AI Governance in Financial Services

# MAS Publishes Industry White Paper on Safeguards for AI Agents in Financial Services

On 3 July 2026, the Monetary Authority of Singapore (**MAS**), together with leading financial institutions (**FIs**) and FinTechs, published an industry white paper titled Safeguards for Agentic Finance at Runtime (**White Paper**) on developing safeguards for artificial intelligence (**AI**) agents in finance.

The White Paper proposes a framework, Safeguards for Agentic Finance at Runtime (**SAFR**), for the governance of AI agents in financial services, defining how agent actions are authorised, how human oversight is activated, and what is recorded at the point of every decision. It also discusses implementation considerations for FIs deploying SAFR in their agentic AI systems. SAFR is jointly developed by the financial industry under MAS' BuildFin.ai initiative, which supports the responsible development and deployment of AI solutions in the financial sector.

## Status of SAFR

The White Paper states that SAFR neither constitutes regulatory guidance or supervisory expectations nor prescribes or anticipates future directions for such guidance or expectations. SAFR is described as an industry reference approach for FIs to implement within their own infrastructure, using their own rule configurations and governance arrangements. Each FI remains responsible for determining how its deployment aligns with applicable supervisory expectations and internal governance requirements.

We summarise below the key elements of the SAFR framework as highlighted in MAS' media release and the White Paper.

## The SAFR Framework

The White Paper provides that before any agentic action can be executed, it should be packaged with the context needed to assess it in a Governance Envelope, which captures:

(a) The action type, and the scope and parameters of the proposed action;

(b) The actual steps the agent executed in arriving at the proposed action; and

(c) Context metadata, including the agent identity, applicable mandate, current account or system state, and operative policy constraints.

The Governance Envelope is then assessed by the four runtime components set out below.

### *Agent Identity*

Under the framework, each proposed action should be bound to a recognised, registered agent, verified against that agent's registry entry before any other evaluation proceeds. The White Paper provides that Agent Identity verification is the condition that allows the Controls Repository to retrieve the applicable controls, as the agent's identity shapes what it may be permitted to do. An envelope that fails this check results in the proposed action being rejected.

### *Controls Repository*

The Controls Repository is the FI's configurable rulebook: the controls, drawn from sources such as organisational policies, regulatory requirements, product rules, and user-provided mandates, against which a proposed action is checked.

The White Paper sets out several categories of controls, each independently configurable. FIs implementing the framework should ensure that configuration of the controls is owned by the FI's risk and compliance leadership. The applicable categories include:

(a) **Authorisation**: Which agents are permitted to act for which principals, who can delegate authority to whom, and to what depth, and which action types each agent class is permitted to initiate.

(b) **Exposure limits**: Per-action and aggregate value threshold. Below the threshold, autonomous actions are permitted. Above it, human review is required or the action is blocked. These mirror existing delegated authority frameworks and should be set to match them.

(c) **Rate limits**: Maximum action rate per time window. Protects against runaway agents, data feed errors, and adversarial prompt injections that cause agents to act at abnormal speed or volume.

(d) **Evidence quality**: Minimum confidence threshold and required evidence for autonomous execution. Actions where the agent's stated confidence falls below the threshold are routed to human review regardless of value.

The White Paper further provides that when defining the controls to determine the thresholds required for escalation to a human reviewer, FIs must assess three operational dimensions to ensure that escalation is substantive rather than nominal:

(a) **Escalation volume**: FIs should assess the volume of actions they can manually assess. An escalation function that generates more reviews than the FI can meaningfully process defeats its own purpose.

(b) **Review turnaround**: Escalations to a human reviewer should include a timeout window. If no decision is made within that period, the action should default to block or be escalated to a senior reviewer. The window should reflect realistic reviewer availability, including overnight and weekend coverage for time-sensitive domains.

(c) **Reviewer authority**: Reviewers must have clear authority to approve, modify, or decline escalated actions. Their decisions should carry the same institutional weight as the original agent decision. Escalating to someone without authority creates the appearance of oversight without making it effective.

### *Disposition Engine*

The Disposition Engine evaluates each in-scope action against the controls retrieved from the Controls Repository and resolves it to one of the possible outcomes. The White Paper provides that the Disposition Engine produces one of four outcomes for every action:

(a) **Deny**: The proposed action violates a hard regulatory or policy constraint, or presents a risk profile above defined thresholds. The action is rejected before execution.

(b) **Escalate**: The proposed action is within scope and below hard constraints, but above the threshold for autonomous execution. The action is held pending human review.

(c) **Auto-Execute**: The proposed action is within scope, below hard constraints, and within defined risk thresholds. The action proceeds without requiring human intervention.

(d) **Observe**: The proposed action is permitted to proceed but is flagged for monitoring. The action executes while a structured observation is logged for subsequent review.

FIs should note that in multi-step agentic workflows, the White Paper provides that this control flow should apply to each agent action independently. An Auto-Execute or Observe outcome at one step carries no authority into the next (i.e., prior authorisation should not carry forward as the agent adapts to intermediate results and changing conditions).

### *Audit Log*

Under the framework, every governance decision produces an entry in the Audit Log, which serves as an immutable, tamper-evident record that captures the complete history of every action an agent takes, and every decision made about it. Once written at the point of decision, an entry cannot be modified by the agent or by any downstream system.

FIs implementing the framework should ensure that each log entry captures the essential elements of a governance decision, including:

(a) The Governance Envelope as submitted;

(b) The mandate against which the action was checked;

(c) The outcome produced by the Disposition Engine;

(d) The specific rules applied and the basis for that outcome; and

(e) The time elapsed at each stage.

The White Paper also sets out two deployment patterns for FIs adopting the framework. Under native integration, the agent is instrumented to emit a Governance Envelope before each proposed action, with the SAFR validator returning an outcome decision before the agent takes any action. This pattern is recommended for new agent deployment. Under gateway integration, a SAFR gateway intercepts outbound API calls at the infrastructure layer, wrapping each call in a Governance Envelope without any changes to the agent code. This pattern is designed for legacy systems, third party agents or existing deployment where modifying agent code is not feasible.

## Use Cases

Industry members have applied the SAFR framework across use cases, including:

(a) **Agent-assisted payments and treasury operations**: Autonomous agents can execute routine transactions within predefined mandates, improving efficiency and reducing operational frictions.

(b) **Wealth management and advisory workflows**: AI agents review documents and generate structured assessments within narrowly scoped task boundaries, supporting faster and more consistent compliance review.

(c) **Client engagement**: AI agents generate client insights and draft materials within approved content boundaries, enabling staff to engage clients more effectively and productively.

## Next Steps

Interested industry partners are invited to join the BuildFin.ai work group to contribute to and help shape subsequent iterations of SAFR.

The Future of Finance Institute (**FFI**) will support future adoption of the SAFR framework through the facilitation of industry pilots and sandbox experimentation. As was recently announced, the FFI will be established to accelerate the adoption of new financial technologies and catalyse innovation in the financial sector, focusing initially on AI and tokenisation. The FFI will help to connect the financial and technology ecosystems and provide shared resources and expertise needed to promote broad-based adoption of these technologies.

If you would like information and/or assistance on the above or any other area of law, you may wish to contact the Partner at WongPartnership whom you normally work with or any of the following Partners:

**Elaine CHAN**
Co-Head – Financial Services Regulatory

**LAM Chung Nian**
Head – Intellectual Property, Technology & Data

**TIAN Sion Yoong**
Partner – Financial Services Regulatory

**Kylie PEH**
Partner – Intellectual Property, Technology & Data

**CHAN Jia Hui**
Partner – Financial Services Regulatory

Connect with WongPartnership.

# WPG MEMBERS AND OFFICES

- contactus@wongpartnership.com

### SINGAPORE

-

WongPartnership LLP
12 Marina Boulevard Level 28
Marina Bay Financial Centre Tower 3
Singapore 018982
t +65 6416 8000
f +65 6532 5711/5722

### CHINA

-

WongPartnership LLP
Shanghai Representative Office
Unit 1015 Link Square 1
222 Hubin Road
Shanghai 200021, PRC
t +86 21 6340 3131
f +86 21 6340 3315

### INDONESIA

-

Makes & Partners Law Firm
Menara Batavia, 7th Floor
Jl. KH. Mas Mansyur Kav. 126
Jakarta 10220, Indonesia
t +62 21 574 7181
f +62 21 574 7180
w makeslaw.com

### MALAYSIA

-

Cheang & Ariff
Advocates & Solicitors
Loke Mansion
273A, Jalan Medan Tuanku
50300 Kuala Lumpur
t +60 3 2691 0803
f +60 3 2693 4475
w cheangariff.com

-

Foong & Partners
Advocates & Solicitors
13-1, Menara 1MK, Kompleks 1 Mont' Kiara
No 1 Jalan Kiara, Mont' Kiara
50480 Kuala Lumpur, Malaysia
t +60 3 6419 0822
f +60 3 6419 0823
w foongpartners.com

### MIDDLE EAST

-

Al Aidarous Advocates and Legal Consultants
Abdullah Al Mulla Building, Mezzanine Suite 02
39 Hameem Street (side street of Al Murroor Street)
Al Nahyan Camp Area
P.O. Box No. 71284
Abu Dhabi, UAE
t +971 2 6439 222
f +971 2 6349 229
w aidarous.com

-

Al Aidarous Advocates and Legal Consultants
Oberoi Centre, 13th Floor, Marasi Drive, Business Bay
P.O. Box No. 33299
Dubai, UAE
t +971 4 2828 000
f +971 4 2828 011

### PHILIPPINES

-

Gruba Law
27/F 88 Corporate Center
141 Valero St., Salcedo Village
Makati City 1227, Philippines
t +63 2 889 6060
f +63 2 889 6066
w grubalaw.com