



Singapore finalises model governance framework for generative AI

3 June 2024

LEGAL UPDATE



In this Update

On 30 May 2024, Singapore published the finalised version of the Model AI Governance Framework for Generative AI (“**Framework**”), following a public consultation on a draft version of the Framework between 16 January 2024 to 15 March 2024.

In this legal update, Head of TMT Lim Chong Kin and Director Cheryl Seah set out the key things you must know about this Framework and how it may affect you.

We also provide a snapshot of the 9 dimensions to build trustworthy generative AI in the framework, with an explanation of the significance of each dimension and the follow-up actions required of policymakers and the industry.

03

#1: WHAT IS THIS FRAMEWORK AND HOW DOES IT FIT IN WITH SINGAPORE’S AI GOVERNANCE POLICIES?

03

#2: WHO DOES THE FRAMEWORK AFFECT? HOW WILL THE FRAMEWORK AFFECT THE OPERATIONS OF COMPANIES DEVELOPING / DEPLOYING / USING GENERATIVE AI TOOLS?

04

#3: WHAT ARE THE KEY CONCEPTS IN THE FRAMEWORK?

06

#4: WHAT DOES THE FRAMEWORK NOT COVER?

07

#5: WHAT DO I NEED TO DO ABOUT THIS FRAMEWORK?

On 30 May 2024, Singapore issued the finalised version of the Model AI Governance Framework for Generative AI (“the Framework”) at the Asia Tech x Singapore Summit, following a public consultation on a draft version of the Framework between 16 January 2024 to 15 March 2024.

The framework sets out 9 dimensions which must be addressed to minimise the risks posed by generative AI and promote its responsible use.

Approximately 70 organisations – ranging from industry players (e.g. Microsoft, OpenAI, Visa), as well as government, research institutions and associations – contributed feedback. There are no significant or conceptual changes from the version issued for public consultation on 16 January 2024. However, further proposals from the contributors have been integrated (e.g. that technical solutions such as digital watermarking to identify AI-generated content should be complemented by enforcement mechanisms).

In this newsflash, we set out the key things you must know about this framework and how it may affect you, across 5 FAQs:

#1: What is this framework and how does it fit in with Singapore’s AI governance policies?

Traditional AI (which makes predictions based on existing/historical data instead of creating new content) – also called discriminative AI – is covered by the [Model AI Governance Framework](#) (issued in 2019 and updated in 2020). To assess the performance of a traditional AI model against a set of internationally recognised AI governance/ethics principles, Singapore has developed [AI Verify](#) (a testing and governance toolkit).

For generative AI (which became a hot issue after the public launch of ChatGPT in November 2022), there was [discussion](#) of its risks back in June 2023, followed by a [second paper](#) in October 2023 that sets out commonly-used tests by the industry to evaluate LLMs, as well as a recommended baseline for LLM evaluation, covering robustness, factuality, propensity to bias, toxicity generation and data governance.

The AI landscape keeps developing, and regulatory guidelines must keep pace with it. This Framework builds on the earlier foundational materials for traditional and generative AI, in order to address generative AI concerns while continuing to facilitate innovation.

#2: WHO DOES THE FRAMEWORK AFFECT? HOW WILL THE FRAMEWORK AFFECT THE OPERATIONS OF COMPANIES DEVELOPING / DEPLOYING / USING GENERATIVE AI TOOLS?

The framework is a roadmap for what action should be taken to address the risks posed by generative AI, to ensure that it is developed and used responsibly. Much of its actions required are targeted at policymakers/governments to implement.

In other words, it does not set out binding obligations as yet on companies developing generative AI models, or building applications on top of them, or who use generative AI tools to generate content. However, it gives companies a clear picture of the areas that the Singapore government will be looking deeper into, like accountability, the use of data to train models, and requiring third-party testing/verification. Hence, this would eventually impact companies developing/deploying AI tools.

#3: WHAT ARE THE KEY CONCEPTS IN THE FRAMEWORK?

To give you a snapshot of the [36-page Framework](#), we have condensed it into a table listing out each of the 9 dimensions covered, with an explanation of why each is significant and the follow-up actions suggested (which apply to both governments/policymakers as well as the industry):

	9 dimensions for trustworthy generative AI	Why is it significant	Action required
1	Accountability of players across the AI chain to end-users, so that end-users are protected	There are many players in the AI ecosystem (e.g. model developers, deployers, cloud service providers who provide platforms on which AI applications are hosted), so the right incentive structure must be put in place for different players to be responsible to end-users	- Responsibility should be allocated in 2 areas – upfront in the development process (ex-ante) as well as when issues are discovered after the AI model is deployed (ex-post) - Suggestion that responsibility should be allocated based on the level of control each player has in the generative AI development chain, tapping on the experience of related industries such as the controls and measures between cloud service providers who provide the base infrastructure and their customers who host applications

	9 dimensions for trustworthy generative AI	Why is it significant	Action required
			Suggestion to take additional measures to protect users – e.g. offering indemnities, updating legal frameworks to make it easier for users to prove damage caused by AI-enabled products and services and no-fault insurance
2	Data used in model training	- The quality of the data used to train a model affects the quality of its output - Some data used for model training is contentious – such as personal data and copyrighted material	- Policymakers should articulate how existing personal data laws apply to generative AI, and promote research into Privacy Enhancing Technologies - Policymakers should foster open dialogue amongst all relevant stakeholders to resolve (in a pragmatic way) copyright issues for data used in AI training
3	Trusted model development and the application deployment on top of the model	Even though end-users have limited visibility over the development process, transparency about this process builds trust and enhances broader awareness and safety over time	- The industry should adopt common safety practices (e.g. Reinforcement Learning from Human Feedback to generate output more aligned with human preferences and values; and Retrieval-Augmented Generation to reduce hallucinations) - Standardise disclosure about models (akin to “food/ingredient labels”) to facilitate comparability across models and incentivise safer model use - Standardise evaluation of generative AI models so that a baseline set of required safety tests are deployed to evaluate both front-end performance and back-end safety
4	Incident reporting	- Even the most robust AI systems are not foolproof - Timely reporting allows for monitoring and remediation, and supports continuous improvement of AI systems	- AI developers should have reporting channels for uncovered safety vulnerabilities in their AI systems and follow up to patch the system - Organisations must report material incidents arising from their use of AI systems (and policymakers must decide on the level of severity of the AI incident that requires reporting to the public/government – e.g. death, serious injury, serious disruption of critical infrastructure)
5	(Third-party) testing and assurance	- Independent verification helps to build trust with end-users 2 key aspects – how to test (testing methodology) and who to carry out the test (to ensure independence)	- In the near term, third-party testing will comprise the same set of benchmarks and evaluation methods used by developers themselves - As the area matures, policymakers and international standards

	9 dimensions for trustworthy generative AI	Why is it significant	Action required
			organisations like ISO/IEC and IEEE to develop common standards for AI testing to ensure quality and consistency
6	Security	With new technology, new threats arise – there are both traditional software security concerns (where there are already approaches to address them), and novel threat vectors against the generative AI model (where solutions are still being developed)	New tools must be developed to address the risks specific to generative AI – e.g. input-moderation tools to detect unsafe prompts and block malicious code
7	Content provenance	Users should be aware that they are interacting with AI-generated content, to reduce risk of misinformation	- Policymakers must work with key parties in the AI content lifecycle (e.g. publishers) to develop and deploy technical solutions like digital watermarking and cryptographic provenance, where appropriate; and these technical solutions must be interoperable - These technical solutions must be complemented by enforcement mechanisms - Policymakers must raise public awareness of content provenance and how to verify the authenticity of content
8	Safety and alignment research & development (R&D)	Safety techniques and evaluation tools to address the potential risks posed by AI models must keep pace with the speed of technological developments	- Investment in R&D, with more AI safety R&D institutes set up to conduct alignment research in tandem with AI companies - Global cooperation among AI safety R&D institutes to optimise limited resources and keep pace with commercial developments
9	AI for the public good	Responsible use of AI should go beyond risk mitigation and actively seek to improve people's lives	- Governments should partner companies and communities on digital literacy initiatives so that all members of society can have access to AI and know how to use it safely and responsibly (e.g. how to identify deepfakes, how to use chatbots safely) - Governments should drive innovation in the industry especially among SMEs through e.g. the use of sandboxes - Upskilling of the workforce and job redesign

	9 dimensions for trustworthy generative AI	Why is it significant	Action required
			Ensuring AI is sustainable – e.g. minimising its carbon footprint

#4: WHAT DOES THE FRAMEWORK NOT COVER?

The framework does not set out definitive positions on issues of privacy, copyright law, etc (e.g. it does not say that it is acceptable to use copyrighted material to train generative AI). Rather, it acknowledges that these areas are important to various stakeholders and there is presently uncertainty, hence it is useful for policymakers to articulate how the laws will apply to generative AI (e.g. by issuing guidelines and codes of practice).

#5: WHAT DO I NEED TO DO ABOUT THIS FRAMEWORK?

Companies developing or deploying generative AI solutions can study the framework to see if there are areas they can look further into, to enhance the safe and responsible use of their AI solution. Some live issues faced by companies include the allocation of liability (where the generative AI solution produces output that is infringing or defamatory), creating acceptable-use policies for users (e.g. if the public interacts with the AI solution), as well as the IP issues when training generative AI models on copyrighted works, or using the generative AI model output.

Please do not hesitate to contact any members of [our Artificial Intelligence & Digital Trust Practice](#), if you require more information about the framework, or how Singapore's AI laws and guidelines presently apply to your business operations.

The content of this article does not constitute legal advice and should not be relied on as such. Specific advice should be sought about your specific circumstances. Copyright in this publication is owned by Drew & Napier LLC. This publication may not be reproduced or transmitted in any form or by any means, in whole or in part, without prior written approval.

Please do not hesitate to contact any members of our **Artificial Intelligence & Digital Trust Practice**, if you require more information about the Framework, or how Singapore's AI laws and guidelines presently apply to your business operations:



Lim Chong Kin

Managing Director, Corporate & Finance
Head, Telecommunications,
Media & Technology

T: +65 6531 4110

E: chongkin.lim@drewnapier.com



Benjamin Gaw

Director, Corporate and Merger &
Acquisitions

T: +65 6531 2393

E: benjamin.gaw@drewnapier.com



David N. Alfred

Director, Corporate & Finance
Co-Head, Data Protection,
Privacy & Cybersecurity Practice

T: +65 6531 2342

E: david.alfred@drewnapier.com



Anastasia Chen

Director, Corporate & Finance

T: +65 6531 4123

E: anastasia.chen@drewnapier.com



Cheryl Seah

Director, Corporate & Finance

T: +65 6531 4167

E: cheryl.seah@drewnapier.com



Albert Pichlmaier

Senior Cybersecurity and Privacy Engineer,
Corporate & Finance

T: +65 6531 4108

E: albert.pichlmaier@drewnapier.com

Drew & Napier LLC

10 Collyer Quay
#10-01 Ocean Financial Centre
Singapore 049315

www.drewnapier.com

T: +65 6535 0733

T: +65 9726 0573 (After Hours)

E: mail@drewnapier.com

 **DREW & NAPIER**