DSCI
PROMOTING DATA PROTECTION

Brief on

PROPOSED ADVISORY GUIDELINES ON USE OF PERSONAL DATA IN GENERATIVE AI, 2026

(SINGAPORE)

Brief: Proposed Advisory Guidelines on Use of Personal Data in Generative AI, 2026 *(Singapore)*

On 2 June 2026, the Personal Data Protection Commission (PDPC), Singapore, with the support of the Infocomm Media Development Authority (IMDA), issued a preliminary draft titled 'Proposed Advisory Guidelines on Use of Personal Data in Generative AI' hereinafter referred to as 'Guidelines'. The Guidelines aim to clarify how the existing Personal Data Protection Act 2012 (PDPA) applies across the Generative AI lifecycle, seeking to balance individuals' right to protect their personal data with organisations' need to use such data for reasonable purposes.

Part I: Introduction and Scope

The Guidelines are advisory and not legally binding; the PDPA and its subsidiary legislation would prevail in case of any inconsistency. The Guidelines should be read alongside the PDPC's existing instruments, including the Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems, the Advisory Guidelines on Key Concepts in the PDPA, the Privacy Enhancing Technologies Adoption Guide, and the Guide to Basic Anonymization.

Definitions introduced:

- **Generative AI Models:** Models, including those trained on large datasets through self-supervision at scale, that display significant generality and can perform a wide range of distinct tasks across systems or applications.
- **Generative AI Systems:** AI systems or applications based on Generative AI Models, for direct use as well as integration into other systems.

Scope:

The Guidelines address three key data protection issues: the collection and use of personal data to develop Generative AI Models; the allocation of data protection responsibilities across the lifecycle; and the handling of individuals' requests concerning their personal data. They are organized around three lifecycle stages - Development, Deployment, and Post-Deployment.

Part II: Development - Collecting and Using Personal Data

Publicly Available Exception:

Web-scraping is recognised as a common means of compiling training datasets. Where such datasets include personal data, organizations may consider relying on the Publicly Available

Exception (Part 2, First Schedule, PDPA) in lieu of seeking consent, subject to the reasonableness test under Section 18.

Digital barriers: The existence of a digital barrier would not, by itself, prevent personal data from being publicly available. The PDPC considers a digital barrier to be any technical or financial measure that meaningfully restricts data access, including:

- Paywalls or subscriptions (hard, metered, or dynamic).
- Registration requirements, such as sign-up processes requiring name, email, and contact details.
- Authentication mechanisms, such as passwords, API keys, or one-time codes.
- Anti-bot tools, such as bot blockers and CAPTCHA systems.

Assessment factors: Where data sits behind a barrier, relevant factors would include the purpose of the barrier (e.g. data monetization), its effect (whether data remains accessible to the public at large), the steps needed to access it, and whether the data is freely available elsewhere.

Illustrative carve-outs: Public registers (even fee-based), metered news paywalls treated as monetization mechanisms, and open-but-registration-gated online forums are cited as examples of data that remains publicly available despite barriers.

Notify-and-assess best practice: Where an organization intends to scrape personal data behind another organization's barrier, the PDPC considers it best practice to notify that organisation. The notified organization may then assess its own disclosure basis, allow access, deny and harden its barriers to signal that data is not public, or escalate unresolved disputes to the Commission. The applicability of the exception is expressly distinct from contractual terms of use, licensing, and general or criminal law.

Consent and Notification Obligations

For User Data - personal data provided to, or generated through use of, an organization's products and services - consent is required unless deemed consent or an exception (e.g. business improvement or research) applies, complemented by the Notification Obligation under Section 20(1).

Key position on notifications:

- **General Notifications are insufficient:** Generic statements citing 'new product development' would not constitute adequate consent for large-scale model training or fine-tuning.
- **AI-Specific Notifications are required:** Organizations would need to provide explicit notice identifying the model functions, the types of personal data used, how that data trains or fine-tunes the model, and how individuals can decline or withdraw consent.

- **Heightened sensitivity:** This is treated as especially important given concerns over sensitive data being reconstructed or disclosed, and over data being repurposed in unanticipated ways such as financial profiling.

Additional Considerations

Organizations are encouraged to practise data minimization and to implement appropriate technical, process, and legal controls where personal data is necessary for development.

Part III: Deployment - Data Protection Responsibilities of Stakeholders

The Guidelines map PDPA roles onto three stakeholders across the technology stack, clarifying that personal data may be handled by multiple parties.

Stakeholder	Definition	PDPA Role(s)	Core Responsibilities
Model Providers	Develop and/or make available Generative AI Models	Organization (own development); Data Intermediary (inference on behalf of users)	Retention Limitation (S.25); Protection Obligation (S.24); document safeguards for downstream reliance
System Providers	Develop and/or make available Generative AI Systems	Organization and/or Data Intermediary	Protection against evolving risks (e.g. prompt injection); periodic security review; share system-level safeguards
System Deployers	Use and/or enable use of Generative AI Systems under their authority	Organization (bears primary responsibility)	Purpose Limitation (S.18); Protection; Accountability

Model Providers: As organizations, they must comply with all PDPA obligations, including where they collect end-user prompts and inputs from downstream systems. As data intermediaries running inference, they must make reasonable security arrangements and are encouraged to document access controls, data residency, and retention policies to support downstream compliance.

System Providers: Relevant where systems are built bespoke or retailed off-the-shelf, but not where systems are developed and deployed in-house. They are expected to periodically review security arrangements as new system types emerge, and to share information on safeguards - such as development-environment controls and data-leakage testing metrics - with downstream deployers.

System Deployers: They bear primary responsibility for ensuring chosen systems can meet PDPA obligations and must secure sufficient information on upstream safeguards. Specific expectations would include:

- **Purpose discipline:** Specifying the intended purpose and the amount of personal data required and not processing data for illegal or harmful purposes.
- **New data sources:** Tracking and safeguarding prompts, outputs, agent or tool activity data, and internal enterprise data, and educating end users on permissible inputs.
- **Documented accountability:** Developing written policies and pre-emptively publishing them to demonstrate compliance.
- **Agentic risk:** Reviewing the sufficiency of safeguards as capabilities evolve, weighing privacy–utility trade-offs, and referring to IMDA's Model AI Governance Framework for Agentic AI.

Part IV: Post-Deployment - Access and Correction Obligations

The Access and Correction Obligations (Sections 21, 22, and 22A) would apply to personal data in the form in which it was collected, used, and disclosed for model and system development, subject to PDPA exceptions such as disproportionate burden.

Acknowledged challenges:

The PDPC recognises present-day difficulties, including the scale of training data, the storage of data as embeddings rather than in traditional repositories, and the technical difficulty of removing specific information from trained models.

Best practices expected where reasonable:

- **Upstream handling:** Verifying accuracy at collection, applying data-cleaning techniques such as de-duplication, and maintaining data provenance records.
- **Case-by-case review:** Acceding to requests where reasonable, with Retrieval-Augmented Generation databases cited as more tractable, and removing inaccurate data before future training runs.
- **Emerging techniques:** Tracking the maturity of, and adopting, measures such as machine unlearning to remove inaccurate personal data.

Part V: Public Consultation Questions

The PDPC invites views on five questions:

Q1. Further examples of digital barriers:

Having set out a functional test rather than a fixed list, the Commission intend to know which further restrictions practitioners encounter that would benefit from the same clarity useful given how quickly access-control technology changes.

Q2. Should organizations provide 'AI-Specific Notifications' in situations other than Generative AI Model training and/or fine-tuning?

The draft recommends organisations to send AI-specific notifications when they train or fine-tune models. This brings up to the question if the same should apply to other instances of usage of personal data in Generative AI.

Q3. Additional best practices for notifications and for access and correction requests:

These paragraphs cover how to write notifications and how to handle access and correction requests. This asks what other useful approaches organizations have found in practice.

(Q4) Further safeguard information that Providers should share downstream:

The Guidelines depend on each layer being open about the safeguards it has in place. The question is if the providers may share if any further safeguard information is required to be provided so others down the line can meet their own obligations.

(Q5) Are there other agent-specific data challenges or risks that would be helpful to clarify at paragraph 9.5?

Further clarifications on AI agents specific data challenges or risks are asked as the Guidelines indicate only their ability to plan and act.

Conclusion

The Guidelines adopt an interpretive, principles-based approach, extending Singapore's technology-neutral PDPA to Generative AI rather than enacting bespoke legislation. They offer operational clarity on web-scraping and the public-availability test, raise the consent threshold for training on User Data, and allocate accountability across a Model Provider, System Provider, and System Deployer taxonomy anchored on the deployer as the primary responsible party. Paired with a frank treatment of access and correction feasibility, the Guidelines signal a pragmatic and technically literate posture toward Generative AI governance.