



Advisory Guidelines on the use of Personal Data in Generative AI
1 July 2026

AsiaDPO is a Singapore registered society of a self-organising peer-to-peer community of Data Protection Officers (DPOs). We are committed to the development and advancement of data protection and privacy domains through a practice-led approach, serving as an expert group with a distinctive voice.

Contact Person

Lindsay Courtney
AsiaDPO

administrator@asiadpo.org

AsiaDPO remains the first and only self-organising, peer-to-peer registered society *by DPOs for DPOs*. AsiaDPO exists to develop and advance practices in data protection and privacy domains, to serve as a trusted resource of industry expertise and to be a distinctive voice for our members. This is the DPO in action, in practice. As DPOs, we are uniquely placed to balance data use and protective outcomes by integrating privacy, accountability and trust into critical organisational capabilities.

We thank the Personal Data Protection Commission (PDPC) and the Infocomm Media Development Authority (IMDA) for the opportunity to provide comments on the Proposed Advisory Guidelines on the use of Personal Data in Generative AI (“Guidelines”). The proposed Guidelines will help organisations strike a balance between the need to protect individuals’ personal data and the responsible development and deployment of Generative AI.

This submission represents anonymised feedback from AsiaDPO members to the proposals in this closed Consultation, in line with AsiaDPO’s constitution, which is designed to increase peer-to-peer collaboration in a safe and open environment.

General Comments

We welcome the proposed Guidelines and the PDPC’s engagement with industry. The comments below are offered from a practice-led perspective, with the aim of helping data protection officers, privacy teams, providers, deployers and organisations apply the Guidelines consistently in practice. Where we suggest a clarification, our intention is to support workable and proportionate compliance, not to reduce the protections the Guidelines provide.

Responses to PDPC’s Questions

Question 1: Are there other examples of “digital barriers” under paragraph 3.4 that would be useful to clarify?

Automated anti-scraping signals as a contextual factor. We recommend that the Guidelines clarify how automated anti-scraping signals, such as a “robots.txt” file (a standard machine-readable instruction by which a website asks automated programs not to collect parts of its content), should be taken into account when assessing the context in which data is made accessible and whether collection is contemplated. Such a signal does not necessarily operate as a digital barrier under paragraph 3.4, and it may not prevent access in a technical sense, but it may nonetheless indicate that the source organisation does not contemplate or permit systematic collection, and so may be relevant both to the public availability assessment and to whether collection and use would be appropriate in the circumstances.

Worked examples applying the factors in paragraph 3.5. The concept of a digital barrier is useful, but its application is fact dependent because registration requirements, subscriptions, authentication mechanisms and similar controls serve different purposes across different platforms. We would propose that the Guidelines provide additional positive and negative worked examples applying the factors in paragraph 3.5, illustrating when a barrier functions primarily as a monetisation or convenience mechanism so that the data remains generally accessible to the public, and when it instead indicates that access is bounded to a defined audience. It would also assist if the Guidelines indicated whether a platform’s intended audience or contextual norms should affect the analysis. In practice, the absence of worked examples in this area can lead organisations to reach inconsistent conclusions on similar facts.

Access limited to a defined audience. A category that would benefit from clarification is data to which access is limited to a defined audience, such as invite-only or closed communities, authenticated customer dashboards, and professional or institutional databases available only to subscribers or registered members. We suggest that audience-limited access should be treated as a relevant factor under paragraph 3.5, rather than as automatically determinative. The relevant question is whether the access arrangement indicates that the data was made available to the public at large, or only to a specific audience for a specific purpose. This approach is consistent with paragraph 3.6(c), which treats large online forums that are open to anyone to join, subject only to registration, as publicly available.

Additional examples of access controls. It would assist if the Guidelines included examples involving geo-blocking or geofencing, enterprise-domain or institutional-network restrictions, and access through proprietary applications, APIs, approved-developer programmes or other scoped technical interfaces. Such controls may restrict access by location, network, application environment or technical interface rather than by individual identity. We suggest that the Guidelines clarify how such controls should be weighed under paragraph 3.5, including where they restrict access only in part or serve purposes such as security, licensing, service delivery or jurisdiction-specific compliance, rather than necessarily indicating that the data is not publicly available.

Point in time at which public availability is assessed. Organisations can find it burdensome, in practice, to continuously re-verify that scraped data remains publicly available. We suggest that the Guidelines confirm, consistent with the Advisory Guidelines on Key Concepts in the PDPA, that whether personal data is publicly available is assessed at the point of collection, and that data does not cease to have been validly collected merely because it later becomes inaccessible. Further worked examples illustrating how the factors in paragraph 3.5 are applied would also be helpful.

Question 2: Should organisations provide “AI-Specific Notifications” in situations other than Generative AI Model training and/or fine-tuning?

Avoiding duplicative notifications where an existing notification is adequate. We suggest that a separate AI-specific notification should not be required in every case where a Generative AI system processes

personal data. Where a Generative AI system is used for an existing, already-notified purpose, such as customer support, administrative support, translation or document summarisation, including where general productivity tools have been enhanced with Generative AI features, and the personal data is not used to train, fine-tune or improve a model, the existing notification should be sufficient, provided it adequately describes the purpose of collection, use and disclosure. In these situations the method of processing has changed, but the purpose has not.

We suggest that an AI-specific notification should instead be expected where the use of Generative AI materially changes the purpose, nature, or consequences of processing, or departs materially from individuals' reasonable expectations. It would assist organisations if the Guidelines identified the kinds of situations in which an AI-specific notification would generally be expected, for example:

- where personal data, including in prompts, uploaded files, outputs or interaction logs are used to train, fine-tune, evaluate or improve an in-house, third party, or system provider's model;
- where personal data is used for retrieval-augmented generation, vector databases, embeddings, memory features or personalisation that persist beyond a single session (these are techniques by which a system stores or re-uses personal data so that it can be drawn on in later interactions); and
- where personal data is used to generate profiles, predictions or decisions that individuals would not reasonably expect.

Recognising existing transparency frameworks. We suggest that in assessing whether the notification obligation is met, the PDPC could recognise that an organisation may already operate under an existing transparency or disclosure framework that achieves a comparable outcome. Some sectoral frameworks already require organisations to explain how decisions supported by AI are made. Where such a framework adequately informs individuals of the relevant purposes, it may be capable of meeting the notification obligation without a separate, dedicated notice, provided this is assessed together with any existing privacy notices and the applicable PDPA consent, deemed consent or exceptions analysis relied on by the organisation. Such recognition should not be treated as automatically displacing the distinct requirements of the PDPA.

Question 3: Are there other best practices that would be useful to highlight at paragraphs 4.6 (substance of notifications) and 10.4 (access and correction requests)?

On the substance of notifications (paragraph 4.6).

We suggest that paragraph 4.6 could, where relevant and practicable, encourage organisations to include two further matters in their notifications: the scope and effect of any option to opt out of or withdraw consent; and whether prompts, uploaded files, outputs, interaction logs or telemetry may be used for training, fine-tuning, evaluation, monitoring or product improvement.

The Guidelines could usefully confirm that the multiple communication channels shown in paragraph 4.6 example are illustrative rather than cumulative requirements, and that organisations may select appropriate channels having regard to context, the user population and the risk profile. Relatedly, while paragraph 4.4 explains that General Notifications are insufficient for large-scale model training or fine-tuning, we suggest it would assist organisations to confirm that AI-related disclosures may otherwise be delivered through a proportionate combination of general policies and contextual product design, so as to inform individuals effectively without creating duplicative notices.

On access and correction requests (paragraphs 10.3 and 10.4).

We would propose that the Guidelines clarify how access and correction requests should be assessed by reference to where personal data is held or retrievable within the Generative AI system or processing flow. Personal data may be held in different forms or locations, including original source records, retrieval-augmented generation repositories (stores of documents that a system can draw on at the time of a query), embeddings (numerical representations derived from data), system logs, fine-tuning datasets, and the outputs a model generates. These locations may differ significantly in whether the personal data is identifiable, searchable, retrievable, and reasonably correctable. Providing non-exhaustive examples or factors relevant to assessing what is reasonable and technically feasible in different contexts would help organisations respond appropriately to access and correction requests, while preserving flexibility.

Separately it would be useful for the Guidelines to clarify the treatment of embeddings. Embeddings are numerical representations derived from data and may not, in all cases, be readily intelligible or directly attributable to an individual. However, depending on how they are generated, stored, linked to source records, or used with other information, embeddings may in some circumstances remain capable of being associated with identifiable individuals. Further guidance on when embeddings may constitute personal data, and what access or correction obligations may reasonably apply to them, would assist organisations in designing appropriate governance and request-handling processes.

It would also assist individuals and organisations if the Guidelines addressed how a request should be handled where the individual interacts with one organisation but the model was developed by another, including whether the request is made to the first organisation, which then relays it to the developer, or whether and in what circumstances the individual may approach the model developer or system provider directly, for example where that party independently collects, uses or controls the relevant personal data.

We suggest that the Guidelines could recognise, as good practice, that an organisation may adopt a documented framework for responding to AI-related access and correction requests. Such a framework would support consistent triage across technical layers, an assessment of technical feasibility and proportionality, and clear communication to the individual about what has been done and what limitations remain. This is best presented as an operational and accountability measure, rather than as a separate legal requirement.

Where it is not technically feasible to extract or correct personal data at the level of a trained model, we respectfully suggest that the Guidelines could confirm a proportionate approach: that an organisation which implements the upstream data-handling measures described in paragraph 10.4(a), together with robust controls on what the model outputs (for example, system instructions or guardrails that prevent the model from producing inaccurate personal data, output filters, post-processing or escalation mechanisms where appropriate to the use case), may be relevant to assessing whether the organisation has taken reasonable steps to address the Correction Obligation in respect of that output. This would recognise the present-day technical limitations described in paragraph 10.3 without setting expectations that organisations cannot meet.

We note that machine unlearning, a developing set of techniques for removing the influence of specific data from a trained model, is referred to in paragraph 10.4(c). We suggest it would be helpful to frame machine unlearning as an emerging technique that may be relevant where it is mature and feasible, rather than as a generally expected measure, so that organisations are not held to expectations that current technology cannot reliably support.

Question 4: Is there additional information on data protection safeguards that would be helpful for Model and System Providers to share with downstream stakeholders?

System Deployers bear primary responsibility for ensuring that they can meet their own obligations under the PDPA when using a Generative AI system. In practice, however, deployers often have limited visibility into proprietary models and limited ability to negotiate the terms on which a model or system is provided, particularly where the provider is located overseas. We suggest that the Guidelines could assist by indicating, in a practical and non-exhaustive way, the categories of information that Model and System Providers could make available to support downstream due diligence, where reasonable and proportionate, without requiring disclosure of source code, model weights, proprietary architecture or sensitive security information, and having regard to the provider's role and the deployer's intended use case. Such categories could include:

- whether downstream prompts, inputs, outputs, uploaded files, logs, embeddings or metadata are used for training, fine-tuning, evaluation, abuse monitoring or product improvement;
- retention timelines;
- processing locations and the use of sub-processors;
- security and Generative AI-specific safeguards;
- the configuration options available to the deployer;
- high level information, where available and relevant to the intended deployment, on the categories and sources of personal data used for model development and the basis on which such data was collected or used, without requiring disclosure of proprietary datasets, detailed provenance records or commercially or security-sensitive information; and

- the support available for handling individual access and correction requests, and for incident response.

We suggest that one particularly useful category is the treatment of interaction data across its full lifecycle, namely whether prompts, uploaded files, interaction data and outputs are processed only to provide the requested service, or are also retained for logging, debugging, abuse monitoring or support, reviewed by personnel, or re-used for evaluation, service improvement, fine-tuning or model development. More standardised disclosure in this area would support consistent due diligence by deployers and meaningful transparency to users.

A due diligence checklist template would also assist deployers, particularly smaller organisations, in obtaining clear declarations or attestations from providers and in conducting a proportionate assessment when procuring a model or system.

Question 5: Are there other agent-specific data challenges or risks that would be helpful to clarify at paragraph 9.5?

We recommend that paragraph 9.5 be expanded to address a broader range of risks that arise as Generative AI systems take on agentic functions, meaning that they plan and carry out multi-step actions across systems with reduced human involvement. From a practice-led perspective, PDPA-relevant risks include privilege escalation (an agent obtaining access beyond what was intended), aggregation of personal data across systems, disclosures made by the agent without specific human direction, drift away from the original purpose, the accumulation of persistent memory or long-lived context, and delegated decision-making. It would assist organisations if the Guidelines provided examples of relevant controls, such as human oversight and escalation thresholds, scoped permissions for the tools an agent may use, action and audit logging, agent identity management, governance for sensitive actions, use of kill switches and broader monitoring.

Where paragraph 9.5 refers to escalation protocols for “complex or high-risk tasks”, it would assist if the Guidelines provided non-exhaustive factors or examples for identifying such tasks. Relevant factors may include the sensitivity and volume of personal data involved, whether external disclosure may occur, whether the action affects an individual’s rights, interests or access to services, whether the action is reversible, and whether human review is practicable before the action is taken. It would be useful for the Guidelines to address cross-step aggregation and autonomous tool or API calls as agent-specific risks. Cross-step aggregation may generate new personal data, profiles, metadata or sensitive inferences, and may cause processing to drift beyond the originally notified purpose. Autonomous tool or API calls may disclose personal data to external services or recipients, particularly where such disclosure was not contemplated in the original notification or purpose assessment. These examples would assist organisations in applying the Purpose Limitation, Notification and Protection Obligations to agentic systems.

We also suggest that the Guidelines could address multi-agent systems, in which several agents interact, as a distinct source of risk. In such systems, personal data may propagate across agents, it can become harder to identify which component was responsible for a given action, context and memory may accumulate, and logging, auditability and human oversight become more complex.

We further suggest that it would assist organisations if the Guidelines clarified how the Accountability Obligation, and in particular the maintenance of an accurate inventory of personal data, can be met where an agent operates with a degree of autonomy and is not given specific instructions about the data-processing pathways it follows or the databases it may access.

Additional Practice-Led Observations

The following observations do not respond directly to any of the five questions, but we offer them because they bear on the practical implementation of the Guidelines.

Public availability should be relevant but not conclusive. Paragraph 3.3 makes the Publicly Available Exception subject to whether a reasonable person would consider the use of the data appropriate in the circumstances. We recommend that the Guidelines make clear that public availability is not, by itself, conclusive of whether collecting and using that data to develop a Generative AI Model is appropriate in the circumstances. Relevant considerations may include the context in which the data was made available, individuals' reasonable expectations, whether individuals had a meaningful choice over publication, the scale of collection, whether the data will be aggregated with other datasets, the sensitivity of the data, the potential for profiling, and the practical difficulty of removing or correcting data after a model has been trained. Further examples illustrating when such use would and would not be appropriate would support consistent assessments.

Personal data obtained from public registers. We suggest that, where personal data is obtained from public registers, public accessibility should not be treated as conclusive for the purpose of developing a Generative AI Model. Organisations should still consider whether collection and use would be appropriate in the circumstances, having regard to the purpose for which the register was made available, whether individuals had a meaningful choice over their inclusion, and whether bulk extraction, aggregation with other datasets, or use for model training or profiling would be appropriate. We note that paragraph 3.6(a) treats registers administered by public agencies, including those accessible on payment of a fee, as examples of publicly available data; our observation concerns the separate question of whether a particular use is appropriate, rather than whether the data is publicly available.

Relationship between the Exception and other legal constraints. We suggest that the clarification in paragraph 3.10 is helpful and could be cross-referenced earlier in the discussion of scraping, so that organisations are reminded that reliance on the Publicly Available Exception under the PDPA is distinct from their obligations under contract, intellectual property law, foreign data protection laws and other legal frameworks. This is particularly relevant where a Singapore-based organisation collects personal

data from overseas sources or personal data relating to individuals outside Singapore, as reliance on the Publicly Available Exception under the PDPA may not address whether foreign privacy or data protection laws apply to the collection or use. It would also assist to confirm that a source organisation is not required to permit scraping merely because data is considered publicly available for PDPA purposes, as it may have separate legal, contractual or security reasons for not doing so.

Notifying the source organisation before scraping. The best practice in paragraphs 3.7 to 3.9 of notifying a source organisation before scraping data behind a digital barrier, raises practical questions, particularly for organisations that may need to engage many sources. We recommend that the Guidelines clarify how a notification should be directed, for example to a published contact point for data, privacy or scraping matters where one exists; what level of detail is expected; and how the absence of a response should be treated. In particular, we suggest it would be helpful to confirm that the absence of a response should not be treated as consent, authorisation, or confirmation that the data is publicly available. We also suggest that whether personal data is publicly available is, under the PDPA, an objective question, and that it would be helpful for the Guidelines to indicate the weight to be given to a source organisation's own view in that assessment.

Consent, deemed consent and exceptions across the lifecycle. We suggest that the Guidelines could clarify how notification and consent operate in multi-party arrangements, where model providers, system providers and enterprise customers are involved and an upstream provider may have no direct relationship with the individual, including which party bears primary accountability. It would also assist to clarify how deemed consent and the exceptions to consent may apply to Generative AI development, so that organisations do not conclude that consent is always required, and to provide examples of how the relevant exceptions may apply to use cases such as fraud detection or risk monitoring, where seeking consent or offering an opt-out could undermine the purpose of the control, while emphasising necessity, proportionality and governance.

Lifecycle data governance. We suggest that the Guidelines could address expectations for personal data already used to train models before the Guidelines take effect, including whether a risk-based remediation approach is acceptable. It would assist to clarify what is expected when consent is withdrawn after training, distinguishing between ceasing future use, excluding the data from future training runs, and modifying an already-trained model where this is feasible and proportionate.

It would also be helpful to confirm that data minimisation under paragraph 5.1 means using only personal data that is adequate, relevant and necessary for the training and functioning of a model, rather than reducing data volumes as such. Relatedly, paragraph 9.2 could encourage deployers to consider, at the outset of each use case, whether personal data in identifiable form is needed. Where a Generative AI system permits open-ended user input, such as through a conversational interface, it would be useful to clarify that organisations may, in appropriate cases, communicate the intended purpose of the system

and instruct users to input only relevant personal data, rather than specify exhaustively every permitted category of input.

On retention, illustrative approaches, governance controls and review cycles would assist, including confirmation that retention for legitimate purposes such as investigations, regulatory compliance or legal holds may remain permissible where necessary and appropriately governed.

Handling of data in deployed systems. We suggest that the Guidelines could clarify the expectations for prompts, uploaded content and outputs in deployed systems, including when they may be stored, reviewed, retained or re-used, and the safeguards expected for segregation and tenant isolation in shared environments. It would assist to distinguish processing that is necessary to deliver the requested service from a provider's re-use of that data for evaluation, improvement or model development, and to confirm whether enterprise customers may restrict such re-use by contract. Guidance on the logging and traceability of AI-generated outputs would support dispute handling, incident investigation and accountability in customer-facing environments.

Stakeholder roles are functional, not fixed. We suggest that the Guidelines could state more explicitly that the roles of Model Provider, System Provider and System Deployer are functional rather than fixed, since a single organisation may perform more than one role and may move between roles depending on the function it performs at a given stage. We also suggest that the responsibility described in paragraph 9.1 could be expressed as a responsibility on System Deployers to take reasonable steps to select, configure and use Generative AI systems so that the deployer can comply with its own obligations under the PDPA, rather than in a way that could be read as requiring the systems themselves to meet those obligations, or as making deployers responsible for upstream conduct that they cannot assess or control.

Scope precision in examples. It would assist organisations if the Guidelines confirmed that the chatbot guidance applies where a chatbot involves Generative AI or other relevant AI capabilities, as distinct from purely scripted or rule-based systems that present different risks. We also suggest that the worked examples should be framed as illustrative and should avoid creating unintended negative signals about broad categories of use cases. For example, it would be helpful to clarify that the example given at paragraph 4.5 is not intended to suggest that financial profiling, financial risk assessment, or fraud detection and prevention are inherently inappropriate. Such use cases may serve legitimate and beneficial purposes. Alternatively, the reference to financial profiling could be replaced with a more neutral example, or with examples across multiple sectors, to avoid the impression that the Guidelines identify financial sector profiling as a special category of concern.

Intra-group development. It would assist if the Guidelines included a worked example covering group-wide AI development, where several entities within a group contribute data to a shared platform for common purposes, clarifying which entity should provide notice, how responsibilities are allocated, whether a central platform team may act as a data intermediary, and how access, correction and withdrawal requests are handled where processing spans multiple entities.

Interoperability with other regulatory frameworks. Where organisations are already subject to sectoral regulations that address similar issues, it would assist if the Guidelines recognise that these frameworks may be relevant to how organisations operationalise their PDPA obligations in Generative AI use cases. This would help organisations avoid duplicative processes while preserving the distinct requirements of the PDPA.

Conclusion

We are supportive of the proposed Guidelines and welcome the PDPC's consultative approach. Our comments are offered constructively, in the shared interest of guidance that is clear, proportionate and workable in practice. A common thread runs through them: that expectations under the Guidelines should be risk-based, proportionate and technology-neutral, and should be made concrete through worked examples and, where possible, illustrative scenarios that help organisations reach consistent conclusions. Recognising that the laws and regulations of today can greatly shape the data protection profession of tomorrow, we would welcome the opportunity to continue our dialogue with the PDPC on any of the matters set out above.