Submission by the Asia Tech Alliance (ATA) on the Singapore Advisory Guidelines on Use of Personal Data in Generative AI

Organisation: Asia Tech Alliance (ATA), www.asiatechalliance.com

Contact person: Lih Shiun Goh, Managing Director, secretariat@asiatechalliance.com

Executive Summary

The Asia Tech Alliance (ATA) welcomes the opportunity to provide input on the Proposed Advisory Guidelines on use of Personal Data in Generative AI. We commend the Commission for taking a pragmatic, principles-based approach that reflects a solid understanding of how generative AI systems operate in practice. In particular, we welcome the recognition that digital barriers do not automatically exclude personal data from being "publicly available"; the clear categorization of responsibilities across Model Providers, System Providers, and System Deployers; the acknowledgement that access and correction requests present genuine technical challenges in the AI context; and the forward-looking reference to agentic AI governance. The advisory, non-binding nature of the Guidelines also provides welcome flexibility as both the technology and industry practices continue to evolve rapidly. ATA supports the overall objectives of the Guidelines and welcomes the Commission's efforts to provide greater regulatory certainty for organisations developing and deploying Generative AI systems.

As Generative AI systems continue to develop rapidly, it is important that regulatory guidance remains sufficiently flexible to accommodate evolving technical architectures, operational practices, and international standards. In this regard, ATA believes that practical implementation guidance, developed through ongoing engagement between regulators, industry, academia, and civil society, can play an important role in supporting effective and proportionate compliance outcomes.

ATA's comments focus on several areas where additional clarification could further enhance the effectiveness of the Guidelines. These include the application of the Publicly Available Exception to large-scale model training activities, the scope of data handling measures for different categories of training data, and transitional considerations relating to AI-specific notification requirements. Our recommendations seek to support the Commission's policy objectives while ensuring that implementation remains operationally feasible across a diverse range of AI development and deployment contexts.

Summary of main comments

- **Calibrating the Publicly Available Exception for model training activities**
    - ATA recommends that the Commission clarify that the expectation to notify host organisations in paragraph 3.7 should apply only where a collecting organisation seeks to override an express, technically-enforced access restriction, rather than to routine, automated collection of publicly accessible web content.

- **Reconsidering and clarifying AI-Specific notification to individuals**
    - ATA recommends that instead of introducing an expectation to provide AI-Specific notifications to individuals, the Commission maintains the existing standard whereby disclosing the use of personal data for model development in privacy policies and terms of service is considered adequate notification.

- **Calibrating upstream data handling measures**
    - ATA recommends that accuracy verification at the point of collection is scoped to structured first-party data, and provides technical considerations to calibrate the feasibility of upstream data handling measures to comply with access and correction obligations.

Positive elements to acknowledge

- **Providing greater regulatory certainty for Generative AI**
    - ATA welcomes the Commission's efforts to provide practical guidance on how existing PDPA obligations apply in the Generative AI context. Greater clarity supports responsible innovation while helping organisations implement data protection safeguards with confidence.

- **Adopting a lifecycle-based approach to AI governance**
    - ATA appreciates the Guidelines' structured approach to addressing data protection considerations across the development, deployment, and post-deployment stages of the Generative AI lifecycle. This reflects the distinct risks and responsibilities that arise at different stages of AI systems.

- **Recognising the role of publicly available data**
    - ATA welcomes the Commission's recognition that publicly available data may continue to play an important role in Generative AI development and appreciates the guidance provided on assessing the applicability of the Publicly Available Exception in modern digital environments.

- **Promoting meaningful transparency and user awareness**
    - ATA supports the Commission's objective of ensuring that individuals are provided with meaningful information about how their personal data may be used in Generative AI development. The focus on transparency and informed decision-making can strengthen trust in AI-enabled products and services.

- **Encouraging practical approaches to access and correction requests**
    - ATA appreciates the Commission's recognition of the technical challenges associated with access and correction obligations in Generative AI systems, while encouraging organisations to adopt reasonable and evolving measures that support individuals' rights.

Key areas for improvement

- **Calibrating the Publicly Available Exception for model training activities (Question 1)**
    - ATA welcomes the Commission's recognition that information does not cease to be publicly available simply because access is subject to certain online restrictions, such as registration requirements or paywalls. We appreciate the guidance provided at paragraphs 3.4 to 3.9 on how organisations should assess whether certain types of data qualify for the Publicly Available exception in these circumstances.
    - ATA notes that the Guidelines introduce a best practice expectation for organisations to notify host organisations before collecting personal data that sits behind certain digital barriers, even where the collecting organisation has assessed that the data remains publicly available.
    - While ATA appreciates the objective of promoting transparency and cooperation between organisations, the development of models often requires the collection of publicly accessible information from a very large

number of online sources. In this context, a general expectation to notify individual host organisations for each source, including those operating basic registration requirements, would introduce significant operational friction with limited additional protection to individuals whose personal data is already publicly accessible.

- ATA also notes that paragraph 3.9 allows host organisations to contest whether their data can be characterised as publicly available. When there is a dispute, the host and collecting organisations should "discuss and agree". ATA is concerned that the notification mechanism at paragraph 3.7, combined with the dispute framework at paragraph 3.9, creates a de facto veto power for hosting organisations. Where a notified organisation can unilaterally "deny access" by asserting that data is not publicly available, with the only recourse being a referral to the Commission, this imposes substantial delay and uncertainty on AI developers. If the collecting organisation has already conducted its own assessment that the information is publicly available — and bears the legal risk of breaching the PDPA if that assessment is wrong — the additional purpose in seeking agreement from the notified organisation is unclear. The site owner may have non-privacy related commercial interests in denying access, which are irrelevant to whether the data is "publicly available" within the meaning of the PDPA.
- In practice, this will likely incentivise hosting organisations to post hoc contest the publicly available characterisation for commercial rather than privacy reasons, generating disputes that the Commission would need to adjudicate at scale. This outcome would be contrary to the objective of supporting responsible innovation for AI development.
- ATA recommends that the Commission clarify that the expectation to notify host organisations in paragraph 3.7 should be applied only where a collecting organisation seeks to override an express, technically-enforced access restriction (such as an authenticated API with rate limits enforced against bots), rather than to routine, automated collection of publicly accessible web content. Basic registration walls, CAPTCHAs, and robots.txt directives should not, in themselves, trigger a notification obligation under paragraph 3.7. Compliance or non-compliance with website terms of service or robots.txt protocols — while relevant to contractual or other legal obligations — are not

related to the data protection analysis of whether data is "publicly available" under the PDPA.

- ATA would also welcome guidance confirming that organisations acting in good faith reliance on the criteria set out in paragraphs 3.3 to 3.6 can rely on the publicly available exception, even where a host organisation subsequently disputes that characterisation. In particular, ATA recommends that the Commission clarify that where an organisation has conducted and documented a reasonable assessment under paragraphs 3.3–3.6 at the time of collection, such documentation will be recognised as a demonstration of due diligence and accountability in the event of a subsequent review.
- ATA notes that the Publicly Available Exception, as currently framed in Part 2 of the First Schedule to the PDPA, was not drafted with large-scale AI model training in mind. While advisory guidelines provide helpful interpretive guidance, they cannot provide the legal certainty that a statutory amendment would deliver. ATA encourages the Commission to recommend to the Ministry that the PDPA be amended to include an express AI development exception, clarifying that the collection, use, and disclosure of personal data that is publicly available on the internet is permitted for the purposes of developing, training, testing, deploying, and improving AI models, without the need for case-by-case digital barrier assessment. Such an amendment would position Singapore as a jurisdiction of choice for frontier AI development and align with the objectives of the National AI Strategy 2.0.

- **Reconsidering and clarifying AI-Specific notification to individuals (Question 2)**
  - ATA supports the Commission's objective at paragraphs 4.3 to 4.4 of ensuring that individuals are able to provide meaningful transparency about how their personal data may be used in connection with model development. The move toward AI-Specific notifications reflects an attempt toward enhanced transparency.
  - However, ATA wishes to highlight that the existing industry practice — a notice in the terms of service or privacy policy that user data may be used to improve AI models and services — is well-established and sufficient. Generative AI is now widely popular and there has been extensive public reporting raising awareness about the volume of data required to train AI

models, reducing the justification for onerous individualised notifications. Singapore would be imposing obligations more onerous than global standards, potentially placing organisations operating locally at a disadvantage without a commensurate uplift in individual protection.

- More fundamentally, the Guidelines at paragraph 4.4 frame AI-Specific notifications exclusively within a consent paradigm, without acknowledging that consent is not the only — and in many cases, not the most appropriate — legal basis for processing personal data for AI model development under the PDPA. Organisations may lawfully process user data for model training in reliance on other legal bases, including the legitimate interests exception, business improvement exception, or research exception.
- In addition, the example in Paragraph 4.6 suggests that users should be provided with updated privacy policies containing a hyperlink to decline consent specifically for AI training, while continuing to use the service. However, requiring organisations to continue providing the service while honouring a selective withdrawal of consent for one specific use (model training) creates significant technical complexity and may be practically impossible to implement where the trained model is integral to the service being provided.
- ATA recommends that the Commission maintains the existing standard whereby disclosure in privacy policies and terms of service is considered adequate notification.
- Should the Commission decide to keep an AI-specific notification requirement, ATA notes that the notification content recommended at paragraph 4.6(a), requiring organisations to disclose the "function(s) of the Generative AI Model", may not be feasible for foundation models, which are general-purpose by design and cannot enumerate specific downstream use cases at the point of training.
- ATA recommends that the Commission clarify that for general-purpose foundation models, a notification describing the broad categories of model capability (e.g., "language understanding and generation") is sufficient, without requiring exhaustive enumeration of all potential applications.

- **Facilitating a transition toward AI-Specific notification requirements (Question 2)**

- Currently, the Guidelines remain silent on how the AI-Specific notification requirement applies to models and datasets that were developed prior to the finalisation of these Guidelines, in reliance on general notifications that were consistent with then-accepted practice. Organisations developed and trained models in reliance on notification practices that were consistent with prevailing PDPA guidance at the time. Clarifying the prospective application of the AI-Specific Notification requirement would support regulatory certainty and facilitate orderly implementation without undermining the Commission's objective of enhancing transparency going forward.
- ATA recommends that the Commission include a provision confirming that the AI-Specific notification requirement will apply prospectively to new training and fine-tuning runs initiated after the final Guidelines take effect, and that it does not invalidate models developed in good faith under previously accepted consent frameworks. A transitional period, accompanied by clear interim compliance guidance, would further support industry in making the adjustments needed in an orderly and sustainable manner.
- ATA recommends that the Commission explicitly confirm that existing models trained under general notification frameworks (e.g., privacy policies disclosing use for “product improvement” or “development of new features”) remain lawfully trained and may continue to be deployed, updated, and served without requiring re-consent from users. Requiring retrospective re-notification and re-consent for billions of data points already incorporated into model weights would be technically infeasible, operationally disproportionate, and inconsistent with the PDPA's deemed consent provisions under Section 15.

- **Calibrating upstream data handling measures (Question 3)**
  - ATA supports the upstream data quality measures outlined at paragraph 10.4(a) as a meaningful contribution to supporting individuals’ access and correction rights in the Generative AI context.
  - However, the proposed best practice of verifying data accuracy at the point of collection carries different practical implications depending on the data source. For structured data collected directly from individuals through an organisation’s own products or services, such verification is feasible and appropriate. For large, unstructured datasets ingested under the Publicly

Available Exception – such as billions of data points from open web content gathered through automated processes – real-time accuracy verification of third-party public information would not be operationally workable at the scale required for pre-training.

- ATA recommends that the Commission clarify that the expectation to verify data accuracy at the point of collection applies to structured, first-party data collected directly from individuals, and does not extend to open web datasets ingested under the Publicly Available Exception. For such large-scale datasets, alternative quality measures — including outlier detection, de-duplication and statistical validation techniques (referenced elsewhere in the Guidelines) — represent the state of the art and are more proportionate than item-level accuracy verification. Compliance with the accuracy obligation should be assessed based on what is reasonable in the circumstances, having regard to the scale and nature of data collection.
- ATA appreciates the Commission's acknowledgment at paragraph 10.3 that significant technical challenges exist in facilitating access and correction requests in the Generative AI context, including the difficulty of identifying specific individuals' data within model embeddings. Fundamentally, trained model weights are mathematical abstractions (statistical parameters) that do not "store" personal data in any retrievable or queryable form. Once personal data has been processed through model training (gradient descent over billions of parameters), the resulting model is a new artifact, not a repository of the input data.
- ATA therefore recommends that the Commission provide guidance confirming that:
    - Access and correction obligations under Sections 21, 22, and 22A apply to identifiable personal data held in structured form (e.g., training datasets, RAG databases, prompt logs), but do not extend to trained model parameters;
    - The "disproportionate burden" exception under paragraph 1(j)(ii) of the Fifth Schedule should be presumptively applicable where an access or correction request relates to data that has been incorporated into model weights, given the current state of the art;
    - The reference to "machine unlearning" at paragraph 10.4(c), while forward-looking, should be framed as aspirational rather than as a

current compliance expectation, given that no validated, scalable machine unlearning technique exists today. We suggest replacing this reference with broader, technology-neutral language such as: "adopt appropriate and demonstrably effective technical measures as they mature," without naming a specific unproven technique. This preserves regulatory flexibility while avoiding implicit obligations around a technology that is not yet fit for purpose;

- Data provenance record-keeping obligations at paragraph 10.4(a)(iii) should be calibrated to what is technically feasible. For web-scale datasets comprising billions of documents, line-item provenance tracking is not operationally workable, and dataset-level documentation (describing sources, time periods, and filtering criteria) should be recognised as sufficient; and
- The expectation at Paragraph 10.4(b) that organisations remove personal data from training datasets before undertaking future training runs presupposes the ability to reliably identify and isolate specific individuals' data within datasets of billions of data points. We recommend that this expectation be qualified to reflect what is reasonable in the circumstances, consistent with the proportionality principle acknowledged at Paragraph 10.

- **Information sharing by Model and System Providers (Question 4)**
  - On the question of what information Model Providers and System Providers should share with downstream stakeholders, ATA supports the principle of transparency. However, we note that the appropriate level of disclosure will vary depending on the nature of the model or system, the sensitivity of the data being processed, and the deployment context. We recommend that the Guidelines maintain their current principles-based approach, allowing providers to determine the appropriate level and format of disclosure rather than imposing a prescriptive list of required information. Commercial confidentiality and proprietary information should also be protected.

- **Dedicated guidance for agentic AI (Question 5)**
  - ATA notes that paragraph 9.5 addresses agentic AI in a single paragraph with a cross-reference to IMDA's Model AI Governance Framework for Agentic AI.

Given the distinct and rapidly evolving data protection challenges posed by agentic systems — including multi-step autonomous data access, tool use, and action-taking — ATA recommends that the Commission develop dedicated, principles-based guidance on agentic AI data protection rather than appending it to guidelines primarily designed for conventional GenAI inference patterns.
    - ATA supports the Commission engaging with industry through future consultations as the technology and deployment patterns mature, before introducing prescriptive agent-specific requirements.

- **Clarifying the Retention Limitation obligation for AI training data**
    - ATA notes that the Guidelines at paragraph 7.2 apply the Retention Limitation Obligation to training data held by Model Providers. While ATA supports responsible data governance, we note that in the AI development context, training data is retained not merely to "serve the original purpose of collection" but to support essential ongoing activities including model reproducibility, safety evaluation, red-teaming, bias auditing, regulatory compliance, and iterative model improvement.
    - ATA recommends that the Commission clarify that retention of training datasets for these purposes constitutes a continuing legitimate business purpose under Section 25, and that periodic review (rather than deletion) is the appropriate standard.

- **Clarifying the Purpose Limitation obligation for general-purpose models**
    - ATA notes that paragraph 9.2 applies the Purpose Limitation Obligation to System Deployers, requiring them to specify the intended purpose of processing. While purpose specification is appropriate for narrow, application-specific deployments, it is inherently difficult to reconcile with the nature of general-purpose foundation models whose value lies precisely in their versatility across tasks.
    - ATA recommends that the Commission provide guidance acknowledging that for general-purpose models, purpose limitation should be assessed at the output/application layer (i.e., by System Deployers in their specific use cases) rather than at the model development or model provision layer.

- **Cross-border data flows and extra-territorial application**
    - ATA notes that the Guidelines do not address how obligations apply to training data sourced from outside Singapore, or to models trained by overseas organisations using data that may incidentally include personal data of Singapore residents. Given the inherently global nature of AI model training, ATA recommends that the Commission provide guidance on extra-territorial scope, mutual recognition frameworks, and how Singapore-based deployers should assess compliance for models trained under other jurisdictions' legal frameworks.

- **Preferring output-focused over input-focused regulation**
    - ATA encourages the Commission to consider, in future iterations of its guidance, a shift toward output-focused regulatory frameworks for Generative AI. Rather than seeking to regulate what data enters a model (inputs), which raises considerable scalability and feasibility challenges with no clear commensurate privacy benefits, the Commission could focus regulatory attention on what a model produces (outputs) — including whether outputs disclose personal data, generate harmful content, or facilitate unlawful activity. This approach is both more proportionate and more effective at addressing actual privacy harms, while preserving the conditions necessary for AI innovation.

Conclusion

ATA welcomes the Commission's efforts to provide timely and practical guidance on how the PDPA applies to Generative AI systems. The principles-based approach in the guidelines reflects a sound understanding of the operational realities of AI development and deployment, and provides a constructive foundation for ongoing regulatory development.

ATA hopes that the comments will assist the Commission in refining the Guidelines to ensure that they remain operationally feasible, technologically proportionate, and aligned with international developments. In particular, ATA encourages the Commission to consider further clarification on the scope of the Publicly Available Exception, the application of AI-specific notification requirements, and the treatment of trained model

parameters under access and correction obligations — areas where additional guidance could meaningfully support responsible and compliant AI development.

ATA looks forward to continued engagement with the Commission and remains committed to working collaboratively with regulators, policymakers, and other stakeholders to support the development of a trusted and innovative AI ecosystem in Singapore.

About the Asia Tech Alliance

The Asia Tech Alliance (ATA) is a regional industry association formed to amplify Asia-Pacific's voice in the global digital economy and champion technology that reflects the region's values, innovation, and potential. Founded by 12 leading global technology companies, the ATA serves as a trusted platform for governments, industry, and communities to co-create human-centric, innovative solutions to the region's most pressing opportunities and challenges. Through policy engagement, collaborative initiatives, and strategic partnerships, the Alliance is committed to shaping an inclusive, agile, and forward-looking digital future for Asia-Pacific.