

SINGAPORE PROPOSES MODEL AI GOVERNANCE FRAMEWORK FOR GENERATIVE AI

Posted on March 1, 2024



Category: [CNPupdates](#)

General disclaimer

This article is provided to you for general information and should not be relied upon as legal advice. The editor and the contributing authors do not guarantee the accuracy of the contents and expressly disclaim any and all liability to any person in respect of the consequences of anything done or permitted to be done or omitted to be done wholly or partly in reliance upon the whole or any part of the contents.

Authors: Bill Jamieson , Tan Wee Liang

Singapore's Infocomm Media Development Authority ("IMDA") and its non-profit wholly-owned subsidiary AI Verify Foundation announced a public consultation on its jointly developed draft Model AI Governance Framework for Generative AI ("Draft GenAI Governance Framework") on 16 January 2024. The Draft GenAI Governance Framework highlights possible areas for future policy interventions relating to generative artificial intelligence ("AI") and provides options for such interventions.

Feedback on the Draft GenAI Governance Framework may be sent to info@aiverify.sg by 15 March 2024. The IMDA and AI Verify Foundation seek international feedback, given that AI governance is a collective international effort. To that end, Minister Josephine Teo announced the Draft GenAI Governance Framework at the World Economic Forum to allow international input and contribute to facilitating international conversations among policymakers, industry, and the research community to enable trusted development globally. The framework is expected to be finalised in mid-2024.

This expanded Draft GenAI Governance Framework updates an existing 2019 framework (last updated in 2020) covering only traditional AI and is intended to develop in parallel with the National AI Strategy 2.0 launched in December 2023. Traditional AI refers to AI models that make predictions by leveraging insights derived from historical data. Typical traditional AI models include logistic regression, decision trees and conditional random fields. In contrast, generative AI is an AI model capable of generating text, images, or other media. They learn the patterns and structure of their input training data and generate new data with similar characteristics. The recent advent of generative AI has reinforced some of the same AI risks (e.g. bias, misuse, lack of explainability) and introduced new ones (e.g. hallucination, copyright infringement, value alignment). This Draft GenAI Governance Framework capitalises on initiatives like a discussion paper by IMDA in 2023 on the risks associated with greater use of generative AI, as well as efforts to guide the safety evaluation of generative AI models and ongoing evaluation tests on AI products.

Summary of the Draft GenAI Governance Framework

The Draft GenAI Governance Framework seeks to propose a systemic, pragmatic and balanced approach to addressing generative AI risks and concerns while continuing to facilitate innovation. The Draft GenAI Governance Framework identifies nine key governance dimensions around areas like accountability, security and testing to facilitate a comprehensive and trusted AI ecosystem – one that enables end-users to use generative AI confidently and safely while allowing space for cutting-edge innovation:

1. **Accountability:** Establishing the proper incentive structure for different players (e.g. model developers, application deployers, and cloud service providers) in the AI system development life cycle to be responsible to end-users.
2. **Data:** Ensuring data quality and pragmatically addressing potentially contentious training data.
3. **Trusted development and deployment:** Enhancing transparency around baseline safety and hygiene measures based on industry best practices in development, evaluation, and disclosure.

General disclaimer

This article is provided to you for general information and should not be relied upon as legal advice. The editor and the contributing authors do not guarantee the accuracy of the contents and expressly disclaim any and all liability to any person in respect of the consequences of anything done or permitted to be done or omitted to be done wholly or partly in reliance upon the whole or any part of the contents.

4. Incident reporting: Implementing an incident management system for timely notification, remediation, and continuous improvements.
5. Testing and assurance: Providing external validation, added trust through third-party testing, and developing common AI testing standards for consistency.
6. Security: Addressing new threat vectors that arise through generative AI models.
7. Content provenance: Providing transparency about the source of content.
8. Safety and alignment R&D: Accelerating R&D through global cooperation among AI safety institutes to improve model alignment with human intention and values.
9. AI for the public good: Responsible AI includes harnessing AI to benefit the public by democratising access, improving public sector adoption, upskilling workers, and developing AI systems sustainably.

Key Takeaways

The Draft GenAI Governance Framework demonstrates Singapore's holistic efforts to enhance AI governance. It sheds light on the perspectives of policymakers in regulating the development and deployment of generative AI systems. The effects of generative AI will be felt in the economy in various ways; it is expected that different sectors will use generative AI in their own ways and find an expression of governance most suited to their circumstances. The testing and validation of AI tools is a potential industry that could grow out of this increased focus on AI governance.

In addition, the Draft GenAI Governance Framework and the high-level principles discussed within are a welcome addition to the tapestry of papers and guidance on generative AI crafted by policymakers worldwide.

General disclaimer

This article is provided to you for general information and should not be relied upon as legal advice. The editor and the contributing authors do not guarantee the accuracy of the contents and expressly disclaim any and all liability to any person in respect of the consequences of anything done or permitted to be done or omitted to be done wholly or partly in reliance upon the whole or any part of the contents.