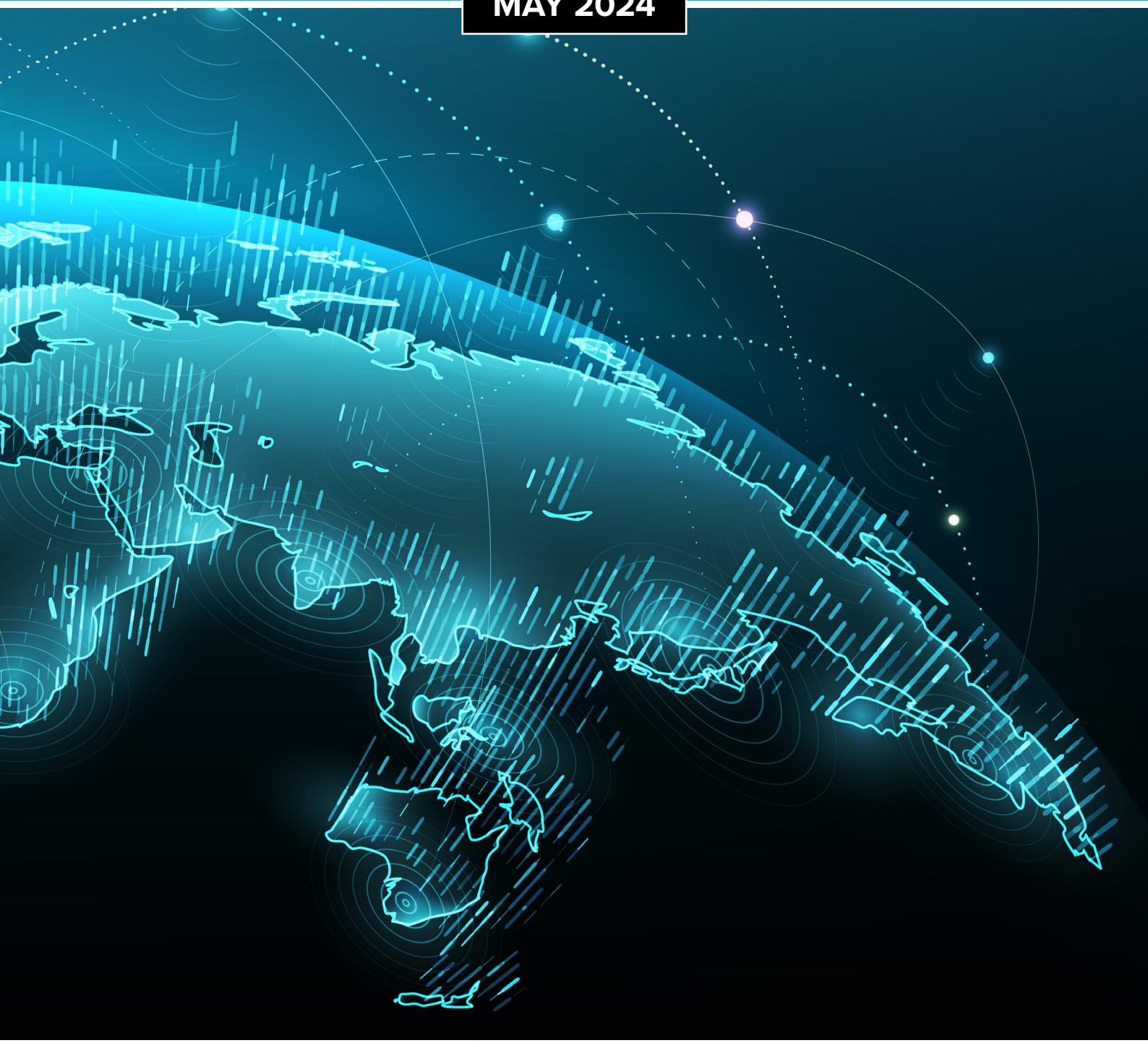
Navigating Governance Frameworks for Generative AI Systems in the Asia-Pacific

MAY 2024



FUTURE OF PRIVACY FORUM | AI

AUTHORED BY

Dominic Paulger
Policy Manager for Asia-Pacific, Future of Privacy Forum

ACKNOWLEDGEMENTS

This paper benefited from review and recommendations by Gabriela Zanfir-Fortuna, Anne J. Flanagan, Rob van Eijk, and Josh Lee Kok Thong. It also benefited from review and contributions from Bianca-Ioana Marcu and Lee Matheson.

FUTURE OF PRIVACY FORUM | CENTER FOR ARTIFICIAL INTELLIGENCE

About Future of Privacy Forum (FPF)

The **Future of Privacy Forum (FPF)** is a non-profit organization that serves as a catalyst for privacy leadership and scholarship, advancing principled data practices in support of emerging technologies. Learn more about FPF by visiting fpf.org.

About FPF's Center for Artificial Intelligence

The **Center for Artificial Intelligence** at the Future of Privacy Forum is dedicated to navigating the complex landscape of AI governance and its intersection with privacy and data protection law. Drawing on expertise from a global Leadership Council comprising industry leaders, academics, civil society, and policymakers, the Center provides sophisticated, practical policy analysis to help organizations align innovation with responsible implementation while meeting evolving regulatory requirements. Learn more about the FPF Center for AI at fpf.org/ai.

TABLE OF CONTENTS

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Across the APAC region, there is increasing interest in both understanding how generative artificial intelligence (AI) systems and large language models (LLMs) work, and exploring approaches to manage these technologies.

Leveraging the Future of Privacy Forum (FPF)'s global work on governance and regulation of AI,[1] FPF's Asia-Pacific (APAC) office commenced a research project on the regulatory and governance landscape for generative AI systems and LLMs in the APAC region in April 2023. The project focuses on 5 jurisdictions:

1. **Australia**
2. **China**
3. **Japan**
4. **Singapore**
5. **South Korea**

This Report is the culmination of that project. It notes that these jurisdictions are at an inflection point in the governance of generative AI systems, with a risk of fragmentation both within and between jurisdictions as regulatory responses diverge.

Section 1 of the Report charts early regulatory responses to generative AI in the 5 APAC jurisdictions. It notes that while most responses to date have favored voluntary guidelines and multi-stakeholder consultations, China has taken a unique approach by enacting binding regulations for generative AI.

This section also posits that as jurisdictions develop their respective generative AI governance frameworks, there are areas of consensus that could inform efforts to address regulatory fragmentation. Specifically, the 5 jurisdictions broadly agree on five identifiable risks posed by generative AI systems, and on certain recommended courses of action to address these risks. These five risks are:

- factual inaccuracies;
- lack of transparency;
- inappropriate use of personal data;
- malicious use of generative AI systems; and
- biased or discriminatory output.

Section 2 examines the broader landscape of existing laws and regulations in the five jurisdictions that may apply to generative AI. It highlights data protection law as a key source of legal obligations for developers and deployers of generative AI systems due to the use of personal data in training these systems.

This section also discusses data protection issues that are relevant for generative AI, such as lawful grounds for collecting and processing personal data, including publicly available personal data, managing data quality, handling personal data breaches, and fulfilling individual rights such as access to, correction, and erasure of personal data.

Section 3 highlights takeaways for policymakers and developers and deployers of generative AI, including but not limited to those in industry, in the APAC region to foster responsible governance of generative AI.

A key takeaway for **policymakers** is the need to counter the risk of regulatory fragmentation across the region. This may include ensuring alignment and interoperability with international policy frameworks, providing guidance on applying existing laws to generative AI, and promoting cross-regulator coordination.

Amongst the key takeaways for **developers and deployers of generative AI** are that robust internal AI governance structures, data management practices, privacy protection processes, security safeguards, and transparency measures are widely recognized building blocks for responsible development and deployment of generative AI systems. Enabling traceability of AI-generated content and clearly indicating its nature are also unanimous recommendations across the early regulatory responses to generative AI in the five jurisdictions.

INTRODUCTION

This Report explores the regulatory and governance landscape for generative AI in the APAC region, focusing on 5 jurisdictions: Australia, China, Japan, Singapore, and South Korea.

This Report observes that policymakers in the APAC region have generally taken a different approach to AI governance to their counterparts in other regions. Whereas several jurisdictions, such as the European Union (EU), have worked on crafting AI specific laws,[2] the APAC region has generally prioritized voluntary frameworks and non-binding guidelines.

Although generative AI has been around in some form for decades, it was only in late 2022 that publicly accessible and consumer-facing generative AI systems entered the market at scale. Policymakers in APAC and around the world are still at a very early stage in developing governance frameworks that are specific to the recent generative AI boom.[3] **Section 1** of this Report charts **early regulatory responses to generative AI** in the five APAC jurisdictions (these responses are summarized in detail in the **Appendix**) and concludes that:

- The majority of policymakers in the five jurisdictions do not currently appear to be looking to enact binding regulations to govern generative AI. A key exception is China, which has enacted technology-specific regulations. South Korea also plans to enact comprehensive AI regulation, but an existing bill does not specifically address generative AI.
- There are areas of consensus in emerging regulatory responses to generative AI among the five jurisdictions in APAC. These areas of consensus could inform efforts by policymakers and industry to increase regulatory interoperability between jurisdictions, as APAC jurisdictions develop their governance approaches for generative AI. Examples of such areas of consensus include identified common risks and recommended measures in the emerging generative AI regulatory frameworks.

Importantly, the development of generative AI systems does not take place in a regulatory vacuum. In the absence of binding regulations that specifically address generative AI, the main sources of legal obligations for generative AI are relevant existing technology-neutral laws and regulations. To that end, **Section 2** of the Report charts existing laws and regulations in the 5 jurisdictions that may apply to generative AI today. Findings include:

- **Data protection law has been a key source of binding legal obligations for generative AI.** This is because datasets containing personal data have been used to train several existing generative AI models, and this has therefore brought the resulting systems within the scope of such law.
- There are **several data protection issues particularly relevant for generative AI** in the 5 jurisdictions, such as the lawful grounds for processing personal data, or the rules applicable to collecting and processing publicly available personal data.
- In APAC (as elsewhere), **data protection authorities have been uniquely placed to take regulatory action** regarding generative AI systems, and some have already taken such action or issued guidance.

Looking to the future, **Section 3** of the Report notes that the APAC region is at an inflection point in the governance of generative AI systems. In particular, the section highlights that there is a risk of regulatory fragmentation surrounding generative AI in APAC because:

- *within* jurisdictions, multiple laws, frameworks, and guidelines may apply to generative AI systems; and
- *between* jurisdictions, regulatory responses to generative AI diverge as each jurisdiction pursues different priorities.

Finally, the Report highlights takeaways for policymakers and developers and deployers of generative AI, to consider in developing governance frameworks for generative AI and addressing the risk of regulatory fragmentation within and between jurisdictions.

Notes

This Report is informed by discussions held during two roundtables organized by FPF APAC in 2023, which sought input on the project from regulators, industry leaders, academics, and civil society representatives from across the APAC region, and beyond.

- The first roundtable was jointly organized with the Personal Data Protection Commission of Singapore (PDPC) and held in person during Singapore's Personal Data Protection Week in July 2023.
- The second roundtable was held virtually in October 2023 to open the discussion to a broader range of stakeholders from across the APAC region.

Both roundtables were held under the Chatham House Rule. FPF sincerely thanks all stakeholders who participated in our roundtables for their participation and insights.

The Report that follows should not be taken to reflect the views of any participant in the roundtables, and any errors are attributable to the author. The Report does not constitute legal advice.

Scope

The Report focuses on responsibilities for private sector organizations that develop and deploy AI under general laws, frameworks, and guidelines that apply to all kinds of organizations. It does not focus on sectoral AI laws or frameworks (e.g., healthcare, financial services, or other similar highly regulated sectors) or consider the responsibilities of public sector bodies.

Further, it also does not focus on: (1) intellectual property issues; (2) environmental, social, and governance risks; (3) competition concerns; and (4) labor force issues.

Definitions

This Report uses the term "**generative AI system**" expansively to include systems built using generative AI models, as well as applications built on top of such models.

This Report uses the term "**governance**" to refer to the set of policies and procedures that seek to ensure that AI technologies are developed, deployed, and used responsibly. This includes both voluntary frameworks and legally binding obligations.

GENERATIVE AI

Generative AI is a subset of AI technology that involves **AI models** – programs that have been trained on a set of data to recognize patterns or make decisions without further human intervention.[4] The models that power generative AI systems are capable of producing content in response to open-ended instructions from users in the form of a "prompt." The responses can appear human-made and are different with each iteration – the same prompt can produce a different output each time it is given to the system, especially given the iterative nature of models that continue to be trained by the data contained in the prompts.

These factors set generative AI apart from earlier "discriminative" AI models[5] that are effective at identifying the distinctions between different classes of data and are well-suited for tasks like pattern recognition, data classification, and prediction.

Although generative AI systems often appear to have human-like creative abilities, it is important to note that in producing content, generative AI models are merely analyzing patterns in their training data and using this analysis to make predictions – for instance, about what word to place next in a sentence or what pixel to place in an image. The models that power generative AI systems were not designed to store and retrieve information with 100% accuracy or verify the accuracy of the information they produce.[6] However, generative AI systems can combine these models with other technological solutions that allow for limited verification.

Modern generative AI systems can accept inputs and produce outputs in a wide range of different "modes," based on the data that they have been trained on.

LLMs are traditionally trained on large amounts of text data and so can accept inputs and produce outputs in text. More recently, several developers have begun releasing "**multimodal**" models, such as OpenAI's GPT-4, Google's Gemini, and Anthropic's Claude, which can accept inputs and produce outputs in both text and image form.

An ongoing area of research focuses on broadening the scope of generative AI models to incorporate additional data modalities, such as 3D environments, video, and audio. For instance, in February 2024, OpenAI introduced its Sora model, which enables users to generate a minute-long video from a short text prompt.[7]

When it comes to considering the policy implications of generative AI, it is helpful to look at how the technology is typically built. A key point that stakeholders raised during FPF's two roundtables on the governance of generative AI systems was the need to consider how generative AI is built across the different layers of its technology stack. We elaborate more on the technology stack for generative AI below.

Infrastructure layer

Generative AI systems run on physical hardware, demanding large capacity server infrastructure systems incorporating incredibly powerful central processing units (CPUs) and graphics processing units (GPUs). As most mainstream generative AI systems require such significant amounts of computational power, these systems often run high-density, environmentally controlled servers in cloud data centers which may be centrally located or distributed across different physical locations.

Model layer

Modern generative AI is built on **new AI models** that are based on complex algorithms that are highly capable and have been found to demonstrate human-like performance on a wide range of tasks. The technological development that enabled the development of these models was the discovery of a new type of neural network algorithm, known as "**transformers**" in 2017.[8] Combined with advances in computing infrastructure, the transformer algorithm allowed for the creation of AI **models** that can be trained efficiently on massive amounts of data.

Further, while much attention has been paid to so-called "general purpose"[9] generative AI models (such as OpenAI's ChatGPT, Google's Gemini, and Anthropic's Claude), there are narrower, use-case specific models that are trained on highly specific forms of data:

- **computer code** (e.g., Codex (OpenAI);[10] AlphaCode (DeepMind);[11] Project Wisdom (IBM)).[12]
- **chemistry data** (e.g., ChemBERTa (University of Toronto);[13] Chemformer (AstraZeneca);[14] MoLFormer (IBM));[15]
- **climate data** (e.g., ClimaX (Microsoft);[16] IBM-NASA geospatial intelligence foundation model);[17] and
- **financial data** (e.g., BloombergGPT (Bloomberg));[18]
- **ancient texts**;[19]
- **protein structures**;[20] and
- **organic molecules.**[21]

Application layer

The majority of users do not interact directly with generative AI models, but rather with **applications** that are built upon these models.

Applications that are built on generative AI models make highly capable AI accessible to the public at large and have the potential to aid humans in conducting numerous tasks like computer coding, content creation, transcription, and translation that used to be laborious or required significant training.

A tipping point for the adoption of applications built on generative AI models was the public release of an LLM-based chatbot, ChatGPT, by OpenAI in November 2022.[22] It was perhaps the first time in history that the public had interacted with an AI application that was both available for widespread consumer use and able to respond to or perform a wide range of different tasks, rather than just specific tasks.[23] Since then, an increasing number of such applications have been released on the market for a wide range of different use cases.

SECTION 1

Regulatory Responses to Generative AI in APAC

Overview

Unlike their counterparts in the EU who have worked on crafting AI-specific laws and hard regulations, such as the **AI Act**, policymakers in the APAC region have generally taken a different approach to AI governance, prioritizing voluntary frameworks and non-binding guidelines.

APAC Frameworks for AI Generally

Prior to 2023, AI-specific governance frameworks in the 5 jurisdictions covered by this Report were mainly limited to voluntary ethical principles and guidelines that apply generally to all forms of AI, including but not limited to generative AI. However, they were drafted before the emergence of modern generative AI systems so do not take into account specific policy considerations that are unique to such technologies. These include:

- **Australia's "AI Ethics Framework" (2019)**;[24]
- **China's "Ethical Principles for New Generation AI" (2021)**;[25]
- **Japan's "Social Principles of Human-Centric AI," (2019)**,[26] **and "Governance Guidelines for Implementation of AI Principles" (2022)**;[27]
- **Singapore's "Model AI Governance Framework" (2020)**;[28] and
- **South Korea's "Human Centered AI Ethics Standards" (2020)**.[29]

Of the 5 jurisdictions, only **South Korea** has been actively working on a comprehensive AI regulation. A draft "**Bill on Fostering AI and Creating a Foundation of Trust**" was tabled in the National Assembly in June 2021, but has not been enacted as of this Report's publication.[30]

Further information on the content of these frameworks may be found in the Appendix.

APAC Frameworks for Generative AI

Following the public release of several major generative AI systems in early 2023, the 5 jurisdictions have taken a wide range of different regulatory responses to generative AI, reflecting their distinct priorities, legal frameworks, and the role they envision for these technologies.

These responses are outlined below. Further information on the approaches adopted by each jurisdiction may be found in the Appendix.

AUSTRALIA

- **"Rapid Response Information Report on Generative AI" (March 2023)**, an influential report commissioned by the Australian Government that provides an overview of the development, regulatory landscape, and potential risks and opportunities of LLMs and foundation models.[31]
- **The eSafety Commissioner's "Tech Trends Position Statement on Generative AI" (August 2023)**, which provides an explainer on generative AI technologies and guidance for industry on minimizing online harm risks when developing and deploying generative AI.[32]
- **A public consultation on "Safe and Responsible AI in Australia"** that included a discussion paper published in June 2023,[33] and an interim response from the Australian government in January 2024.[34]
- **"Digital Platform Regulators Forum Working Paper on LLMs" (October 2023)**, a paper examining the regulatory implications of LLMs across several regulatory domains, including privacy and online safety.[35]

CHINA

- **"Regulations on the Administration of Deep Synthesis of Internet Information Technology"** (**Deep Synthesis Regulations**) **(January 2023)**, a set of binding regulations applying to deep-fakes and other forms of synthetic media.[36]
- **"Interim Measures for the Management of Generative AI Services" (Interim Generative AI Measures) (August 2023)**, a more detailed regulation outlining state policy principles for generative AI and establishing obligations on service providers throughout the lifecycle of a generative AI system.[37]
- **"Basic Security Requirements for Generative AI Services" (February 2024)**, a technical standard that outlines technical requirements for complying with the Interim Generative AI Measures.[38]

JAPAN

- **"Notice Regarding Cautionary Measures on the Use of Generative AI Services" (June 2023)**, a short guideline on complying with Japanese data protection law when using LLM chatbots, issued by the Personal Information Protection Commission (PPC) together with an **enforcement decision against OpenAI**.[39]
- **"Guidelines for AI Business Operators" (April 2024)**, a set of draft guidelines that aim to update Japan's voluntary framework in response to advanced AI, including generative AI.[40]

SINGAPORE

- **"Generative AI: Implications for Trust and Governance" (June 2023)**, a discussion paper that outlines proposals for policymakers and business leaders to build a trusted, responsible global ecosystem for adopting generative AI.[41]
- **"Proposed Model AI Governance Framework for Generative AI" (January 2024)**, a draft policy framework that builds on the earlier discussion paper that highlights potential regulatory actions and governance measures various stakeholders could adopt to enhance generative AI trust and safety.[42]

SOUTH KOREA

- The Personal Information Protection Commission (PIPC)'s **enforcement decision against Open AI (March 2023)**.[43]
- **"Policy Direction for Safe Use of Personal Information in the AI Era" (August 2023)**, which outlines measures that the PIPC will take in response to emerging privacy challenges from AI and provides preliminary guidance on protecting privacy when developing and deploying AI systems.[44]

Comparison Shows a Wide Spectrum of Policy Responses

This subsection of the Report (1) compares the policy responses to generative AI in the 5 jurisdictions; summarizes (2) risks from generative AI identified by these policy responses; and (3) measures proposed by these policy responses to govern generative AI.

The majority of the above policy responses to generative AI are voluntary governance frameworks or early efforts by policymakers to shape the future direction of governance of generative AI. To date, only China has enacted legally binding regulations. That said, these policy responses mark an evolution away from earlier, principle-based or thematic frameworks to more comprehensive frameworks that increasingly recognize that different responsibilities may arise at the different stages of the AI lifecycle.

Many of these policy responses also identify specific **risks** arising from generative AI and, in some cases, propose **measures** that developers and deployers could adopt to improve their governance of generative AI.

These are covered in more detail below in the subsections on risks identified and measures recommended by policymakers in the 5 jurisdictions.

While the diversity of these responses precludes a fully like-for-like comparison, a high-level comparison of these approaches reveals a spectrum of policy strategies, from voluntary guidelines and international collaboration, to the development of comprehensive national legislation and actions by various regulators.

Jurisdictions Differ in the Forms and Legal Effect of Their Policy Responses to Generative AI

A key difference between jurisdictions is in whether they prioritize voluntary norms and bottom-up multi-stakeholder frameworks or binding top-down regulations for governing generative AI.

Broadly, Australia, Singapore, and Japan have favored **multi-stakeholder consultations** involving government, industry, academia, and civil society. Their aim is to develop **internationally aligned frameworks** and **voluntary guidelines** that enable responsible innovation in generative AI while mitigating risks:

- Australia's approach has prioritized domestic public consultation, leveraging expert reports, and inter-agency coordination to establish a risk-based framework.
- Japan has based its approach on international collaboration, notably through its G7 presidency in 2023.
- Singapore has sought to bring together local, regional, and international stakeholders to collaborate on developing a bespoke governance framework for generative AI, as well as on AI governance testing for generative AI systems (and other AI technologies more broadly).

In contrast, China has adopted a more prescriptive approach by enacting two sets of **binding, technologically specific regulations** to govern

generative AI and related issues, such as synthetic media. These aim to align the provision of generative AI-powered services to China's national interests and principles and impose obligations on providers of services using generative AI and technologies that use AI to create or modify media.

South Korea is charting a hybrid course: while its Ministry of Science and ICT has been developing comprehensive national AI legislation, its data protection authority, the PIPC, has concurrently focused on issuing detailed guidance and establishing programs to enable AI innovation while managing privacy risks.

Jurisdictions Differ in How Their Policy Responses to Generative AI Allocate Roles and Responsibilities

Another key difference concerns the allocation of roles and responsibilities within emerging governance frameworks. Policy responses to generative AI in the 5 jurisdictions are increasingly recognizing that different responsibilities arise at different stages of the lifecycle of an AI system, including development and deployment.

However, a comparison of emerging governance frameworks across the 5 jurisdictions reveals that although the frameworks recognize these different responsibilities, they do not always clearly identify the different roles associated with them.

For instance, some frameworks clearly differentiate between the responsibilities of developers, deployers, and users at different stages of the AI lifecycle. Frameworks in this category include Japan's Guidelines for AI Business Operators and to a lesser extent, Singapore's Proposed Model AI Governance Framework for Generative AI.

Others, however, identify responsibilities that apply at different stages of the AI lifecycle but use the generic term "service providers" for all such responsibilities, without differentiating between those who could be categorized elsewhere as developers or deployers. Frameworks in this category include the Tech Trends Position Statement on Generative AI by Australia's eSafety Commissioner and China's generative AI regulations.

Jurisdictions Differ as to Which Entities Have Issued Responses to Generative AI

Within jurisdictions, a further difference is in which agencies or branches of government have been leading efforts to govern generative AI.

AUSTRALIA

The Australian government has taken a multi-agency approach. The **Department of Industry, Science and Resources** (DISR) leads public consultations and framework development.

Australia's data protection authority, the **Office of the Australian Information Commissioner** (OAIC)'s role in governance of generative AI has largely been confined to its participation in the **Digital Platform Regulators Forum** (DP-REG), and it has not issued any generative AI-specific guidance to date.

By contrast, another DP-REG member, the **eSafety Commissioner,** has played a far more active role by providing guidance to industry on mitigating risks from generative AI, albeit with a focus on online safety.

CHINA

China's approach is centralized. China's cyberspace regulator, the **Cyberspace Administration of China** (CAC) has been responsible for issuing all binding regulations. Technical standards are set by the cybersecurity standards body, known as **TC260**.

JAPAN

Japan has prioritized international collaboration through the G7 Hiroshima AI Process. These efforts have involved the **Ministry of Foreign Affairs**, the **Ministry of Internal Affairs and Communications**, the **Digital Agency**, and the **Ministry of Economy, Trade and Industry** (METI).

Domestically, **METI** has played a central role in developing AI governance frameworks, while the data protection authority, the **PPC**, has issued preliminary guidance on the privacy implications of the use of LLM chatbots.

SINGAPORE

Singapore's combined infocommunications, media, and data protection regulator, the **Infocomm Media Development Authority** (IMDA) has led collaborative efforts, partnering with industry to establish the AI Verify Foundation and engaging stakeholders through initiatives like the proposed generative AI governance framework.

SOUTH KOREA

The **Ministry of Science and ICT** (MSIT) has been leading the development of a comprehensive AI bill.

The data protection authority, the **PIPC**, has taken a proactive stance on data protection and enforcement actions against non-compliant AI practices.

Data protection authorities are often in a unique position relative to other agencies or branches of government as they have an existing legal mandate to regulate the processing of personal data. This

enables them to regulate generative AI systems that are trained on or otherwise use personal data. In APAC, data protection authorities in Japan and South Korea have been most active in addressing generative AI's privacy and data protection implications. Both jurisdictions issued guidance for businesses on complying with data protection laws when using services like ChatGPT. Both have also pursued enforcement actions against OpenAI over ChatGPT's handling of personal data.

Common Risks from Generative AI Identified by Policymakers in the 5 Jurisdictions

As policymakers in the 5 jurisdictions covered by this Report seek to understand and respond to this evolving landscape, they have highlighted several potential risks from generative AI in their various regulatory responses.

Given the rapid pace of advancement in generative AI capabilities, it is fair to assume that not all potential risks posed by generative AI have been identified at this stage. However, an analysis of APAC policymakers' responses to generative AI indicates an emerging consensus around key risks. In particular, policymakers in all 5 of the jurisdictions covered by this Report identified the following risks from existing generative AI systems:

- the potential for generative AI systems to produce **factual inaccuracies**;[45]
- **lack of trust and transparency**;
- **inappropriate use of personal data** to train generative AI models;
- **malicious use** of generative AI systems, including to spread misinformation and disinformation; and
- generation of **biased or discriminatory content**.

Factual Inaccuracies

Policymakers highlighted risks associated with generative AI's tendency to produce factual inaccuracies. Several regulatory responses raise concerns that such inaccuracies can mislead or even harm people – for instance, by making defamatory statements that harm individuals' reputations or providing ineffective health advice.

This risk arises because generative AI models are probabilistic rather than deterministic – they generate output by predicting what item should come next in a sequence (e.g., a word in a sentence or a pixel in an image) based on the data that they have been trained on. Text-based generative AI applications may therefore sometimes produce statements that are grammatically correct but factually inaccurate based on probabilities assigned to information in their training data.

While the models can make highly accurate predictions, they have no grounding in the real world outside of their training data. This means that, by default, these systems are unable to verify the information they produce and may fail to understand the context of that information. For instance, the information may not have been fact-checked, may reflect deliberate misinformation posted online, or may not have been intended to be factually accurate, as is the case for creative works like fiction, poetry, or humor.[46] Even when trained on accurate statements, LLMs can still occasionally assign high probabilities to factually inaccurate statements due to their reliance on statistical patterns.[47]

Lack of Trust and Transparency

Policymakers highlighted lack of transparency as a risk for generative AI systems. Transparency is viewed as a multifaceted issue spanning:

- documenting model and system capabilities, training data, limitations, and intended uses;
- explaining organizational policies; and
- clearly indicating when people are interacting with AI systems or AI-generated content.

Crucially, policymakers in all 5 of the jurisdictions have closely linked transparency with accountability and have highlighted that without insight into how these systems work and what data they were trained on, it becomes difficult to understand outputs, apportion liability, seek redress, anticipate safety risks, and establish effective safeguards.

Potential transparency gaps identified include lack of clarity around personal data use, incomplete transparency on model capabilities and limitations, and inadequate information for stakeholders to make informed decisions and establish effective safeguards.

Inappropriate Use of Personal Data

Policymakers highlighted that the training of generative AI models on personal data may give rise to data protection and privacy risks.

Notably, policymakers raised particular concerns where such data was obtained through "**scraping**" – the automated extraction of data from the internet. For instance, in South Korea, the PIPC's Policy

Direction for Safe Use of Personal Information in the AI Era highlights that generative AI systems that have been trained on "scraped" personal data may effectively process personal data in ways unanticipated by data subjects i.e. to train Generative AI models, thereby potentially increasing the scale of privacy infringements.

This aligns with broader international trends. For instance, in August 2023, data protection authorities from 12 jurisdictions, including Australia, released a **joint statement on data scraping and the protection of privacy**.[48] The statement outlines privacy risks related to data scraping, including targeted cyberattacks, identity fraud, monitoring, profiling, unauthorized political or intelligence gathering, and unwanted direct marketing. The statement also emphasizes that even publicly available personal data remains subject to data protection laws, with obligations applying to both data scrapers and operators of platforms hosting the data.

Other risks arise from the tendency of generative AI models to "**memorize**" specific phrases, sentences, or even longer passages from their training data and reproduce this information in their outputs.[49] This can lead systems inadvertently to leak personal data, as well as confidential information, that they have been trained on.[50] This data security risk may also extend to personal data that users input into a generative AI system, such as an LLM chatbot, if the system retains that data for further training.

This risk is discussed in further detail under Section 2: Existing Laws.

Malicious Use

Policymakers recognized the risk of users and threat actors attempting to circumvent safeguards designed to prevent the malicious use of generative AI systems. This practice, commonly known as "**jailbreaking**," involves manipulating prompts to bypass restrictions and make AI perform undesired tasks.[51] Research has shown that it can be relatively easy to evade safeguards established by generative AI service providers by slightly altering the prompt. For instance, a generative AI system might be prohibited from explaining how to rob a bank *per se* but may still provide the prohibited information if asked to write a one-act play about how to rob a bank, or to explain how to rob a bank "for educational purposes."[52]

Policymakers have highlighted the following as potential malicious uses for the technology, including:

- **Fraud**: AI-generated content can be used to facilitate "phishing" attacks, where threat actors deceive individuals into disclosing sensitive information by posing as trustworthy entities.
- **Promotion of hate, discrimination, and abuse**: Generative AI can be manipulated to spread harmful ideologies and promote abusive behavior.
- **Harmful content**: In the absence of proper guardrails, AI systems may inadvertently provide inappropriate or harmful information, potentially causing harm to users or facilitating unlawful acts.
- **Abusive and illegal content**: Recent advancements in image generation models have enabled the creation and dissemination of abusive and illegal content, such as the non-consensual creation of synthetic media featuring the likenesses of real persons in compromising situations, and the creation of synthetic child sexual abuse material.
- **Malware creation**: Generative AI models have the potential to be used in crafting malicious software code across various programming languages for cybercrime purposes.
- **Misinformation and disinformation**: Most policymakers highlighted the risk that generative AI could be used maliciously to increase the scale and effectiveness of misinformation and disinformation campaigns.

Bias and Discrimination

Statistically, it is likely that any data set will contain some bias. Policymakers have therefore identified the risk that biases in the data used to train generative AI systems may cause these systems to produce outputs that amplify these biases and encourage discrimination.

Broadly, policymakers highlight two high-risk forms of bias that can arise in generative AI training data. **Historical bias** refers to patterns of harmful stereotypes and negative attitudes towards certain groups being reflected in the training data due to historical depictions of those groups.[53] **Representation bias** occurs when certain groups are over or underrepresented in the datasets.[54] Both types of bias in a training data set can potentially result in discriminatory output which can be harmful to individuals. In the case of generative AI, where the algorithm is constantly being trained by its human reviewers, this harm may be further amplified.

A related issue is "**toxicity**."[55] Where the training dataset contains negative, discriminatory, offensive, or excessively insensitive perspectives, a generative AI system may produce outputs that reflect such perspectives, even without direct human influence. When it comes to determining what forms of content are toxic, such assessment depends on the context and is often highly subjective. Additionally, some forms of content may be offensive even if they do not use blatantly inflammatory language, as is the case with coded language or "dog whistles."

Measures Recommended by Policymakers in the 5 Jurisdictions to Govern Generative AI Vary in Nature, but Share Some Commonalities

In addition to identifying potential risks from generative AI, policymakers in all 5 jurisdictions covered by this Report have begun identifying possible measures that developers and deployers of generative AI could implement to address these potential risks.

However, policymakers have done so in a range of different contexts, for different purposes, and with different levels of granularity.

Policymakers in Australia (DISR) and Singapore have focused on **proposing, and seeking feedback on, potential measures** that could be included in future AI guidelines or regulation. These proposals tend to be drafted in very broad terms, and businesses are not necessarily expected to adopt them.

Further, policymakers in Australia (eSafety Commissioner), Japan, and South Korea have focused on issuing **voluntary guidelines recommending specific measures** for industry to implement. These guidelines target different groups, including:

- all businesses that develop, deploy, and use AI (Japan);
- those businesses who process personal data in the development, deployment, and use of generative AI (South Korea);
- those whose development, deployment, and use of generative AI has implications for users' online safety (Australia).

These guidelines outline recommended measures in detail but are still sufficiently open-ended and flexible that different businesses could adapt them to their specific needs.

Policymakers in China have focused on **enacting legally binding regulations requiring service providers to implement measures** in relation to specific technologies.

The diversity of these responses precludes a fully like-for-like comparison. However, a broad survey of regulatory responses to generative AI in the 5 jurisdictions covered in this Report indicates that there are some emerging areas of substantial consensus.

In particular, all 5 of the jurisdictions highlighted the need for organizations to develop their own **internal AI governance and risk management policies** and **provide transparency** by publishing documentation (e.g., model cards), AI governance policies, and transparency reports.

The survey also indicates further areas of consensus in measures highlighted by at least 4 of the 5 jurisdictions. These include:

- Conducting **impact assessments** to identify and mitigate harm.
- Implementing measures to **manage data quality** to mitigate against harmful biases.
- Developing **privacy management programs** and disclosing a privacy policy.
- Deploying **security measures**, including:
  - Assessing security risks;
  - Conducting security testing on systems before deployment;
  - Monitoring systems after deployment;
  - implementing measures to address identified risks and vulnerabilities;
  - Reporting security incidents;
  - Implementing security controls to address risks to both physical security and cybersecurity;
  - Sharing information regarding risks and best practices.
- Developing and implementing **measures to indicate that content is AI-generated**, such as watermarking, labeling, and other authentication and provenance mechanisms.

Interestingly, many of these measures overlap significantly with recommendations made by the G7 in the "**Hiroshima AI Process Comprehensive Policy Framework**" – one of the most detailed international frameworks for governance of advanced AI systems, including generative AI, that has been released at the international level to date. As policymakers in the APAC region continue to develop their national-level generative AI governance frameworks, there is scope for international alignment along these lines.

SECTION 2

Existing Laws in the 5 Jurisdictions Likely Relevant to Generative AI

The majority of the jurisdictions covered by this Report currently lack AI-specific laws, such as the EU's AI Act and their regulatory responses to generative AI generally have not been legally enforceable.

This means that for the time being, the main source of binding legal obligations governing generative AI systems in all jurisdictions covered by this Report (except China) remains existing, technology-neutral laws. Further, even for jurisdictions like China that have begun to enact regulations to address specific risks presented by generative AI, existing laws will remain relevant to matters that fall outside of the scope of these generative AI-specific regulations.

Identifying all relevant laws and obligations across the 5 jurisdictions is beyond the scope of this Report and would depend on the specific circumstances of how organizations develop, deploy, and use generative AI.

The table below provides a **snapshot of existing legal frameworks** in the 5 jurisdictions covered by this Report that may be relevant to risks from generative AI identified by policymakers and summarized in the previous section, from consumer protection law to criminal law.

Importantly, the mapping in the table below does not include data protection and privacy law, whose cross-cutting applicability to generative AI will be explored in detail in the following sub-section.

Mapping of Existing Legal Frameworks in the 5 Jurisdictions that are Relevant to Generative AI, in addition to Data Protection Law

POTENTIAL HARM CAUSED BY GENERATIVE AI SYSTEMS	AUSTRALIA	CHINA	JAPAN	SINGAPORE	SOUTH KOREA
Producing factually inaccurate or misleading output	**Consumer protection law** Competition and Consumer Act: Australian Consumer Law **Civil remedies** Common law of contract and tort (e.g., negligence, defamation, misrepresentation) **Professional regulation** (finance, medical, legal sectors)	**Consumer protection law** Law on the Protection of Consumer Rights and Interests **Civil law** (contract, tort, defamation, etc.) Civil Code **Professional regulation** (finance, medical, legal sectors)	**Consumer protection law** Act against Unjustifiable Premiums and Misleading Representations **Civil law** (contract, tort, defamation, etc.) Civil Code **Professional regulation** (finance, medical, legal sectors)	**Consumer protection law** Consumer Protection (Fair Trading) Act **Civil remedies** Common law of contract and tort (e.g., negligence, defamation, misrepresentation) Defamation Act **Professional regulation** (finance, medical, legal sectors)	**Consumer protection law** Framework Act on Consumers Act on Fair Labeling and Advertising Act on the Regulation of Terms and Conditions **Civil law** (contract, tort, defamation, etc.) Civil Act **Professional regulation** (finance, medical, legal sectors)
Misplaced human reliance on AI-generated content	**Civil remedies** Common law of contract and tort (e.g., negligence) **Professional regulation** (finance, medical, legal sectors)	**Civil law** (contract, tort, etc.) Civil Code **Professional regulation** (finance, medical, legal sectors)	**Civil law** (contract, tort, etc.) Civil Code **Professional regulation** (finance, medical, legal sectors)	**Civil remedies** Common law of contract and tort (e.g., negligence) **Professional regulation** (finance, medical, legal sectors)	**Civil law** (contract, tort, etc.) Civil Act **Professional regulation** (finance, medical, legal sectors)

POTENTIAL HARM CAUSED BY GENERATIVE AI SYSTEMS	AUSTRALIA	CHINA	JAPAN	SINGAPORE	SOUTH KOREA
Causing physical economic, or psychological harm	**Criminal law** Criminal Code **Consumer protection law** Competition and Consumer Act, Australian Consumer Law **Civil remedies** Common law of contract and tort. **Sectoral laws** (finance, medical, legal sectors) **Online Safety** Online Safety Act	**Criminal law** Criminal Law **Consumer protection law** Law on the Protection of Consumer Rights and Interests **Civil law** (contract, tort, etc.) Civil Code Tort Liability Law **Sectoral laws** (finance, medical, legal sectors)	**Criminal law** Penal Code **Consumer protection law** Consumer Product Safety Act **Civil law** (contract, tort, etc.) Civil Code **Sectoral laws** (finance, medical, legal sectors)	**Criminal law** Penal Code **Consumer protection law** Consumer Protection (Fair Trading) Act **Civil remedies** Common law of contract and tort **Sectoral laws** (finance, medical, legal sectors) **Online safety law** Online Safety Code Online Criminal Harms Act 2023	**Criminal law** Criminal Act **Consumer protection law** Framework Act on Consumers Act on the Regulation of Terms and Conditions **Civil law** (contract, tort, etc.) Civil Act **Sectoral laws** (finance, medical, legal sectors) **Online content law** Telecommunications Business Act Network Act
Creating biased or discriminatory content	**Civil remedies** Common law of contract and tort. **Anti-discrimination law** Racial Discrimination Act Sex Discrimination Act Disability Discrimination Act Age Discrimination Act	**Criminal law** Criminal Law **Civil law** (contract, tort, etc.) Civil Code **Content regulation** Regulations on Ecological Governance of Internet Information Content	**Criminal law** Penal Code **Civil law** (contract, tort, etc.) Civil Code **Anti-Discrimination law** Act on the Promotion of Efforts to Eliminate Unfair Discriminatory Speech and Behavior Against Persons Originating from Outside Japan	**Criminal law** Penal Code **Civil remedies** Common law of contract and tort. Protection from Harassment Act 2014. **Anti-Discrimination law** Maintenance of Religious Harmony Act	**Criminal law** Criminal Act **Civil law** (contract, tort, etc.) Civil Act
Creating/ spreading disinformation or misinformation	**Criminal law** Criminal Code **Laws against the spread of online disinformation and misinformation** Australian Code of Practice on Disinformation and Misinformation (voluntary)[56] *Combatting Misinformation and Disinformation Bill* **Civil remedies** Common law of tort (e.g., defamation)	**Criminal law** Criminal Law, Articles 221, 243 **Laws against the spread of online disinformation and misinformation** Regulations on Ecological Governance of Internet Information Content **Civil law** (contract, tort - defamation, etc.)	**Criminal law** Penal Code **Civil law** (contract, tort - defamation, etc.) Civil Code	**Criminal law** Penal Code Defamation Act **Laws against the spread of online disinformation and misinformation** Protection from Online Falsehoods and Manipulation Act 2019 **Civil remedies** Common law of tort (e.g., defamation). Defamation Act	**Criminal law** Criminal Act **Online content law** Telecommunications Business Act Network Act **Civil law** (contract, tort - defamation, etc.) Civil Act

POTENTIAL HARM CAUSED BY GENERATIVE AI SYSTEMS	AUSTRALIA	CHINA	JAPAN	SINGAPORE	SOUTH KOREA
Bullying and harassment	**Criminal law** Criminal Code, Sections 474.17 and 474.17A (using a carriage service to menace, harass or cause offense) **Civil remedies** Common law of tort **Online Safety** Online Safety Act, Parts 5 (cyber-bullying material targeted at an Australian child), 6 (non-consensual sharing of intimate images), 7 (cyber-abuse material targeted at an Australian adult)	**Civil law** (e.g. tort) Civil Code **Anti-harassment regulation** Law on the Protection of Women's Rights and Interests *Draft law on cyberbullying*	**Criminal law** Penal Code **Civil law** (e.g. tort) Civil Code **Anti-harassment regulation** Equal Opportunity Act Act on the Promotion of Efforts to Eliminate Unfair Discriminatory Speech and Behavior Against Persons Originating from Outside Japan	**Criminal law** Penal Code **Civil remedies** Common law of tort **Online safety law** Online Safety Code **Anti-harassment regulation** Protection from Harassment Act 2014	**Criminal law** Criminal Act **Civil law** (e.g. tort) Civil Act **Online content law** Telecommunications Business Act Network Act
Fraud, including phishing	**Criminal law** Division 134 (Obtaining property or a financial advantage by deception) **Civil remedies** Common law of tort. **Online Safety** Online Safety Act	**Criminal law** Criminal Law, Articles 266, 287 Law on Combating Telecom and Online Fraud **Civil law** (e.g. tort) Civil Code	**Criminal law** Act on the Prohibition of Unauthorized Computer Access, Articles 6, 7, 12 Penal Code, Article 161-2 **Civil law** (e.g. tort) Civil Code	**Criminal law** Penal Code, ss 416, 419, 170 Computer Misuse Act 1993, ss 3-4 **Civil remedies** Common law of tort. **Online safety law** Online Criminal Harms Act 2023	**Criminal law** Criminal Code **Civil law** (e.g. tort) Civil Act **Online content law** Telecommunications Business Act Network Act
Malware generation	**Criminal law** Criminal Code, Vol 2, Part 10.7 (computer offenses)	**Criminal law** Criminal Law, Articles 285-287 Cybersecurity Law, Article 27	**Criminal law** Penal Code, Article 161-2, 168-2, 168-3, 234-2, 246-2	**Criminal law** Computer Misuse Act 1993, s 5 Computer Misuse and Cybersecurity (Amendment) Act 2017, s 8B(1)(b)	**Criminal law** Criminal Act

POTENTIAL HARM CAUSED BY GENERATIVE AI SYSTEMS	AUSTRALIA	CHINA	JAPAN	SINGAPORE	SOUTH KOREA
Creation/ distribution of abusive material » Extremist content » Child abuse » Non-consensual intimate images	**Criminal law** Criminal Code, Divisions 80 (urging violence and advocating terrorism or genocide), 471, D (Offences relating to use of carriage service for child abuse material), 474 (telecommunications offences) **Civil remedies** Common law of contract and tort **Data protection law** Privacy Act **Online Safety** Online Safety Act, Parts 6 (non-consensual sharing of intimate images), 8 (material that depicts abhorrent violent conduct), 9 (online content scheme)	**Civil law (e.g. tort)** Civil Code **Data protection law** Personal Information Protection Law **Online safety law** Regulations on Ecological Governance of Internet Information Content	**Criminal law** Penal Code Act on Regulation and Punishment of Acts Relating to Child Prostitution and Child Pornography, and the Protection of Children **Civil law** (e.g. tort) Civil Code **Data protection law** Act on the Protection of Personal Information	**Criminal law** **Civil remedies** Common law of contract and tort. **Data protection law** Personal Data Protection Act **Online safety law** Online Safety Code Online Criminal Harms Act 2023	**Criminal law** Criminal Act **Civil law** (e.g. tort) Civil Act **Data protection law** Personal Information Protection Act **Online content law** Telecommunications Business Act Network Act
Age-Inappropriate Content	**Online Safety** Online Safety Act, Parts 9 (online content scheme)	**Online Safety** Law on the Protection of Minors Regulations on Ecological Governance of Internet Information Content	**Online Safety** The Act on Development of an Environment that Providers Safe and Secure Internet Use for Young People	**Online Safety law** Online Safety Code Online Criminal Harms Act 2023	**Online content law** Telecommunications Business Act Network Act Youth Protection Act

Data Protection

Though numerous existing laws and regulation in the 5 jurisdictions may apply to generative AI, data protection law is likely to be one of the main sources of binding legal obligations for generative AI, considering the horizontal applicability of the rules to any "processing of personal data," the fact that there are dedicated supervisory authorities to enforce it, and the common use of personal data in training generative AI models.

While only some of the 5 jurisdictions reviewed in this Report have laws specifically designed to address areas like online safety or misinformation, all have data protection laws. These include:

- **Australia**'s Privacy Act, which gives effect to the Australian Privacy Principles (APPs).[57]
- **China**'s Personal Information Protection Law (PIPL).[58]
- **Japan**'s Act on the Protection of Personal Information (APPI).[59]
- **Singapore**'s Personal Data Protection Act (PDPA).[60]
- **South Korea**'s Personal Information Protection Act (PIPA).[61]

Further, these laws have generally already been implemented. Regulatory authorities have been established to enforce data protection laws in all of the 5 jurisdictions, and many have also issued detailed guidance on compliance.

This subsection of the Report raises several considerations that:

- policymakers can think through when examining how their respective data protection laws apply to generative AI, and
- developers and deployers of generative AI systems can note when complying with data protection laws in the 5 jurisdictions.

Not every instance of operating a generative AI system will involve the processing of personal data. However, many generative AI systems do so, especially where the large datasets used to train the generative AI model contain personal data, and/or the system collects personal data that users have entered into it, and uses this data to further train the underlying model.

Insofar as generative AI models or applications process personal data, they may be subject to obligations under personal data protection law. Failure to comply with these obligations may give rise to penalties or other sanctions from data protection authorities.

There have already been several important decisions by data protection authorities in Italy, Japan, and South Korea concerning OpenAI's provision of services to users through its LLM chatbot, ChatGPT.[62]

Common issues across these decisions included:

- Lack of legal authority to process personal data to train a generative AI model.
- Failure to adequately inform data subjects about the processing of their personal data, including failing to provide information in languages other than English.
- Processing personal data in violation of data protection principles, such as data quality and data minimization.
- Failing to provide mechanisms for data subjects to exercise rights, such as correction of their data or opting out of processing of their data to train the GPT model.
- Failing to verify the ages of users and obtain parental consent for use of ChatGPT by minors.

Further, data protection authorities in **Japan** and **South Korea** have issued guidance that specifically addresses the application of data protection laws to generative AI**.**

Separately, there have also been statements from multiple data protection authorities at the international level that identify issues under existing data protection and privacy laws that may arise from the development and deployment of generative AI. These statements include:

- **the G7 data protection and privacy authorities' statement on generative AI (June 2023)**
- **the Global Privacy Assembly's Resolution on Generative AI (October 2023); and**
- **the joint data protection authorities' (DPAs) statement on data scraping (August 2023)**.

Legal Authority to Process Personal Data to Train Generative AI Models

One of the biggest challenges for organizations to comply with existing data protection laws in the context of developing and deploying generative AI is ensuring that these organizations have legal authority (or 'lawful ground,' or 'legal basis') to process personal data to train generative AI models.

All data protection laws require organizations to fulfill certain criteria (such as obtaining **consent** from data subjects, or establishing that the processing is **necessary** for a specified purpose) before organizations have legal authority to process personal data.

Previous work by FPF has identified two main challenges for organizations that process personal data in multiple jurisdictions in APAC.[63] These include: (1) the lack of consistency between jurisdictions in the

available legal bases for processing personal data; and (2) the lack of alternative legal bases to consent (such as "legitimate interests") that can be relied on to process personal data in a variety of different circumstances.

This work found that as a result of these issues, organizations that process personal data in multiple APAC jurisdictions would likely have to build their compliance frameworks around consent, as this legal basis was the main "common denominator" for data protection laws in the APAC region, especially for sensitive personal data.

The most suitable legal basis for processing personal data will likely depend on the circumstances in which the organization obtained the dataset for training a generative AI model. These circumstances include:

- **Collecting data** through "**web crawls**" of publicly available web pages containing personal data (also commonly referred to as "scraping");
- **Collecting data from end-users of generative AI applications to refine the underlying model**. Generative AI applications may also process personal data if end-users input personal data via a prompt, and the application retains that data (e.g., to further train the underlying AI model); and
- **Reusing an existing dataset** that contains personal data.

COLLECTION OF PERSONAL DATA

Training Datasets Obtained through "Web Crawls"

Modern generative AI systems, particularly LLMs, rely heavily on training data derived from large-scale "crawls" of the internet.[64] This involves employing automated programs to systematically navigate and extract information from websites.

For instance, several common, publicly available datasets for training AI systems are based on the "**Common Crawl**" – a massive compilation of publicly available websites collected regularly since 2008.[65] These include:

- **Colossal Clean Crawled Corpus** (C4), a cleaned version of the Common Crawl prepared by AllenAI that has been used to train, among others, Meta's LLaMa model.[66]
- **LAION-400M**[67] **and LAION-5B**:[68] Two datasets containing, respectively, 400 million and 5 billion pairs of image and text data prepared by the **Large-scale Artificial Intelligence Open Network** (LAION) that was used to train, among others, Stability AI's Stable Diffusion model.

These massive datasets are then fed into the LLM, where it learns the underlying statistical relationships and patterns within human language. This allows the AI to develop the ability to generate human-like text, translate languages, and perform other complex natural language processing tasks.

As web crawls may, by their nature, capture personal data that has been posted online, organizations would likely need to establish legal authority to process that personal data by fulfilling the requirements for a legal basis under relevant data protection laws.

However, there are practical issues with doing so. Processed data may have been made public without those individuals' knowledge or consent; it may also qualify as sensitive under various data protection laws, particularly if such data has been leaked online. Further, given the size of crawled datasets, such data may potentially relate to millions of data subjects worldwide.

Given these factors, it may not be possible to rely on **consent** to process personal data in training datasets obtained through web crawls. First, it would not be feasible to identify the data subjects whose personal data is present in the dataset. Second, even if an organization was able to identify such data subjects, the organization may not have the necessary contact information to seek their consent as it has no prior relationship with them.

More suitable legal bases within the 5 jurisdictions studied are those which permit the processing of personal data without consent if:

- the personal data is **publicly available**;
- the processing is necessary for a **legitimate interest** of the organization or a third party; or
- the processing is for **research purposes**.

Consent and Sensitive Personal Data: As discussed above, it is likely infeasible to obtain consent from data subjects in this situation. Another challenge is that data protection laws in 4 of the 5 jurisdictions covered by this Report (Australia, China, Japan, and South Korea) require consent to process **sensitive personal data**.

3 of these 4 jurisdictions (Australia, China, and South Korea) do not provide alternatives to consent that would apply to the use of personal data to train a generative AI model. This could prevent organizations operating in these jurisdictions from using web crawled datasets for this purpose or otherwise expose them to the risk of enforcement actions from data protection authorities.

Publicly Available Personal Data: Data protection laws in 3 of the 5 jurisdictions (China, Japan, and Singapore) expressly provide legal bases that permit the processing of publicly available personal data without consent. Organizations would likely have little difficulty in complying with the relevant provisions in Japan's APPI and Singapore's PDPA to process personal data for the purpose of training a generative

AI model, as these provisions appear to only require that the personal data is available to the public.

However, organizations may encounter difficulties relying on the relevant provision in China's PIPL due to the safeguards required. In particular, it may be difficult to argue that the processing of a data subject's publicly available personal data for the purpose of training a generative AI model would not have a significant impact on the data subject's rights and interests, given that internationally, there have been several high-profile cases where generative AI systems trained on publicly available data have produced factual inaccuracies that are potentially defamatory. For instance, in April 2023,

- The Washington Post reported that ChatGPT had falsely claimed that a law professor had been accused of sexual harassment and cited a non-existent Washington Post article as the source of the information.[69] The newspaper came out to say that there was no such article.
- An Australian mayor threatened to sue OpenAI for defamation after ChatGPT falsely claimed that he had been convicted of bribery and imprisoned.[70]

Further, the relevant provision of the PIPL is of limited benefit in this situation as it does not apply to sensitive personal data.

Legitimate Interests: Data protection laws in 2 of the 5 jurisdictions (Singapore and South Korea) permit collection and use of personal data without the data subject's consent if the collection or use is in the legitimate interests of the organization. However, the relevant provision of South Korea's PIPA notably does not apply to sensitive personal data.

It would be possible for an organization to argue that the development or refinement of a generative AI system is in the legitimate interests of a developer.

However, under both laws, the organization would need to establish that this interest outweighs the rights and interests of the data subject. In the absence of guidance from data protection authorities, organizations may be reluctant to take the legal risk of relying on this interpretation. Evidence that the organization has implemented safeguards to prevent material harms to data subjects from the processing of their personal data, such as potentially defamatory AI-generated content, would likely help to bolster the organization's case that its interest outweighs the impact on the data subject.

Research Purposes: Of the data protection laws in 5 jurisdictions covered by this Report, only Japan's APPI provides an exception to consent requirements for collecting and using personal data for research purposes (the relevant provision of Singapore's PDPA applies only to use of personal data for this purpose).

Theoretically, a business could rely on this provision to process personal data to train a generative AI model provided that it collaborates with an academic institution, and one of the purposes for processing the personal data is academic research. However, these requirements may limit the value of this provision where a generative AI is trained for solely commercial purposes.

A detailed summary of relevant provisions in the data protection laws of the 5 jurisdictions is presented in the table below.

Jurisdiction	Summary of Relevant Provisions
Australia	The Privacy Act does not contain any provisions that specifically authorize the processing of personal data that is publicly available, or processing for legitimate interests or research purposes. Rather, in these situations, organizations would need to comply with the Privacy Act's requirements for collecting personal data from sources other than the data subject. In particular, the organization would have to establish that: » the data is reasonably necessary for one of its functions or activities (APP 3.1); » the collection of the personal data is by lawful and fair means (APP 3.5); » it is unreasonable or impracticable to collect personal data directly from data subjects (APP 3.6) If the personal data constitutes "**sensitive personal information**," APP 3.3 requires that the organization also obtain the data subject's **consent** for the collection of personal data. This requirement is subject to exceptions. However, none of these exceptions would generally apply to the use of personal data to train an AI model.

Jurisdiction	Summary of Relevant Provisions
China	The most relevant legal basis under the PIPL is **reasonable processing of publicly available personal data**. Article 13(6) of the PIPL permits data controllers to process publicly available personal data to a reasonable extent without the data subject's consent if: » the data subject personally disclosed the data; or » the data was otherwise legally disclosed. However, this provision is subject to two safeguards. Data controllers may not rely on this provision if: » the data subject **expressly refuses** the processing; or » processing of the publicly available data may have a **significant impact on an individual's rights and interests** (Article 27). In these cases, the data controller would have to seek **consent** from the data subject. Such consent is only valid if it is voluntarily given, explicit, and fully informed (Article 14). **Consent** is also required for the processing of **sensitive personal data** (Article 29). Additionally, the data controller must inform data subjects of why it is necessary to process such data, and what impact such processing may have on their rights and interests (Article 30).
Japan	For routine business uses of personal information, the default rule under the APPI is that businesses must **specify a legal purpose** for which they will use personal information (known as the "**purpose of use**") (Articles 17 and 19). Businesses must inform a data subject of the purpose of use either before or upon acquiring the data subject's personal information and must update the data subject if the purpose of use changes (Article 21). However, these requirements are subject to exceptions, including where informing the data subject of the purpose of use would harm the business's rights or legitimate interests or where the purpose of use is already clear in the circumstances. By default, **consent** is required for: » processing of personal information **beyond the scope necessary to achieve the purpose of use** (Article 18(2)); or » processing of "**sensitive personal information**" (Article 20(2)). However, there are **exceptions** to these requirements for **research purposes** and **publicly available data**. **Research purposes**: A business would not need to obtain consent for processing of sensitive personal information if: » the business: • obtains such information from an academic research institution; and • processes that information jointly with an academic research organization at least partially for the purposes of academic research, and » there is no risk that the processing will unjustly infringe on the data subject's rights and interests. **Publicly available data**: Additionally, the business would not need to comply with the consent requirements for processing sensitive personal information if that information is **open to the public** by a person identifiable by that information, a national government organ, a local government, an academic research institution, or other body permitted by regulations.

Jurisdiction	Summary of Relevant Provisions
Singapore	The PDPA authorizes organizations to process a data subject's personal data without consent if the processing satisfies the requirements for any of the **exceptions to consent** in the First and Second Schedules to the PDPA (Sections 13 and 17). Relevant exceptions to consent include: » Processing of personal data that is **publicly available**. » **Legitimate interests** (First Schedule, Part 3). **Publicly available data**: Part 2 of the First Schedule to the PDPA permits the collection, and use of "publicly available" personal data about an individual, without that individual's consent. Such data is considered "publicly available" if it is generally available to the public. This includes 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 (Section 2(1)). According to guidelines from Singapore's PDPC, organizations may rely on this exception if the personal data was publicly available at the time it was collected and do not need to verify whether the data is still publicly available at the time it is used (Advisory Guidelines on Key Concepts in the PDPA, paragraph 12.87).[71] **Legitimate interests**: Part 3 of the First Schedule to the PDPA permits an organization to collect and/or use personal data if the collection and/or use is in the legitimate interests of the organization or a third party. To rely on this provision, the organization must establish that the legitimate interest outweighs any adverse effect on the individual by: » conducting a risk assessment; and » implementing reasonable measures to address any risks of adverse effects identified in the assessment.
South Korea	The most relevant legal basis under the PIPA is **legitimate interests**. Article 15(1)(6) of the PIPA permits organizations to collect and use personal data without the data subject's consent if the collection and use is necessary to achieve a legitimate interest of the organization, and that legitimate interest clearly takes precedence over the data subject's rights. Additional safeguards apply to this provision. The collection and use must be significantly related to the legitimate interest of the organization and must be within a reasonable scope. If the organization is unable to rely on this provision, it would have to obtain **consent** from the data subject. **Consent** would also be required for collection and use of **sensitive personal data** (Article 23).

Collecting Data From End-Users of Generative AI Applications to Refine the Underlying Model

Generative AI applications may collect data from prompts given to the system, which are then used to further train and refine the system. Some of these prompts may contain personal data, in which case, the application would be collecting personal data for a specific purpose and so, would be subject to the obligations of a "data controller" (or equivalent) under relevant data protection laws.

Compared with the previous scenario, where data is scraped from publicly available websites, there would be fewer issues with obtaining consent from users because the organizations would have a relationship with data subjects who use its services.

The table below presents a detailed summary of relevant legal bases for processing personal data in this scenario under the data protection laws of the 5 jurisdictions covered by this Report.

Notably, several data protection authorities' enforcement decisions against OpenAI specifically addressed this scenario (see above). These decisions all emphasized the need to obtain informed consent from users for the collection and use of personal data from prompts to further train a generative AI model and provide data subjects with the right to opt out of such collection and use.

Further, the Italian *Garante*'s preliminary order also found that OpenAI could **not** rely on a legal basis under the GDPR which allows the processing of personal data without consent, where that processing is **necessary for the performance of a contract**.

Jurisdiction	Summary of Relevant Provisions
Australia	In situations where organizations collect personal data directly from data subjects, APP 3.1 requires the organization, before collecting personal data, to establish that the data is reasonably necessary for one of its functions or activities. The collection of the personal data must also be by lawful and fair means (APP 3.5). **Consent for collection of sensitive personal data**: APP 3.3 requires that if the personal data constitutes "sensitive personal information," the organization must also obtain the data subject's **consent** for the collection of personal data. This requirement is subject to exceptions, but none of these exceptions would generally apply to use of personal data to train an AI model.
China	The most relevant legal basis under the PIPL is **consent**. In this situation, a data controller could rely on **consent** (Article 13(1). In order to be valid under the PIPL, the consent must be voluntarily given, explicit, and fully informed (Article 14). **Consent** is also required for processing of **sensitive personal data** (Article 29). Additionally, the data controller must inform data subjects of why it is necessary to process such data, and what impact such processing may have on their rights and interests (Article 30).
Japan	For routine business uses of personal information, the default rule under the APPI is that businesses must **specify a legal purpose** for which they will use personal information (known as the "**purpose of use**") (Articles 17 and 19). Businesses must inform a data subject of the purpose of use either before or on acquiring the data subject's personal information and must update the data subject if the purpose of use changes (Article 21). However, these requirements are subject to exceptions, including where informing the data subject of the purpose of use would harm the business's rights or legitimate interests or where the purpose of use is already clear in the circumstances. By default, **consent** is required for: » processing of personal information **beyond the scope necessary to achieve the purpose of use** (Article 18(2)); or » processing of "**sensitive personal information**" (Article 20(2)). PPC Japan's "Notice regarding Cautionary Measures on the Use of Generative AI Services" highlights that under Japanese data protection law, service providers should: » provide collection notices with a clear statement of the purpose(s) for which the data is collected and processed and » obtain consent from users before processing their sensitive personal information.

Jurisdiction	Summary of Relevant Provisions
Singapore	The most relevant legal bases under the PDPA are **consent** and **legitimate interests**. The PDPA authorizes organizations to collect a data subject's personal data if the organization obtains **consent** or if the collection satisfies the requirements for any of the exceptions to consent in the First and Second Schedules to the PDPA (Sections 13 and 17). **Consent**: In this situation, an organization could rely on express consent (Section 14). However, the PDPA also permits consent to be deemed under certain circumstances. A relevant circumstance to this situation is **deemed consent by notification** (Section 15A). To rely on this provision, an organization must take reasonable steps to bring to the individual's attention: » the organization's intention to process the data subject's personal data; » the purpose for which the organization will process the data; and » a reasonable period and procedure for the data subject to object to the proposed processing. The organization must also conduct an **impact assessment** to determine the likely impact of the processing on the data subject, and take steps to mitigate potential risks. **Legitimate interests**: Part 3 of the First Schedule to the PDPA permits an organization to collect and/or use personal data if the collection and/or use is in the legitimate interests of the organization or a third party. To rely on this provision, the organization must establish that the legitimate interest outweighs any adverse effect on the individual by: » conducting a risk assessment; and » implementing reasonable measures to address any risks of adverse effects identified in the assessment.
South Korea	The most relevant legal bases under the PIPA are **consent**, and **legitimate interests**. **Consent**: Article 15(1) of the PIPA permits organizations to collect and use personal data if they obtain consent from the data subject for such collection and use. Under Article 15(2) of the PIPA, when obtaining consent, the organization must inform the data subject of: » The purpose for the collection and use of the personal data. » Details of the personal data that will be collected. » The period during which the data will be retained and used. » The data subject's right to withhold consent, and the consequence of exercising the right. Under Article 15(3) of the PIPA, once the organization has obtained consent, it may use the personal data for any purpose which is within the scope reasonably related to the initial purpose for which the data was collected. **Legitimate interests**: Article 15(1)(6) of the PIPA permits organizations to collect and use personal data without the data subject's consent if the collection and use is necessary to achieve a legitimate interest of the organization, and that legitimate interest clearly takes precedence over the data subject's rights. Additional safeguards apply to this provision. The collection and use must be significantly related to the legitimate interest of the organization and must be within a reasonable scope.

REUSE OF EXISTING DATASETS

This situation assumes that an organization has legally collected personal data for a specific purpose (the **primary purpose**) but intends to use this data for a new purpose (**secondary purpose**).

In this scenario, the organization would need to ensure that it has the legal authority to use the personal data for the secondary purpose of training a generative AI model. This would depend on the legal basis relied upon to collect and use the data for the primary purpose.

It is reasonable to assume that in a business context, the legal basis to process the data for the primary purpose would likely be consent. If so, it is possible that the organization may be able to rely on this consent, if the secondary purpose is within the scope of or closely related to the primary purpose.

In other cases, the organization would have to:

» obtain **fresh consent**;

» fulfill the requirements for an alternative legal basis or exception to consent, such as **legitimate interests**; or

» anonymize the data to take it out of the scope of data protection law.

The table below presents a detailed summary of relevant legal bases for processing personal data in this scenario under the data protection laws of the 5 jurisdictions covered by this Report.

Compared with the scenario of a training dataset from a web crawl, it may be easier for the organization to obtain fresh consent, as the organization may already have established communication channels with data subjects when it sought consent to use the personal data for the primary purpose.

Jurisdiction	Summary of Relevant Provisions
Australia	The Privacy Act has specific requirements for secondary use of personal data. Specifically, APPs 6.1 and 6.2 require that if an organization holds personal data that was collected for a primary purpose, the organization may only use that data for a secondary purpose if: » the data subject **consents** to the use of their personal data for a secondary purpose, or » the secondary purpose is (directly*) **related to the primary purpose**, if the data subject would **reasonably expect** the organization to use the personal data for the secondary purpose. * For sensitive personal information.
China	Data controllers may be able to **rely on the original consent** if one of the stated purposes for processing included training of generative AI models. If not, data controllers would likely have to obtain **fresh consent** pursuant to Article 14 of the PIPL, which requires data controllers to obtain fresh consent if the purpose for processing personal information changes.
Japan	By default, Article 18(2) of the APPI requires a business to obtain the data subject's **consent** to process personal information for a secondary purpose unless such processing is within the scope necessary to achieve the primary purpose. However, this is subject to an **exception for academic research**. Businesses do not need to obtain consent to process personal data for a secondary purpose if that secondary purpose at least partially includes the purpose of academic research. To rely on this exception, the business would also have to provide the personal information to an academic research institution (or equivalent) for processing and ensure that there is no risk that the processing will unjustly infringe on the data subject's rights and interests.

Jurisdiction	Summary of Relevant Provisions
Singapore	In order to use a data subject's personal data for a secondary purpose, an organization could rely on the data subject's **consent** or a relevant exception to consent. Relevant exceptions to consent in this context may include: » **legitimate interests**; » **business improvement purposes**; and » **research purposes**. **Consent for secondary purposes**: After an organization has obtained consent to collect a data subject's personal data for a primary purpose, the organization must **notify the data subject of a secondary purpose** before using the personal data for that secondary purpose (Sections 18 and 20). If the organization collected the personal data without the data subject's consent, then the organization must either obtain fresh consent to use the personal data for a secondary purpose, or satisfy an exception to consent for use of the data. **Legitimate interests**: Part 3 of the First Schedule to the PDPA permits an organization to use personal data if the use is in the legitimate interests of the organization or a third party. To rely on this provision, the organization must establish that the legitimate interest outweighs any adverse effect on the individual by conducting a risk assessment and implementing reasonable measures to address any risks of adverse effects identified in the assessment. **Business improvement purposes**: Division 2, Part 2 of the Second Schedule to the PDPA permits an organization to use personal data for various "business improvement purposes" including improving or enhancing goods and services and developing new goods and services. To rely on this provision, the organization must establish that: » the purpose for processing cannot reasonably be achieved without the use of the personal data in an individually identifiable form; and » a reasonable person would consider the use of the personal data for that purpose to be appropriate in the circumstances. **Research purposes**: Division 3, Part 2 of the Second Schedule to the PDPA permits an organization to use personal data for a research purpose if the following conditions are met: » the research purpose cannot reasonably be accomplished unless the personal data is used in an individually identifiable form; » there is a clear public benefit to using the personal data for the research purpose; » the results of the research will not be used to make any decision that affects the individual; and » in the event that the results of the research are published, the organization publishes the results in a form that does not identify the individual.
South Korea	The most relevant legal bases under the PIPA are **consent** and **legitimate interests**. **Consent for secondary purposes**: The organization may rely on the original consent to the extent that the secondary purpose is within the scope of the primary purpose. If the secondary purpose is outside of the scope of the primary purpose, the organization would have to obtain fresh consent pursuant to Article 18 of the PIPA. **Legitimate interests**: Article 15(1)(6) of the PIPA permits organizations to collect and use personal data without the data subject's consent if the collection and use is necessary to achieve a legitimate interest of the organization, and that legitimate interest clearly takes precedence over the data subject's rights. Additional safeguards apply to this provision. The collection and use must be significantly related to the legitimate interest of the organization and must be within a reasonable scope.

Data Protection Principles

DATA MINIMIZATION

Data minimization is a commonly found principle in data protection laws internationally that pertains to the limitation of collection and use of personal data to only what is necessary for a specified purpose.

However, as noted previously, training generative AI models often requires large datasets to learn patterns and generate realistic outputs. These datasets, especially those obtained through "web crawls", may contain significant amounts of personal data which is not strictly necessary for the training of the system, but may be difficult to remove from the dataset.

The principle of data minimization is found in some form in the data protection laws of all 5 jurisdictions covered by this report. It is stated explicitly in the data protection laws of China and South Korea and is implicit in the laws of Australia, Japan, and Singapore.

Jurisdiction	Summary of Relevant Provisions
Australia	The Privacy Act does not expressly recognize the principle of data minimization. However, collection of personal data is subject to a standard of reasonable necessity or relevance. Under APPs 3.2 and 3.3, an organization may only collect personal data if the data is reasonably necessary for, or directly related to, one or more of the organization's functions or activities.
China	Article 6 of the PIPL requires that the collection of personal data must be limited to the smallest scope necessary to achieve the purpose for processing the data. This provision also expressly prohibits excessive collection of personal data.
Japan	The APPI does not expressly recognize the principle of data minimization. However, Article 18(2) of the APPI prohibits businesses from processing personal data beyond the scope necessary to achieve the **purpose of use**, unless they obtain the data subject's consent in advance or satisfy other conditions (see above).
Singapore	The PDPA does not expressly recognize the principle of data minimization. However, Section 18 of the PDPA, which limits collection of personal data for purposes that a reasonable person would find inappropriate, may prevent excessive collection of personal data to some extent.
South Korea	Articles 3(1) and 16(1) of the PIPA require controllers to collect the minimum personal data necessary to fulfill the purpose for processing the data. Further, Article 3(6) of the PIPA requires controllers to minimize the possibility of infringing data subjects' privacy when processing their personal data.

PURPOSE LIMITATION

The principle of purpose limitation requires that personal data can only be collected for specified, explicit, and legitimate purposes and not further processed in a manner incompatible with those purposes.

When training generative AI models on large datasets containing personal data, that personal data may have been collected for other original purposes unrelated to AI training. This discrepancy may complicate compliance with data protection law, as repurposing data for AI training may be deemed incompatible with the initial purpose, potentially requiring additional legal bases, such as consent (see above).

Data protection laws in all 5 of the jurisdictions expressly recognize this principle.

Jurisdiction	Summary of Relevant Provisions
Australia	The Privacy Act implements the principles of purpose limitation in two major respects. Firstly, collection of personal data is limited to purposes which relate to an organization's functions or activities. Under APPs 3.2 and 3.3, an organization may only collect personal data if the data is reasonably necessary for, or directly related to, one or more of the organization's functions or activities. Secondly, an organization may only use or disclose the data for a primary purpose and must satisfy certain conditions before it may use or disclose the data for any other purpose. Specifically, APPs 6.1 and 6.2 require that if an organization holds personal data that was collected for a primary purpose, the organization may only use that data for a secondary purpose if: » the data subject **consents** to the use of their personal data for a secondary purpose, or » the secondary purpose is (directly*) **related to the primary purpose**, if the data subject would **reasonably expect** the organization to use the personal data for the secondary purpose. * For sensitive personal information.
China	Articles 5 and 6 of the PIPL require that processing of personal data must have a clear and reasonable purpose, be directly related to that purpose, and should use a method that has the minimum impact on data subjects' rights and interests.
Japan	The APPI implements the principle of purpose limitation by requiring businesses to identify a **purpose of use** for personal data (Articles 17 and 19). Businesses must obtain the data subject's consent or satisfy other conditions (see above) before using the data for any purpose that is beyond the scope necessary to achieve the purpose of use (Article 18(2)).
Singapore	The PDPA expressly recognizes the principle of purpose limitation. Section 18 of the PDPA only permits organizations to collect, use or disclose personal data about an individual only for purposes that: » a reasonable person would consider appropriate in the circumstances; and » the individual has been informed of, if applicable.
South Korea	Article 3(1) of the PIPA requires data controllers to identify the purpose for processing personal data. Articles 3(2) and 18(1) require controllers to process personal data in an appropriate manner to the extent necessary to fulfill that purpose and not use the data beyond such purposes. Further, Article 3(6) of the PIPA requires controllers to minimize the possibility of infringing data subjects' privacy when processing their personal data.

FAIRNESS

The data protection principle of fairness requires that personal data be processed in a way that is fair and lawful, and respects individual rights.

As discussed in Section 1, when training generative AI models on large datasets, there are risks that the datasets contain biases, inaccuracies, or underrepresentation of certain demographics that may lead these systems to produce output that is biased, discriminatory, or toxic. Such output would certainly contravene the principle of fairness in data protection.

However, in practice, ensuring fairness becomes very complex when using massive datasets, especially those obtained through internet scraping to train generative AI systems. Further, the scale of such datasets may make it challenging to ensure that all data has been collected fairly and lawfully.

Data protection laws in all 5 of the jurisdictions contain some form of requirement that processing of personal data must be fair. While South Korea's PIPA expressly requires fairness in personal data processing, Australia, China, Japan, and Singapore all have implied fairness requirements based on provisions on lawful data collection, good faith, respect for autonomy, and reasonableness standards.

Jurisdiction	Summary of Relevant Provisions
Australia	The Privacy Act does not expressly recognize the principle of fairness. However, under APP 3.5, an organization may only collect personal data by means that are fair and lawful.
China	While the PIPL does expressly recognize the principle of fairness, it only applies this principle to data controllers who provide important internet platform services involving a huge number of users and complicated business types (Article 58). However, the principle of fairness is implicit in other general provisions of the PIPL. Specifically, Article 5 of the PIPL requires that the processing of personal data should be undertaken in good faith and should not involve vitiating factors, such as misrepresentation, fraud, or coercion. Further, Article 6 of the PIPL requires that personal data should be processed in a manner that has the minimum impact on data subjects' rights and interests.
Japan	The APPI does not expressly recognize the principle of fairness. However, Article 3 of the APPI provides a basic principle that personal data should be processed prudently and with respect for the autonomy of data subjects. Further, Article 19 of the APPI prohibits businesses from using personal data in any way that could provoke or induce an unjust act, and Article 20(1) of the APPI prohibits businesses from acquiring personal data by deception or other wrongful means.
Singapore	The PDPA does not expressly recognize the principle of fairness. However, Section 18 of the PDPA subjects purposes for processing personal data to a reasonableness standard. This may serve to prohibit unfair uses of personal data.
South Korea	Article 3(1) of the PIPA requires controllers to collect personal data fairly. Further, Article 3(6) of the PIPA requires controllers to minimize the possibility of infringing data subjects' privacy when processing their personal data.

Personal Data Breaches

Where generative AI models have been trained on datasets containing personal data, these models may generate content that discloses personal data in ways that may cause material or mental harm to data subjects.

For instance, in September 2022, a California-based AI artist found that photographs from her private medical records had been included in a training dataset that was scraped from the internet and had been used to train several image generation models, including Stable Diffusion.[72]

This risk arises from certain features of the transformer architecture which powers many generative AI models today.[73] These models learn by exposure to large datasets and capture the statistical patterns present in the data. In doing so, the model may "**memorize**" information that it was trained on, meaning that the model reproduces specific phrases, sentences, or even longer passages from its training data.[74]

Data protection laws generally require data controllers to secure personal data that is within their control. However, the nature of foundational models raises unique issues, as they may repeat personal data from their training datasets due to the "memorization" issue (see above), either through unintended operation of the system or in response to a malicious prompt that exploits a vulnerability in the AI system. This may lead to unintended disclosure of personal data.

Jurisdiction	Summary of Relevant Provisions
Australia	APP 11 requires organizations to take reasonable steps to: » protect the personal data that they hold from misuse, interference, loss, or unauthorized access, modification, or disclosure; and » proactively delete or de-identify personal data they hold, if data is no longer necessary for any purpose for which it was processed (subject to exceptions for certain legal obligations).
China	The PIPL outlines several operational measures that data controllers must implement to prevent unauthorized access to, breach, tampering or loss of any personal data (Article 51).
Japan	Article 23 of the APPI requires businesses to take necessary and appropriate measures to manage the security of personal data, including preventing leaks, loss, or damage.
Singapore	Section 24 of the PDPA requires organizations to protect personal data in their possession or under their control by, among other provisions, making reasonable arrangements to prevent unauthorized access, collection, use, disclosure, copying, modification or disposal, or similar risks.
South Korea	Article 29 of the PIPA requires controllers to adopt such technical, managerial, and physical measures as are necessary to ensure the safety of personal data and prevent the loss, theft, unauthorized disclosure, forgery, alteration of, or damage to, the data.

Data breaches may also trigger obligations to notify applicable data protection authorities and data subjects. In the latter case, the issue of lack of individualized relationship between the AI operator and the user may create compliance and enforcement difficulties.

Jurisdiction	Summary of Relevant Provisions
Australia	An organization is required to prepare a statement to the OAIC as soon as practicable after discovering an "eligible data breach," (Section 26WK), i.e., an unauthorized access to, or disclosure or loss of personal data, and a reasonable person would conclude that this is likely to result in serious harm to the data subject (Section 26WE). The organization must then notify affected data subjects as soon as practicable after making the statement to the OAIC (Section 26WL). These requirements are subject to exceptions.
China	Article 57 of the PIPL requires organizations to immediately adopt remedial measures and notify the CAC and affected data subjects in the event that a leak, distortion, or loss of personal data has, or might have, occurred. Organizations are permitted not to notify affected data subjects if measures to address the data breach are effective in mitigating harm to data subjects.
Japan	Article 26 of the APPI requires businesses to notify the PPC of any incident involving the security of personal data if the incident is likely to cause harm to the data subject's rights and interests. According to the PPC Order, this notification must be given within 3 to 5 days.[75] Businesses must also "promptly" inform affected data subjects of the breach, unless it is difficult to do so, and the business has implemented necessary measures to protect the data subjects' rights and interests. The PPC has not provided further clarification on the timelines for notifying data subjects.
Singapore	Section 26D of the PDPA requires organizations to notify the PDPC within 3 calendar days of assessing that a data breach has occurred and: » results in, or is likely to result in, significant harm to an affected individual; or » is, or is likely to be, of a significant scale. According to guidelines issued by the PDPC, organizations are expected to complete the above assessment within 30 calendar days (Advisory Guidelines on Key Concepts in the PDPA, paragraph 20.4).[76] Organizations must also notify affected data subjects of the breach in any manner that is reasonable in the circumstances. This requirement is subject to exceptions. In particular, organizations are not required to notify affected data subjects if the organization implemented measures prior to the breach that would render it unlikely that the breach would result in significant harm to the data subject.

Jurisdiction	Summary of Relevant Provisions
South Korea	Article 34 of the PIPA requires controllers to notify the PIPC and affected data subjects "without delay" in the event of a data breach. This requirement does not appear to be subject to exceptions. According to guidelines issued by the PIPC, controllers should notify the PIPC and/or the Korea Internet and Security Agency within 72 hours if the breach involves the personal data of 1,000 or more data subjects, sensitive personal data or unique identifiers, or illegal and unauthorized access to personal data.[77]

Quality of Data

Data protection laws in all 5 of the jurisdictions covered in this Report require organizations to maintain the quality of personal data.

An important consideration in complying with data protection laws in the context of generative AI is that data scraped from the internet may contain personal data that is inaccurate.

Relying on this data and using it for further data processing may conflict with obligations under applicable data protection laws to ensure that personal data is accurate and up to date.[78] For instance, one of the grounds on which the *Garante* temporarily banned ChatGPT in Italy was that it processed inaccurate personal data in violation of Article 5 of the GDPR.[79] One of the *Garante*'s conditions for lifting the temporary ban was that OpenAI provide a tool for data subjects to request rectification or deletion of their data.[80]

Jurisdiction	Summary of Relevant Provisions
Australia	APP 10 requires organizations to take reasonable steps to ensure that: » the personal data that they collect is **accurate**, **up-to-date**, and **complete**; » the personal data that they **use** is accurate, up-to-date, complete, and relevant, having regard to the purpose for which the data will be used.
China	Article 8 of the PIPL requires data controllers to ensure the quality of personal data and avoid adverse impacts on the rights and interests of individuals caused by inaccurate and incomplete personal data.
Japan	Article 22 of the APPI requires businesses to endeavor to keep the content of personal data accurate and up to date, within the scope necessary to achieve the purpose of use.
Singapore	Section 23 of the PDPA requires organizations to make reasonable efforts to ensure that the personal data that they collect is accurate and complete if the organization is likely to use the data to make decisions that affect the data subject or disclose the personal data to another organization.
South Korea	Article 3(3) of the PIPA requires a controller to ensure that personal data is accurate, complete, and up to date to the extent necessary in relation to the purposes for which the personal information is processed.

Rights to Modification and Erasure of Personal Data

All 5 jurisdictions minimally recognize the rights to access and correction of personal data. A further 3 (China, Japan, and South Korea) also provide a right to erasure.

However, giving effect to these rights may be challenging in the context of generative AI.

From a technical perspective, once personal data has been input into generative AI models, effectively managing and tracking its usage becomes a complex, if not challenging, task due to how generative AI systems process information and store/replicate data across various systems.[81]

From a legal compliance perspective, where a generative AI model was trained on a web crawl dataset, it may also be challenging to give effect to the rights of potentially millions (if not billions) of data subjects whose personal data is included in these datasets.

SECTION 3

Summary of Findings and Key Takeaways for APAC

The regulatory landscape for generative AI in APAC is changing at a fast pace. However, based on an analysis of the existing state of generative AI governance in 5 key APAC jurisdictions, this Report has identified some important considerations for policymakers and for deployers and developers of generative AI in the APAC region. Below, we distill the key takeaways for these stakeholders, taking into account the generative AI-specific frameworks, documents, and guidance discussed in Section 1 and detailed in the Appendix.

Takeaways for Policymakers

Takeaway 1: Alignment and interoperability are needed to counter potential policy fragmentation across the region.

A core finding of the Report is that notwithstanding commonalities in certain aspects, **there is a lack of a coordinated approach to generative AI policy** both **within and between** the 5 jurisdictions covered by this Report. This is perhaps unsurprising given the diversity in these 5 jurisdictions (as in the wider APAC region), as well as the lack of mechanisms for supranational coordination compared with other regions, such as Europe.

However, if policymakers continue to develop frameworks to govern generative AI within legislative and national silos, there is a **risk of fragmentation** in the development of these frameworks. Such fragmentation may increase the costs and complexity of compliance across jurisdictions in APAC. This may in turn hinder investment in, and adoption of, potentially valuable technologies at scale, preventing society from reaping the benefits in productivity and innovation from these technologies. It may also create a situation where levels of personal data protection are inconsistent across jurisdictions in the APAC region.

This Report has also identified:

- **A lack of regulatory certainty** (in some areas) **around how existing frameworks apply to AI systems**. The extent to which these laws and rules apply to AI systems is often a matter of legal interpretation, in need of specific regulatory guidance particularly where there are tensions between the nature of processing personal data through Generative AI systems and existing rules.
- **Lack of coordination between legal frameworks (within *and* between jurisdictions)**. Where multiple laws and rules apply to the same issue, there is a risk that their requirements may overlap or even contradict one another. This may create further legal uncertainty, as it may not be clear which laws apply or take precedence in the event of a conflict, or unnecessary layers of regulation.
- **Inconsistency in regulatory responses**. Regulators may not have the same powers to address AI systems that fall within their mandate. This may result in different, and possibly conflicting, responses in different sectors.

It is therefore important for **policymakers to ensure alignment and interoperability** with other leading international frameworks when crafting regulatory responses to generative AI.

Most jurisdictions covered in this Report share the same fundamental aims and have been adopting an incremental approach to AI governance premised on voluntary guidance and consultations. There appears to be an emerging consensus around the risks posed by existing generative AI systems and measures to address them. This emerging consensus could form the basis for regional and international discussions. There are also emerging frameworks at the international level, such as the **G7's Hiroshima AI Process Comprehensive Policy Framework**, that could aid these discussions.

During FPF's roundtables for this project, several stakeholders emphasized the need for a **common taxonomy** of key terms like "generative AI," "foundation models," and "large language models," that aligns with established regional and global definitions. In this regard, policymakers can benefit from **aligning terminology with emerging international standards** that are being developed in fora including the International Standards Organization (ISO), the Institute of Electrical and Electronics Engineers Standards Association (IEEE SA), the Organization for Economic Cooperation and Development (OECD), and the G7, and encouraging use of standardized terms by all stakeholders. Doing so will aid the establishment of robust standards, guidelines, and regulations for different applications of generative AI.[82]

Takeaway 2: Guidance on the application of existing laws to generative AI should be provided to support legal certainty.

In the absence of AI-specific regulation, **existing technology-neutral laws will continue to be the main source of legal obligations that govern generative AI**. In particular, data protection law plays a major role as all jurisdictions have enacted such laws, and the presence of personal data in training datasets used to train current generative AI models, combined with the consumer-facing nature of many generative AI models, provides data protection authorities with a regulatory lever to govern generative AI.

However, as these laws were not drafted with generative AI in mind and pre-date the current generative AI boom, it would be beneficial if relevant authorities could provide **guidance on how these laws apply to generative AI**. This recommendation is especially relevant to data protection law, as **Section 2 of this Report** has identified several areas of potential ambiguity in how existing data protection laws apply to generative AI systems.

Further, while some DPAs in the 5 jurisdictions have begun issuing guidance on the application of data protection law to certain kinds of AI systems (such as recommendation and decision-making systems[83]), to date, only DPAs in Japan and South Korea have issued guidance on the application of their respective data protection laws to generative AI specifically, and this guidance is still preliminary.

In developing this guidance, and despite the leadership of several data protection authorities, collaboration among relevant regulators within a jurisdiction, including any government body with regulatory authority that may be relevant to generative AI, is essential to ensure that each jurisdiction's approach to generative AI governance is consistent across regulatory domains and avoids the risk of regulatory fragmentation within that jurisdiction. Ideally, relevant regulators should coordinate on priorities and approaches and work together to identify potential gaps in existing frameworks, proposing targeted reforms as necessary.

In particular, it may be helpful to look not only for gaps in existing frameworks but also overlaps where multiple legislative or regulatory frameworks may govern the same issue. Such overlaps may complicate compliance, especially if requirements are contradictory.

Takeaways for Industry including Developers and Deployers of Generative AI Systems

As of early 2024, there is an emerging body of voluntary guidance issued from the 5 APAC jurisdictions studied, outlining good practices that developers and deployers of generative AI could consider adopting in their approaches to govern this technology.

This subsection of the Report summarizes these practices, building on the commonalities identified in **Section 1** and serving as a resource for developers and deployers of generative AI systems thinking through generative AI governance in APAC or comparing existing approaches in APAC with legally binding requirements in the EU and US.

Takeaway 3: All five jurisdictions recognize developing internal AI governance and risk management policies as a good practice.

As shown from our survey of early regulatory responses to generative AI outlined in Section 1, policymakers in APAC have highlighted that a good practice before developing or deploying a generative AI system is to design a robust internal AI policy and strategy to encourage the organization to foster a culture of responsible innovation.

Existing AI governance frameworks in the 5 jurisdictions point to the following as relevant factors to consider:

- Assessing the organization's AI proficiency.[84]
- Setting governance principles and goals.[85]
- Integrating ethical guidelines, risk management protocols, and compliance measures.[86]
- Ensuring compliance with existing laws and guidelines.[87]
- Conducting risk and impact assessments to systematically evaluate potential harms to guide mitigation efforts.[88]
- Documenting risk and impact assessments to facilitate transparency and build organizational accountability.[89]
- Clearly allocating responsibilities within the organization, including potentially establishing an AI taskforce or committee to coordinate efforts.[90]
- Training employees in the design, function and implementation of AI systems.[91]
- Regularly reviewing governance structures and measures to ensure alignment with objectives and address evolving risks.[92]

Takeaway 4: Effective governance is essential to mitigate model bias and discriminatory outputs from generative AI systems.

When developing generative AI systems, organizations must prioritize data governance, including implementing good data practices, evaluating training data sources, and evaluating model output for representativeness. Identifying potential biases in the model is key to mitigate against the risk of discriminatory or harmful output.

Existing AI governance frameworks in the 5 jurisdictions have highlighted the following potential measures to mitigate bias and prevent or discriminatory outputs from generative AI systems:

- Thoroughly evaluating training data sources for representativeness and potential biases.[93]
- Documenting data provenance to enable traceability and accountability.[94]
- Regularly auditing data quality across dimensions like accuracy, completeness, and relevance.[95]
- Proactively conducting bias assessments and ethical reviews of training data.[96]
- Moderating and redacting problematic content from training data.[97]
- Fine-tuning models after initial training to reduce harmful outputs.[98]
- Employ output filtering techniques to catch and block biased generations.[99]
- Leverage bias detection tools during data preprocessing.[100]
- Continuously monitoring and updating datasets with human oversight.[101]

Takeaway 5: Ensuring privacy by design in the development and deployment of generative AI systems can build public trust.

When developing underlying generative AI models and deploying generative AI-based systems and applications, organizations can benefit from adopting a "Privacy by Design" approach that builds in data protection safeguards from the earliest stages and at regular intervals thereafter. This can help build public trust in the technology.

Potential privacy-preserving measures identified or recommended in generative AI-specific policy documents across the five jurisdictions include:

- Minimizing collection and use of personal data.[102]
- Redacting or anonymizing personal data in training datasets.[103]
- Conducting Data Protection Impact Assessments (DPIAs).[104]
- Ensuring legal compliance for data collection and usage.[105]
- Developing and publishing a privacy policy to address the organization's use of personal data in training an AI model.[106]
- Obtain informed consent from users and provide mechanisms for users to opt out of collection or use of their personal data to train generative AI models.[107]

A limited yet important privacy enhancing technique that can offer an alternative to personal data to train generative AI models is the use of **synthetic data**: artificial data generated from original data by an AI model that has been trained to reproduce the characteristics and structure of the original data.[108]

Such data can potentially be used for pre-training, fine-tuning, and testing AI models,[109] and preliminary research has found that models trained on synthetic data achieved over 90% of the quality of models trained on real datasets.[110]

According to the Confederation of European Data Protection Organisations (CEDPO), potential benefits of synthetic data include enhanced privacy by minimizing the use of personal data, better data quality through "near-perfect" labeling, reduced costs, and reduced cybersecurity attack surfaces. However, synthetic data is not synonymous with anonymous data and carries a risk of reidentification.[111] The use of other Privacy Enhancing Technologies, such as differential privacy, in combination with synthetic data, could mitigate the risks of reidentification but may not completely remove it.[112]

Takeaway 6: Implementing safety and security measures is paramount for safer generative AI systems.

Implementing appropriate safety and security measures helps to ensure that generative AI systems are used safely and responsibly, minimizing the potential for misuse or harm to users and third parties.

Existing AI governance frameworks in the 5 jurisdictions have highlighted the following potential safety measures:

- Conducting risk and impact assessments, prompt testing and design, and ongoing evaluation.[113]
- Adding friction points, such as educative prompts or inappropriate content detection, when users attempt to generate content.[114]
- Implementing age-appropriate design with effective age verification measures, limiting content generation for underage users to age-appropriate material.[115]
- Implementing policies and processes to detect malicious actors or harmful content, testing models for potential misuse and putting safeguards in place to prevent harmful content generation.[116]

They also highlight the following potential security measures:

- Engaging in thorough testing and evaluation processes to mitigate risks.[117] This could include voluntary certifications, audits, and third-party assessments,[118] as well as "crowdsourcing" the detection of vulnerabilities in open-source models.[119]
  - Testing and evaluation processes could also include "**red teaming**"[120] – a practice where an authorized security team pretends to be attackers and tries to break into an organization's systems to test its security.[121]
- Establishing channels to share information regarding risks and best practices, including incident reporting.[122]
  - This includes, but is not limited to, complying with notifiable data breach requirements in data protection laws (see above).

Takeaway 7: All five jurisdictions recognize that providing meaningful transparency in the development and deployment of generative AI systems is essential.

As discussed in Section 1, transparency is a multi-faceted concept that is closely related to accountability. In particular, it allows scrutiny of potential harms, calibrates user expectations, and ultimately nurtures public trust in generative AI technologies.

Existing AI governance frameworks in the 5 jurisdictions have highlighted the following potential measures to facilitate meaningful transparency in the development and deployment of generative AI systems:

- Publishing clear organizational policies covering user safety, privacy, terms of use, content guidelines, and impact assessments.[123]
- Providing notices on data collection purposes, factual inaccuracies, and dissuading sharing of personal and/or confidential information.[124]
- Enhancing context-appropriate explainability and interpretability to clarify how models function and arrive at outputs. This could include:
  - Maintaining comprehensive documentation on data provenance, design choices, training procedures, performance metrics, and ethical evaluations.[125]
  - Utilizing model cards, system cards, and value alignment cards to present technical details in an accessible manner.[126]
  - Clarifying models' capabilities, limitations, and intended or prohibited uses.[127]
- Disclosing transparency reports.[128]
- Providing mechanisms for stakeholders to request further information, provide feedback, and seek redress.[129]

While full transparency may be impossible, an important consideration is ensuring that appropriate explanations are tailored to the needs of different stakeholders, which may include regulators, downstream providers of services that use generative AI models, and end-users.[130]

In addition, technical explanations of algorithms may not be useful to the general public. It may be helpful instead to focus on real-world impacts rather than solely technical inner workings to improve meaningful consent.

Takeaway 8: Indicating that content is AI-generated and enabling traceability are unanimously included in the generative AI frameworks studied.

Policymakers in the 5 jurisdictions were unanimous in highlighting the need for mechanisms to enable stakeholders, including regulators and the general public, to identify content as AI-generated.[131] This is closely related to transparency but also has implications for safety and security.

Industry is already working on technology to embed digital labels or watermarks in AI-generated content indicating that the content was generated by their system. For instance, the Coalition for Content Provenance and Authenticity (C2PA) is developing an open industry standard using cryptography to embed digital signatures and ownership details into AI-generated content.[132]

However, solutions like these may not be suitable for AI-generated text, since text can be more easily separated from metadata. While statistical watermarking and other techniques for text are emerging,[133] this area is still in early development.[134]

Noting that the technology to accomplish this is still at an early stage of development, organizations could consider implementing measures to make AI-generated outputs detectable, such as:

- Digital labeling or watermarking indicating AI-generated provenance, whether visible markers or embedded metadata.[135]
- Exploring statistical watermarking techniques tailored for text data.[136]
- Coordinating efforts to imprint subtle "fingerprints" in training data or model architectures that enable detection of AI-generated output.

APPENDIX

The Report's Appendix can be accessed electronically by scanning the QR code below:

NOTES

ENDNOTES

1 See, for example, *The Spectrum of Artificial Intelligence – An Infographic Tool* (December 14, 2020), available at https://fpf.org/blog/the-spectrum-of-artificial-intelligence-an-infographic-tool/, *The Spectrum of AI – Companion to the FPF AI Infographic* (updated July 18, 2023), available at https://fpf.org/blog/newly-updated-report-the-spectrum-of-artificial-intelligence-companion-to-the-fpf-ai-infographic/, and *Generative AI for Organizational Use: Internal Policy Checklist* (July 13, 2023), available at https://fpf.org/resource/fpf-releases-generative-ai-internal-policy-checklist-to-guide-development-of-policies-to-promote-responsible-employee-use-of-generative-ai-tools/

2 The **Appendix to this Report** includes brief summaries of the EU's AI Act and the US' Executive Order on the Safe, Secure, and Trustworthy Development of Artificial Intelligence.

3 In November 2022, OpenAI launched ChatGPT, a consumer-facing generative AI tool that can create predictive content based on natural language input. Since then, a number of competitors have emerged in the space in both the business-to-consumer and business-to-business markets, as well as a number of more sophisticated generative AI tools.

4 https://www.ibm.com/topics/ai-model

5 Jebara, T. (2012). *Machine learning: discriminative and generative* (Vol. 755). Springer Science & Business Media.

6 Weidinger, L. et al. (2021), Ethical and social risks of harm from Language Models, https://arxiv.org/abs/2112.04359, page 22.

7 https://openai.com/sora

8 Vaswani, A. et al. (2017), *Attention Is All You Need*, https://arxiv.org/abs/1706.03762.

9 Triguero, I, et al. (2024). "General Purpose Artificial Intelligence Systems (GPAIS): Properties, definition, taxonomy, societal implications and responsible governance." *Information Fusion* 103 (2024): 102135.

10 https://openai.com/blog/openai-codex

11 https://deepmind.google/discover/blog/competitive-programming-with-alphacode/

12 https://research.ibm.com/blog/ai-for-code-project-wisdom-red-hat

13 Chithrananda, S. et al. (2020) *ChemBERTa: Large-Scale Self-Supervised Pretraining for Molecular Property Prediction*. https://arxiv.org/abs/2010.09885

14 Irwin, R. et al. (2022) "Chemformer: a pre-trained transformer for computational chemistry." *Machine Learning: Science and Technology* 3(1). https://iopscience.iop.org/article/10.1088/2632-2153/ac3ffb

15 https://research.ibm.com/blog/molecular-transformer-discovery

16 https://www.microsoft.com/en-us/research/group/autonomous-systems-group-robotics/articles/introducing-climax-the-first-foundation-model-for-weather-and-climate/

17 https://newsroom.ibm.com/2023-08-03-IBM-and-NASA-Open-Source-Largest-Geospatial-AI-Foundation-Model-on-Hugging-Face

18 https://www.bloomberg.com/company/press/bloomberggpt-50-billion-parameter-llm-tuned-finance/

19 Assael, Y. et al. (2022). "Restoring and attributing ancient texts using deep neural networks." *Nature* 603, 280–283. https://www.nature.com/articles/s41586-022-04448-z

20 Al Quraishi, M. (2021). "Machine learning in protein structure prediction," *Current Opinion in Chemical Biology* 65 1–8, https://doi.org/10.1016/J.CBPA.2021.04.005; Rives, A. et al. (2021). "Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences.: *Proceedings of the National Academy of Sciences* 118, 15 (2021). https://doi.org/10.1073/pnas.2016239118.

21 Rothchild, A. et al. (2021). *C5T5: Controllable Generation of Organic Molecules with Transformers*. https://arxiv.org/abs/2108.10307

22 https://openai.com/blog/chatgpt

23 Mollick, E. (2022), "ChatGPT is a Tipping Point for AI" *Harvard Business Review.* https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai

24 https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework

25 https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html. An unofficial English translation is available at https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/

26 https://www8.cao.go.jp/cstp/ai/ningen/ningen.html. English translation available at https://www8.cao.go.jp/cstp/ai/humancentricai.pdf

27 https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_2.pdf

28 https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework

29 Available at https://openresearch-repository.anu.edu.au/bitstream/1885/277585/1/SKAI_31.pdf. No authoritative English language translation is available.

30 https://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_Y2B1M0R6G2I2P1B0V2X9H4Z0X3M3J2

31 https://www.chiefscientist.gov.au/sites/default/files/2023-06/Rapid%20Response%20Information%20Report%20-%20Generative%20AI%20v1_1.pdf

32 https://www.esafety.gov.au/industry/tech-trends-and-challenges/generative-ai

33 https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/Safe-and-responsible-AI-in-Australia-discussion-paper.pdf

34 https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/safe-and-responsible-ai-in-australia-governments-interim-response.pdf

35 https://dp-reg.gov.au/publications/working-paper-2-examination-technology-large-language-models

36 http://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm

37 http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

38 https://www.tc260.org.cn/upload/2024-03-01/1709282398070082466.pdf

39 https://www.ppc.go.jp/files/pdf/230602_kouhou_houdou.pdf

40 https://www.meti.go.jp/press/2024/04/20240419004/20240419004.html

41 https://aiverifyfoundation.sg/downloads/Discussion_Paper.pdf

42 https://aiverifyfoundation.sg/downloads/Proposed_MGF_Gen_AI_2024.pdf

43 https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=9055#LINK

44 https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=9083

45 Commonly referred to as "hallucinations."

46 *Supra* n 6.

47 *Ibid*.

48 https://ico.org.uk/media/about-the-ico/documents/4026232/joint-statement-data-scraping-202308.pdf

49 Tirumala, K. et al. (2022). "Memorization Without Overfitting: Analyzing the Training Dynamics of Large Language Models" *36th Conference on Neural Information Processing Systems* (NeurIPS 2022). https://arxiv.org/pdf/2205.10770.pdf

50 *Supra* n 6, page 19.

51 Burgess, M. (2023). "The Hacking of ChatGPT Is Just Getting Started" *Wired* https://www.wired.com/story/chatgpt-jailbreak-generative-ai-hacking/; Oremus, W. (2023). "The clever trick that turns ChatGPT into its evil twin" *The Washington Post.* https://www.washingtonpost.com/technology/2023/02/14/chatgpt-dan-jailbreak/

52 *Supra* n 23.

53 OECD (2023), "AI language models: Technological, socio-economic and policy considerations", *OECD Digital Economy Papers*, No. 352, OECD Publishing, Paris, https://doi.org/10.1787/13d38f92-en, page 34.
54 *Ibid*.
55 *Supra* n 6, page 15.
56 https://digi.org.au/disinformation-code
57 https://www.legislation.gov.au/Latest/C2022C00361
58 https://www.gov.cn/xinwen/2021-08/20/content_5632486.htm. A nonbinding English translation is available at http://en.npc.gov.cn.cdurl.cn/2021-12/29/c_694559.htm.
59 https://elaws.e-gov.go.jp/document?lawid=415AC0000000057. An unofficial English translation is available at https://www.japaneselawtranslation.go.jp/en/laws/view/4241/en
60 https://sso.agc.gov.sg/Act/PDPA2012?
61 https://www.law.go.kr/법령/개인정보보호법. An unofficial English translation is available at https://elaw.klri.re.kr/kor_service/lawView.do?hseq=53044&lang=ENG
62 Zanfir-Fortuna, G. (2023) *How Data Protection Authorities are De Facto Regulating Generative AI,* https://fpf.org/blog/how-data-protection-authorities-are-de-facto-regulating-generative-ai/.
63 Paulger, D. (2022). *Balancing Organizational Accountability and Privacy Self-Management in Asia-Pacific.* https://fpf.org/blog/new-report-promotes-accountability-based-approach-to-data-protection-in-the-apac-region/
64 https://www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/
65 https://commoncrawl.org/overview
66 https://paperswithcode.com/dataset/c4
67 https://laion.ai/blog/laion-400-open-dataset/
68 https://laion.ai/blog/laion-5b/
69 Verma, P. and Oreumus, W. (2023). "ChatGPT invented a sexual harassment scandal and named a real law prof as the accused." *The Washington Post.* https://www.washingtonpost.com/technology/2023/04/05/chatgpt-lies/
70 Kaye, B. (2023). "Australian mayor readies world's first defamation lawsuit over ChatGPT content" *Reuters.* https://www.reuters.com/technology/australian-mayor-readies-worlds-first-defamation-lawsuit-over-chatgpt-content-2023-04-05/
71 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
72 Edwards, B. (2023) "Artist finds private medical record photos in popular AI training data set." *Ars Technica.* https://arstechnica.com/information-technology/2022/09/artist-finds-private-medical-record-photos-in-popular-ai-training-data-set/
73 *Supra* n 6, page 20.
74 *Supra* n 49.
75 https://www.ppc.go.jp/personalinfo/legal/leakAction/
76 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
77 https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS217&mCode=D010030000&nttId=10059
78 Confederation of European Data Protection Organisations, (2023). *Generative AI: The Data Protection Implications*, https://cedpo.eu/generative-ai-the-data-protection-implications/, pages 4-5.
79 https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9870847#english
80 https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9874702#english
81 *Supra* n 78, page 6.
82 World Economic Forum (2023). *The Presidio Recommendations on Responsible Generative AI*. https://www.weforum.org/publications/the-presidio-recommendations-on-responsible-generative-ai/, page 3.
83 See, for example, the Personal Data Protection Authority of Singapore's Advisory Guidelines on use of Personal Data in AI Recommendation and Decision Systems, available at https://www.pdpc.gov.sg/guidelines-and-consultation/2024/02/advisory-guidelines-on-use-of-personal-data-in-ai-recommendation-and-decision-systems
84 **China**: Ethical Principles for New Generation AI, Section II. **Japan**: Governance Guidelines for Implementation of AI Principles (Ver 1.1), page 15.
85 **Australia**: eSafety Commissioner Position Statement, pages 6, 29. **China**: Ethical Principles for New Generation AI, Section II. **Japan**: Governance Guidelines for Implementation of AI Principles (Ver 1.1), page 17. **Singapore**: Model AI Governance Framework, page 17. **South Korea**: Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
86 **China**: Ethical Principles for New Generation AI, Section III.
87 **China**: Ethical Principles for New Generation AI, Sections II, III; Deep Synthesis Regulations, Articles 4-6; Interim Generative AI Measures, Article 7. **Japan**: Governance Guidelines for Implementation of AI Principles (Ver 1.1), page 17. **South Korea**: Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
88 **Australia**: Rapid Research Information Report, page 12; Safe and Responsible AI in Australia Discussion Paper, pages 40-41; eSafety Commissioner Position Statement, pages 6, 29. **China**: Ethical Principles for New Generation AI, Section II. **Japan**: Governance Guidelines for Implementation of AI Principles (Ver 1.1), page 9; Guidelines for AI Business Operators, Recommendation D-2. **Singapore**: Model AI Governance Framework, page 29. **South Korea:** Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
89 **Australia**: Safe and Responsible AI in Australia Discussion Paper, pages 40-41. **Singapore**: Model AI Governance Framework, page 29.
90 **Australia**: Government's Interim Response to Safe and Responsible AI in Australia Consultation, page 20; eSafety Commissioner Position Statement, page 29. **China**: Ethical Principles for New Generation AI, Section II; Deep Synthesis Regulations, Articles 7-8. **Singapore**: Model AI Governance Framework, pages 21-23.
91 **Australia**: Safe and Responsible AI in Australia Discussion Paper, pages 6, 40-41; Government's Interim Response to Safe and Responsible AI in Australia Consultation, page 20; eSafety Commissioner Position Statement, page 8. **Japan**: Governance Guidelines for Implementation of AI Principles (Ver 1.1), page 27.
92 **Singapore**: Model AI Governance Framework, page 29.
93 **Australia**: eSafety Commissioner Position Statement, page 6. **Japan**: Guidelines for AI Business Operators, Recommendation D-2, D-3. **Singapore**: Model AI Governance Framework, page 36.
94 **Singapore**: Model AI Governance Framework, pages 37-38.
95 **China**: Basic Security Requirements, Section 5. **Japan**: Guidelines for AI Business Operators, Recommendation D-2, D-3. **Singapore**: Model AI Governance Framework, pages 37-38.
96 **Australia**: eSafety Commissioner Position Statement, page 6. **Japan**: Guidelines for AI Business Operators, Recommendation D-2, D-3. **Singapore**: Model AI Governance Framework, page 24.
97 **China**: Basic Security Requirements, Section 5.
98 **Australia**: eSafety Commissioner Position Statement, page 7. **China**: Basic Security Requirements, Section 6. **Japan**: Guidelines for AI Business Operators, Recommendation P-3.

99 **Australia**: eSafety Commissioner Position Statement, pages 7-8. **Japan**: Guidelines for AI Business Operators, Recommendation P-3.
100 **Australia**: eSafety Commissioner Position Statement, page 6.
101 **Singapore**: Model AI Governance Framework, page 40.
102 **Australia**: eSafety Commissioner Position Statement, page 6. **China**: Interim Generative AI Measures, Article 11. **South Korea:** Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
103 **South Korea**: Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
104 **Australia**: Government's Interim Response to Safe and Responsible AI in Australia Consultation, page 20. **Japan**: Guidelines for AI Business Operators, Appendix 3, Section 4. **South Korea**: Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
105 **China**: Interim Generative AI Measures, Articles 7, 11.
106 **China**: Basic Security Requirements, Section 7. **Japan**: Guidelines for AI Business Operators, Recommendations D-2, P-4. **South Korea**: Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
107 **Australia**: Rapid Research Information Report, page 12; eSafety Commissioner Position Statement, pages 18, 30-32. **China**: Deep Synthesis Regulations, Article 14; Interim Generative AI Measures, Articles 7, 11; Basic Security Requirements, Section 7. **Singapore**: Model AI Governance Framework, page 56. **South Korea**: Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
108 *Supra* n 78, page 21-22. See also UK Information Commissioner's Office, *G7 DPAs' Emerging Technologies Working Group use case study on privacy enhancing technologies*, available at https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/privacy-enhancing-technologies/case-studies/g7-dpas-emerging-technologies-working-group-use-case-study-on-privacy-enhancing-technologies/
109 Competition and Markets Authority (UK), (2023). *AI Foundation Models: Initial Report*, https://www.gov.uk/government/publications/ai-foundation-models-initial-report, page 32.
110 https://lmsys.org/blog/2023-03-30-vicuna/
111 CEDPO, Generative AI: The Data Protection Implications (16 October 2023), page 21-22.
112 *Supra* n 109, page 33. See also UK ICO, G7 DPAs' Emerging Technologies Working Group use case study on privacy enhancing technologies,
113 **Australia**: eSafety Commissioner Position Statement, page 8. **China**: Deep Synthesis Regulations, Article 10; Interim Generative AI Measures, Article 17; Basic Security Requirements, Section 9. **Japan**: Guidelines for AI Business Operators, Recommendation D-2, D-5. P-2, P-5. **Singapore**: AI Verify Discussion Paper, page 22-24; Proposed Model Governance Framework for Generative AI, page 10.
114 **Australia**: eSafety Commissioner Position Statement, page 8.
115 **Australia**: eSafety Commissioner Position Statement, page 29. **China**: Interim Generative AI Measures, Article 10; Basic Security Requirements, Section 7.
116 **Australia**: eSafety Commissioner Position Statement, page 8. **China**: Deep Synthesis Regulations, Article 10; Interim Generative AI Measures, Article 14; Basic Security Requirements, Section 7. **Singapore**: Proposed Model Governance Framework for Generative AI, page 16.
117 **Australia**: Government's Interim Response to Safe and Responsible AI in Australia Consultation, pages 13, 20; eSafety Commissioner Position Statement, pages 8, 29. **China**: Deep Synthesis Regulations, Articles 10, 15; Basic Security Requirements, Section 6. **Japan**: Guidelines for AI Business Operators, Recommendation D-2, D-5. P-2, P-5. **Singapore**: AI Verify Discussion Paper, page 22-24; Proposed Model Governance Framework for Generative AI, pages 11-12.
118 **Australia:** Safe and Responsible AI in Australia Discussion Paper, pages 40-41. **Singapore**: AI Verify Discussion Paper, page 24; Proposed Model Governance Framework for Generative AI, pages 15-16.
119 **Singapore**: AI Verify Discussion Paper, page 24.
120 **Australia**: eSafety Commissioner Position Statement, pages 8, 29. **Singapore**: AI Verify Discussion Paper, page 25; Proposed Model Governance Framework for Generative AI, page 11.
121 https://csrc.nist.gov/glossary/term/red_team
122 **Australia**: Government's Interim Response to Safe and Responsible AI in Australia Consultation, page 20. **China**: Deep Synthesis Regulations, Article 10. **Japan**: Guidelines for AI Business Operators, Recommendation D-5. **Singapore**: AI Verify Discussion Paper, pages 24-25; Proposed Model Governance Framework for Generative AI, pages 13-14. **South Korea:** Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
123 **Australia**: eSafety Commissioner Position Statement, page 31. **Japan**: Governance Guidelines for Implementation of AI Principles (Ver 1.1), pages 39-43; Guidelines for AI Business Operators, Recommendation P-7. **South Korea:** Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
124 **Australia**: eSafety Commissioner Position Statement, pages 17, 30. **Japan**: Guidelines for AI Business Operators, Recommendation P-7.
125 **Australia**: Safe and Responsible AI in Australia Discussion Paper, pages 40-41; Government's Interim Response to Safe and Responsible AI in Australia Consultation, page 20. **Japan**: Guidelines for AI Business Operators, Recommendation D-6, D-7, P-6. **Singapore**: AI Verify Discussion Paper, page 21.
126 **Australia**: Safe and Responsible AI in Australia Discussion Paper, pages 40-41: eSafety Commissioner Position Statement, pages 7, 23, 31. **China**: Basic Security Requirements, Section 7.
127 **Australia**: Safe and Responsible AI in Australia Discussion Paper, pages 40-41; eSafety Commissioner Position Statement, pages 9, 30. **China**: Interim Generative AI Measures, Article 15. **South Korea**: Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
128 **Australia**: eSafety Commissioner Position Statement, page 31. **South Korea**: Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
129 **Australia**: Safe and Responsible AI in Australia Discussion Paper, pages 40-41; eSafety Commissioner Position Statement, pages 9, 30. **China**: Interim Generative AI Measures, Article 15. **South Korea**: Policy Direction for Safe Use of Personal Information in the AI Era, Section IV.
130 **Australia**: eSafety Commissioner Position Statement, page 23. **Singapore**: Model AI Governance Framework, pages 44-45.
131 **Australia**: Government's Interim Response to Safe and Responsible AI in Australia Consultation, pages 13, 20; eSafety Commissioner Position Statement, pages 9, 29-30. **China**: Deep Synthesis Regulations, Articles 16-17, Interim Generative AI Measures, Article 12. **Japan**: Guidelines for AI Business Operators, Appendix 3. **Singapore**: AI Verify Discussion Paper, page 21; Proposed Model Governance Framework for Generative AI, page 17.
132 https://c2pa.org/
133 *Supra* n 109, page 94.
134 *Ibid*; See also Rosenblatt, B (2023). "Google And OpenAI Plan Technology To Track AI-Generated Content" *Forbes.* https://www.forbes.com/sites/billrosenblatt/2023/07/22/google-and-openai-plan-technology-to-track-ai-generated-content/?sh=348a1ece131b
135 **Australia**: Government's Interim Response to Safe and Responsible AI in Australia Consultation, pages 13, 20; eSafety Commissioner Position Statement, pages 9, 29-30. **China**: Deep Synthesis Regulations, Articles 16-17; Interim Generative AI Measures, Article 12. **Singapore**: AI Verify Discussion Paper, page 21; Proposed Model Governance Framework for Generative AI, page 17.
136 **Singapore**: AI Verify Discussion Paper, page 21; Proposed Model Governance Framework for Generative AI, page 17.

1350 EYE STREET NW | SUITE 350 | WASHINGTON, DC 20005 FPF.ORG | info@fpf.org