



**D** DREW & NAPIER

New governance  
framework for  
generative AI:  
Singapore's IMDA  
seeking feedback  
internationally

17 January 2024

**LEGAL  
UPDATE**

# In this Update

On 16 January 2024, Singapore launched its proposed Model AI Governance Framework for Generative AI (“**Framework**”) at the World Economic Forum in Davos, Switzerland.

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

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

## 03

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

## 04

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

## 06

FAQ #4: WHAT DOES THE FRAMEWORK NOT COVER

## 07

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

## **FAQ #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.

## **FAQ #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 the 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.

There are also portions in the Framework that seek the industry's views (e.g. on a baseline level of transparency on the safety measures taken

when developing AI models, such as publishing an overview of the training data used, evaluation results, and the model’s known risks/limitations and measures taken to address those risks – see pages 10 and 11 of the Framework for details; or on incident reporting – see pages 13 and 14 of the Framework for details). Companies would do well to respond to the call for feedback to shape the eventual standards.

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

To give you a snapshot of the [22-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):

	<b>9 dimensions for trustworthy generative AI</b>	<b>Why is it significant</b>	<b>Action required</b>
1	<b>Accountability</b> of players across the AI chain to end-users, so that end-users are protected	<ul style="list-style-type: none"> <li>• 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), and allocation of responsibility amongst them is important</li> </ul>	<ul style="list-style-type: none"> <li>• Suggestion that responsibility should be allocated based on the level of control each player has in the generative AI development chain</li> <li>• 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</li> </ul>
2	<b>Data</b> used in model training	<ul style="list-style-type: none"> <li>• The quality of the data used to train a model affects the quality of its output</li> <li>• Some data used for model training is contentious – such as personal data and copyrighted material</li> </ul>	<ul style="list-style-type: none"> <li>• Policymakers should articulate how existing personal data laws apply to generative AI, and promote research into Privacy Enhancing Technologies</li> <li>• Policymakers should foster open dialogue amongst all relevant stakeholders to resolve (in a pragmatic way) copyright issues for data used in AI training</li> </ul>
3	<b>Trusted model development and the application deployment on</b>	<ul style="list-style-type: none"> <li>• Even though end-users have limited visibility over the development process, transparency about this builds trust and enhances</li> </ul>	<ul style="list-style-type: none"> <li>• 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</li> </ul>

	<b>top of the model</b>	broader awareness and safety over time	<p>Retrieval-Augmented Generation to reduce hallucinations)</p> <ul style="list-style-type: none"> <li>Standardise disclosure about models (akin to “food/ingredient labels”) to facilitate comparability across models and incentivise safer model use</li> <li>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</li> </ul>
4	<b>Incident reporting</b>	<ul style="list-style-type: none"> <li>Even the most robust AI systems are not foolproof</li> <li>Timely reporting allows for monitoring and remediation, and supports continuous improvement of AI systems</li> </ul>	<ul style="list-style-type: none"> <li>AI developers should report safety vulnerabilities in their AI systems and then patch the system (i.e. act pre-emptively)</li> <li>Organisations must report 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)</li> </ul>
5	<b>(Third-party) testing and assurance</b>	<ul style="list-style-type: none"> <li>Independent verification helps to build trust with end-users</li> <li>2 key aspects – how to test (testing methodology) and who to carry out the test (to ensure independence)</li> </ul>	<ul style="list-style-type: none"> <li>Policymakers and international standards organisations like ISO/IEC and IEEE to develop common standards for AI testing to ensure quality and consistency</li> </ul>
6	<b>Security</b>	<ul style="list-style-type: none"> <li>With new technology, new threats arise – e.g. prompt attacks can be injected through the model architecture, allowing attackers to exfiltrate sensitive information and model weights</li> </ul>	<ul style="list-style-type: none"> <li>New testing 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</li> </ul>
7	<b>Content provenance</b>	<ul style="list-style-type: none"> <li>Users should be aware that they are interacting with AI-generated content, to reduce risk of misinformation</li> </ul>	<ul style="list-style-type: none"> <li>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</li> </ul>

			<p>appropriate; and these technical solutions must be interoperable</p> <ul style="list-style-type: none"> <li>• Policymakers must raise public awareness of content provenance and how to verify the authenticity of content</li> </ul>
8	<b>Safety and alignment research &amp; development (R&amp;D)</b>	<ul style="list-style-type: none"> <li>• There is always more to be done to improve model safety, since technological developments do not cease</li> </ul>	<ul style="list-style-type: none"> <li>• Investment in R&amp;D, with more AI safety R&amp;D institutes set up to conduct alignment research in tandem with AI companies</li> <li>• Global cooperation among AI safety R&amp;D institutes to optimise limited resources and keep pace with commercial developments</li> </ul>
9	<b>AI for the public good</b>	<ul style="list-style-type: none"> <li>• Responsible use of AI should go beyond risk mitigation and actively seek to improve people's lives</li> </ul>	<ul style="list-style-type: none"> <li>• Governments should partner companies and communities on digital literacy initiatives (e.g. how to identify deepfakes, how to use chatbots safely)</li> <li>• Governments should drive innovation in the industry especially among SMEs through e.g. the use of sandboxes</li> <li>• Upskilling of the workforce and job redesign</li> <li>• Ensuring AI is sustainable – e.g. minimising its carbon footprint</li> </ul>

#### **FAQ #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).

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

Companies can study the Framework in greater detail and give their feedback/input to IMDA at [info@aiverify.sg](mailto:info@aiverify.sg) by 15 March 2024. The email header should state: “Comments on the Proposed Model Governance Framework for Generative AI”.

*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 proposed Framework, or how Singapore's AI laws and guidelines presently apply to your business operations, or if you require our assistance in drafting any feedback to the IMDA:



**Lim Chong Kin**

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

T: +65 6531 4110

E: [chongkin.lim@drewnapier.com](mailto:chongkin.lim@drewnapier.com)



**Benjamin Gaw**

Director, Corporate and Merger &  
Acquisitions

T: +65 6531 2393

E: [benjamin.gaw@drewnapier.com](mailto: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](mailto:david.alfred@drewnapier.com)



**Anastasia Chen**

Director, Corporate & Finance

T: +65 6531 4123

E: [anastasia.chen@drewnapier.com](mailto:anastasia.chen@drewnapier.com)



**Cheryl Seah**

Director, Corporate & Finance

T: +65 6531 4167

E: [cheryl.seah@drewnapier.com](mailto:cheryl.seah@drewnapier.com)



**Albert Pichlmaier**

Senior Cybersecurity and Privacy Engineer,  
Corporate & Finance

T: +65 6531 4108

E: [albert.pichlmaier@drewnapier.com](mailto:albert.pichlmaier@drewnapier.com)

**Drew & Napier LLC**


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

[www.drewnapier.com](http://www.drewnapier.com)

T: +65 6535 0733

T: +65 9726 0573 (After Hours)

E: [mail@drewnapier.com](mailto:mail@drewnapier.com)

 **DREW & NAPIER**