Data Protection and Artificial Intelligence

Aug 2024

Introduction

Following the recent advancements in Artificial Intelligence (AI) technology, the use of AI in workplaces is an ongoing trend among many organisations. To support this development, the Office of the Privacy Commissioner for Personal Data (PCPD) issued a paper titled "Artificial Intelligence: Model Personal Data Protection Framework (2024)" in June 2024 to provide institutions with relevant measures regarding personal data protection.

The Framework covers measures in the following 4 areas: **Establish AI Strategy and Governance**, **Conduct Risk Assessment and Human Oversight**, **Execute Customisation of AI Models and Implementation** and **Management of AI Systems and Foster Communication and Engagement with Stakeholder**.

The Data Protection Principles under the Personal Data (Privacy) Ordinance

Generally, organisations should comply with the six **Data Protection Principles** (DPPs) in Schedule 1 of the **Personal Data (Privacy) Ordinance** (PDPO) when handling personal data in the process of procuring, implementing and using AI solutions. The six DPPs are:

1. Purpose and manner of collection
2. Accuracy and duration of retention
3. Use of data
4. Data security
5. Openness and transparency
6. Access and correction

Establish AI Strategy and Governance

It is recommended for organisations to maintain an internal AI Governance Strategy, which consists of:

Strategy: Common elements may include setting out the **functions** and **scopes** of the AI system, ensuring appropriate technical infrastructure such as processing tools, and reviewing the strategies based on feedback and emerging laws.

Governance issues: Governance issues arise especially when third parties are engaged in the customisation or buying of AI. Possible aspects concern:

- the 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
- the plan for continuously scrutinising changing landscape
- the plan for monitoring
- managing and maintaining AI solution
- the evaluation of AI suppliers.

Internal governance structure: A governance committee is recommended with a **cross-functional team** led by **C-level executives**, with the participation of **senior management** and **interdisciplinary**. Sufficient resources, expertise and authority should be given to effectively establish an **internal reporting system** in the case of any system failure or data protection or ethical concerns. Outsourcing external experts for independent AI and ethics advice may also be applicable.

Firms may provide adequate training to relevant personnel to enhance the overall knowledge, skills and awareness on AI utilisation in the working environment.

Conduct Risk Assessment and Human Oversight

Comprehensive risk assessment is necessary for organisations to systematically identify, analyze and evaluate the risks regarding AI systems. Risk assessments should be conducted by a **cross-functional team** and **properly documented** to identify potential risks, and appropriate measures should be adopted in accordance.

Risk factors: Common factors that determine the risk levels in an AI system are:

- the requirements of the six **DPPs** under the PDPO
- **volume**, **sensitivity** and **quality** of data
- the **security** of personal data
- the **probability of privacy risks** and potential **severity** of its impacts
- other ethical factors such as the adequacy of mitigation measures

Level of Human Oversight: A risk-based approach should be adopted to determine the corresponding level of human oversight so as to minimize residual risks. Levels of human oversight determined should be proportionate to the level of relevant risks:

- **High risk levels**: "**human-in-the-loop**"; where human actors retain control of the decision-making process
- **Mid risk levels**: "**human-in-command**"; where human actors make use of the output of AI, oversee its operation and intervene when necessary
- **Low risk levels**: "**human-out-of-the-loop**"; where AI operates without human intervention)) is engaged.

Organisations could also request the AI supplier to provide information and explanation about the AI output to ensure adequate human oversight is established.

Risk Mitigation Trade-offs: Organisations to strike a balance between conflicting criteria when seeking to mitigate risks. In such circumstances, organisations are urged to consider the context in which they are deploying the AI system, and document the rationale for their decisions.

Execute Customization of AI Models and Implementation and Management of AI Systems

The customisation and management process can be dissected into the 3 following steps:

Data preparation and management: Companies should comply with requirements under the PDPO, such as minimising the amount of personal data involved, ensuring the quality of the data, and properly documenting the handling of data.

Customization and implementation:

During the customisation process, it is suggested that organisations:

- **Validate the customization of AI solutions** in respect of privacy obligations and ethical requirements
- **Test the AI system** for errors to ensure its reliability, robustness and fairness
- Perform rigorous **User Acceptance Tests**

During implementation, **a holistic approach** to the security testing of AI systems is recommended. Organizations should observe industry best security practices in maintaining code and managing security risks, and pay due attention to security advisories and alerts.

To ensure system and data security, organizations should establish and implement:

- measures that **minimize the risk of attacks** against machine learning models
- internal guidelines on **the acceptable input** and **the permitted/prohibited prompts** to be entered into the AI system
- multiple **layers of mitigation** that prevent system errors or failures
- **contingency plans** that promptly suspend AI and trigger fallback solutions whenever necessary
- mechanisms that ensure the **transparency** of the operations of the system, and enable **traceability** and **auditability** of the system's output

Management and continuous monitoring: High risk systems require frequent and stringent monitoring and reviewing. For example, organizations may conduct re-assessments of the system to identify and address new risks, as well as monitor AI models for any "model drift" or "model decay".

Additionally, organizations should establish an **AI Incident Response Plan** to monitor and address incidents that may occur. The plan should encompass elements to define, monitor, report, investigate and recover from the incident.

Periodic internal audits should also be conducted to ensure that the use of AI is up to date with relevant policies and strategies of the organization.

Foster Communication and Engagement with Stakeholders

Effective communication should be maintained with the stakeholders of the organization to ensure transparency and to respect the data subjects' rights.

Communication with data subjects: When personal data is involved, organizations must, in accordance with the PDPO, communicate to the data subject:

- the **purpose** for which the data are used
- the **classes of persons** to whom the data may be transferred
- the organization's **policies and practices** in relation to personal data in the customization and use of AI

Data subjects have the right to submit **data access** and **data correction requests**, which organizations are obliged to handle thoroughly. It should also be considered whether an option for individuals to **opt-out** is feasible.

Communication with other stakeholders: Apart from the data subject, the organization also has the obligation to effectively communicate with other stakeholders such as **staff**, **individual customers** and **regulators**.

Feedback channels: Organizations should provide channels for both internal staff and external stakeholders (e.g. customers) to **provide feedback**, **seek explanation** or **request human intervention**.

Language and Manner: Elements of the AI system and its risks should be disclosed and explained clearly in **plain laymen language** upon evaluating the stakeholder's comprehension, needs, and possible adverse impacts to security and legitimacy.

Conclusion

The recommendations mentioned above are by no means exhaustive and additional appropriate measures should be adopted when procuring, implementing, and using AI solutions. It is expected that Hong Kong would safely develop into an innovation and technology hub and the digital economy would expand in the Greater Bay Area under the guidance of this Framework.

Please **reach out to us** if you have any questions.

23/F Pico Tower, 66 Gloucester Road, Hong Kong • 香港告士打道 **66** 号笔克大厦 **23** 楼

T: +852 3170-1088 • **E: CONTACT.HK@KEMPLLP.COM** • **WWW.KEMPLLP.COM**