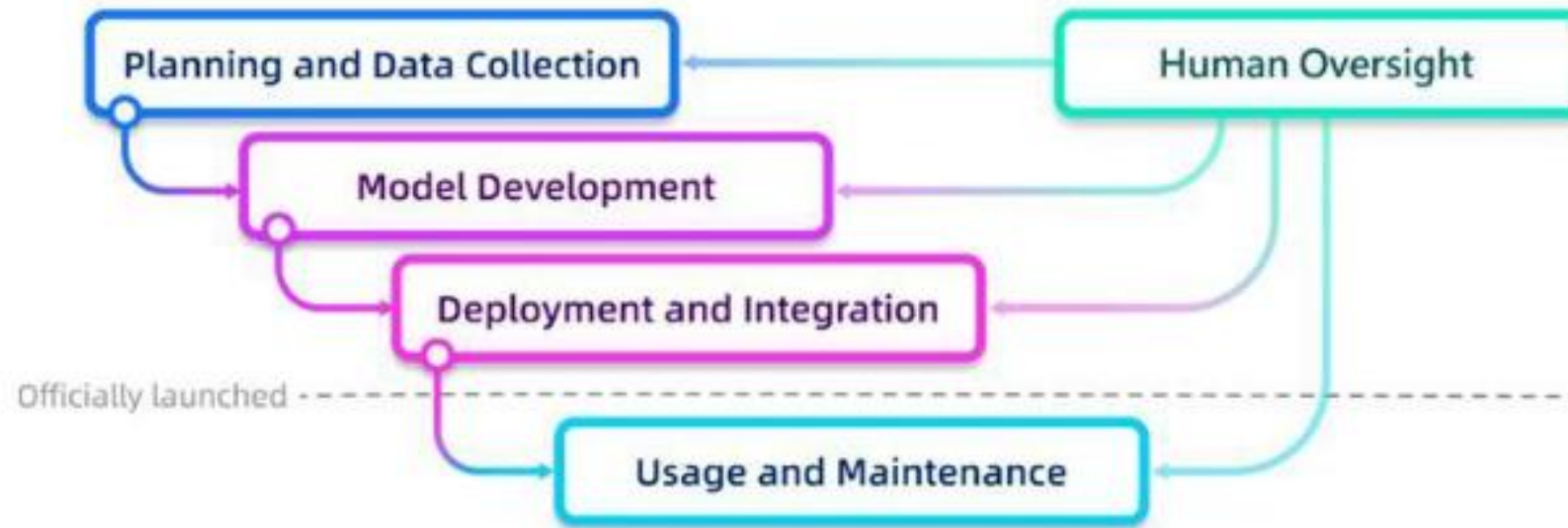


Figure 2. Lifecycle of Generative AI Model

# Digital Policy Office: Human Oversight

## 2.1.3.5 Human Oversight

Appropriate human oversight is critical for ensuring the trust and accountability framework of generative AI systems. The degree of human oversight should be based on the impact of different stages (e.g., data collection, model training, and output generation). The greater the impact, the stronger the need for human oversight. Models can be categorised based on the level of human oversight:

- **Collaborative Generative AI Models:** For models used in less impactful decision-making scenarios, human judgment can complement AI. These systems typically require limited human oversight due to smaller data volumes.
- **Human-Dominated Models:** When collaborative models are insufficient, human-dominated models are used. These rely primarily on human decision-making and operations, with AI serving as an auxiliary tool.

# Industry-Oriented Trustworthy AI Principles in Hong Kong



# Digital Policy Office: Finance

- **Finance:** The financial sector should strengthen fairness in the use of generative AI. When providing recommendation or decision-support services, all potential candidates should have an equal opportunity to be recommended, and necessary mechanisms should be in place to prevent human manipulation. Where possible, the financial sector should consider restricting interference with recommendation weights through manual settings, model training interventions, or other means. Where applicable, financial institutions like banks may need to customise models to align with specific user requirements. The financial sector should consider disclosure of information as much as practical and user choice should be provided to help users understand the working mechanisms, effects, and potential negative impacts of generative AI. If possible, users should be able to opt into using generative AI knowingly, and they should have the ability to terminate related services at anytime. Instead of opt-in, the HKMA circular “Consumer Protection in respect of Use of Generative Artificial Intelligence” dated 19 August 2024 states that customers should be provided with the option to opt out of using GenAI and request human intervention on GenAI-generated decision at their discretion as far as practicable, during the early stage of deploying customer-facing GenAI applications. Where an opt-out option cannot be provided for some reasons, banks should provide channels for customers to request for review of the GenAI-generated decisions.

# Digital Policy Office: Healthcare

- **Healthcare:** When using generative AI to assist in diagnosis, extreme caution should be exercised. Users must be explicitly informed that generated content may contain errors or fabrications, and such content should not be used directly as diagnostic reports but should be reviewed by licensed professionals as a reference. Personal data protection is crucial, and generative AI should adhere to the principle of data minimisation when collecting sensitive information such as identity details, biometric data, medical conditions, and patient histories. The purpose, usage, and processing methods of collected data must be clearly communicated to individuals, and explicit consent must be obtained before collection. Measures should be implemented to prevent data leaks during collection, transmission, processing, and storage. Additionally, data collected for healthcare purposes should not be repurposed for insurance, job recommendations, or other industries.



# Digital Policy Office: Legal

- Legal:
  - In the legal industry, the accuracy and reliability of generative AI outputs must be ensured, and generated content should include citations that trace back to the original legal texts. AI-generated content should not be used directly as legal documents but should serve as a reference after review by legal professionals. To protect personal data, sensitive legal cases involving trade secrets or private information should not be processed using public AI services that lack security and confidentiality guarantees.

# Digital Policy Office: Education

- **Education:** The education sector should regulate the use and scope of generative AI rather than outright banning students from using it. However, students should obtain teacher approval before using AI in their coursework, and AI-generated content should be clearly identifiable to prevent misuse that violates academic integrity. When teachers use generative AI in teaching, they must ensure that the generated content is truthful, accurate, and consistent in both textual and visual representation. If AI is used for grading assignments and exams, final results should always be reviewed by human educators.

# Digital Policy Office: Journalism

- **Journalism:** When using generative AI for news gathering, the principles of truthfulness, objectivity, and impartiality must be upheld. A diverse range of information sources should be used as input, and both technical and procedural safeguards should be implemented to minimise factual inaccuracies, misleading content, or distortions caused by model hallucinations. It is recommended that AI-generated news content include source attributions to facilitate manual verification. Generated content must undergo fact-checking and full editorial review before being published. Journalistic ethics must be strictly adhered to, and generative AI should not be used to create fabricated text, images, or audio-visual content that misrepresents facts or infiltrates news reports in any form.

# Digital Policy Office: Retail

- Retail: Retail businesses using generative AI for product recommendations, dynamic pricing, or customer service must ensure that algorithmic outcomes remain fair and transparent across different customer segments and geographic locations, maintaining market fairness. When collecting customer preferences and purchasing behaviour, businesses must comply with data protection and privacy regulations while conducting personalised marketing and promotions. To address customer concerns or confusion regarding generative AI, human support and real-time response mechanisms should be available to safeguard consumer rights.

# Digital Policy Office: Logistics

- **Logistics:** In transportation scheduling and intelligent route planning, generative AI should rely on reliable and up-to-date traffic and geographic data to minimise bias risks. In logistics processes such as shipping, warehousing, and delivery, any handling of personal addresses or consumer habits must incorporate appropriate encryption and access control measures to prevent data breaches. If industrial robots, such as automated sorting arms, are used in conjunction with generative AI, regular security and stability assessments should be conducted to prevent collisions or accidents.

# Digital Policy Office: Industry

- **Industry:** In industrial process monitoring and production line optimisation, AI models should be trained with high-quality, rigorously validated datasets to ensure accurate system judgments and predictions. If predictive maintenance or automated fault diagnosis functions are introduced, AI-generated results must be reviewed by supervisory engineers or quality control personnel. Strong security measures must be in place, particularly when handling confidential formulas/know-hows, or other trade secrets to prevent technology/confidential information leaks.