DREW & NAPIER

Singapore Introduces Model Governance Framework for Agentic AI

26 January 2026

LEGAL UPDATE

In this Update

On 22 January 2026, Singapore published the Model AI Governance Framework for Agentic AI ("**MGFA**"), ensuring that Singapore's AI regulatory strategy keeps pace with the latest technological developments. AI agents can take autonomous actions, adapt to new information, and interact with other agents and systems to complete tasks on behalf of humans. While this promises greater efficiency, the increased capabilities of AI agents also introduce new risks that we must manage.

In this legal update, Head of TMT Lim Chong Kin and Director Cheryl Seah explore key issues in the MGFA: (a) what is agentic AI; (b) what are the risks associated with its use; (c) what steps/best practices does the MFGA recommend for organisations deploying agentic AI to manage the risks (including the potential allocation of liability across developers, vendors and customers); and (d) what can organisations do now.

What is Agentic AI?

Agentic AI systems are systems that can plan across multiple steps to achieve specified objectives, using AI agents. Unlike generative AI, which requires a prompt to produce content, agentic AI can take actions, adapt to new information, and interact with other agents and systems to complete tasks autonomously.

The MGFA focuses on agents built on language models. In summary, the language model functions as the "brain" of the agent, where it will receive instruction (from a human user) that can be in natural language and not in code, and analyse how to execute the instructions to achieve the objective. It will then (a) draw on repositories of information stored and available to the language model; as well as (b) draw on tools (e.g. calculator app, calendar app, machine, web crawlers, etc.) linked to the language model to complete the task. If the language model is the "brain", the tools are the "arms and legs" that have received signals from the "brain", enabling the agent to perform actions and interact with other systems, such as controlling a device, or performing a transaction.

There are AI agents that are "deterministic", meaning that they produce consistent and identical outputs when given the same inputs, so their performance is predictable. There are also AI agents that are "non-deterministic", where they produce variable outputs even with the same input. This lack of predictability affects the level of confidence that can be placed in the agent's output, and how much human oversight is then required.[1]

It is common for an agentic system to have multiple agents working together, where each agent specializes in a certain task and work in parallel with other agents. The design of each agent will affect its limits and capabilities on 2 fronts:

- Action-space (impacting authority and capabilities): The range of actions the agent can execute are determined by the tools it is permitted to access (e.g. sandboxed tools that cannot affect any other system, or an organisation's internal system, or external systems through third-party APIs) and the transactions it can execute (e.g. being able to read and retrieve information only rather than write to and modify data in a system).

[1] See section 2.1 of the World Economic Forum's "Agents in Action: Foundations for Evaluation and Governance" white paper (Nov 2025) https://reports.weforum.org/docs/WEF_AI_Agents_in_Action_Foundations_for_Evaluation_and_Governance_2025.pdf, referenced in footnote 3 of the MGFA.

- Autonomy (impacting decision-making): The degree to which the agent can decide when and how to act towards a goal is determined by its instructions (e.g. is it instructed to follow an SOP or can it exercise its own judgment) and level of human involvement in the agentic system.

Why does Agentic AI pose additional risk?

While agents are already transforming workplaces through coding assistants, customer service agents, and enterprise productivity workflows, their greater capabilities bring forth new risks. Some risks are already known to us and managed, such as the risks arising from the language models the agentic system is built on. However, these existing risks can manifest themselves differently, or compound themselves, due to the added capabilities of AI agents.

For example, a hallucination from an AI agent may result in a wrong plan to compete a task, and this mistake escalates where the initial wrong output is passed on to other agents - if an inventory figure from one agent is erroneous, a downstream agent might use this data and place an order for excessive or insufficient stock.

The MGFA highlights that organisations must be aware of these 5 risk categories:

Risk Category	Description
Erroneous actions	Incorrect actions such as an agent fixing appointments on the wrong date (resulting in harm due to delayed treatment) or producing flawed code (leading to exploitable security vulnerabilities).
Unauthorised actions	The agent takes actions outside its permitted scope or authority, like acting without escalating for human approval despite the SOP requiring it.
Biased or unfair actions	The agent's actions result in unfair outcomes across different profiles and demographics (e.g. in hiring decisions, or vendor selection).
Data breaches	Exposure or manipulation of sensitive data, including PII or confidential information such as trade secrets and customer lists. For example, a security breach where attackers exploit the agent to reveal private information, or the agent discloses such sensitive information as it fails to recognize such information is sensitive.
Disruption to connected systems	Agents causing disruption to the systems they are connected to when they are compromised or malfunctioning, e.g. deleting a production codebase or overwhelming external systems with requests.

What best practices for organisations does the MGFA recommend to manage risk?

The MGFA establishes four key dimensions that organisations deploying agentic AI must address. This would apply whether the organisation is developing the AI agent in-house or using third-party developed agentic solutions.

Dimension 1: Assess and bound the risks upfront

Organisations should adapt their internal structures and processes to account for new risks from agents due to their adaptive and autonomous nature, and their ability to directly impact the environment they operate in through taking actions (via the tools they are connected to).

Step 1: Is this use case suitable for agent deployment based on the likelihood and impact of the risk:

Organisations must conduct risk identification and assessment as the first step when considering whether to use agentic AI. The MGFA sets out the factors to consider when assessing risk, where it is a function of both:

(1) "Impact" (severity of impact if the risk manifests)	(2) "Likelihood" (probability of the risk manifesting)
• Level of tolerance of error in the domain and use case in which the agent is being deployed to. • Whether the agent can access sensitive data, such as personal information or confidential data. • Whether the agent can access external systems. • Whether an agent can only read from or modify the data and systems it has access to. • Reversibility of agent's actions.	• Agent's level of autonomy – a higher level of autonomy can result in higher unpredictability and increased likelihood of error. • How complex the task is (number of steps, level of analysis required at each step) – where a higher level of complexity increases unpredictability and thus the likelihood of error. • Agent's level of access to external systems and who the external systems are maintained by, as exposure to external systems makes the agent more vulnerable to prompt injections and cyberattacks.

Step 2: After selecting an appropriate agent use case, further bound the risks by defining limits and permission policies for each agent:

Design Boundaries: Organisations should limit the scope of impact of their agents by designing appropriate boundaries at the planning stage. This includes:

- Defining policies giving agents only the minimum tools and data access needed to complete tasks.
- Defining SOPs for agentic workflows that the agent must follow rather than granting the agent the freedom to define every step, to improve consistency and reduce unpredictability.
- Designing mechanisms to take agents offline and limit potential scope of impact when they malfunction.

Agent Identity Management: Organisations should extend identity management to agents to track individual agent behaviour and establish who (e.g. which employee or department) is accountable for each agent. Each agent should have its own unique identity, and authorisation should ensure that human users cannot set permissions for agents greater than what the human user is authorised to do.

Dimension 2: Make Humans Meaningfully Accountable

The MGFA makes it clear that "the organisations that deploy agents and the humans who oversee them remain accountable for the agents' behaviours and actions". This is despite the fact that agents can operate autonomously, and multiple stakeholders are involved across the agent lifecycle (both within and outside of the organisation).

Clear Allocation of Responsibilities Within the Organisation: Within the organisation, responsibilities should be allocated to different teams depending on their expertise, across the agent lifecycle:

- Key decision-makers such as board members, department leaders (setting goals, defining permitted use cases, governance approach).
- Product teams (design, implementation, testing, user education).
- Cybersecurity teams (protecting agentic systems from cyber threats, managing security measures, incident response).
- Any individuals using the output of AI agents (ethical use, compliance with organisational policies).

Clear Allocation of Responsibilities Outside the Organisation: This may be achieved through contractual means, as against the agentic AI

system provider that provided the agentic system to the organisation. In turn, the agentic AI system provider would have upstream contractual arrangements with the model developers whose models it built the agents on, as well as other software and tool providers whose products are integrated.

Design checkpoints for meaningful human oversight: Organisations should define significant checkpoints or action boundaries requiring human approval, especially before sensitive actions are executed. These include high-stakes actions and decisions, irreversible actions (e.g. permanently deleting data, sending communications, making payments), outlier or atypical behaviour, and user-defined boundaries based on the user's risk appetite.

Audit Human Oversight: Organisations should also implement measures to ensure continued effectiveness of human oversight, including training humans to identify common failure modes (e.g. inconsistent agent reasoning or the agent referring to outdated policies) and regularly auditing the effectiveness of human approvals since humans are susceptible to automation bias.

Dimension 3: Implement Technical Controls and Processes

Organisations should implement technical measures across the agent lifecycle to ensure the AI agents operate safely and reliably.

During Development: Implement controls during the design and development phase to mitigate identified risks, in addition to baseline software and LLM controls. For example –

- Where it comes to planning and reasoning, prompt the agent to summarise its understanding and request clarification from the user before proceeding.

- Where it comes to tools, configure strict input formats and apply principle of least privilege, limiting the tools available to each agent.

- Where it comes to protocols (a standardised way for an agent to communicate with tools and other agents), use standardised protocols where applicable, whitelist trusted servers and sandbox any code execution.

Before Deployment: Test agents for safety and reliability to ensure they will work as expected and that controls are effective. The starter kits for testing of LLM-based apps can be used as language models are a core component of agents, but new risks must also be covered, such as overall task execution accuracy, policy compliance, tool calling accuracy, and reactions to errors and edge cases. Testing should cover entire

agent workflows, individual and multi-agent behaviour, and occur in realistic environments.

During and After Deployment: Gradually roll out agents into production to control risk exposure (e.g. to trained users first, with restricted tools initially, in lower-risk internal systems first). Continuously monitor and log agent behaviour, establishing reporting and failsafe mechanisms for failures or unexpected behaviours.

Dimension 4: Enable End-User Responsibility

Organisations cannot solely rely on their agentic system developers for successful deployment of agentic AI – they must also require their end-users to use agentic AI responsibly. An "end user" is the person that interacts with the agentic system, or who integrates the agent into their work processes. Therefore, organisations must provide sufficient information to the end users to enable their responsible use of the agentic system.

Transparency: Users should be informed upfront that they are interacting with an agent and not a human. They should also be informed of the agent's capabilities (e.g. what actions it can take, what user's data it can access), who to escalate the issue to if they suspect the agent is malfunctioning, as well as the user's own responsibilities (such as whether he/she must double-check all information provided by the agent).

Education and Training: Organisations whose employees are integrating agents into their workflows (e.g. coding assistants) should additionally provide training covering foundational knowledge on agents (use cases, prompting best practices), and how to have effective oversight of agents (common failure modes, common user mistakes). As agents take over entry-level tasks, organisations should ensure employees retain core skills through sufficient training and work exposure.

What can organisations do now?

While the MGFA does not impose binding legal obligations, it provides a clear picture of Singapore's regulatory direction for agentic AI and establishes industry best practices. Organisations can start with the following:

1. **Conduct a gap analysis:** Review current AI governance policies against the MGFA's four dimensions to identify areas requiring enhancement.

2. **Review vendor contracts:** Ensure contracts with agentic AI providers, model developers, and tooling providers adequately address security arrangements, performance guarantees, data protection, and liability allocation.

3. **Establish internal governance:** Work with technical teams to define appropriate use cases, set agent boundaries, establish human approval checkpoints, and implement monitoring mechanisms.

4. **Develop user policies:** Create acceptable use policies for employees using agentic AI and ensure appropriate training is provided.

5. **Monitor developments:** The MGFA is described as a living document that will be continuously updated and refined. Organisations should monitor for updates and prepare for potential future regulatory developments.

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.

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