Artificial Intelligence:
Model Personal Data Protection Framework —
Protecting Personal Data Privacy in the Use of Agentic AI

PCPD
HK

PCPD.org.hk

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

30

支持機構：
Supporting Organisations:

中華人民共和國香港特別行政區政府
數字政策辦公室
Digital Policy Office
The Government of the Hong Kong Special Administrative Region of the People's Republic of China

香港應用科技研究院
Hong Kong Applied Science and Technology Research Institute

Introduction

A new form of artificial intelligence (AI), "agentic AI", has emerged from the background of increasing AI prevalence. Unlike conventional AI chatbots, agentic AI can operate with a high degree of autonomy. It offers considerable potential for enhancing productivity but also increases the risks to personal data privacy. Organisations must remain vigilant and strengthen their oversight when deploying and using agentic AI.

This guidance aims to assist data users, including organisations and individuals, in understanding the personal data privacy risks associated with the use of agentic AI. It also aims to provide practical recommendations for safeguarding personal data privacy in accordance with the Personal Data (Privacy) Ordinance (Cap. 486) ("PDPO"). It serves as a supplementary guidance to the *Artificial Intelligence: Model Personal Data Protection Framework* ("Model Framework"), which is published by the Office of the Privacy Commissioner for Personal Data and remains generally applicable to the use of agentic AI.

What is Agentic AI?

Agentic AI refers to an intelligent system with the capabilities of autonomous perception, memory, decision-making, interaction and execution[1]. At its core, agentic AI is built upon foundation models (e.g. large language models) that are integrated with additional tools, including databases, memory and computer operating systems.

Compared with conventional AI chatbots, which are generally used to generate content such as responses to questions, document summaries or code, agentic AI has a more versatile functionality. It can act autonomously on behalf of the user to execute multi-step tasks according to pre-defined workflows, including handling emails, making restaurant reservations and settling payments. Multiple AI agents may be configured to collaborate within an agentic system to improve performance by allowing specialised agents to work in parallel.

Importantly, AI agents are not legal persons; organisations, as data users[2], remain accountable for complying with the requirements of the PDPO, including the six Data Protection Principles ("DPPs") under the PDPO insofar as personal data is collected, held, processed or used by an AI agent.

---

1 National Technical Committee 260 on Cybersecurity of Standardization Administration of China, Practical Guidance of Cybersecurity Standards – Security Guidelines for the Deployment and Use of AI Agents (2026), https://www.tc260.org.cn/portal/article/2/71c613fd3db34b6a9da62f25b8219733

2 Section 2(1) of the PDPO defines a "data user" as a person who, either alone or jointly or in common with other persons, controls the collection, holding, processing or use of the personal data.

Risks of Agentic AI

From the perspective of the protection of personal data privacy, agentic AI generally poses higher risks than ordinary AI chatbots. For instance:

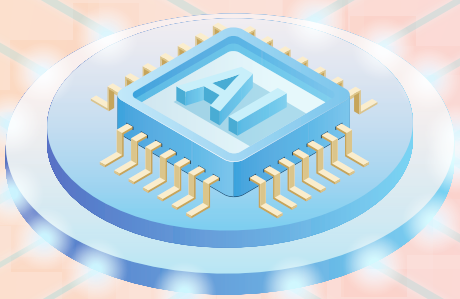
- **Extensive Access:** Agentic AI typically operates with higher default access rights, enabling it to access files, emails, account credentials and browser-stored contents. Unless stringent access restrictions are imposed, agentic AI may access a vast amount of personal data (e.g. personal data stored in memory), increasing the risks of unauthorised access, reproduction of personal data by third parties, or accidental erasure where the agentic AI misinterprets user commands.

- **System Vulnerabilities:** Agentic AI may be granted a high level of access to multiple systems and data sources. Consequently, vulnerabilities in its system design or safety controls could pose significant risks to the security of personal data.

- **Vulnerable Plugins or Skills:** For agentic AI that allows users to install plugins or skills, any such plugins or skills that have not undergone a rigorous security review may contain malicious code. Hackers may exploit such vulnerabilities to gain unauthorised access and take over users' accounts or seize control of the entire computer system, leading to the leakage of personal data.

- **Function Creep:** As agentic AI often aggregates and recombines personal data from multiple sources and some agentic AI systems may use the data to train large language models, there is a risk that such data may be used for purposes beyond the original purpose for which the data was collected (or a directly related purpose).

- **Multi-agent Risks:** The use of multi-agent systems may cause inaccurate personal data (whether acquired from external sources or generated through hallucination by the AI model) to be cascaded across AI agents through their interactions, which may lead to unfair or harmful outcomes for the data subjects. Security risks may also be amplified. In the absence of proper technical and organisational measures, the complicated flows of personal data may also render it difficult for data subjects to exercise their right to access or correct their personal data.



How to Address these Risks?

Given the elevated personal data privacy risks associated with the use of agentic AI, organisations should adopt a cautious and responsible approach to its use and be mindful of their obligations to respect individuals' privacy. The following recommendations set forth some practical considerations regarding the responsible use of agentic AI and for ensuring compliance with the requirements under the PDPO.

1 Avoid the excessive or arbitrary collection of personal data for use by agentic AI

According to DPP 1(1), personal data shall only be collected for a lawful purpose directly related to a function or activity of the data user, and the collected data shall be necessary and adequate but not excessive for such purpose. Organisations should adhere to this data minimisation principle and ensure that they are not collecting personal data for use in agentic AI arbitrarily or solely on the basis that the data might be useful in the future, especially where sensitive personal data is involved (e.g. Hong Kong Identity Card numbers, bank account numbers and passwords). Access control should be considered by establishing and implementing clear rules to ringfence the information and systems that an agentic AI may access for a specific purpose.

2 Be transparent about the use of agentic AI when processing personal data

Enhancing transparency is required to promote trust in the use of agentic AI. To comply with the notification requirements under DPP 1(3)[3], organisations should be transparent about their use of agentic AI when processing personal data, for instance, by including such information in their Personal Information Collection Statements. In accordance with DPP 5, organisations should also clearly state the kind of personal data held, as well as the main purposes for which the data is or is to be used in their Privacy Policy Statements, and shall take all practicable steps to ensure that a person can ascertain these policies and practices. To further enhance transparency, organisations may also consider stating in their Privacy Policy Statements the region(s) where personal data is stored or processed and the relevant arrangements.

3 Ensure the accuracy of personal data processed by agentic AI

Organisations shall take all practicable steps to ensure the accuracy of personal data processed by agentic AI, in accordance with DPP 2. These may include adopting approaches such as chain of thought[4], retrieval-augmented generation[5], context-specific fine-tuning[6] and human review to reduce the risk of generating outputs containing hallucinated or inaccurate personal data.

4 Set appropriate retention periods

In accordance with DPP 2(2), organisations using agentic AI systems that maintain conversation histories, cache data or long-term memory containing personal data shall take all practicable steps to ensure that the personal data is not kept longer than is necessary for the fulfillment of the purpose (including any directly related purpose) for which the data is or is to be used. Organisations are recommended to prescribe maximum retention periods for personal data and to implement measures to erase personal data that is no longer required.

---

[3] DPP 1(3) requires data users to take all reasonably practicable steps to inform the relevant data subjects of the purpose for which the data is to be used, the classes of persons to whom the data may be transferred, whether the data subject supplies the data obligatorily or voluntarily, and the data subject's rights to request access to and correction of their personal data.

[4] According to Google Research (the research division of Google), chain of thought is a prompting method used to improve the reasoning abilities of language models by decomposing multi-step problems into intermediate steps.

[5] Retrieval-augmented generation changes the capabilities of a large language model by adding an information retrieval system that provides grounding data to improve the performance of the large language model in specific use cases or domains.

[6] Fine-tuning is the process of taking AI models trained on large and general datasets and updating or adapting them using other specific data for a specific purpose or need.

5 Ensure that personal data is not used for a new purpose without consent

Pursuant to DPP 3, personal data shall not be used for a new purpose[7] without the prescribed consent of the data subject. To ensure that agentic AI does not use personal data beyond the scope of the original purpose of collection without the data subject's prescribed consent, organisations should clearly delineate the purposes for which the data will be collected and processed by agentic AI and the circumstances under which human oversight is required.

6 Ensure the security of personal data in agentic AI systems

To comply with DPP 4 and protect personal data in agentic AI systems against unauthorised or accidental access, processing, erasure, loss or use, organisations may take the following security measures when using agentic AI, and adopt a "human-in-the-loop"[8] approach where the decisions made by agentic AI may have a significant impact on individuals:

(a) Use the latest official versions of agentic AI from official channels and avoid using third-party or outdated versions to reduce the risks of data breach incidents arising from unpatched system vulnerabilities;

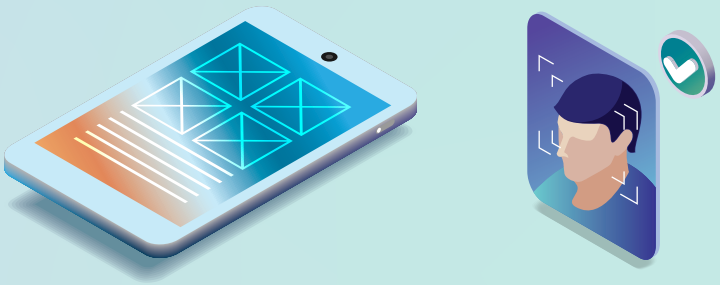
(b) Adopt adequate measures to ensure system and data security, such as separating the runtime environment of agentic AI from local devices or servers, strengthening network controls, strictly managing Internet-facing surfaces, limiting access rights and establishing effective protection mechanisms;

(c) Install and use plugins or skills cautiously by verifying that the relevant programmes are the latest official versions that have undergone security verification to ensure their security, reviewing the programmes to check for embedded malicious code, and refraining from using them if their security cannot be ascertained;

(d) Grant the minimum access rights necessary to complete the tasks at hand, and avoid granting administrator account rights to agentic AI; consider setting access rights to agentic AI for different departments within the organisation; carefully consider the nature and sensitivity of the personal data involved and refrain from providing personal data to agentic AI arbitrarily, especially when confidential or sensitive personal data (e.g. identification documents, bank account numbers, passwords) is involved;

(e) Adopt large language model guardrails by incorporating a comprehensive set of technical, procedural and governance mechanisms at each stage of (namely, before, during and after) the deployment of the model to reduce the risk of leakage of personal data; and

(f) Establish mechanisms to enable the traceability and auditability of agentic AI, such as equipping the system with logging capabilities to record AI agents' operations, operation timestamps, tools used, data accessed, and reasoning of decisions, etc.



[7] Under DPP 3(4), "new purpose" means any purpose other than the purpose (or a directly related purpose) for which the data was to be used at the time of collection.

[8] "Human-in-the-loop" refers to human actors retaining control of the decision-making process to prevent and/or mitigate errors or improper output and/or decisions made by AI.

7 Uphold data access and correction rights

The complex data flows within an agentic AI system may pose difficulties in identifying, locating and retrieving all personal data held with respect to an individual or in tracing the source(s) of inaccurate personal data. Therefore, organisations should select agentic AI systems that adopt the principles of "privacy-by-design" and "privacy-by-default" to ensure that the systems support the exercise of data access and data correction rights under DPP 6. Organisations should ascertain whether agentic AI service providers have incorporated data protection principles into their system designs, established robust data governance mechanisms and embedded risk management measures by default.

8 Conduct continuous risk assessments

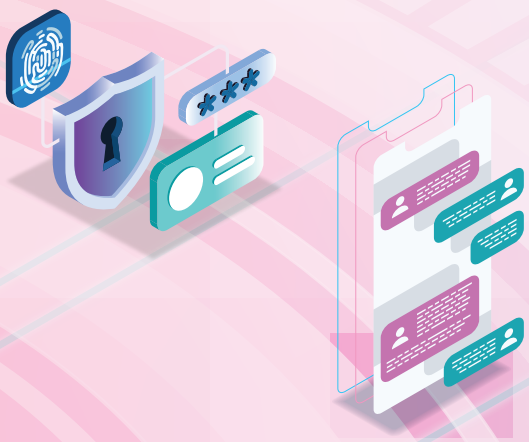
Before using agentic AI, organisations should test its safety and reliability. During its use, organisations should continuously assess the associated personal data privacy risks and adopt appropriate governance and supervision measures proportionate to the purposes of the use of agentic AI, the nature and sensitivity of the personal data involved, the reversibility of the operations and the potential impact on individuals, as well as remain vigilant to any requests by the agentic AI to execute high-risk operations. If the decisions made by agentic AI may have a significant impact on individuals, organisations should adopt a "human-in-the-loop" approach to retain final control over the decision-making process to prevent and mitigate AI-induced errors and automation bias. This includes high-stakes, irreversible or atypical actions, such as approving the transmission of personal data to third parties, permanently deleting personal data or modifying system configurations.

9 Assign clear responsibilities and provide training

Organisations should establish an internal governance structure equipped with sufficient resources, expertise and decision-making authority and should allocate responsibilities across the organisation to ensure that personal data privacy is adequately protected throughout the implementation and use of agentic AI. When engaging an external vendor, organisations should adopt contractual or other means to ensure that the data processor complies with the relevant requirements regarding data retention (DPP 2) and security (DPP 4) under the PDPO. Adequate training should be provided to all relevant personnel to raise their awareness of the personal data privacy risks associated with the use of agentic AI and their ability to manage such risks.

Recommendations 1, 3, 4, 5 and 6 above also apply to individual users who process personal data when using agentic AI.

To facilitate implementation of the recommendations in this guidance, organisations and individual users may refer to the Security Checklist in Annex A, which sets forth the practical steps that users may take to safeguard personal data privacy when using agentic AI.



Annex A: Security Checklist[9]

Checklist	Applicable to	
	Organisations	Individuals
Evaluation Stage		
☐ Delineate the purposes for which personal data will be collected and processed by agentic AI	❖	❖
☐ Acquire a comprehensive understanding of the relevant knowledge about agentic AI	❖	❖
☐ Choose agentic AI with built-in security safeguards	❖	❖
☐ Do not use agentic AI that poses significant and unreasonable security risks	❖	❖
☐ Review the recent developments of the agentic AI project intended for deployment	❖	❖
Preparation Stage		
☐ Clearly state the relevant information concerning the use of agentic AI to process personal data in the Personal Information Collection Statement and Privacy Policy Statement	❖	
☐ Obtain the agentic AI installation files from official channels and verify their authenticity and integrity	❖	❖
☐ Select an appropriate deployment method and implement corresponding security measures	❖	❖
☐ Select an appropriate large language model for operating the agentic AI	❖	❖
☐ Enhance security protection through security tools, security services, etc.	❖	❖
☐ Establish a robust internal governance structure and provide adequate training to relevant personnel	❖	
Deployment Stage		
☐ Do not deploy agentic AI using one-click deployment or other automated scripts from unverified sources	❖	❖
☐ Configure security settings for the Internet-facing surfaces of the agentic AI	❖	❖
☐ Enable and configure the logging function of the agentic AI	❖	❖
☐ Check the trustworthiness of plugin sources before installation	❖	❖
☐ Grant the agentic AI minimum rights; do not grant it operating system administrator rights	❖	❖
☐ Restrict the scope of information and systems that the agentic AI can access	❖	❖
☐ Define high-risk operations in advance and manage them using the security mechanisms of the agentic AI	❖	❖
Use Stage		
☐ Strengthen security safeguards for the use of skills	❖	❖
☐ Strengthen security protections for personal and sensitive data	❖	❖
☐ Access online information services or participate in online social media discussions with agentic AI while remaining in compliance with the Personal Data (Privacy) Ordinance	❖	❖
☐ Review the installation and configuration	❖	❖
☐ Regularly review the necessity and security of agentic AI's public-facing interfaces	❖	❖
☐ Regularly back up important data within the agentic AI operating environment	❖	❖
☐ Regularly review and manually manage the contents of long-term memory file records	❖	❖
☐ Revoke rights for sensitive actions in a timely manner; regularly remove unused skills and conversation records	❖	❖
☐ Monitor alert messages and carry out verification and follow-up actions to eliminate security risks promptly	❖	❖
☐ Keep abreast of security advisories related to agentic AI and take corresponding measures promptly	❖	❖
Uninstall Stage		
☐ Stop using the agentic AI before uninstallation	❖	❖
☐ Export or transfer data that needs to be retained before uninstallation	❖	❖
☐ Clean up the operating environment	❖	❖

[9] The Security Checklist set out in Annex A was prepared with reference to the Practical Guidance of Cybersecurity Standards – Security Guidelines for the Deployment and Use of AI Agents, published by the National Technical Committee 260 on Cybersecurity of Standardization Administration of China in July 2026.

Tel : **2827 2827**

Fax : **2877 7026**

Address : **Unit 1303, 13/F, Dah Sing Financial Centre, 248 Queen's Road East, Wanchai, Hong Kong**

E-mail : **communications@pcpd.org.hk**

PCPD Website
pcpd.org.hk

Download this Publication

Disclaimer

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.

August 2026