LEGAL UPDATE

NAVIGATING THE FUTURE: SINGAPORE'S PROPOSED MODEL AI GOVERNANCE FRAMEWORK FOR GENERATIVE AI

1. There is no question that generative artificial intelligence (AI that can be used to generate content such as text, audio, videos, and other media) is one of the most promising and interesting fields of technology. Since its release in 2022, ChatGPT has amazed the world with its uncanny ability to give life-like (and often but not always useful!) responses to all sorts of questions and input – from the profound, to the mundane and the abrasive.

2. Generative AI is the future. It has boundless potential to improve everyone's lives. But it also has great potential for harm. For example, deepfake videos can and have been used by bad actors to spread misinformation and untruths. ChatGPT is also often used by professionals and students who do not realise that ChatGPT is a chatbot and not necessarily an information resource – possibly to their detriment. A US lawyer was fined and reprimanded by a court for filing a legal brief created by ChatGPT and containing false case citations. The misuse of generative AI (whether deliberate or inadvertent) is a serious cause for concern.

3. Consequently, regulators have sought to introduce frameworks and rules to regulate generative AI.

4. Before the rise of generative AI, Singapore's Infocomm Media Development Authority ("**IMDA**") and the Personal Data Protection Commission ("**PDPC**") had, in 2019, released the existing Model AI Governance Framework (the "**Existing Framework**").

5. However, given the new issues which have arisen as a result of developments in generative AI (e.g. hallucination, copyright infringement, and value alignment), the AI Verify Foundation ("**AVF**") and IMDA have, on 16 January 2024, issued a draft Model AI Governance Framework for Generative AI in Singapore (the "**Draft Framework**"). According to the press release by IMDA, the objective of the Draft Framework is to holistically address new issues that have emerged with generative AI.

6. The Draft Framework "seeks to set forth a systematic and balanced approach to address generative AI concerns while continuing to facilitate innovation".

Overview of the Draft Framework

7. The Draft Framework proposes 9 dimensions to "foster a trusted ecosystem":

20 February 2024

For any queries relating to this article, please contact

Tan Tee Jim, S.C.
tanteejim@leenlee.com.sg

Basil Lee
basillee@leenlee.com.sg

Authors:

Tan Tee Jim, S.C.
Basil Lee
Chee Kai Hao
Poon Chong Ming

Lee & Lee
25 North Bridge Road
Level 7
Singapore 179104
Tel: +65 6220 0666

a) Accountability

The Draft Framework seeks to ensure that there is an appropriate allocation of responsibility within the AI development chain by putting in place the right incentive structure for different players (such as model developers, application developers and cloud service providers) to be responsible towards end users. Responsibility may be allocated both upfront in the development process and post-deployment.

b) Data

As data is at the core of model and application development in generative AI, the Draft Framework calls for policymakers to: (1) articulate how existing personal data laws apply to generative AI in a manner that still protects the rights of individuals; and (2) engage in open dialogue with relevant stakeholders to develop a legal framework to address copyright-related issues arising in generative AI. To ensure that only trusted data sources are used, AI developers are encouraged to ensure data quality such as by using only trusted data sources, using data quality control measures, and adopting best practices in data governance.

c) Trusted Development and Deployment

To ensure transparency and disclosure in the development and deployment of the AI model, while balancing legitimate considerations such as safeguarding business and proprietary information, safety best practices need to be implemented by model developers and application developers across the AI development lifecycle, in particular around “development” (e.g. having baseline safety practices), “disclosure” (e.g. standardising disclosure of information), and “evaluation” (standardised model safety evaluations).

d) Incident Reporting

To ensure timely notification and remediation, establishing structures and processes to enable incident monitoring and reporting is key. This includes vulnerability reporting and incident reporting.

e) Testing and Assurance

The Draft Framework encourages the development of a third-party testing ecosystem to bring independent verification to AI models as it provides transparency and builds greater trust with end-users. As to “how” to test, the Draft Framework considers that third-party testing will be more effective if there are standardised testing methodologies, and as such encourages the setting of common benchmarks and methodologies with shared tooling to facilitate testing across different models. In the long run, mature AI testing could be formalised through standards organisations for harmonised third-party testing. With regards to “who” to perform the testing, industry bodies and governments could join efforts to build a qualified pool of third-party, and to eventually implement an accreditation mechanism.

f) Security

New threat vectors that arise through generative AI models should be distinguished from conventional software security threats and addressed by adapting and refining the "security-by-design" concept for use in generative AI, and developing new security safeguards.

g) Content Provenance

Given the difficulty to distinguish between AI-generated and original content today (e.g. deepfake) and the harms of misinformation, transparency about where and how content is created helps to inform end-users when consuming online content. Technical solutions such as digital watermarking (embedding information within content) and cryptographic provenance (tracking and verifying the digital content origin and any edits made, with the records cryptographically protected) are suggested. However, their use must be complemented by policies that are carefully designed to enable practical use in the right context, including allowing end-users to verify content authenticity, standardising the types of edits to be labelled for AI-generated contents and raising awareness regarding content provenance amongst end-users.

h) Safety and Alignment Research & Development (R&D)

As present safety techniques and evaluation tools do not address all potential risks, our capacity to align and control generative AI must keep up with them. The Draft Framework urges the acceleration of R&D in model safety and alignment such as through global cooperation and the setting up of AI safety institutes. Further, practical steps may be taken to enhance the speed of translation and application of new R&D insights such as through understanding and systematically mapping the various research directions and methods that have emerged. These areas of research include the so-called "forward alignment" and "backward alignment" of AI models.

i) AI for Public Good

Generative AI has the potential to drive growth and productivity while empowering people and growth globally. The Draft Framework sets out four "concrete touchpoints" where AI can have beneficial and long-term effects: democratising access to generative AI for all members of society, public service delivery, upskilling of workforce to harness AI effectively as well as to navigate job transformations and transitions, as well as sustainability.

Commentary

8. While the Existing Framework and Draft Framework are non-binding (as opposed to laws or regulations), they provide a good indication of how AI will be regulated in Singapore. Sector-specific regulators will no doubt pay close attention to the updated framework in formulating their policies. Parties developing and deploying generative AI applications would therefore do well to pay heed to these developments in order to react quickly to any changes in regulations.

9. Indeed, local agencies have already begun to or are looking to enact AI-related regulations. For example, the Monetary Authority of Singapore has launched Project MindForge to develop a risk framework for the use of generative AI in financial sectors, while the PDPC on 31 August 2023 concluded a public consultation for its Proposed Advisory Guidelines on Use of Personal Data in Recommendation and Decision Systems.

10. A public consultation on the Draft Framework is presently ongoing and will close on 15 March 2024. Interested parties should provide their comments before the deadline.

11. If you have any question on any aspect of the Draft Framework, please contact our Mr. Tan Tee Jim, SC (tanteejim@leenlee.com.sg) or Mr. Basil Lee (basillee@leenlee.com.sg).

About Lee & Lee

Lee & Lee is one of Singapore's leading law firms being continuously rated over the years amongst the top law firms in Singapore. Lee & Lee remains committed to serving its clients' best interests, and continuing its tradition of excellence and integrity. The firm provides a comprehensive range of legal services to serve the differing needs of corporates, financial institutions and individuals. For more information: visit www.leenlee.com.sg.

The following partners lead our departments:

Kwa Kim Li Managing Partner kwakimli@leenlee.com.sg	Quek Mong Hua Litigation & Dispute Resolution quekmonghua@leenlee.com.sg	Owyong Thian Soo Real Estate owyongthiansoo@leenlee.com.sg
Tan Tee Jim, *S.C.* Intellectual Property tanteejim@leenlee.com.sg	Adrian Chan Corporate adrianchan@leenlee.com.sg	Louise Tan Banking louisetan@leenlee.com.sg