



PERSONAL DATA
PROTECTION COMMISSION
S I N G A P O R E

PROPOSED ADVISORY GUIDELINES ON USE OF PERSONAL DATA IN GENERATIVE AI

Issued 2 June 2026

Supported by:



TABLE OF CONTENTS

| | |
|--|-----------|
| PART I: INTRODUCTION AND SCOPE | 3 |
| 1 Introduction | 3 |
| 2 Scope of the Advisory Guidelines | 4 |
| PART II: DEVELOPMENT – COLLECTING AND USING PERSONAL DATA TO DEVELOP GENERATIVE AI MODELS | 4 |
| 3 Publicly Available Exception | 4 |
| 4 Consent and Notification Obligations | 7 |
| 5 Additional Considerations | 9 |
| PART III: DEPLOYMENT - DATA PROTECTION RESPONSIBILITIES OF GENERATIVE AI STAKEHOLDERS | 9 |
| 6 Generative AI Stakeholders | 9 |
| 7 Model Providers | 10 |
| 8 System Providers | 11 |
| 9 System Deployers | 12 |
| PART IV: POST DEPLOYMENT - ADDRESSING INDIVIDUAL REQUESTS ABOUT PERSONAL DATA | 14 |
| 10 Access and Correction Obligations | 14 |
| PART V: PUBLIC CONSULTATION QUESTIONS | 16 |

PART I: INTRODUCTION AND SCOPE

1 Introduction

- 1.1 The Personal Data Protection Act 2012 (“**PDPA**”) governs the collection, use and disclosure of personal data by organisations in a manner that jointly recognises the right of individuals to protect their personal data and the need for organisations to leverage personal data for reasonable purposes.
- 1.2 This Part clarifies how the PDPA applies to Artificial Intelligence (“**AI**”) contexts in line with the Act’s objectives. The Personal Data Protection Commission’s (“**Commission**”) Advisory Guidelines on use of Personal Data in AI Recommendation and Decision Systems offers foundational guidance on (i) when organisations can rely on exceptions to consent under the PDPA (e.g. for business improvement or research purposes) for AI system development, testing and monitoring; (ii) how organisations can be transparent about whether and how their AI systems use personal data; and (iii) how third-party service providers may support organisations in implementing AI systems.
- 1.3 “**Generative AI Models**” are models, including those trained on a large amount of data using self-supervision at scale, that display significant generality and can competently perform a wide range of distinct tasks regardless of the way the models are placed on the market and that can be integrated into a variety of systems or applications. “**Generative AI Systems**” are AI systems or applications based on Generative AI Models, both for direct use as well as for integration in other AI systems.
- 1.4 These Advisory Guidelines (“**Guidelines**”) clarify how personal data can be used in Generative AI Models and/or Systems. The Guidelines also build on and should be read in conjunction with the Commission’s Advisory Guidelines on use of Personal Data in AI Recommendation and Decision Systems, Advisory Guidelines on Key Concepts in the PDPA, Privacy Enhancing Technologies Adoption Guide as well as the Guide to Basic Anonymisation.
- 1.5 The Guidelines are advisory in nature, are not legally binding on the Commission or on any other party, and do not constitute legal advice. They neither modify nor supplement in any way the legal effect and interpretation of any laws cited, including, but not limited to, the PDPA and any subsidiary legislation issued under the PDPA. The provisions of the PDPA and any subsidiary legislation will prevail over these Guidelines in the event of any inconsistency. These Guidelines should not be construed to limit or restrict the Commission’s administration and enforcement of the PDPA.

2 Scope of the Advisory Guidelines

- 2.1 The Guidelines address some of the key data protection issues relating to Generative AI today. These include the (i) collection and use of personal data to develop Generative AI Models; (ii) allocation of data protection responsibilities across the Generative AI lifecycle; and (iii) handling of individuals' requests concerning the processing of their personal data for Generative AI.
- 2.2 The Guidelines are organised according to the typical stages of the Generative AI lifecycle as follows:

| Section | Stage of Generative AI Lifecycle | Topics |
|----------|---|--|
| Part II | Development: Collecting and using personal data to develop Generative AI Models | <ul style="list-style-type: none"> Publicly Available Exception Notification and Consent Obligations Protection Obligation |
| Part III | Deployment: Processing personal data in deployed Generative AI Models and/or Systems | <ul style="list-style-type: none"> Retention Limitation Obligation Protection Obligation Purpose Limitation Obligation Accountability Obligation |
| Part IV | Post-Deployment: Addressing individuals' requests about personal data | <ul style="list-style-type: none"> Access and Correction Obligations |

PART II: DEVELOPMENT – COLLECTING AND USING PERSONAL DATA TO DEVELOP GENERATIVE AI MODELS

3 Publicly Available Exception

- 3.1 A large amount of data is needed to develop Generative AI Models, including during the pre-training and fine-tuning stages. Web-scraping, the automated process of extracting large amounts of online data, is a common means of compiling datasets for such purposes. Where web-scraped datasets include personal data, organisations may consider relying on the “**Publicly Available Exception**” in lieu of seeking consent.
- 3.2 This refers to Part 2 of the First Schedule to the PDPA, which enables organisations to collect, use or disclose, without consent, personal data about an individual that is publicly available. Publicly available personal data means personal data that is generally available to the public (i.e. can be obtained or accessed with few or no restrictions), including personal data which can be observed by reasonably expected

means at a location or an event at which the individual appears and that is open to the public.¹

3.3 Where personal data forms part of online data that is publicly accessible without any restrictions, organisations may rely on the Publicly Available Exception to collect the data to develop a Generative AI Model. This remains subject to a reasonable person considering the use of personal data to develop the Generative AI Model appropriate in the circumstances.²

3.4 However, where personal data forms part of online data that is placed behind a **digital barrier**, additional considerations will apply as follows, which organisations should consider before relying on the Publicly Available Exception. The Commission considers a digital barrier to be any technical and/or financial measure that meaningfully restricts data access in whole or in part. Examples of digital barriers include:

- a) Paywalls (e.g. hard, metered, dynamic) or subscriptions;
- b) Registration requirements (e.g. sign-up processes requiring information like the user's name, email address and contact details);
- c) Authentication mechanisms (e.g. passwords, Application Programming Interface ("API") keys, one-time codes); and
- d) Tools, systems or configurations that detect and prevent automated programs from accessing website data (e.g. AI bot blockers, Completely Automated Public Turing tests to tell Computers and Humans Apart).

3.5 While the existence of a digital barrier *per se* does not prevent personal data from being publicly available,³ much depends on the circumstances and facts of each case. In the Generative AI context, relevant factors to consider include:

- a) The purpose of the digital barrier (e.g. to enable data monetisation);
- b) The effect of the digital barrier (e.g. whether the online data remains accessible to the public at large or only a specific group of persons);
- c) The steps needed (e.g. number, complexity) to access the personal data; and

¹ See S. 2 of the PDPA and paragraphs 12.84 to 12.94 of the Advisory Guidelines on Key Concepts in the PDPA.

² See S. 18 of the PDPA and paragraph 13.4 of the Advisory Guidelines on Key Concepts in the PDPA.

³ See paragraphs 12.85 to 12.86 of the Advisory Guidelines on Key Concepts in the PDPA.

- d) Whether the personal data can be accessed without any restrictions from other online sources.
- 3.6 The Commission considers the following to be examples of publicly available data, even though such data sits behind digital barriers:
- a) Registers administered by public agencies intended to be generally available. This includes registers where the data is provided only after the payment of a fee as it is still obtainable by any member of the public;
 - b) News or media websites that allow access to a limited number of articles before requiring payment or a subscription as these paywalls function as monetisation mechanisms rather than substantive barriers to access; and
 - c) Large online forums that are open to anyone to join but require users to provide registration details as a condition of access and participation. Such requirements are not so complex or burdensome to suggest the data is not generally available.
- 3.7 Where an organisation intends to scrape personal data behind digital barrier(s) put up by another organisation (having assessed that such data is publicly available), the Commission considers it best practice for the collecting organisation to notify the other organisation of its intention to do so. This affords the notified organisation the opportunity to separately consider whether disclosure would fall within circumstances or be for purposes where consent is not required under the PDPA (e.g. the data is publicly available). Such best practices also ensure that when an organisation collects personal data about an individual from another organisation without the individual's consent, the organisation provides sufficient information regarding the purpose of collection.⁴
- 3.8 Ensuring that both collecting and notified organisations consider personal data behind digital barrier(s) to be publicly available is especially important in the development of Generative AI Models because of the technical difficulties with removing or correcting data post-training and the ability to deploy Generative AI Models across a wide range of use cases.
- 3.9 Where the notified organisation considers that the personal data in its possession or control is publicly available, it should allow the collecting organisation to access and collect the data. However, where the notified organisation considers that the data is not publicly available, it should inform the collecting organisation and deny access. The notified organisation may also consider enhancing the barriers implemented to

⁴ See also paragraph 12.37 of the Advisory Guidelines on Key Concepts in the PDPA.

make explicit that the data is not publicly available. In such cases where the collecting organisation disputes the notified organisation's assessment and the parties cannot agree after discussion, it may refer the matter to the Commission.

- 3.10 For the avoidance of doubt, the applicability of the Publicly Available Exception and compliance with the PDPA are distinct from organisations' contractual obligation(s) to respect online terms of use or service or licensing agreements, and requirements under general law, including criminal law.

4 Consent and Notification Obligations

- 4.1 Another key source of data used to develop Generative AI Models is personal data provided by an individual to an organisation, or created in the course of or as a result of the individual's use of the organisation's products or services ("**User Data**").

- 4.2 Unless deemed consent or relevant exceptions to consent apply⁵, consent is required for the use of User Data for the development of Generative AI Models ("**Consent Obligation**").⁶ This requirement is complemented by the **Notification Obligation**, which requires that individuals be notified of the purpose of the intended use of their personal data when their consent is sought for such use.⁷ Among other things, Section 20(1) of the PDPA requires an organisation to inform the individual of:

- a) The purposes for the collection, use and disclosure of their personal data, on or before collecting the personal data; and
- b) Any other purpose of the use or disclosure of the personal data of which the individual has not been informed under paragraph (a), before use or disclosure of personal data for that purpose.

- 4.3 There are two possible approaches by which organisations currently notify and seek consent for the use of User Data for the development of Generative AI Models. The first is to provide individuals with a general statement on the purpose of processing. These statements may cite the use of personal data for "new product development" without specifying AI or Generative AI Model development. These are referred to as "**General Notifications**". The second method is to provide an explicit statement that the purpose of use includes AI and/or Generative AI Model development ("**AI-**

⁵ For example, the business improvement and research exceptions. See paragraphs 4.1 to 6.4 of the Advisory Guidelines on the Use of Personal Data in AI Recommendation and Decision Systems.

⁶ See S. 13 of the PDPA.

⁷ See S. 14(1) read with S. 20(1) of the PDPA. See also paragraphs 9.2 to 9.3 of the Advisory Guidelines on the Use of Personal Data in AI Recommendation and Decision Systems.

Specific Notifications”). AI-Specific Notifications may also be accompanied by an explicit mechanism that allows individuals to decline or withdraw their consent.

- 4.4 The Commission is of the view that General Notifications are an insufficient means of obtaining consent to use User Data for the development of Generative AI, specifically, for large-scale AI model training and/or fine-tuning. Organisations must provide AI-Specific Notifications for this purpose. The *raison d’être* for the **Consent** and **Notification Obligations** is to enable individuals to provide *meaningful consent*. While notifications need not be overly technical or detailed, individuals must be able to understand the type(s) of personal data and how it will be used to train and/or fine-tune AI models, as well as the function(s) of these models.
- 4.5 This is especially important in the Generative AI context as Generative AI Models are deployed across a wide range of use cases. Individuals may rightfully have concerns over their (i) personal data of a more sensitive nature being exposed, reconstructed and/or disclosed to third parties from Generative AI Models; or (ii) personal data being used for purposes they never anticipated (e.g. developing a model to conduct financial profiling).
- 4.6 Organisations are encouraged to provide information on the following, to the extent practicable, in crafting their AI-Specific Notifications⁸:
- a) The function(s) of the Generative AI Model that require the use of personal data (e.g. text-to-speech function that converts written text into spoken audio);
 - b) A clear description of the type(s) of personal data that will be used to develop the Generative AI Model (e.g. voice data);
 - c) How personal data will be used to train and/or fine-tune the Generative AI Model (e.g. voice data will be used to train the model to recognise speech patterns); and
 - d) How individuals can decline or withdraw consent to the use of personal data for AI training (e.g. by providing step-by-step instructions or an easily accessible opt-out mechanism)

Organisations are also encouraged to make their notifications clearly visible (e.g. via an in-product-pop-up or dedicated web page).

⁸ This incorporates the existing guidance at paragraphs 9.5 to 9.8 of the Advisory Guidelines on use of Personal Data in AI Recommendation and Decision Systems.

Example of appropriate notification to obtain consent for Generative AI Model training:

A social media platform provider wants to notify individuals that it will be using User Data to develop an AI model that generates text and images, and which will be integrated into its platform. The provider:

- Updates its privacy policy to include a statement that reads “When you interact with our AI-enabled features, the text, images and audio that you submit may be used for product improvement purposes, including the training and running of our AI models. This will enable our AI-enabled features to generate more realistic images and speech patterns.”;
- Produces a privacy web page to supplement the updated privacy policy;
- Provides an in-platform privacy update that users can read and acknowledge; and
- Provides an email update to platform users on its updated privacy policy which also includes a hyperlink that allows users to decline consent.

In this case, the social media platform provider is considered to have provided appropriate AI-Specific Notification.

5 Additional Considerations

- 5.1 Notwithstanding the above, organisations are encouraged to practice data minimisation when developing Generative AI Models to minimise unnecessary risks. Where personal data is necessary for Generative AI development, organisations are reminded to implement appropriate technical, process and legal controls for data protection. Organisations may wish to refer to the Commission’s Advisory Guidelines on use of Personal Data in AI Recommendation and Decision Systems for further guidance.⁹

PART III: DEPLOYMENT - DATA PROTECTION RESPONSIBILITIES OF GENERATIVE AI STAKEHOLDERS

6 Generative AI Stakeholders

- 6.1 Generative AI involves multiple layers in the technology stack, which may result in personal data being handled by various stakeholders. Generative AI Systems also have diverse deployment architectures that use and/or generate new categories of data sources. Considering these complexities, this Part clarifies the key roles and

⁹ See paragraphs 7.2 to 7.12 of the Advisory Guidelines on use of Personal Data in AI Recommendation and Decision Systems.

responsibilities of Generative AI stakeholders in safeguarding personal data. Besides the obligations highlighted in this Part, stakeholders are reminded to take necessary steps to ensure compliance with their other obligations under the PDPA.

6.2 The Commission identifies relevant stakeholders to include:

- a) “**Model Providers**” who develop and/or make available Generative AI Models for distribution and use¹⁰;
- b) “**System Providers**” who develop and/or make available Generative AI Systems for distribution and use; and
- c) “**System Deployers**” who use and/or enable the use of Generative AI Systems under their authority.

7 Model Providers

Model Providers as Organisations

- 7.1 Where Model Providers process personal data to develop and deploy Generative AI Models, they are considered organisations under the Act and must comply with all obligations under the PDPA. This extends to situations where Model Providers collect and use personal data from downstream systems (e.g. end-user prompts, inputs) for model development.
- 7.2 As organisations, Model Providers are reminded to pay particular attention to their “**Retention Limitation Obligation**”. Section 25 of the PDPA requires an organisation to cease to retain its documents containing personal data, or remove the means by which personal data can be associated with particular individuals when the purpose for which the data was collected is no longer being served, and retention is no longer necessary for legal or business purposes.
- 7.3 The Commission accepts that it may be necessary for Model Providers to preserve their training data to develop or enhance future models. Where such data includes identifiable personal data, it is good practice for Model Providers develop and make available a data retention policy that includes the rationale for retaining data for a longer period of time. Model Providers should also regularly review the personal data in their possession to determine if such data is still needed.¹¹ This guidance similarly applies to System Providers and Deployers who retain personal data to develop and deploy Generative AI Systems.

¹⁰ This would also include platforms that host multiple models and service providers.

¹¹ See paragraphs 18.5 to 18.8 of the Advisory Guidelines on Key Concepts in the PDPA.

Model Providers as Data Intermediaries

- 7.4 Model Providers may process personal data on behalf of downstream users, as part of service(s) provided to users by System Deployers. For example, Model Providers may run inference¹² to deliver real-time outputs to user inputs or host data on their infrastructure. In this context, they are considered data intermediaries under the Act and must comply with applicable PDPA provisions. This includes protecting personal data in their possession or under their control by making reasonable security arrangements (“**Protection Obligation**”).¹³
- 7.5 In demonstrating compliance with their Protection Obligation, it is good practice for Model Providers to document and make available the measures they have taken to safeguard personal data from downstream sources. For example, a Model Provider may document their data access controls as well as data residency and retention policies. This information will also support System Providers and Deployers in meeting their PDPA obligations by helping them to determine (i) the adequacy of their system-level security arrangements; (ii) whether there has been unauthorised access and modification; and (iii) whether such access and modification is a notifiable data breach.

Example: Model Provider X licenses its Generative AI Model to enterprises for automating customer support responses. Model Provider X assesses it may have to process customer messages containing personal data (e.g. names, contact details and account information) during the inference process. Model Provider X adopts the following measures:

- Implements role-based access controls to limit who can view and process customer data during inference;
- Publishes a data storage info sheet which includes information on where customer data is stored and processed; and
- Adopts a data retention policy which specifies when inference data will be deleted.

8 System Providers

- 8.1 This section is relevant for System Providers who are engaged to develop bespoke, customisable Generative AI Systems or who develop and retail commercial off-the-shelf systems. It is not relevant for entities that develop and deploy Generative AI Systems in-house. Where System Providers process personal data as part of their own datasets to develop systems, they are organisations. Where they process data

¹² Inference is the process where a trained Generative AI Model generates new outputs by reasoning and making predictions on new data.

¹³ See S. 24 of the PDPA. Data intermediaries are also subject to the Retention Limitation Obligation.

on behalf of downstream deployers (e.g. to customise systems for specific use cases, in delivering their Software as a Service (“SaaS”)) they may also take on the role of data intermediaries.

System Providers as Data Intermediaries

- 8.2 As Generative AI Systems diversify, this introduces more security and data protection risks for System Providers (e.g. prompt injection attacks, which can be exacerbated by unclear boundaries between control and user data planes). To comply with their Protection Obligation, System Providers are expected to periodically review the need for additional security arrangements as they develop and make available new types of systems.
- 8.3 Like Model Providers, it is good practice for System Providers to share information on the system-level safeguards they have implemented with downstream deployers, to facilitate the protection of personal data that will be processed by their systems. Examples of relevant information include:
- a) Data security and protection measures around the development environment (e.g. access controls, input or output filters, privacy enhancing technologies¹⁴); and
 - b) Testing and performance metrics (e.g. likelihood of data leakage).

Example: System Provider Y builds a Generative AI assisted legal summarisation API that law firms can integrate into internal workflows to summarise case documents. To help its clients understand how the API protects against data leakage and what risks remain, System Provider Y documents the technical safeguards it has implemented such as input redaction filters and encryption in transit and at rest for data exchanged with the API. It also provides information on its internal testing procedures and metrics related to data leakage including the rate at which personal identifiers are detected and redacted before summarisation.

9 System Deployers

System Deployers as Organisations

- 9.1 System Deployers may develop and deploy Generative AI Systems in-house, in which case they take on the roles of Model and/or System Providers, or procure and use systems as a service (e.g. through SaaS platforms, API-based offerings). Regardless, System Deployers bear primary responsibility for ensuring that the Generative AI

¹⁴ See the Privacy Enhancing Technologies Adoption Guide and Guide to Basic Anonymisation.

Systems they have chosen to use can meet their obligations under the PDPA. When procuring systems, they must therefore ensure that they have sufficient information on upstream safeguards to conduct a holistic assessment.

- 9.2 Another responsibility is ensuring that the personal data processed by their chosen systems is for relevant purposes. Section 18 of the PDPA limits the collection, use or disclosure of personal data about an individual only for purposes and to the extent that a reasonable person would consider appropriate in the circumstances (“**Purpose Limitation Obligation**”). While Generative AI Systems can perform a variety of tasks, System Deployers should be disciplined about specifying the intended purpose of processing and amount of personal data required for this. They are also reminded that personal data should not be processed for illegal and/or harmful purposes.¹⁵

Example: Organisation A wants to deploy a document triaging system in its medical records department. Pre-deployment, it clearly defines the system’s purpose as being to support the administrative prioritisation of medical documents and identifies essential personal data required. Organisation A also configures access controls so that the system cannot query or retrieve data outside its purpose (e.g. it has no access to other databases such as billing systems or patient contact lists).

- 9.3 Pursuant to the Protection Obligation, System Deployers must also safeguard personal data in their possession or under their control. This includes new categories of data sources that are collected through their systems (e.g. end-user prompts, inputs and generated outputs, agent or tool activity data, internal enterprise data). System Deployers must track and designate responsibilities over new data sources and implement corresponding safeguards. This includes educating their end users, whether internal or external, on the specific types of data that should be input into their systems (e.g. only pre-defined categories of personal data) and how data will be processed following collection.
- 9.4 System Deployers are further encouraged to develop clear written policies and document processes in relation to the safeguards undertaken. Pre-emptively making such policies available (e.g. on their website) will also demonstrate accountability in compliance with the PDPA.¹⁶

Example: Organisation B deploys a human resources (“HR”) system to manage its employee engagement, training and performance reviews. In addition, employees can query how

¹⁵ See paragraph 13.4 of the Advisory Guidelines on Key Concepts in the PDPA.

¹⁶ See S. 12(d) of the PDPA and paragraphs 10.2 to 10.4 of the Advisory Guidelines on use of Personal Data in AI Recommendation and Decision Systems.

organisational HR policies apply to them using an in-system chatbot. The system is overseen by Organisation B's HR unit. Organisation B adopts the following measures to safeguard personal data processed by the system:

- Designates its HR unit as responsible for safeguarding the prompts and outputs generated through the system's chatbot;
- Develops guidelines on the types of personal data that can be included in prompts;
- Implements measures to scan prompts for common personal identifiers and limit access to prompt logs to authorised maintenance personnel; and
- Documents these arrangements in internal AI governance guidelines, which are also disseminated to all employees.

9.5 Lastly, as AI risks continue to evolve, System Deployers should regularly review the sufficiency of their safeguards. This is especially relevant where their Generative AI Systems have agentic functionalities. While the definition of AI agents remains unsettled, common features include independent planning and action taking across multiple steps to achieve user-defined objectives. As these enhanced capabilities can exacerbate data protection risks, System Deployers should carefully consider the privacy-utility trade-offs when scoping their agentic use cases. For further guidance on managing agentic risks, organisations may wish to refer to IMDA's Model AI Governance Framework for Agentic AI.

Example: Organisation C deploys an agentic AI assistant to enhance productivity across its sales team. The assistant helps employees streamline customer engagement by providing personalised sales recommendations, automating follow-up tasks and generating insights from customer data. Organisation C assesses that the AI assistant will need more personal data to interpret inputs and complete tasks effectively, which may result in unauthorised data access, use, disclosure etc. It decides to enhance existing its protection safeguards by:

- Implementing additional role-based data controls to restrict the assistant's file and network access;
- Implementing data classification systems that tag personal data of a more sensitive nature and requiring stricter handling measures for such data; and
- Introducing escalating protocols to humans for complex or high-risk tasks.

PART IV: POST DEPLOYMENT - ADDRESSING INDIVIDUAL REQUESTS ABOUT PERSONAL DATA

10 Access and Correction Obligations

- 10.1 Sections 21, 22 and 22A of the PDPA set out the rights of individuals to request for access to and correction of their personal data in the possession or under the control of organisations, and the corresponding obligations of organisations to respond to such requests (“**Access and Correction Obligations**”).
- 10.2 In the Generative AI context, the Access and Correction Obligations apply to personal data in the form(s) in which it has been collected, used and disclosed for model and system development. Organisations must accede to individual requests unless an exception under the PDPA applies.¹⁷ For example, organisations may choose not to (i) provide access to personal data where the burden or expense of doing so would be **unreasonable** to the organisation or disproportionate to individual interests; or (ii) correct personal data where the organisation is satisfied on **reasonable grounds** that the correction should not be made.¹⁸
- 10.3 The Commission recognises that there are present-day challenges to facilitating Generative AI-related access and correction requests. These could be due to (i) the massive amounts of data used to develop Generative AI Models and/or Systems, which can make it difficult to identify, verify and correct data of specific individuals; (ii) the nature of Generative AI Models and/or Systems (e.g. training data is not stored in a traditional repository but as embeddings, User Data is temporarily held in context windows); and (iii) other technical limitations (e.g. difficulty in removing specific information from models).
- 10.4 Notwithstanding the above, organisations are expected to adopt the following best practices to support compliance with their Access and Correction Obligations where reasonable:
- a) Adopt upstream data handling measures such as (i) verifying data accuracy at the point of collection; (ii) implementing data cleaning techniques like de-duplication and outlier detection; and (iii) maintaining data provenance records to document the lineage of training data;
 - b) Review access and correction requests on a case-by-case basis and accede where reasonable (e.g. where the request refers to personal data stored in a Retrieval-Augmented Generation database). Organisations are also encouraged to ensure that personal data, including inaccurate data, is removed from training datasets before they undertake future AI training runs; and
 - c) Track the maturity of and adopt appropriate technical measures, (e.g. machine unlearning) to remove inaccurate personal data from models and systems.

¹⁷ See S. 21(2), S. 22(2) and the Fifth and Sixth Schedules to the PDPA.

¹⁸ Refer to paragraph 1(j)(ii) under the Fifth Schedule to the PDPA, and S. 22(4) of the PDPA.

PART V: PUBLIC CONSULTATION QUESTIONS

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

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

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

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

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

END OF DOCUMENT

DRAFT