FSB FINANCIAL STABILITY BOARD

# Sound Practices for Responsible Adoption of Artificial Intelligence (AI): Consultation report

## Response to Consultation

JPMorgan Chase

1. **Do you agree with the benefits and risks of AI adoption by financial institutions described in this report? Are there any substantive benefits or risks not covered?**

2. **Are the sound practices sufficiently comprehensive and clear to enable financial institutions' responsible AI adoption?**

3. **Do the sound practices strike an appropriate balance between managing risks relating to all forms of AI, and addressing some of the risks relating to emerging and new complex forms of AI, such as GenAI and agentic AI?**

4. **Are the sound practices sufficiently flexible to accommodate and address newer types of AI and responsible adoption over time?**

5. **Do the case studies in this report sufficiently highlight how different types of financial institutions can benefit from responsible AI adoption? Are there additional case studies for inclusion in the report? If so, please provide such case studies, particularly for nonbanks.**

6. **Do the case studies in this report provide actionable insights for financial institutions in their responsible AI adoption?**

7. **Are the definitions in the glossary clear and aligned with industry sound practices, including recent developments in AI?**

8. **Are there any comments you would like to make on this report? Please fill in.**

JPMorganChase appreciates the opportunity to respond to the Financial Stability Board's consultation on sound practices for responsible adoption of AI. We also value the FSB's engagement with the public in developing the sound practices, including through roundtables held in connection with the consultation.

We are strongly aligned with the FSB's approach to expressing sound practices that are outcomes-focused, risk-based, and proportionate to the use and impact of the AI system. In our experience, we have found it most effective to leverage existing governance and risk management frameworks, and to proactively evolve the application of these frameworks in

a targeted manner to reflect the specific and differentiated risks of AI systems, including with agentic AI.

Across the financial sector, banks are moving deliberately in evaluating and deploying AI, including emerging agentic capabilities. As banks explore more transformative use cases, our focus is on ensuring that any deployment remains safe, secure, and governed appropriately. Our comments therefore focus on the control and governance framework we are building today to support more advanced agentic AI capabilities in the future.

In particular, JPMorganChase is actively adapting our governance framework to address material risks of agentic AI systems, and has identified the following non-exhaustive practices for governance: maintaining attribution and traceability of agent actions; constraining agent activity; defining agent identity, delegated authority, and least-privilege access; applying zero-trust and runtime enforcement controls; establishing escalation boundaries for higher-risk actions; conducting evaluation, validation, and continuous assurance; implementing circuit breakers, resiliency controls, and human override mechanisms; and assigning clear human accountability. These governance practices, which are more fully described later in the response, broadly align with the concepts included in the FSB's consultation and are intended to support a risk-based and scalable approach to responsible agentic AI governance.

As a threshold matter, firms should conduct appropriate risk assessments to determine which controls are necessary for a given agentic AI system. Agentic AI may create or amplify risks such as unauthorized or erroneous actions, sensitive-data exposure, disruption to connected systems, and difficult-to-monitor multi-step behavior, which should also be considered as part of the risk assessment. Different agentic systems have different capabilities, connectivity to systems and data, levels of autonomy, business impact, and risk profiles; controls and oversight should be calibrated proportionate to those attributes. Accordingly, sound practices should not be read to require a uniform set of controls across all agentic systems, but rather to support firms in identifying and applying controls that are appropriate to the risks presented by the particular use case.

In addition, many of the governance practices described throughout the paper assume firms have access to sufficient information and technical capabilities to evaluate, validate and oversee agentic AI systems. Where firms deploy third-party agentic AI systems, effective implementation of these governance practices may, in some cases, depend on capabilities, transparency, and information provided by third-party technology providers. Establishing policy expectations for third-parties to maintain baseline minimum controls and provide adequate transparency and information-sharing to their customers could more effectively accelerate safe adoption and improve consistency across the sector.

Finally, many agent-governance practices remain nascent and are not yet uniformly implemented across industry deployments. Descriptions of sound practices for governing agentic AI should be sufficiently concrete to support responsible implementation, while avoiding overly prescriptive technical approaches and acknowledging that some controls may depend on evolving third-party capabilities. This approach should allow firms to adapt controls proportionately as technology, standards, tooling, and operational practices mature.

Agentic governance considerations for responsible adoption

We set forth below a non-exhaustive set of governance practices that may help firms structure practical oversight of agentic AI systems. We are happy to continue the dialogue with you and other stakeholders to share our experiences and lessons learned.

1. Agent Attribution and Traceability

Firms should adopt approaches that enable actions taken by an AI agent to be attributed and distinguished from actions taken by humans or traditional software. Attribution will be particularly important where agentic systems can autonomously plan actions, invoke tools, interact with enterprise systems, or operate with limited human intervention. Attribution includes both visibility and traceability of agent actions and is foundational to the other preventative controls discussed below.

Visibility may involve assigning and documenting individual identifiers to AI agents to facilitate monitoring and identity management. Depending on the use case, this may include visibility into the person, software service or other agent that invoked, approved, configured, or supervised the agent, as well as records that support accountability for decisions and actions taken through agentic workflows. While some attribution practices may resemble employee-related controls, firms should focus on the capabilities needed to support attribution, accountability, and risk-based oversight, rather than framing AI agents as "synthetic employees."

Traceability enables firms to monitor actions taken by an agent and the relevant basis for those actions. Depending on the use case, this may include visibility into retrieval flows, memory access, orchestration logic, tool invocation, execution plans, triggering conditions, and authorization decisions.

This traceability is used to support operational monitoring, business impact tracking, and anomaly detection, including detection of access to resources not specified in the AI agent's scope of authority. The level of attribution and traceability should be proportionate to the risks of the agentic deployment, based on the firm's risk assessment.

2. Environmental Constraints and Containment

Firms should consider whether environmental constraints or containment mechanisms are appropriate. Where warranted, these measures may include constraining the agents' execution environment, checking against the agent's scope of authority, and limiting interaction with external systems, including APIs, data sources, ICT systems, or other agents, to those that are safe and necessary for the agent's approved purpose.

Such constraints can help limit unintended propagation across systems and align an agent's operating environment with its approved purpose. For higher risk or higher autonomy deployments, firms may evaluate preventive structural controls, such as segmentation, policy adjudication, and containment boundaries. The appropriate control posture should be proportionate to the agent's risk profile, system architecture, and operational feasibility, and should not be read to require any specific technical control in all cases.

3. Agent Identity, Delegated Authority, and Least Privilege

Firms should recognize that agents are not merely non-human service accounts executing static instructions, but may operate as decision-capable actors under delegated, scoped authority, selecting which tools to invoke, which systems to access, and which actions to take. Identity and access management for agents should therefore extend beyond authentication to include context-aware authorization: the central question is not only who may access a system, but what action is permitted, under what circumstances, under whose authority, and with what evidence.

Based on the agent's risk profile and intended use, firms should consider scoping an agent's identity, access, and action space to the least privilege necessary for its approved purpose, with clear ownership and an agent identity model appropriate to the use case. Agents may function as independent identities, as representatives acting on behalf of individuals, or as privileged system actors, each with different implications for authorization and accountability.

4. Zero Trust and Runtime Enforcement

Firms should consider applying zero-trust principles to agentic systems in a manner proportionate to risk, avoiding implicit trust in the agent, the user, retrieved content, tool outputs, downstream systems, or other agents, and instead evaluating in context whether a proposed action remains within the agent's approved purpose, delegated authority, and policy constraints. The distinct role of this practice is preventive: authorizing and constraining actions in-flight, before they take effect.

Because agent behavior may materialize at execution time rather than being fully fixed at design, build-time and pre-deployment controls alone may be insufficient for agents with higher degrees of autonomy or risk. In those cases, firms should consider authorization and constraint at the point of action. As agents increasingly invoke or rely on other agents, firms should likewise consider avoiding implicit trust between agents; interoperable standards that allow participating systems to establish who is acting, under what scope, and with what authorization will be important to supporting sound practice across the ecosystem.

5. Escalation Boundaries

Firms should consider defining risk-based escalation boundaries, informed by the firm's risk assessment, that identify when an agent should stop, seek confirmation, or defer to a human rather than act independently. Boundaries should account for the consequences, reversibility, sensitivity, and external impact of any given action. Depending on the risk profile, boundaries may include actions that are prohibited outright, actions that require human approval or dual authorization, and actions that require additional controls.

Human-in-the-loop oversight remains an important safeguard, particularly for irreversible or high-impact actions. However, because human review may have practical limits in agentic and multi-agent contexts, firms should employ complementary safeguards as appropriate, including preventive controls, runtime enforcement, and automated escalation mechanisms.

6. Evaluation, Validation, and Continuous Assurance

Firms should evaluate and, as appropriate, validate AI agents both before deployment and during operational use in a manner proportionate to the agent's risk profile. This review generally should go beyond the underlying model, and assess integrity of the broader workflow, including tool invocation, retrieval processes, memory use, orchestration logic, and interactions with other components. Assessments should consider criteria such as safety, reliability, resilience, performance, explainability where relevant, and compliance with applicable policies.

For use cases with a higher risk profile, firms should consider additional testing methods to determine whether an AI agent can be induced to act outside its defined purpose, authorization boundaries, or policy constraints. Testing should also assess how agents interact with other agents, tools, APIs, databases, and enterprise systems.

Pre-deployment testing alone may not be sufficient, particularly for use cases with higher capabilities or risk profile. Firms should also implement runtime monitoring and adaptive controls to detect and respond to emerging risks or component outages experienced during operation. These risks may include indirect prompt or context injection, tool-output manipulation, persistent memory poisoning, and multi-hop workflow manipulation.

Depending on the use case, testing, monitoring, and evaluation should focus on key intermediate steps, as well as the final outcome. It should identify patterns that may indicate scope creep or misuse—for instance, if the agent queries databases beyond those necessary for the task or accesses customer records more broadly than required.

Such controls help ensure that agentic systems continue to operate safely, reliably, and within approved boundaries over time.

7. Circuit Breakers, Resiliency, and Human Override

Firms should consider maintaining capabilities to halt, suspend, constrain, override, or revoke an agent's access, proportionate to risk. For higher-risk or higher-autonomy agents, firms should consider complementing these capabilities with operational resiliency measures that address how the agent behaves when it cannot safely complete a task, when a dependency is unavailable or degraded, or when continuing the workflow could compound errors. Firms should assess, proportionate to risk, whether an agent can recognize degraded conditions, avoid repeated or compounding retries, escalate appropriately, and enter a controlled failure mode. Where appropriate, agents should be capable of pausing, reverting to a known safe state, handing off to a human or alternate process, or terminating activity without data loss or unintended consequence.

Resiliency assessments should consider the agent's material dependencies, including model providers, orchestration services, tool servers or connectors, APIs, skills, data sources, and other systems needed to execute the workflow. Firms should consider how the agent and the broader process would respond if one of these dependencies becomes unavailable, returns degraded or inconsistent responses, changes behavior due to a model or system update, or cannot provide the level of reliability required for the use case.

In the future, to the extent an agentic AI system is used in processes supporting critical or essential services, firm's existing resiliency practices should extend to the agentic workflow

and its material dependencies, including but not limited to MCP servers, tools, skills, and models. This may include defined thresholds for pausing or constraining agent activity, fallback or manual processes, recovery expectations, and escalation paths. These mechanisms should be calibrated to the agent's autonomy, role in the business process, customer or market impact, and feasibility of intervention.

8. Named Human Accountability

Firms should assign clear human accountability for agentic systems and workflows, with the accountability model calibrated to the risk assessment. Accountability of an agentic system refers to identifiable human ownership of the system, its approved use, its control environment, and material decisions or actions taken through the workflow. The appropriate accountability model may vary depending on the firm's governance structure, the agent's role, the degree of autonomy involved, and the risk profile of the use case. As agentic systems evolve, firms may appropriately adapt controls related to identity, access, supervision, and lifecycle management, but governance should remain grounded in accountable human ownership and risk-based oversight.

Firms should also prioritize effective oversight of AI systems, calibrated to the risk assessment. Oversight is distinct from accountability: it concerns the mechanisms used to monitor, review, challenge, escalate, or intervene in AI-driven activity. Human oversight will be an important component, particularly for higher-risk use cases, but effective oversight need not always require real-time human intervention whenever an AI system produces undesired content. Depending on the context, oversight may include preventive controls, approval workflows, automated monitoring, exception-based alerts, sampling and testing, post-event review, escalation procedures, or human-in-the-loop intervention. For AI systems with lower risk profiles, periodic human review or spot-checking may be appropriate; for AI systems with higher risk profiles, more continuous oversight may be necessary.

When evaluating the effectiveness of oversight, firms should review patterns of approval, modification, escalation, and override. Such review can help identify ineffective oversight, including potential rubber-stamping, recurring overrides that suggest misalignment, or situations where the selected oversight model should be recalibrated. To enable accountability, firms should preserve delegated-lineage records sufficient to reconstruct who authorized material actions, under which scope, and how authority evolved across single-hop or multi-hop workflows where relevant.

Proportionality and implementation considerations

We encourage the FSB to adopt a proportionate, risk-based approach to agentic governance that recognizes the role of firm-specific risk assessments in determining the appropriate controls. As agentic AI capabilities evolve, responsible adoption will require proportionate practices that help firms define accountability, constrain autonomy, enforce authorization boundaries, preserve traceability, and maintain mechanisms to monitor, halt, or override agentic behavior where appropriate. Guidance should make clear that firms may calibrate governance expectations based on autonomy, use case, potential customer or market impact, reversibility of actions, and operational context.

The FSB should also recognize that certain practices discussed above remain emerging. Firms should be encouraged to evaluate these practices progressively, prioritize higher-risk use cases, and maintain flexibility as standards, tooling, and supervisory expectations mature.

Given the pace of change, any regulatory guidance should remain operationally grounded through continued engagement with financial institutions, technology providers, standards bodies, and other relevant stakeholders. Guidance that identifies illustrative practices, rather than fixed technical requirements, would better support responsible innovation while allowing firms to adapt controls as the technology evolves.

Conclusion

Agentic AI presents significant opportunities for firms, customers, and the financial system, but it may also shift certain risks from bounded model outputs toward execution-time governance. Responsible adoption can be supported by proportionate practices that help firms define accountability, constrain autonomy, enforce authorization boundaries, preserve traceability, and maintain mechanisms to monitor, halt, or override agentic behavior where appropriate.

We welcome continued engagement with the FSB and other stakeholders to support secure, resilient, and responsible agentic AI deployment across the financial sector.

JPMorganChase

To:
Ms. Michelle W. Bowman
Chair
Financial Stability Board Standing Committee on Supervisory and Regulatory Cooperation

Ms. Ho Hern Shin
Lead
Financial Stability Board SRC Workstream on Artificial Intelligence

Cc:
Mr. John Schindler
Secretary General
Financial Stability Board

July 22, 2026

**Re: Response to the Financial Stability Board Consultation on Sound Practices**

JPMorganChase appreciates the opportunity to respond to the Financial Stability Board's consultation on sound practices for responsible adoption of AI. We also value the FSB's engagement with the public in developing the sound practices, including through roundtables held in connection with the consultation.[1]

We are strongly aligned with the FSB's approach to expressing sound practices that are outcomes-focused, risk-based, and proportionate to the use and impact of the AI system. In our experience, we have found it most effective to leverage existing governance and risk management frameworks, and to proactively evolve the application of these frameworks in a targeted manner to reflect the specific and differentiated risks of AI systems, including with agentic AI.

Across the financial sector, banks are moving deliberately in evaluating and deploying AI, including emerging agentic capabilities. As banks explore more transformative use cases, our focus is on ensuring that any deployment remains safe, secure, and governed

[1] We note that a number of trade associations and other industry groups have submitted comments in response to the consultation. JPMorganChase supports many of the comments and requests for clarification raised by those organizations.

appropriately. Our comments therefore focus on the control and governance framework we are building today to support more advanced agentic AI capabilities in the future.

In particular, JPMorganChase is actively adapting our governance framework to address material risks of agentic AI systems, and has identified the following non-exhaustive practices for governance: maintaining attribution and traceability of agent actions; constraining agent activity; defining agent identity, delegated authority, and least-privilege access; applying zero-trust and runtime enforcement controls; establishing escalation boundaries for higher-risk actions; conducting evaluation, validation, and continuous assurance; implementing circuit breakers, resiliency controls, and human override mechanisms; and assigning clear human accountability. These governance practices, which are more fully described later in the response, broadly align with the concepts included in the FSB's consultation and are intended to support a risk-based and scalable approach to responsible agentic AI governance.

As a threshold matter, firms should conduct appropriate risk assessments to determine which controls are necessary for a given agentic AI system. Agentic AI may create or amplify risks such as unauthorized or erroneous actions, sensitive-data exposure, disruption to connected systems, and difficult-to-monitor multi-step behavior, which should also be considered as part of the risk assessment. Different agentic systems have different capabilities, connectivity to systems and data, levels of autonomy, business impact, and risk profiles; controls and oversight should be calibrated proportionate to those attributes. Accordingly, sound practices should not be read to require a uniform set of controls across all agentic systems, but rather to support firms in identifying and applying controls that are appropriate to the risks presented by the particular use case.[2]

In addition, many of the governance practices described throughout the paper assume firms have access to sufficient information and technical capabilities to evaluate, validate and oversee agentic AI systems. Where firms deploy third-party agentic AI systems, effective implementation of these governance practices may, in some cases, depend on capabilities, transparency, and information provided by third-party technology providers. Establishing policy expectations for third parties to maintain baseline minimum controls

[2] For example, an agentic system with read-only access – such as one that summarizes internal research reports and routes them to relevant analysts – presents materially different risks than a customer-facing agent that ingests unvetted customer communications, accesses customer's personal identifiable information, and can take actions or communicate with the outside world.

and provide adequate transparency and information-sharing to their customers could more effectively accelerate safe adoption and improve consistency across the sector.[3]

Finally, many agent-governance practices remain nascent and are not yet uniformly implemented across industry deployments. Descriptions of sound practices for governing agentic AI should be sufficiently concrete to support responsible implementation, while avoiding overly prescriptive technical approaches and acknowledging that some controls may depend on evolving third-party capabilities. This approach should allow firms to adapt controls proportionately as technology, standards, tooling, and operational practices mature.

**Agentic governance considerations for responsible adoption**

We set forth below a non-exhaustive set of governance practices that may help firms structure practical oversight of agentic AI systems.[4] We are happy to continue the dialogue with you and other stakeholders to share our experiences and lessons learned.

**1. Agent Attribution and Traceability**

Firms should adopt approaches that enable actions taken by an AI agent to be attributed and distinguished from actions taken by humans or traditional software. Attribution will be particularly important where agentic systems can autonomously plan actions, invoke tools, interact with enterprise systems, or operate with limited human intervention. Attribution includes both visibility and traceability of agent actions and is foundational to the other preventive controls discussed below.

Visibility may involve assigning and documenting individual identifiers to AI agents to facilitate monitoring and identity management. Depending on the use case, this may include visibility into the person, software service or other agent that invoked, approved, configured, or supervised the agent, as well as records that support accountability for decisions and actions taken through agentic workflows. While some attribution practices may resemble employee-related controls, firms should focus on the capabilities needed to support attribution, accountability, and risk-based oversight, rather than framing AI agents as "synthetic employees."

---

[3] The FSB, in its 2023 toolkit on "Enhancing Third-Party Risk Management and Oversight" notes the application of "direct requirements or expectations on financial sector critical service providers" as one tool to identify and manage systemic risks from third-party dependencies.

[4] When we refer to an "AI agent," we generally mean a software implementation where AI autonomously (i.e., without explicit human instruction) determines the use of tools, access to systems, or information retrieval as needed to accomplish a broader goal.

Traceability enables firms to monitor actions taken by an agent and the relevant basis for those actions. Depending on the use case, this may include visibility into retrieval flows, memory access, orchestration logic, tool invocation, execution plans, triggering conditions, and authorization decisions.

This traceability is used to support operational monitoring, business impact tracking, and anomaly detection, including detection of access to resources not specified in the AI agent's scope of authority. The level of attribution and traceability should be proportionate to the risks of the agentic deployment, based on the firm's risk assessment.

**2. Environmental Constraints and Containment**

Firms should consider whether environmental constraints or containment mechanisms are appropriate. Where warranted, these measures may include constraining the agents' execution environment, checking against the agent's scope of authority, and limiting interaction with external systems, including APIs, data sources, ICT systems, or other agents, to those that are safe and necessary for the agent's approved purpose.

Such constraints can help limit unintended propagation across systems and align an agent's operating environment with its approved purpose. For higher-risk or higher-autonomy deployments, firms may evaluate preventive structural controls, such as segmentation, policy adjudication, and containment boundaries. The appropriate control posture should be proportionate to the agent's risk profile, system architecture, and operational feasibility, and should not be read to require any specific technical control in all cases.

**3. Agent Identity, Delegated Authority, and Least Privilege**

Firms should recognize that agents are not merely non-human service accounts executing static instructions, but may operate as decision-capable actors under delegated, scoped authority, selecting which tools to invoke, which systems to access, and which actions to take. Identity and access management for agents should therefore extend beyond authentication to include context-aware authorization: the central question is not only who may access a system, but what action is permitted, under what circumstances, under whose authority, and with what evidence.

Based on the agent's risk profile and intended use, firms should consider scoping an agent's identity, access, and action space to the least privilege necessary for its approved purpose, with clear ownership and an agent identity model appropriate to the use case. Agents may function as independent identities, as representatives acting on behalf of

individuals, or as privileged system actors, each with different implications for authorization and accountability.

**4. Zero Trust and Runtime Enforcement**

Firms should consider applying zero-trust principles to agentic systems in a manner proportionate to risk, avoiding implicit trust in the agent, the user, retrieved content, tool outputs, downstream systems, or other agents, and instead evaluating in context whether a proposed action remains within the agent's approved purpose, delegated authority, and policy constraints. The distinct role of this practice is preventive: authorizing and constraining actions in-flight, before they take effect.

Because agent behavior may materialize at execution time rather than being fully fixed at design, build-time and pre-deployment controls alone may be insufficient for agents with higher degrees of autonomy or risk. In those cases, firms should consider authorization and constraint at the point of action. As agents increasingly invoke or rely on other agents, firms should likewise consider avoiding implicit trust between agents; interoperable standards that allow participating systems to establish who is acting, under what scope, and with what authorization will be important to supporting sound practice across the ecosystem.[5]

**5. Escalation Boundaries**

Firms should consider defining risk-based escalation boundaries, informed by the firm's risk assessment, that identify when an agent should stop, seek confirmation, or defer to a human rather than act independently. Boundaries should account for the consequences, reversibility, sensitivity, and external impact of any given action. Depending on the risk profile, boundaries may include actions that are prohibited outright, actions that require human approval or dual authorization, and actions that require additional controls.

Human-in-the-loop oversight remains an important safeguard, particularly for irreversible or high-impact actions. However, because human review may have practical limits in agentic and multi-agent contexts, firms should employ complementary safeguards as appropriate, including preventive controls, runtime enforcement, and automated escalation mechanisms.

[5] These practices may be relevant to where agentic systems route work to human users as well. See the discussion in "8. Human Accountability."

**6. Evaluation, Validation, and Continuous Assurance**

Firms should evaluate and, as appropriate, validate AI agents both before deployment and during operational use in a manner proportionate to the agent's risk profile. This review generally should go beyond the underlying model, and assess integrity of the broader workflow, including tool invocation, retrieval processes, memory use, orchestration logic, and interactions with other components. Assessments should consider criteria such as safety, reliability, resilience, performance, explainability where relevant, and compliance with applicable policies.

For use cases with a higher risk profile, firms should consider additional testing methods to determine whether an AI agent can be induced to act outside its defined purpose, authorization boundaries, or policy constraints. Testing should also assess how agents interact with other agents, tools, APIs, databases, and enterprise systems.

Pre-deployment testing alone may not be sufficient, particularly for use cases with higher capabilities or risk profile. Firms should also implement runtime monitoring and adaptive controls to detect and respond to emerging risks or component outages experienced during operation. These risks may include indirect prompt or context injection, tool-output manipulation, persistent memory poisoning, and multi-hop workflow manipulation.

Depending on the use case, testing, monitoring, and evaluation should focus on key intermediate steps, as well as the final outcome. It should identify patterns that may indicate scope creep or misuse—for instance, if the agent queries databases beyond those necessary for the task or accesses customer records more broadly than required.

Such controls help ensure that agentic systems continue to operate safely, reliably, and within approved boundaries over time.

**7. Circuit Breakers, Resiliency, and Human Override**

Firms should consider maintaining capabilities to halt, suspend, constrain, override, or revoke an agent's access, proportionate to risk. For higher-risk or higher-autonomy agents, firms should consider complementing these capabilities with operational resiliency measures that address how the agent behaves when it cannot safely complete a task, when a dependency is unavailable or degraded, or when continuing the workflow could compound errors. Firms should assess, proportionate to risk, whether an agent can recognize degraded conditions, avoid repeated or compounding retries, escalate appropriately, and enter a controlled failure mode. Where appropriate, agents should be

capable of pausing, reverting to a known safe state, handing off to a human or alternate process, or terminating activity without data loss or unintended consequence.

Resiliency assessments should consider the agent's material dependencies, including model providers, orchestration services, tool servers or connectors, APIs, skills, data sources, and other systems needed to execute the workflow. Firms should consider how the agent and the broader process would respond if one of these dependencies becomes unavailable, returns degraded or inconsistent responses, changes behavior due to a model or system update, or cannot provide the level of reliability required for the use case.

In the future, to the extent an agentic AI system is used in processes supporting critical or essential services, a firm's existing resiliency practices should extend to the agentic workflow and its material dependencies, including, but not limited to, MCP servers, tools, skills, and models. This may include defined thresholds for pausing or constraining agent activity, fallback or manual processes, recovery expectations, and escalation paths. These mechanisms should be calibrated to the agent's autonomy, role in the business process, customer or market impact, and feasibility of intervention.

**8. Named Human Accountability**

Firms should assign clear human accountability for agentic systems and workflows, with the accountability model calibrated to the risk assessment. Accountability for an agentic system refers to identifiable human ownership of the system, its approved use, its control environment, and material decisions or actions taken through the workflow. The appropriate accountability model may vary depending on the firm's governance structure, the agent's role, the degree of autonomy involved, and the risk profile of the use case. As agentic systems evolve, firms may appropriately adapt controls related to identity, access, supervision, and lifecycle management, but governance should remain grounded in accountable human ownership and risk-based oversight.

Firms should also prioritize effective oversight of AI systems, calibrated to the risk assessment. Oversight is distinct from accountability: it concerns the mechanisms used to monitor, review, challenge, escalate, or intervene in AI-driven activity. Human oversight will be an important component, particularly for higher-risk use cases, but effective oversight need not always require real-time human intervention whenever an AI system produces undesired content. Depending on the context, oversight may include preventive controls, approval workflows, automated monitoring, exception-based alerts, sampling and testing, post-event review, escalation procedures, or human-in-the-loop intervention. For AI systems with lower risk profiles, periodic human review or spot-checking may be

appropriate; for AI systems with higher risk profiles, more continuous oversight may be necessary.

When evaluating the effectiveness of oversight, firms should review patterns of approval, modification, escalation, and override. Such review can help identify ineffective oversight, including potential rubber-stamping, recurring overrides that suggest misalignment, or situations where the selected oversight model should be recalibrated. To enable accountability, firms should preserve delegated-lineage records sufficient to reconstruct who authorized material actions, under which scope, and how authority evolved across single-hop or multi-hop workflows where relevant.

**Proportionality and implementation considerations**

We encourage the FSB to adopt a proportionate, risk-based approach to agentic governance that recognizes the role of firm-specific risk assessments in determining the appropriate controls. As agentic AI capabilities evolve, responsible adoption will require proportionate practices that help firms define accountability, constrain autonomy, enforce authorization boundaries, preserve traceability, and maintain mechanisms to monitor, halt, or override agentic behavior where appropriate. Guidance should make clear that firms may calibrate governance expectations based on autonomy, use case, potential customer or market impact, reversibility of actions, and operational context.

The FSB should also recognize that certain practices discussed above remain emerging. Firms should be encouraged to evaluate these practices progressively, prioritize higher-risk use cases, and maintain flexibility as standards, tooling, and supervisory expectations mature.

Given the pace of change, any regulatory guidance should remain operationally grounded through continued engagement with financial institutions, technology providers, standards bodies, and other relevant stakeholders. Guidance that identifies illustrative practices, rather than fixed technical requirements, would better support responsible innovation while allowing firms to adapt controls as the technology evolves.

**Conclusion**

Agentic AI presents significant opportunities for firms, customers, and the financial system, but it may also shift certain risks from bounded model outputs toward execution-time governance. Responsible adoption can be supported by proportionate practices that help firms define accountability, constrain autonomy, enforce authorization boundaries,

preserve traceability, and maintain mechanisms to monitor, halt, or override agentic behavior where appropriate.

We welcome continued engagement with the FSB and other stakeholders to support secure, resilient, and responsible agentic AI deployment across the financial sector.