1 July 2026

BSA COMMENTS ON PUBLIC CONSULTATION ON PROPOSED ADVISORY GUIDELINES ON USE OF PERSONAL DATA IN GENERATIVE AI

Submitted Electronically to the Personal Data Protection Commission

The Business Software Alliance (**BSA**)[1] appreciates the opportunity to comment on the public consultation on the proposed Advisory Guidelines (Guidelines) on the Use of Personal Data in Generative AI by the Personal Data Protection Commission (**PDPC**). BSA is the global trade association of the enterprise software industry. Our members are leaders in AI, cybersecurity, cloud computing, and other cutting-edge technologies. We value our longstanding engagement with the PDPC and welcome the opportunity to share industry perspectives on the responsible use of personal data in Generative AI.

Summary of BSA's Comments

BSA recommends that the Guidelines:

1. Clarify that the Publicly Available Exception is determined under the Personal Data Protection Act (**PDPA**) independently of contractual, intellectual property, and other non-privacy considerations, and remove the proposed notification process.
2. Provide additional guidance on circumstances where consent is not required, including deemed consent and the business improvement exception, and adopt a risk-based approach to consent and notification for generative AI.
3. Retain the model provider, system provider, and system deployer framework, and recognise that standardised transparency mechanisms are sufficient to satisfy information-sharing obligations across the AI supply chain.
4. Clarify that access and correction obligations should be assessed in light of technical feasibility and system architecture, and avoid implying that organisations must adopt any particular technical approach to comply.

Digital Barriers and the Publicly Available Exception

BSA appreciates the PDPC's efforts to provide greater clarity on the application of the Publicly Available Exception in the context of web-scraped data (e.g., information accessed from the Internet by automated means) used for generative AI development. We welcome the recognition

[1] BSA's members include: Adobe, Alteryx, Amadeus, Amazon Web Services, Asana, Atlassian, Autodesk, Avalara, Bentley Systems, Box, Cisco, Cloudflare, Cohere, Cohesity, Dassault Systemes, Databricks, Datadog, Docusign, Dropbox, Elastic, EY, Graphisoft, HubSpot, IBM, Kyndryl, MathWorks, Microsoft, Notion, Okta, OpenAI, Oracle, PagerDuty, Palo Alto Networks, PTC, Rubrik, Salesforce, SAP, ServiceNow, Shopify Inc., Siemens Industry Software Inc., TrendAI, TriNet, Veeam, Workday, Zendesk, and Zoom Communications Inc.

that the existence of a digital barrier does not automatically mean that data is not publicly available. The factors set out in paragraphs 3.5 and 3.6 provide useful guidance for organisations assessing whether personal data remains publicly available despite the presence of a digital barrier.

Some digital barriers may take forms beyond technical access controls. For example, geolocation-based access restrictions — such as region-locking, IP-based filtering, or territorial content restrictions — may limit or deny access to online data based on a user's geographic location, including by preventing access from outside a designated jurisdiction or market. Similarly, attribution and licensing conditions may operate as digital barriers where access to or use of data is made conditional on requirements such as crediting the data source, displaying specified licensing terms, or including a reference to the applicable licence in a product, application, or website that incorporates or reproduces the data.

We have significant concerns, however, with the Guidelines' treatment of digital barriers in paragraphs 3.7 to 3.9. We are particularly concerned with the statement that it is "best practice" for an organisation seeking to collect publicly available personal data behind a digital barrier to notify the organisation controlling the data, which may then make its own assessment as to whether the data is publicly available and potentially deny access.

We caution against this approach, which conflates the privacy issue at the core of this section of the Guidelines with a range of other non-privacy issues. The determination of whether personal data is "publicly available" under the PDPA should remain an assessment based on the relevant facts and circumstances and the Commission's existing guidance. The proposed notification process blends this personal data protection analysis with separate issues, such as contractual restrictions, website terms of use, licensing arrangements, or intellectual property rights. While these issues may be relevant to whether data may be accessed or used under other areas of law, they are distinct from the question of whether personal data is publicly available under the PDPA.

In particular, the Guidelines could create uncertainty by effectively allowing the organisation controlling access to data to make a de facto unilateral determination of whether the Publicly Available Exception applies under PDPA. Further, paragraph 3.7 suggests that notification helps ensure that organisations provide sufficient information regarding the purpose of collection when collecting personal data from another organisation. However, where an organisation has determined that the Publicly Available Exception applies, the collection, use, or disclosure of the personal data is already permitted under the PDPA without requiring consent. The notification process is therefore not necessary to support consent- or notification-related obligations.

Recommendation: BSA recommends that the PDPC clarify that the determination of whether personal data is "publicly available" remains an assessment under the PDPA based on the relevant facts and circumstances, independent of contractual terms, intellectual property rights, or other non-privacy considerations. We also strongly recommend removing the reference to notification. If the notification process is retained, the Guidelines should make clear that it is intended solely to facilitate communication between organisations and does not create a right for the notified organisation to determine whether the Publicly Available Exception applies.

Consent and Notification Requirements

BSA appreciates the PDPC's efforts to provide guidance on the application of consent and notification obligations in the context of generative AI. Given the increasing adoption of generative AI across a wide range of use cases, greater clarity in this area would be valuable for organisations seeking to comply with the PDPA.

As an initial matter, we encourage the PDPC to provide additional guidance on circumstances in which consent is not required under the PDPA. The Guidelines focus primarily on situations in which organisations must obtain consent for the use of personal data in generative AI development. However, organisations would also benefit from clearer guidance and practical examples illustrating how existing provisions of the PDPA, including deemed consent and the business improvement exception, may apply in the context of generative AI development and deployment. This can ensure that companies are not over-reliant on consent and are able to process personal data in line with the PDPA for a range of purposes without burdening individual data subjects with additional consent requests.

We have three main concerns with the Guidelines' discussion of AI-specific notifications and consent.

First, this section of the Guidance appears to apply to all generative AI models. It extends to any "large-scale AI model training and/or fine-tuning," a threshold that is undefined (e.g., whether it is based on volume of training data, number of data subjects, or model size/weights). This has the practical effect of treating all generative AI training as requiring consent, without recognising that companies may process personal data in such contexts based on other legal bases or existing consent exceptions (e.g., business improvement and research) recognised under the PDPA. Generative AI models are usually trained on very large datasets and used for a wide range of purposes, with very different implications or risks for individuals. For example, a generative AI model can be used to generate formatted templates from text prompts, which does not create the type of risks that may be posed by a generative AI model used to produce medical advice. Similarly, generative AI models may be trained and fine-tuned to draft emails, generate meeting summaries, or produce recipes, which present very different risks when compared to generative AI that supports consequential decisions related to lending, employment, or healthcare. The Guidelines should not require the same safeguards across all generative AI training but instead recognise that many of these types of processing may appropriately rely on existing consent exceptions recognised under the PDPA.[2]

Second, the Guidelines appear to assume that the purpose of processing personal data is the training of a generative AI model itself. In practice, the goal of training a generative AI model is to achieve a specific purpose. For example, an organisation may fine tune a generative AI model to

[2] For example, the PDPC's prior guidance interpreting the PDPA recognises that companies may rely on legitimate interests to create a fraud detection model, without seeking consent from individual data subjects. *See* PDPC, Advisory Guidelines on Key Concepts in the Personal Data Protection Act (Revised 29 April 2026) page 64, *available at* https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/advisory-guidelines/ag-on-key-concepts/advisory-guidelines-on-key-concepts-in-the-pdpa-17-may-2022.pdf. Companies should be able to rely on legitimate interests for this processing whether the fraud detection model involves generative AI model or any other method, AI-related or not. We are concerned the current Guidelines undermine that outcome, both for this use and for other uses that currently rely on legitimate interests.

generate document templates, summarise information, or provide customer support. In other contexts, a model may be fine tuned to support consequential activities such as medical recommendations or financial assessments. Organisations that train such generative AI models in these scenarios are therefore processing personal data to achieve these other purposes.[3] The Guidelines should not treat training a generative AI model as the ultimate purpose, but should recognise that generative AI may be used to achieve a variety of purposes that create different risk profiles and may require different safeguards.

The cybersecurity sector illustrates this point. Organisations train AI-based threat detection systems using network telemetry data, such as DNS queries and network traffic logs, that may incidentally contain personal data. This processing is undertaken to detect malware or other vulnerabilities and prevent cyberattacks. Requiring consent for it would be operationally infeasible and would undermine the threat intelligence that protects the very individuals whose data is involved.

Third, the proposed AI-specific notification requirements are overly prescriptive, particularly if applied broadly across all generative AI use cases. Transparency is important, but notifications are most effective when they help individuals understand meaningful information about how their personal data is being used. The Guidelines currently view general disclosures as insufficient for obtaining consent and instead require highly specific notices for all generative AI model training, including details on: (1) the function of a generative AI model, (2) the type of personal data used to train the generative AI model, (3) how personal data will be used, and (4) how individuals can decline or withdraw consent. This sort of highly specific notice whenever personal data is used in connection with generative AI model training may inundate individuals with notices and contribute to notification fatigue. This ultimately reduces the usefulness of disclosures. Instead, notification requirements should be proportionate to the likely risk to individuals.

Recommendation: BSA recommends that:

- The Guidelines should provide additional guidance and examples of circumstances in which consent is not required under the PDPA, including the application of deemed consent and the business improvement exception for generative AI training.
- The Guidelines should adopt a risk-based approach that recognises the diversity of generative AI use cases and avoids imposing uniform consent and notification expectations across all forms of generative AI development and training.

Roles and Responsibilities in the AI Supply Chain

BSA welcomes the PDPC's recognition that generative AI systems involve multiple stakeholders with distinct roles and responsibilities. We particularly appreciate the effort to distinguish between model providers, system providers, and system deployers. These actors each play an important role in the AI value chain and have access to different information, capabilities, and risk mitigation measures. Clarifying their respective responsibilities helps ensure that obligations are allocated to the entities best positioned to fulfil them.

[3] *See, e.g.*, BSA TechPost, A Legitimate Interest in AI Training (Dec. 13, 2024), *available at* https://techpost.bsa.org/2024/12/13/a-legitimate-interest-in-ai-training.

We also welcome the recognition that these stakeholder categories may overlap with existing roles under the PDPA, namely organisations and data intermediaries. In practice, an entity's role will vary depending on the context of processing and the services it provides. For example, a company may act as a model provider in one context, a system provider in another, and a system deployer for its own internal use cases. Similarly, a stakeholder's status as an organisation or data intermediary may depend on the specific processing activities being undertaken. The Guidelines appropriately recognise these distinctions and provide a useful foundation for analysing responsibilities across the generative AI lifecycle.

While the consent and notification requirements are rightfully placed on organisations, as defined in the PDPA, the level of detail and information (i.e., the types of personal data, how they will be used to train and/or fine-tune AI models or systems, and the function(s) of these models or systems) that the organisations are required to provide to end-users under the AI-specific notification may not be practicable in certain contexts.

We encourage the PDPC to recognise that information-sharing obligations to downstream stakeholders can be met through standardised transparency mechanisms, such as published AI and privacy datasheets and contractual instruments including data processing addenda rather than requiring bespoke disclosures for each downstream relationship. Contractual flow-down to sub-processors, including restrictions on generative AI subcontractors using customer data for their own training or retaining session data post-delivery, provides an effective and auditable supply chain accountability mechanism. The Guidelines should treat these mechanisms as sufficient and avoid implying that individualised, per-relationship disclosures are required.

As the AI ecosystem continues to evolve, the Guidelines may benefit from further clarification regarding the role of system providers, which often integrate foundation models into downstream products and services. In practice, a system provider may adapt or integrate models created by other organisations for specific use cases while implementing additional safeguards, controls, and functionality. We encourage PDPC to keep these roles in mind as you consider responsibilities across the AI value chain.

Recommendation: BSA supports the model provider, system provider, and system deployer framework and encourages the PDPC to recognise that standardised transparency mechanisms are sufficient to satisfy information-sharing obligations to downstream stakeholders across the generative AI supply chain.

Access and Correction Requests

BSA appreciates the PDPC's recognition that there are present-day challenges associated with facilitating access and correction requests in the generative AI context. We agree that the scale of training datasets, the architecture of modern AI systems, and current technical limitations may make it difficult or impossible to identify, access, correct, or remove specific personal data used in model development.

In particular, the Guidelines should recognise that some organisations intentionally do not retain training data after model development or training has been completed. Such practices may be adopted as a privacy-enhancing measure to reduce the amount of personal data retained by an

organisation. In these circumstances, organisations may not be able to identify, retrieve, correct, or remove specific personal data from historical training datasets or from a trained model because the data may no longer be retained. The Guidelines should also clarify that organisations are not expected to undertake actions that are technically infeasible or impossible as a result of privacy-protective data management practices.

This applies when personal data is pseudonymised before training or when training relies on de-identified or synthetic data, making it technically impossible to map an access or correction request to a specific training record. The Guidelines should confirm that access and correction obligations do not apply where personal data has been de-identified, pseudonymised, or deleted before or during training.

We also encourage the PDPC to exercise caution in describing specific technical measures as best practices. For example, paragraph 10.4(a) refers to maintaining data provenance records to document the lineage of training data. While such measures may be appropriate in some contexts, their feasibility and utility may vary considerably depending on the nature, scale, and architecture of the AI system. More broadly, the Guidelines should recognise that technical approaches for identifying, tracing, modifying, or suppressing information within AI systems continue to evolve and may not be feasible, effective, or appropriate in all circumstances.

More generally, organisations should be encouraged to adopt reasonable and proportionate measures to support compliance with access and correction obligations, taking into account the capabilities and limitations of the underlying technology. The Guidelines should avoid creating an expectation that organisations can routinely identify and remove specific personal data from trained models.

Recommendation: BSA recommends that the PDPC clarify that compliance with access and correction obligations should be assessed in light of technical feasibility and the specific architecture of the generative AI system. The Guidelines should avoid creating an expectation that organisations adopt any specific technical approach to identify, trace, modify, remove, or suppress information within AI systems, particularly where such approaches may not be feasible, effective, or appropriate across different technologies and use cases.

Conclusion

BSA appreciates the opportunity to provide comments on the proposed Advisory Guidelines on the Use of Personal Data in Generative AI. We hope that our comments will assist the PDPC in developing practical guidance for organisations using personal data in the context of generative AI. We look forward to continuing our engagement with the PDPC on these issues and remain available as a resource as the Guidelines are finalised.

Yours sincerely,

Wong Wai San
Director, Policy – APAC