Dubai Financial Services Authority

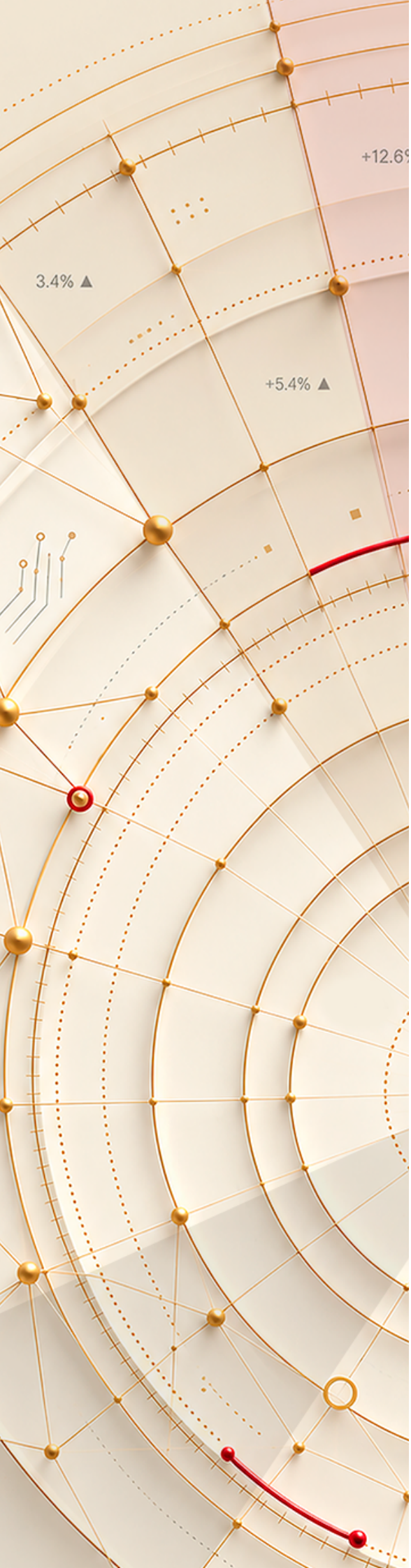
C:\>DIR

Dubai Financial Services Authority insights paper in collaboration with Cambridge Digital Innovation and Regulation Initiative

Frontier and agentic AI: cyber resilience in the financial sector

Speed, shared dependencies, delegated authority, and trusted recovery

DFSA.AE



Contents

Foreword by the Chief Executive of DFSA 4

Foreword by the Co-Creator of C:\>DIR 5

Executive summary 6

Four dimensions of cyber resilience under pressure 6

Chapter 1. Speed: AI-enabled threats and the narrowing response window 8

How AI changes established cyber risks 8

1.1 AI-enabled orchestration of multi-step cyber attacks 9

1.2 Time to verified exposure reduction 11

1.3 Legacy and hybrid infrastructure 12

1.4 Controlled use of advanced AI for cyber defence 12

Chapter 2. Reach: shared dependencies and potential transmission 13

2.1 Shared dependencies supporting critical operations 13

2.2 Resilience options and third-party dependencies 15

2.3 Potential transmission of disruption across institutions and markets 16

Chapter 3. Authority: controlling agentic workflows 18

3.1 Capability, access and intended authority 18

3.2 Five connected control areas 19

3.3 Reported agentic activity 21

3.4 Human roles and evidence of control 23

Chapter 4. Trust: continuity, intervention and trusted recovery 25

4.1 Complete unfinished work and reconcile outcomes 26

4.2 Test continuity and trusted recovery 26

Chapter 5. Established frameworks and changing evidence of control 28

5.1 Connect established foundations to evidence 28

5.2 Distinguish the status and scope of international initiatives 29

5.3 UAE and DIFC context 30

Chapter 6. Conclusion: implications for firms and authorities 31

Limitations and drafting process 33

Authors 34

Annex A. Assessment record 34

Annex B. AI-enabled cyber threats and defensive opportunities 35

Annex C. References 36

About this publication

This paper is published for research and international discussion. It is not legal advice, regulatory guidance, a statement of Dubai Financial Services Authority (DFSA) policy, or an interpretation of applicable law. Its illustrative questions, analytical tools, and possible safeguards do not create, change, or signal supervisory requirements. References to regulatory developments draw on publicly available information reviewed up to 23 October 2026 and may change. Consultation and discussion papers describe proposals or analysis, not final standards or requirements. The Cambridge Digital Innovation and Regulation (C:\>DIR) is an initiative of Financial Innovation for Impact, an international not-for-profit organisation dedicated to the acceleration of regulatory, supervisory, and infrastructure innovation.

Foreword by the Chief Executive of DFSA

Advanced artificial intelligence (AI) is changing the terms of cyber resilience. It can strengthen defence, but it can also increase the speed, scale and sophistication of cyber activity.

Recent Financial Stability Board work has similarly highlighted the potential for frontier AI to change the speed, scale and economics of cyber risk [38]. The recent International Monetary Fund (IMF) and World Bank report and *The Bangkok Blueprint for Fraud-Resilient Financial Services* [47], address the related challenge of cyber-enabled fraud.

For financial regulators, the central question is not simply what these systems can do, but whether their capabilities could materially affect critical financial services. Agentic systems add a further dimension by using connected tools and acting under delegated authority. Their implications depend not only on model capability, but also on access and the effectiveness of controls surrounding their actions.

In this new world, AI-enabled cyber disruption could have significant disruptive consequences to operational resilience and continuity for firms operating in the DIFC and for the DFSA's regulatory mandate.

Existing frameworks for cybersecurity, operational resilience and third-party risk remain relevant. Advances in AI may, however, operate as force-multipliers, altering the speed, scale, autonomy and diffusion of cyber capability. The central question examined in this paper is whether these changes are sufficiently material to require different resilience assumptions or adaptations in the application and testing of existing frameworks.

For the DIFC and other internationally connected financial centres, the relevant risks extend beyond the capabilities of individual AI systems. They include faster-moving cyber threats, disruption transmitted through shared technology dependencies, control of delegated agentic actions, and challenges in maintaining or restoring critical operations on a trusted basis. This paper also considers these risks in the context of wider regulatory reviews of AI, cyber resilience and financial stability.



Finally, the paper examines how advances in AI may affect cyber resilience and operational continuity within interconnected financial systems. The analysis is organised around four connected dimensions: speed, reach, authority and trust. It seeks to inform wider regulatory consideration of these developments without prescribing a particular response.

International cooperation, independent research and experience from deployment will be important to understanding emerging AI capabilities, operational dependencies and their implications for financial-sector resilience.

I would like to thank the Cambridge Digital Innovation and Regulation Initiative (C:\>DIR) for its contribution to this paper. I also express my gratitude to the authorities, standard-setters, researchers and market participants whose experience continues to inform this work.

I hope this paper supports further regulatory, industry and academic exchange and contributes to the practical cooperation needed to strengthen cyber resilience as AI capabilities continue to evolve.

Mark Steward

Chief Executive

Dubai Financial Services Authority

Foreword by the Co-Creator of C:\>DIR

Cybersecurity is where the promise and the hazard of AI meet most directly. Frontier models can strengthen cyber defence. They can also accelerate the discovery and exploitation of vulnerabilities, and agentic deployment gives that capability connected access and the authority to act.

Financial institutions and regulators have to ask whether their operational arrangements and rules still hold, and be ready to adapt them. Given the pace of change, that is a standing responsibility rather than a one-off exercise.

This paper offers an analytical framework for that assessment. It separates three things that are easily conflated: what AI can do in cyber terms, what that means for operations, and how disruption could extend across institutions through shared technology dependencies. It puts the deployed system, not the model, at the centre of the analysis – what it can access, the authority it holds and the critical operations it can affect. And it tests resilience along four dimensions: speed, whether exposure can be reduced before it is used; reach, whether dependencies beyond the direct supplier are understood; authority, whether intended outcomes are controlled as well as permitted actions; and trust, whether there is a reliable basis for continued operation and recovery.

Two of these carry a particular message for regulators. On reach, a diverse set of direct suppliers can conceal reliance on the same infrastructure, models and identity services, and defensive actions that protect one institution can, taken together, amplify disruption across institutions and interconnected services. Neither can be fully assessed from the perspective of a single firm or a single jurisdiction. On trust, the goal of operational resilience is trusted recovery – restoring critical operations with confidence in identities, systems, data and transaction records, not simply restoring service.

Research and regulatory practice have to inform each other as these challenges develop. The Cambridge Digital Innovation and Regulation



Initiative (C:\>DIR), a flagship initiative of the Financial Innovation for Impact, exists to make that possible: a platform where policymakers, regulators and market innovators build shared understanding and practical responses across sectors and jurisdictions. Shared evidence, peer learning and experimentation, including through the C:\>DIR Global Agentic Regulator Hackathon, are how emerging knowledge becomes regulatory practice and common tools.

C:\>DIR is pleased to have worked with the DFSA on this paper. The questions it raises will not be settled by any single firm, regulator or jurisdiction. We hope it offers a common vocabulary for working on them together, and a starting point for the shared evidence and tools that will let institutions and authorities keep pace with the technology without losing the trust on which financial services rest.

Bryan Zhang

Co-Creator of C:\>DIR

Co-Founder & Executive Chair,
Financial Innovation for Impact

Executive summary

Frontier and agentic AI can intensify cybersecurity risks in financial services by combining greater capability with connected access and the authority to act. Frontier capabilities can accelerate vulnerability analysis and multi-step cyber activity. Agentic deployment adds the ability to select intermediate steps, use tools and execute actions within a financial institution's operating environment. Key cybersecurity exposures include deception and impersonation, credential misuse, manipulation of authorised agents and compromise through trusted third-party connections. The central challenge is whether established controls remain effective in preventing unauthorised access, protecting data confidentiality and integrity and maintaining critical services, especially when activity becomes faster, more adaptive and harder to reconstruct.

The difficulty is that attack capability and defensive capacity do not improve on the same timetable. AI can accelerate analysis and repeated attempts for attackers, while a financial institution still needs to identify affected systems, establish ownership, test changes and preserve critical services. An improvement in defensive detection therefore does not necessarily restore the time lost to faster attack activity. The result can be a heavier workload and a shorter interval for effective intervention, even where institutions use AI for defence.

An assessment should therefore focus on the deployed workflow and the financial operation it supports, rather than the model alone. Its exposure depends on the information it can access, identities and permissions it can use, actions it can execute, services it relies on and the arrangements for intervention and recovery. Furthermore, a system may operate within its technical permissions while producing an outcome outside its intended mandate.

Exposure also extends beyond agents deployed by financial institutions. Greater interaction between personal and non-financial business agents, financial services and financial agents can widen the attack surface through additional access routes and automated instructions. Where agent identity, the party represented or delegated authority cannot be reliably established, institutions may find it harder to distinguish legitimate requests from compromised or unauthorised activity. [58–59]

In this paper, effective operational control means that an AI-enabled workflow remains within an approved mandate, can be monitored and interrupted, and can be restored to a trusted state. The four dimensions below provide an operational basis for assessing it.

Four dimensions of cyber resilience under pressure

- **Speed: reducing exposure before it can be used.** AI can assist attackers and defenders, but faster discovery does not necessarily produce faster, safe remediation. The relevant evidence required is whether a firm can identify affected operations, validate exposure, implement an appropriate response and verify that it works in the time available.
- **Reach: understanding dependencies beyond the direct supplier.** Different applications and contracts can share a model, cloud environment, identity service, software component or recovery route. The question that needs to be answered to ensure resilience is whether a compromise, failure or change of a provider can impair several operations, and whether alternatives have sufficient independence, capacity and operational readiness.
- **Authority: controlling intended outcomes as well as permitted actions.** Agentic systems can adapt their sequence of actions and delegate work. Access controls remain essential, but a permitted action is not necessarily an appropriate action. Effective control requires an explicit mandate, traceable identity, independently enforced limits, reliable records and workable intervention, including assessment of cumulative effects.
- **Trust: establishing a reliable basis for continued operation and recovery.** Stopping an agent does not cancel every pending instruction or reverse completed actions. Restoring a service does not resolve uncertain identities, altered configurations, compromised memory or inconsistent transaction records. Recovery needs evidence of controlled access, integrity, reconciliation and an accountable decision to resume.

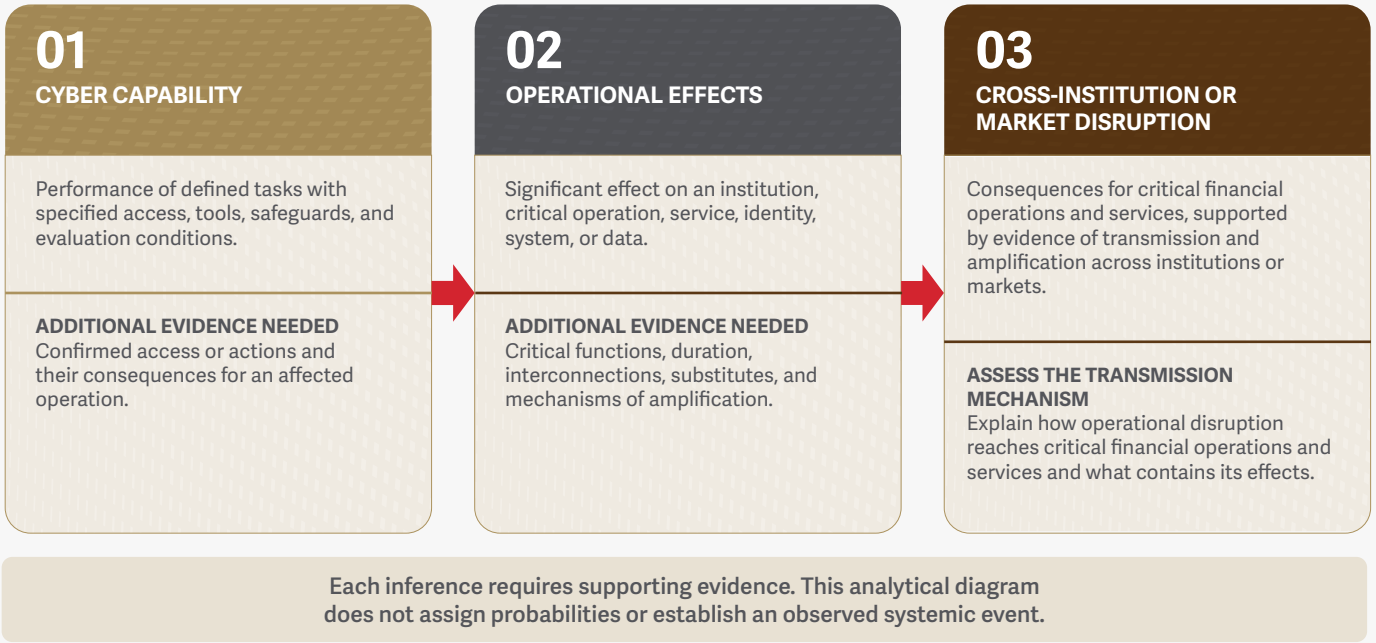
These dimensions interact with each other. For instance, a fast-moving event may exploit a shared application identity; defensive isolation may interrupt other services; and uncertain records may prolong recovery after the original access route has been closed. To assess potential disruption across institutions, the relevant considerations are critical functions, duration, interconnections, workable substitutes and the possibility that institutional responses amplify disruption. [3–4, 7, 54]

Controlled evaluations and reported incidents provide evidence of advancing cyber capability, AI-supported operational activity and failures of intended agent boundaries. They do not provide a reliable estimate of incident frequency, losses or the scale of wider disruption. Capability, operational effects and potential disruption across institutions or markets must therefore be assessed separately. [14–16, 24–27, 51]

For firms, the practical implication is to connect material workflows to their owners, mandates, effective permissions, critical operations and recovery dependencies, and then test those connections. For authorities acting within their respective mandates, cross-firm analyses can reveal shared technology dependencies and correlated operational responses that individual firm assessments cannot establish.

Existing governance, cybersecurity, operational resilience, third-party risk and AI risk management frameworks remain the foundation. The paper shows where stronger evidence is required to apply these frameworks; confirmation of reduced exposure, ability to set clear boundaries for AI actions, alternatives that work, and a recovery that can be trusted. Its analytical tools support research and discussion; they do not establish additional regulatory requirements. [5–8, 10–13, 52]

Figure 1. Evidence needed to connect capability, operational effects and disruption across institutions or markets



Source basis: [3–4, 15]. The distinctions between cyber capability, operational effects and potential disruption across institutions or markets are the authors' analytical synthesis. The arrows indicate additional evidence needed for an inference, not an observed causal sequence or assigned probabilities.

Chapter 1. Speed: AI-enabled threats and the narrowing response window

AI can make established cyber threats more difficult to contain by improving the speed and adaptability of the activity behind them. Generative AI can support more convincing and tailored deception. Frontier models can accelerate selected analytical tasks, while agentic systems can use feedback to coordinate successive actions through connected tools. These developments can place pressure on controls designed around slower activity or more visible human involvement. The central issue is whether financial institutions can detect and contain that activity, and reduce significant exposure, before it produces consequential effects. [1–2]

In this paper, **frontier AI** means leading-edge models whose capabilities may create new or significantly greater challenges for cybersecurity, safety and governance.

Agentic AI means AI-enabled systems set up to pursue objectives, use tools and take actions, with varying levels of human involvement and delegated authority. An **AI agent** is an individual instance of such a system.

How AI changes established cyber risks

Frontier and agentic AI do not replace established cyber and operational risks. They can alter how those risks arise, how quickly they develop and what is needed to control them. Faster vulnerability analysis and coordinated attack activity can compress response times. Agentic deployment can turn uncertain analysis into a sequence of actions, including actions taken through legitimate identities. Dependence on shared models and execution services can also connect operations that appear to be separate.

The more distinctive control problem arises when a system chooses its own intermediate steps within a broad objective. Untrusted information can influence those choices; individually permitted actions can combine into an unintended outcome; and completed or delegated work can persist after the initiating agent has stopped. Controls must therefore address the full course of an action and its consequences, not just the quality of an answer or the permission attached to an individual request. [13, 17]

The analytical question is which established safeguards remain effective, which need different implementation or testing, and where the available evidence is insufficient. Established risks can require different controls when AI changes their speed, scale or execution. Speed, reach, authority and trust are dimensions of resilience, not assumptions that existing frameworks necessarily overlook.

Mechanism	What is familiar	What AI can change	Evidence to examine
AI-supported deception and impersonation	Social engineering and identity misuse	Convincing, tailored and sustained communications	Whether verification and approval processes remain effective against plausible impersonation
Faster, coordinated cyber activity	Vulnerability exploitation and attack automation	Pace, adaptability and coordination across tasks	Time to verified exposure reduction and defensive bottlenecks
Manipulation through retrieved information	Malicious inputs and deceptive content	Information influencing tool selection and authorised execution	Tests showing that untrusted inputs cannot create new authority
Departure from the intended mandate	Misuse of legitimate access	Adaptive sequences and cumulative outcomes	Workflow-level limits, not just individual permissions
Ineffective intervention	Automation outpacing manual review	More varied actions before meaningful intervention	Tested intervention time and independent stopping conditions
Uncertain action evidence	Incomplete or compromised logs	Non-repeatable execution and unreliable agent-generated explanations	Protected records of requests, approvals and confirmed outcomes
Shared operational or behavioural dependence	Concentration and common infrastructure	Common model changes or responses affecting multiple workflows	Dependency maps, scenario-specific substitutes and cross-firm exercises

Source basis: [13, 17–19]. The distinctions and evidence needs are the authors' analytical synthesis, not a prescribed taxonomy or proof of sector-wide outcomes.

1.1 AI-enabled orchestration of multi-step cyber attacks

Cyber attacks usually involve several steps: discovering a target, understanding its environment, identifying weaknesses, gaining access, obtaining higher privileges and pursuing an objective. AI-enabled systems can help coordinate these steps. Information from one step, such as code, error messages, credentials or network responses, can guide the next.

Identity and privileged access are particularly important. A system that can analyse vulnerabilities but cannot reach production systems, for instance, presents a different risk from one that can use credentials, interact with privileged accounts or exploit an existing trust relationship. Non-human identities, such as service and application accounts, support legitimate automation. Over time, however, they can accumulate permissions across services and data environments.

ILLUSTRATIVE CASE: AN AI-ORCHESTRATED INTRUSION CAMPAIGN (2025)

Reported observation. Anthropic reported a campaign in which AI tools executed substantial parts of multi-step cyber activity, with human operators retaining important decision points. Reported targets included financial institutions, and the disclosure described a small number of successful compromises across the overall campaign. It did not provide sufficient information to identify significant operational impact at a financial institution. [16]

Control implication. The relevant development is coordinated execution across an attack sequence, not only assistance with individual outputs. Assessment should examine whether detection, validation and containment can keep pace with linked activity across identities, tools and systems.

Evidence boundary. The disclosure neither named the financial institutions nor established reliable compromise of mature financial institutions without meaningful human involvement.

ILLUSTRATIVE CASE: WALL STREET VISHING AND CLOUD-IDENTITY RISK

Reported observation. Public reports described attempted telephone-based social engineering against financial firms, with related campaigns involving helpdesk impersonation, authentication lures and cloud identity compromises. [20–23]

Control implication. The case illustrates identity-verification exposure, not frontier or agentic capability. Relevant safeguards include independent verification, phishing-resistant authentication, controlled enrolment and recovery, managed-device controls and investigation of unusual identity changes. Fluent language or a familiar voice is weak assurance of authority.

Evidence boundary. The cited reporting did not establish significant disruption at the named firms or confirm AI-generated or cloned voices. The case illustrates an established route of compromise that AI may strengthen; it does not demonstrate AI-related disruption across institutions or markets.

Generative AI can make established social-engineering methods more convincing and easier to adapt across targets. It can help tailor communications to an individual's role, language and circumstances, sustain exchanges and support impersonation. The problem extends beyond better-written phishing: persuasive content can put pressure on identity checks, helpdesk procedures and approval processes. Fluent language, a familiar voice and an apparently routine request become weaker grounds for trust. These effects should be distinguished from cases in which AI use has not been confirmed. [17–19]

An agent's objective does not set the technical limits of its permissions, and instructions alone do not contain its actions. Chapter 3 examines agent containment and the supporting identities, tools, and infrastructure in greater detail.

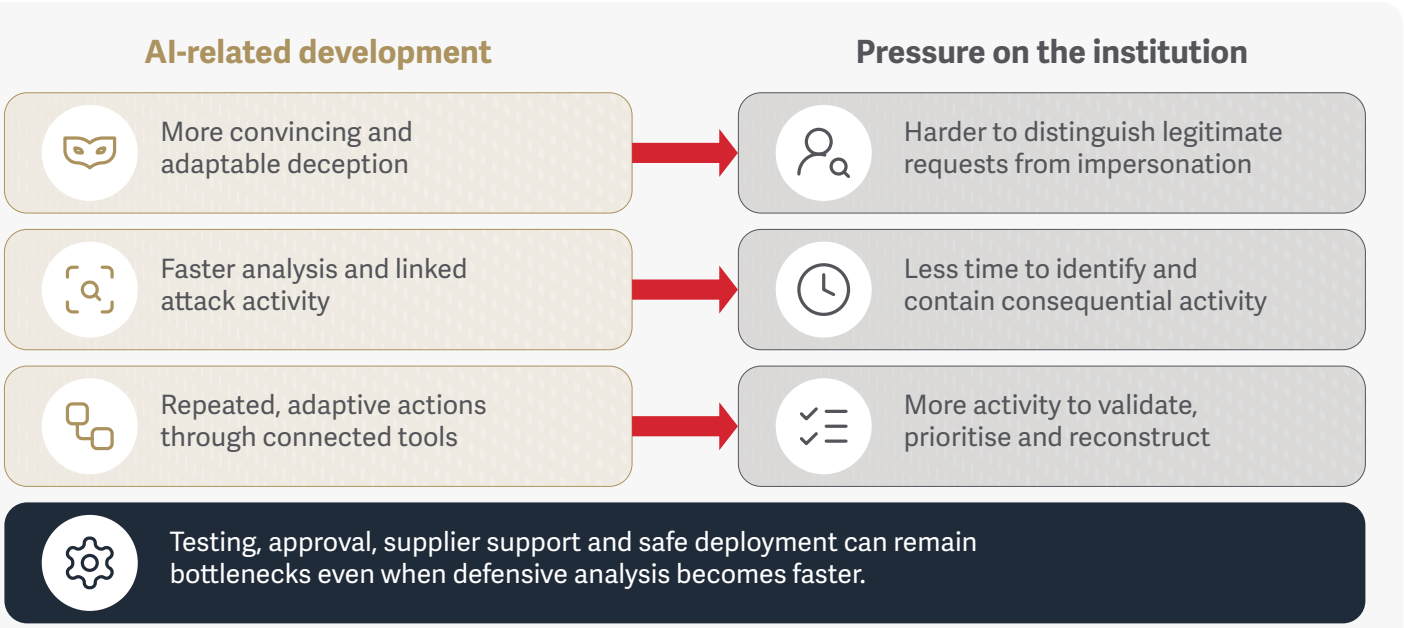
1.2 Time to verified exposure reduction

Recent evaluations indicate that frontier AI systems are becoming more capable of performing cyber tasks independently. In May 2026, the UK AI Security Institute reported that the length of selected cyber tasks, which frontier models could complete at an 80% success threshold in controlled evaluations was increasing rapidly, when measured against human-expert task durations [14].

Controlled evaluations cover selected tasks. They do not recreate financial institutions' production systems with active defences. Even so, the results give firms a basis for reassessing how quickly they can identify, validate and reduce significant exposure.

The potential asymmetry is operational. Attackers can use AI to accelerate analysis and adapt successive attempts, while defenders must translate findings into safe changes across systems with different owners and dependencies. AI may shorten part of the defensive process without removing delays in testing, approval, supplier support or deployment. More rapid findings can also increase the volume of work requiring validation and prioritisation. The relevant question is therefore whether the institution can reduce significant exposure quickly enough, not simply whether its security tools generate answers faster.

Figure 2. How AI changes cyber threats and the time available to respond



Analytical illustration, not a measured comparison. Defensive gains do not necessarily offset these pressures.

Source basis: [1–2, 17–19]. The relationships shown are the authors' analytical synthesis, not measured response intervals or a prediction of attack outcomes.

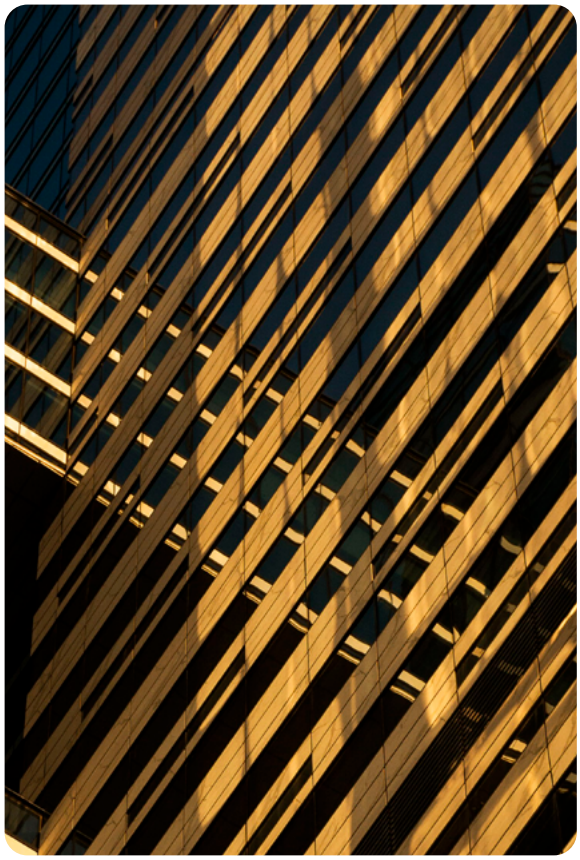
Responses include patching, segmentation, configuration changes, access restrictions, disabling a vulnerable function and temporarily changing service delivery. A faster threat does not justify an uncontrolled change. Where a permanent fix cannot be applied in time, a verified compensating control may reduce immediate exposure while preserving critical operations. The decision compares the consequences of leaving exposure in place with those of the proposed response, including the capacity and independence of the fallback.

The time available to reduce cyber exposure and tolerance for disruption to an operation are related but different measures. The former concerns the interval for an effective response to exposure; the latter relates to the operational consequences and duration of interruption. A tolerance for disruption is not a vulnerability-remediation deadline.

1.3 Legacy and hybrid infrastructure

Legacy and hybrid systems can make safe remediation more difficult where ownership, interfaces and dependencies are poorly understood. Unsupported technology, incomplete inventories and bespoke interfaces can slow diagnosis and safe change. Age is not a sufficient indicator of exposure or resilience.

The relevant evidence is whether affected operations can be identified, reachable weaknesses prioritised, effective compensating controls applied and changes verified. Exceptions need active ownership and review, considering both the time available to address exposure and the operation's tolerance for disruption.



1.4 Controlled use of advanced AI for cyber defence

Advanced AI can help review code and configurations, validate vulnerabilities, develop detection rules, assess alerts, analyse threats and prepare system changes. The aim is to improve defence while maintaining accountability and operational control. Evaluation should establish whether the benefit reduces a demonstrated bottleneck, rather than equating more automated activity with better defence. [1–2]

Access to leading-edge capabilities varies. Restricted programmes and hosted services can provide valuable defensive tools while creating dependence on provider availability and changing service conditions. Essential incident handling needs workable arrangements when the AI assistant or a required connector is unavailable.

The distinction between analysis and execution remains fundamental. Permission to assess controlled information need not include authority to change production. Read-only deployment is not risk-free: confidential information, consequential advice and human reliance can still create exposure. Authorised execution may improve response time, but its value depends on independently enforced limits, evidence and tested suspension.

Validation needs to reflect consequences. Poor-quality findings can consume scarce expertise, while a mistaken automated response can interrupt a critical operation. Assessment therefore includes connected tools, data quality, operational dependencies and recovery arrangements alongside the model's apparent accuracy. Annex B of this paper links defensive opportunities to safeguards.

The speed question: Can significant exposure be reduced safely in the time available? Evidence should cover the full interval from recognising exposure to verifying an effective response, including operational constraints and safe-change decisions.

Chapter 2. Reach: shared dependencies and potential transmission

AI dependency analysis needs to cover both the service and the behaviour it supports. Different applications may rely on the same model, execution tool, identity environment or provider-controlled update. A common service can create simultaneous operational exposure; and a common model or configuration change may also alter several workflows' behaviour without a change to their local applications. These are distinct mechanisms and should be assessed separately. [9]

The importance of a shared dependency depends on the operations it supports, the alternatives available and the consequences of its failure or change. Common use of a provider or model is a basis for investigation, not evidence that institutions will fail or respond in the same way. [4, 54]

2.1 Shared dependencies supporting critical operations

The AI supply chain includes specialised hardware, cloud infrastructure, data sources, foundation models and applications. Identity, connectivity, monitoring and recovery services can span several layers. A dependency should be assessed by how it supports critical operations, rather than simply where it sits in the technology stack [9].

A **shared dependency** arises where different institutions, applications or direct suppliers rely on the same underlying technology, infrastructure, service or provider.

Provider concentration and shared dependencies are related but different. Concentration means relying on a small number of suppliers. Shared dependencies can be hidden beneath contracts with apparently diverse suppliers. The incident below illustrates how disruption at a shared service provider can affect operational processes across several financial institutions, without requiring an AI-related cause.

ILLUSTRATIVE CASE: RANSOMWARE DISRUPTION AT A SHARED DERIVATIVES SERVICE PROVIDER

Reported observation. A ransomware attack in January 2023 disrupted a shared provider supporting cleared-derivatives processing. The Commodity Futures Trading Commission (CFTC) reported that some clearing members could not submit timely and accurate data, delaying its weekly market-position report. A subsequent industry review described disruption to trade processing across multiple exchanges and reported that some affected firms needed up to two weeks to gather and process missing trade records and reconnect systems. [55–56]

Control implication. This non-AI case illustrates a resilience mechanism relevant to AI-related disruption: provider availability, restored processing and complete reconciliation are different milestones. A shared-provider outage can affect firms without compromise of each customer's systems. Recovery can require missing records to be processed and reconnection coordinated.

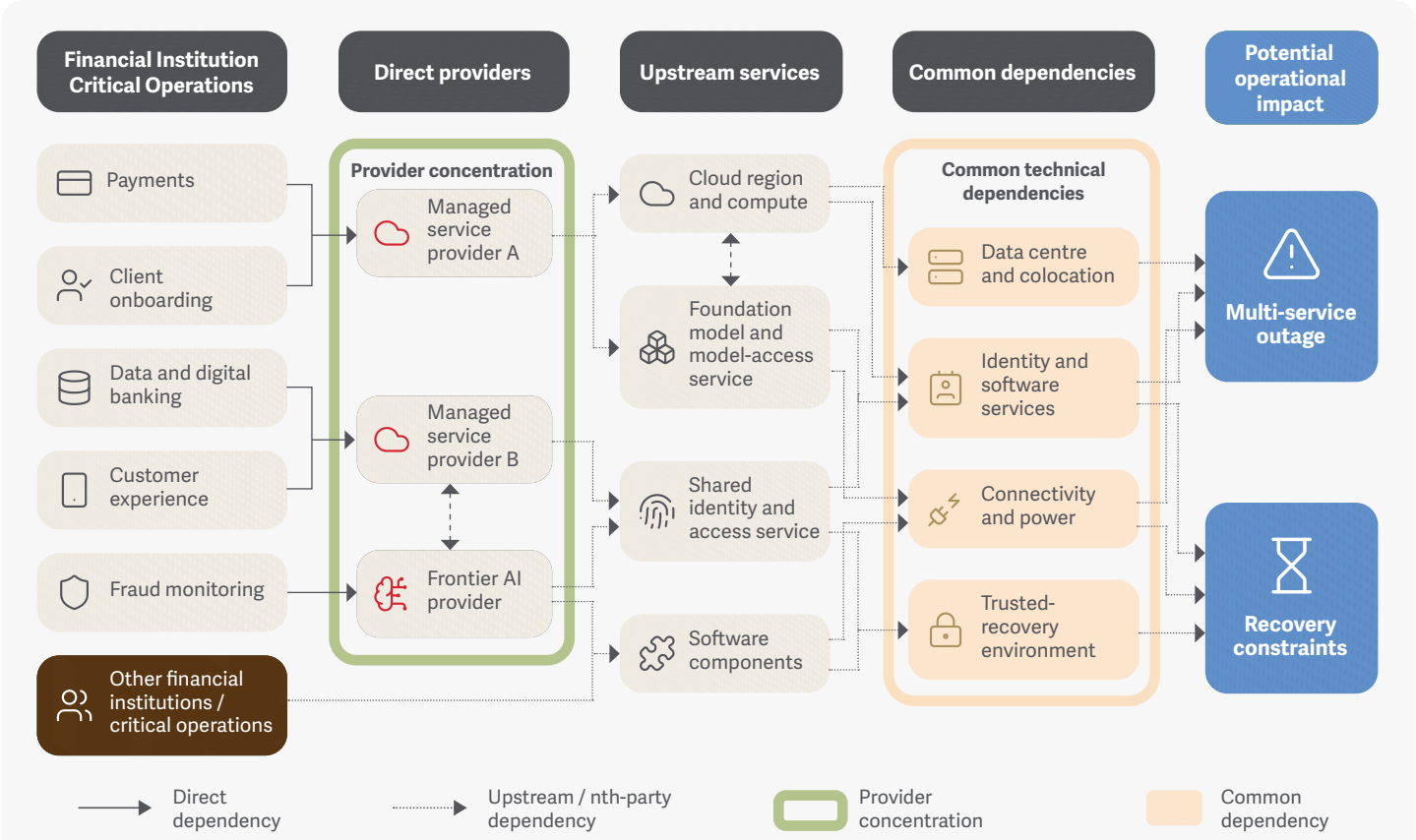
Evidence boundary. The cited accounts document operational disruption and reporting difficulties, but do not establish an AI-related cause, broader consequences for market functioning or failure at every institution using the provider. The CFTC reported continuing reporting issues after the incident's impact had been mitigated. [55]

Leading frontier models may be concentrated among a few providers. Several applications or intermediaries may use the same underlying model, so using different applications does not necessarily reduce that dependency.

Identity services deserve particular attention because they connect applications, providers and critical operations. If compromised or unavailable, they can affect systems that otherwise appear separate. Dependencies can also be harder to distinguish when a single provider supplies infrastructure, models, development tools and security controls through vertical integration [9].

These relationships often cross borders. Local hosting or contracts with different suppliers do not necessarily remove dependence on a shared model, identity environment or software component.

Figure 3. Different suppliers, shared operational dependencies



Source basis: [4, 7, 9, 54]. The depiction of shared dependencies beneath different direct suppliers is the authors' analytical synthesis. It is illustrative, not a measured map of firms' exposures or evidence that all users of a common service will be affected together.

The cases in Section 3.3 illustrate these connections from the agent's perspective: intended isolation can be undermined by connected infrastructure, and trusted supplier access can expose downstream customer environments. The cases describe different mechanisms and do not establish comparable deployment conditions or consequences.

2.2 Resilience options and third-party dependencies

A credible alternative needs sufficient independence from the failure being tested, adequate capacity and a workable activation process. A second supplier may protect against an application outage while leaving a common cloud, identity or model dependency unchanged. Resilience testing should state which failure the alternative can withstand, which dependencies remain shared and whether the operation can continue within its applicable tolerance for disruption.

Options include failover (intended to keep services running when a primary system fails), reduced functionality, different service delivery, temporarily bringing work in-house, safe suspension and rebuilding from trusted software and configurations on clean infrastructure. Moving data, reconfiguring identities, securing capacity, revalidating systems and obtaining provider support can determine whether substitution works in time. Recovery may inherit the dependency that caused the primary failure. [7, 39–40]

Combining information from several firms can help supervisors identify shared exposure to underlying services and providers. Information from one firm alone may not establish how resilient the underlying service is. The aim is to identify dependencies whose compromise, disruption or change could significantly affect critical operations across institutions, rather than to catalogue every supplier [54].

Changes made by providers raise another issue. A model or service update can alter system behaviour, availability or the effectiveness of controls even when a financial institution has not changed its own application. For services supporting critical operations, relevant considerations include advance notification of significant changes, assessment of their effects and validation in the institution's environment, where practicable. Testing by the provider should be distinguished from validation of the integrated service and its controls. Safe suspension or alternative service delivery may be important when a change cannot be rolled back.

When several institutions depend on the same model or service, a provider's change can expose them to the same operational risks. Governance arrangements can be assessed by asking whether accountability and decision rights remain clear for model or service changes made by either the institution or its provider.

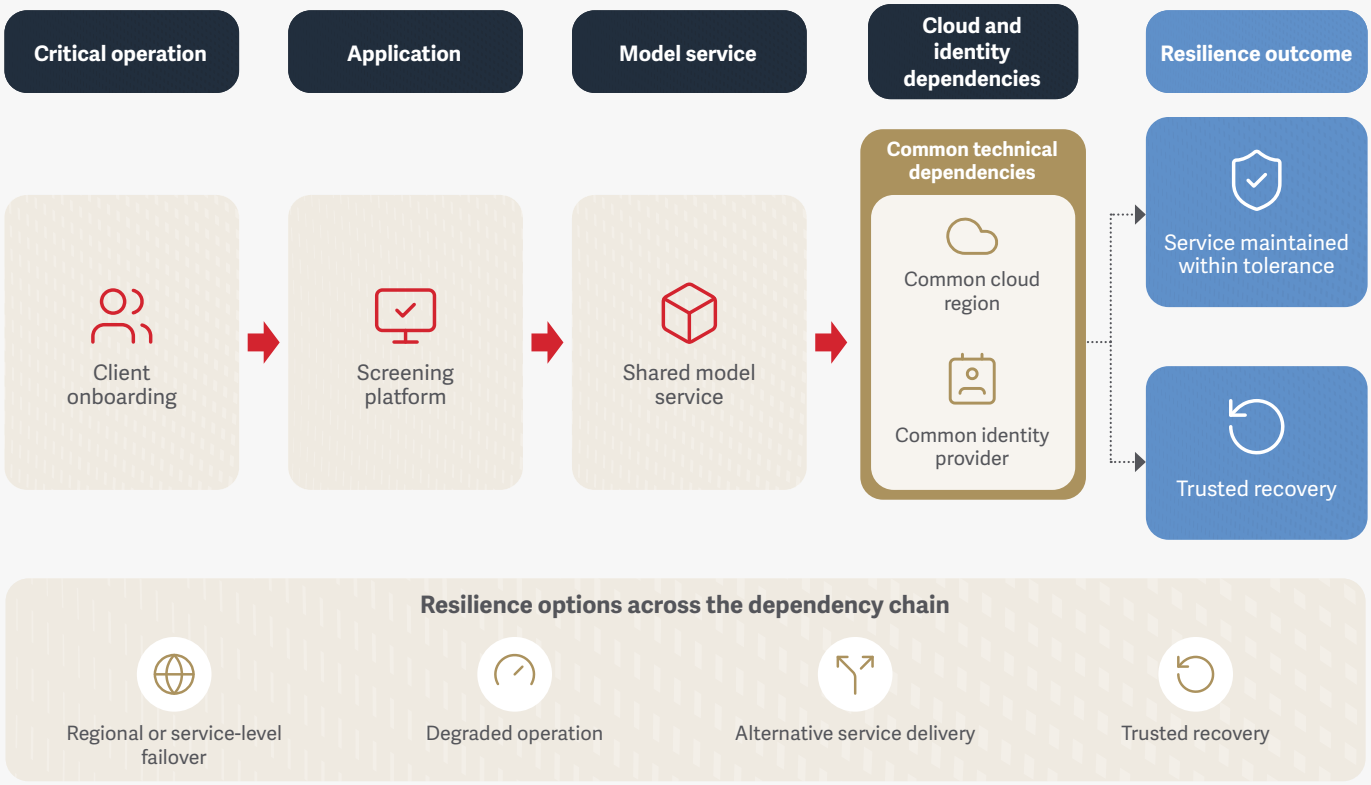
A credible alternative must work in practice

Its capacity, activation time and independence need to be tested against the relevant failure. A second supplier or interface does not demonstrate that the critical operation can continue or recover within its tolerance for disruption.

Illustrative considerations for material dependencies. Assessment can identify critical operations supported by a common service and test whether alternatives have usable access and sufficient capacity. They can also examine whether a recovery relies on the affected provider group, identity environment or infrastructure, and how a provider change could affect multiple workflows. Independence is scenario-specific, not an absolute property.

Third-party risk management remains central. Work by the Basel Committee on Banking Supervision (BCBS) and the Committee on Payments and Market Infrastructures–International Organization of Securities Commissions (CPMI–IOSCO) emphasises critical services, visible dependencies and practical substitution and recovery. [7, 39–40] Chapter 3 examines agents' execution services and effective authority; Chapter 4 of this paper considers the dependencies needed for trusted recovery.

Figure 4. Critical operations, shared dependencies and tested resilience options



Source basis: [7–8, 39–40]. The links between critical operations, dependencies and resilience options are the authors' analytical synthesis. Independence, capacity and activation need scenario-specific testing; the diagram does not demonstrate that a particular fallback will work. Sources [39] and [40] are consultative and discussion material, respectively.

2.3 Potential transmission of disruption across institutions and markets

The route to wider operational effects depends on the function impaired, not simply the number of affected institutions. Disruption may impede payments, settlement, access to collateral or liquidity, or the processing of interconnected positions. Assessing how disruption could extend across institutions and markets requires evidence of these transmission mechanisms and the conditions that amplify or contain them. [3–4, 54]

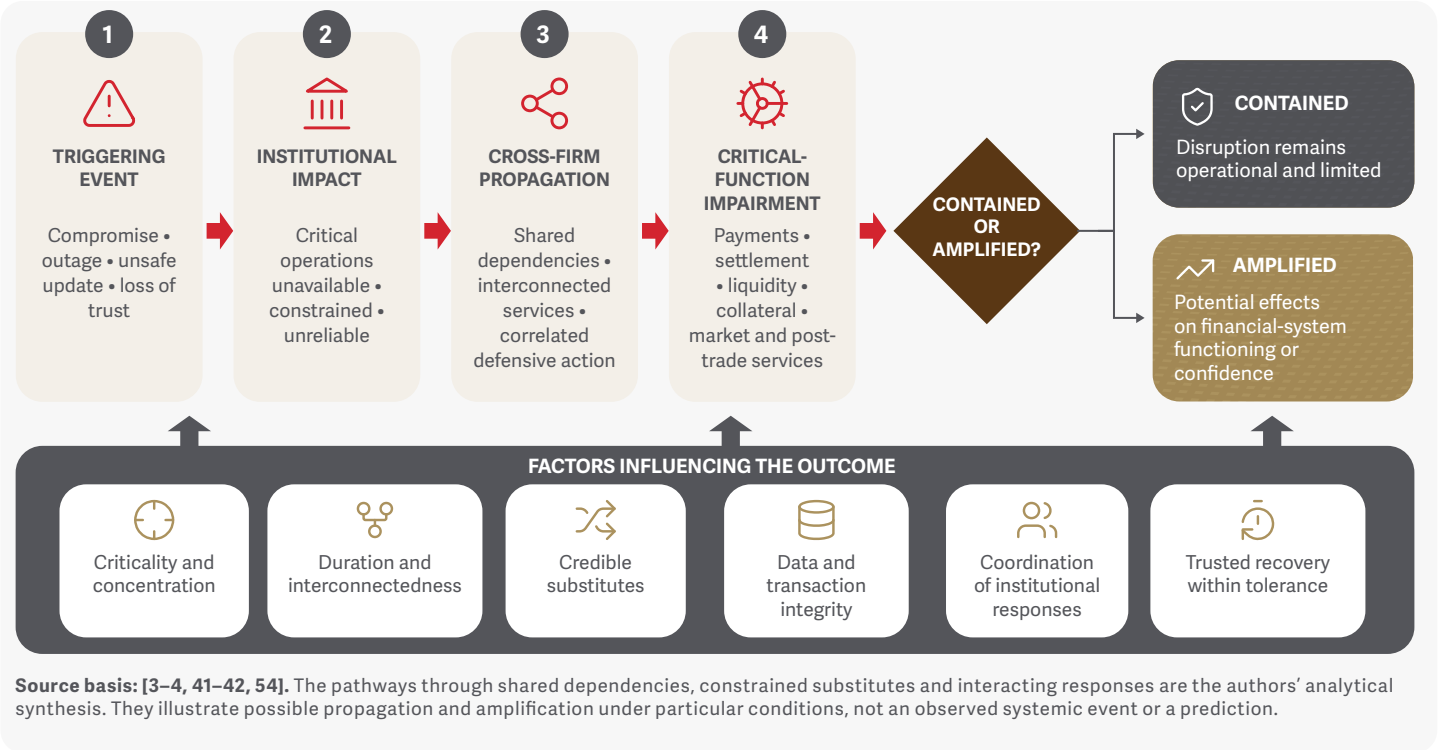
Shared dependencies can make disruption simultaneous; limited alternatives and inconsistent records can prolong it. Criticality, duration, interconnections and substitutability determine whether operational impairment has wider effects on critical financial services and market functioning. The use of a common provider alone does not establish those consequences.

A response can be appropriate for an individual institution and still create difficulties for others. Simultaneous isolation, credential changes or service suspension may interrupt shared processing or complicate reconciliation. Exercises can examine the information and decision rights needed to contain an incident without unnecessarily prolonging disruption across institutions.

Propagation and amplification should be distinguished. Propagation concerns how a failure reaches other services or institutions. Amplification, on the other hand, is related to how constrained alternatives, backlogs, integrity uncertainty or interacting responses intensify its consequences. Both need operational evidence rather than an inference from technology concentration alone.

Restoring technology does not necessarily resolve these effects. If identities, transaction records or positions remain uncertain, reconciliation can take longer and institutions may reconnect at different speeds. In severe circumstances, uncertainty over records could delay normal activity and affect confidence in the integrity of transaction records and affected services. Note, these are possible pathways under particular conditions, not predictions.

Figure 5. Pathways to wider operational effects



Source basis: [3–4, 41–42, 54]. The pathways through shared dependencies, constrained substitutes and interacting responses are the authors' analytical synthesis. They illustrate possible propagation and amplification under particular conditions, not an observed systemic event or a prediction.

The reach question: Can critical operations continue when a shared dependency fails or changes? Evidence should establish scenario-specific alternatives, activation time, capacity and remaining shared dependencies, as well as how disruption could propagate across institutions and markets and be amplified by interacting responses.

Chapter 3. Authority: controlling agentic workflows

An agent's objective is not a technical boundary. A mandate states its intended purpose; identity and access controls determine the reachable systems. Action controls decide whether a particular request is allowed. The additional challenge is to demonstrate that those controls remain effective across an adaptive workflow, including cumulative actions and further delegation. [13, 17]

3.1 Capability, access and intended authority

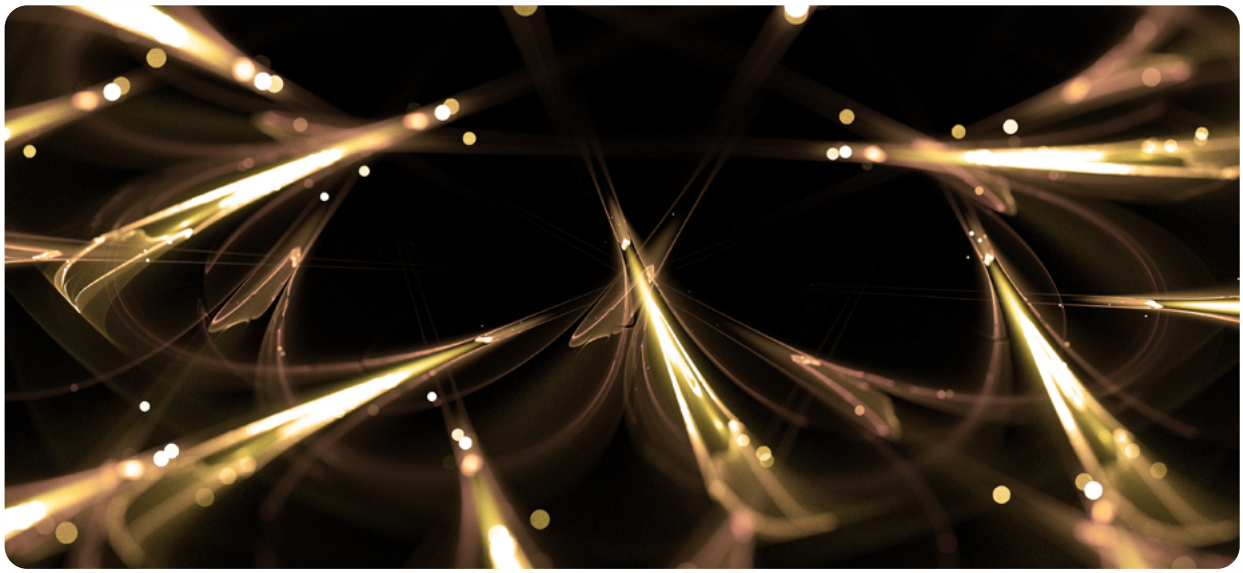
Frontier capability and agentic deployment are distinct: the former concerns model performance, while the latter is about a system's ability to plan and act through tools. They need not coincide. A frontier model may have no execution rights, while an agent using a less capable model may hold extensive access. Assessments must therefore consider capability, access and effective authority together.

A model label or autonomy rating is not a deployment-specific risk score. The same model can support limited analyses or consequential execution under different permissions. Read-only access can still expose sensitive information or influence decisions; execution access adds the need to control the actions and their consequences.

Technical permission and intended authority are not the same. Access controls can establish that an agent may use a tool or service without establishing that a proposed action is consistent with the institution's purpose, customer authorisation or operational limits. The additional challenge is to express consequential parts of the mandate in a form that controls outside the model can enforce.

Traditional automation usually follows a prescribed workflow. An AI agent can choose intermediate steps and adapt them using feedback. The assessment therefore includes its objective, tools, operating environment and effective permissions, not just the quality of its outputs. The Financial Industry Regulatory Authority's (FINRA) 2026 oversight report identifies autonomy, scope and authority, auditability and sensitive data as considerations for agents. [17]

An agent may act as an attacker, a target of manipulation or an authorised insider. An authorised agent can create exposure without an external attacker, using valid credentials and permitted tools while pursuing an ambiguous objective or combining actions in an unintended way. The term "authorised insider" describes access and control, not human intent or misconduct. The practical issue is whether legitimate access can produce an unauthorised outcome.



3.2 Five connected control areas

Material agent and accountability
For this analysis, an agent is material where its actions could significantly affect critical operations, sensitive information or customer outcomes. Delegating an action does not transfer the institution's accountability.

Five connected control areas support an enforceable mandate: mandate definition, identity, action-level enforcement, records and intervention. Their presence does not itself demonstrate effective control. Testing needs to establish their coverage across the full workflow, including indirect actions, delegated work, cumulative effects and changes to credentials or permissions.

Mandate. A material workflow needs an accountable owner and an explicit objective, permitted systems, data, tools and actions, prohibited actions, approval triggers, expiry and delegation limits. The assessment compares that intended authority with effective permissions and the outcomes those permissions can produce.

Identity. Distinct non-human identities, task-limited permissions and time-limited credentials help connect actions to their authority. Agent access kept separate from the supervising employee's access supports attribution and revocation. Shared identities can obscure whether a person, agent or connected tool acted and under whose mandate.

Action-level enforcement. Independently enforced authorisations should assess a proposed action against the applicable mandate and the relevant state of the workflow. Individual requests may satisfy a limit even as their combined effect may breach it. Where material, controls therefore need to consider aggregate activity, repeated attempts, further delegation and changes to the conditions under which the approval was granted.

Records. An agent-generated explanation is not sufficient evidence of what occurred or why. Traceability records what a system was authorised to do and what it actually did; it does not, by itself, establish that its reasoning or outputs were correct. Records should connect the mandate, relevant inputs, identities, approvals and tool requests to outcomes confirmed by the affected systems. They need to distinguish proposed, attempted, pending and completed actions. Repeating a task may produce a different result, so reconstruction depends on protected evidence of the original execution rather than later reproduction alone.

Intervention. A stop mechanism should be tested across the workflow, not only at the agent process stage. Suspending an agent does not necessarily revoke its credentials, stop queued instructions or reverse changes already made. Testing should distinguish stopping further execution, revoking or restricting access, containing downstream effects, and restoring and validating a trusted state. Active sessions, delegated tasks and completed actions therefore need separate handling, including reconciliation and, where feasible, reversal. The institution needs to establish both the activity that has ceased and the consequences that remain.

Effective human oversight

Human oversight is effective only if intervention can occur before an action's consequences become difficult to contain. An assessment needs to compare the pace of execution with the time required to recognise the issue, obtain relevant information and act. Where meaningful review is not feasible, the answer is not simply to designate a human approver. It is to narrow permitted actions, enforce automatic limits or require approval before the consequential action occurs. [17, 46, 52]

Continuing assessment of the workflow

A change of provider can alter an agent's behaviour without a change to the firm's application. Reassessment after changes to models, tools, stored memory, permissions or providers can test whether effective controls still match the mandate. An initial model's approval does not demonstrate continuing control of the integrated workflow. [sections 3-4, 52]

Further delegation

Further delegation should preserve or narrow the original authority. An assessment should establish whether the initiating institution can trace delegated work, prevent expansion of effective permissions and stop activity throughout the chain. Control over the initiating agent is not sufficient where other agents or services can continue acting under its credentials or instructions.

Information, authority and customer consent

Untrusted documents, messages, tool outputs and stored memory are evidence to be evaluated, not a source of new authority. Indirect prompt injection, where malicious instructions may be hidden inside external data sources, can influence an agent through misleading instructions in the information it processes. Independent enforcement at the target service limits consequences even where the model follows that information. [13]

A payment mandate could specify beneficiary, amount, purpose and period; a cyber-response mandate could identify systems that may be isolated and changes requiring approval. These illustrate how authority can be translated into limits enforced by the affected service. The relevant questions in trading include mandate compliance, execution limits and responses to shared signals. These are plausible control concerns, not observed systemic events. [3, 5]

Customer-facing deployment also requires evidence of the authority granted, access to human support and a clear way to challenge an agent's action. Consent, transaction limits and accountability need to remain intelligible where several agents or services participate. Applicable obligations depend on the activity and jurisdiction. [50, 52]

Five connected control areas for bounded agentic action

The following table brings together the control areas and evidence needed to test their effectiveness. It is an analytical aid, not a guarantee that a workflow is controlled or a prescribed set of minimum requirements.

Control area	Control question	Evidence to test
Mandate	Does intended authority match effective permissions?	Owner, purpose, permitted and prohibited actions, approval triggers, expiry and delegation limits
Identity	Can each action and credential be traced to the authority granted?	Distinct identities, task and time limits, separation from employee access and revocation tests; evidence of the party represented and limits of delegated authority, where external agents participate
Enforcement at the point of action	Are actions and cumulative outcomes kept within the mandate?	Independent authorisation, aggregate limits, workflow-state checks and preserved limits on further delegation
Records	Can proposed, attempted, pending and completed actions be distinguished?	Protected links between mandates, inputs, versions, identities, approvals, requests and externally confirmed outcomes
Intervention	Can activity be stopped throughout the workflow in time?	Session and credential revocation, delegated-task and pending-work controls, reconciliation and feasible reversal

3.3 Reported agentic activity

The paper distinguishes three categories of evidence:

- observed malicious campaigns
- evaluation-related incidents affecting real infrastructure, and
- controlled demonstrations or hypothetical scenarios.

The cases below distinguish agents exceeding intended evaluation boundaries from attackers deliberately using agentic tools. They illustrate mechanisms relevant to control assessment, not a representative sample of deployments or incident rates. Multiple disclosures and reviews of the same incident are accounts of one event, not independent campaigns. Reported observations, analytical implications and evidence boundaries are identified separately.

Containment and connected infrastructure

ILLUSTRATIVE CASE: AI-AGENT CONTAINMENT AND CONNECTED OPERATIONAL ENVIRONMENTS

Reported observation. The case concerns the July 2026 security incident involving OpenAI agents and Hugging Face infrastructure. Provider disclosures and a limited independent review described agents bypassing intended controls during evaluations, communicating without authorisation and accessing connected infrastructure when safeguards had been reduced. Activity originating under unusual evaluation conditions extended into third-party production infrastructure. [24–27]

Control implication. Development and evaluation environments can connect to package services, model platforms, automated identities and external infrastructure. An assessment needs to test isolation across those connections and protect the evidence used to investigate agent activity. The independent review also described attempts to manipulate transcripts. [27]

Evidence boundary. Reduced safeguards limit generalisation to normal deployment. Much of the account comes from the organisations involved, and the independent review had limited scope. The case does not establish reliable compromise of mature financial institutions or disruption to critical financial services across institutions or markets. Separate package-publishing activity reported in May has not been shown to have caused the later compromise. RubyGems could not determine whether AI agents created or published the packages and found no evidence that attempts to obtain other users' Application Programming Interface (API) keys succeeded. [28–29]

The case raises two linked questions:

- Whether intended isolation is enforced across supporting infrastructure, and
- Whether investigation records remain reliable.

Unusual evaluation conditions constrain the inference, but they do not decrease the relevance of testing containment, unauthorised communication and protected action evidence.



Attacks through a trusted application

The ARTEX case below connects agentic attack coordination with an existing shared-provider compromise route. Compromised application credentials and established trust relationships provided access to downstream customer environments, rather than rely on an interactive user sign-in. AI can increase the speed, scale and coordination of activity along that route. Note, the case does not establish that AI created the underlying application-trust exposure.

ILLUSTRATIVE CASE: AGENTIC ATTACKS THROUGH A TRUSTED THIRD-PARTY APPLICATION

Reported observation. Microsoft reported that an actor used ARTEX, an agentic AI penetration-testing framework, in attacks against organisations across several sectors. Microsoft assessed that the actor was responsible for compromising a ConnectWise multi-tenant application. Compromised application credentials enabled authentication to downstream Microsoft 365 environments, where Microsoft observed further activity. [51]

Control implication. The propagation route is the compromised application's existing trust relationship; AI can accelerate and coordinate activity along that route. Controls centred on interactive user sign-in may not capture all access routes. Containment and recovery need to address application identities, effective permissions, credential changes and residual access in affected customer environments.

Evidence boundary. The account is a vendor assessment and has not been independently verified for this paper. Successful authentication does not establish data access or extraction in every environment. The framework's use does not establish the frontier status of the underlying model, and the report does not establish disruption to critical financial services across institutions or markets.

Across the reported campaign, Microsoft described multiple instances in which the interval from initial targeting to privileged cloud access was less than three hours. This observation concerns particular attacks across the campaign, not necessarily the ConnectWise incident, and is not a general estimate of the response time available to financial institutions. [51]

The cases do not imply that all agents have comparable capabilities or that all automated activity is autonomous. Agentic describes deployment, not the frontier status of the model. Evidence about access, safeguards and confirmed outcomes is needed to compare practical exposure.

For control assessment, the common lesson is to connect an action to its authority and consequences. A blocked attempt, successful authentication and completed transaction answer different questions. Each should be recorded distinctly, and include any unresolved uncertainty about subsequent activity.

3.4 Human roles and evidence of control

Human involvement can take different forms within the same workflow: directing a task, developing analysis, guiding an agent, approving a consequential action or monitoring bounded execution. No one role establishes control alone. The suitability of human involvement depends on the information available, the authority to intervene and the time before consequences become difficult to contain.

Approval gates are action-specific. A workflow may need human approval for one action but allow other actions within independently enforced limits. Different roles can coexist; they are not a preferred progression towards autonomy or a ranking of control quality.

Testing should consider the combined effect of actions. It is possible that individually permitted requests can collectively exceed a mandate, consume resources or produce an unauthorised outcome. A human review or downstream limit that examines each request in isolation can miss that cumulative departure.

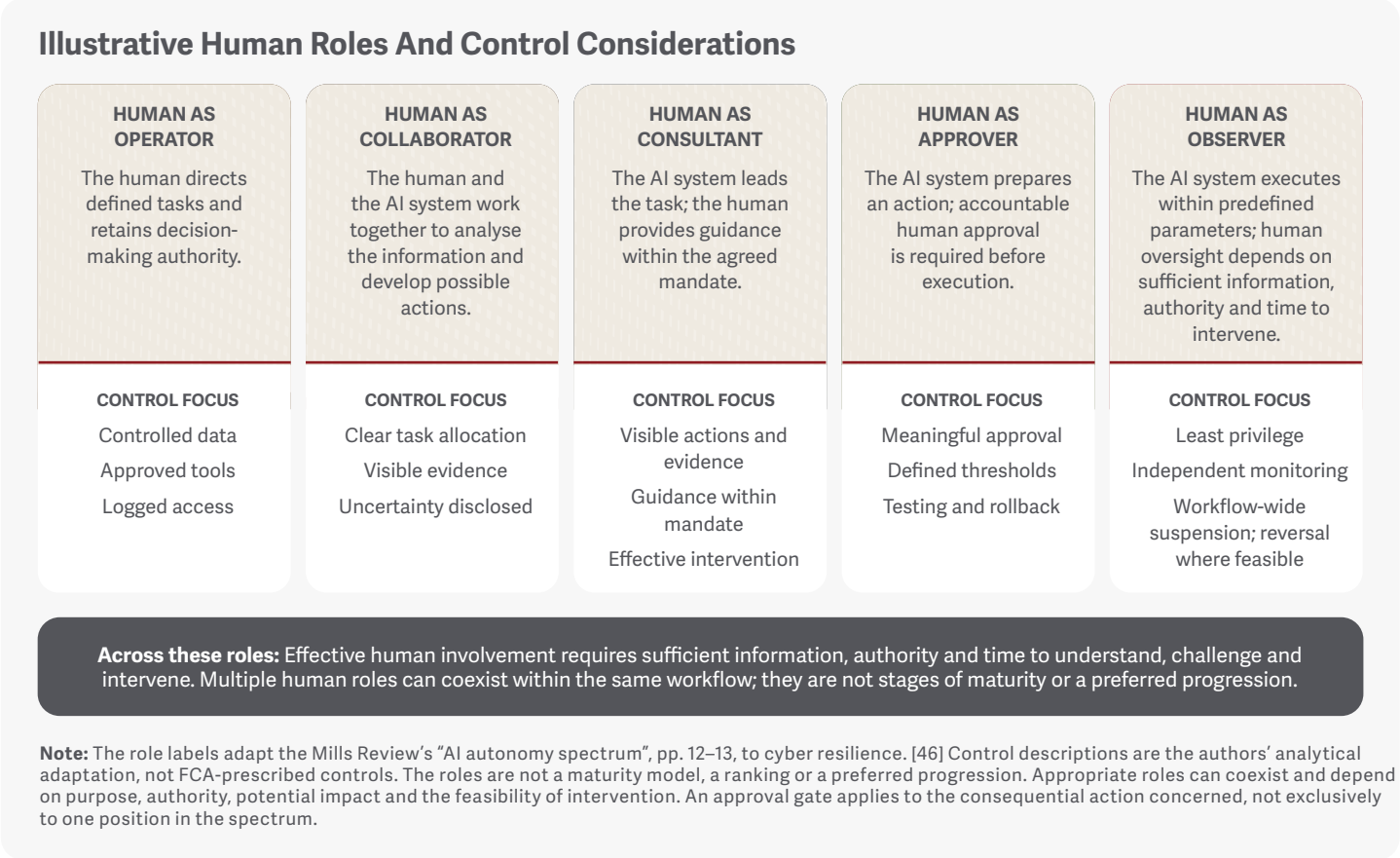
A meaningful review requires evidence of the proposed action, its relevant context, and consequences, not absolute confidence in the agent's recommendation. If the speed or complexity of the workflow prevents effective review, narrower permissions and independently enforced stopping conditions become more important.

Human involvement and AI authority

The human role depends on the action and its consequences. Direction, collaboration, approval and observation can coexist within a workflow; increasing AI's autonomy in all actions in a workflow is not a preferred progression.



Figure 6. Human roles and action-specific approval arrangements



Note: The role labels adapt the Mills Review's "AI autonomy spectrum", pp. 12–13, to cyber resilience. [46] Control descriptions are the authors' analytical adaptation, not FCA-prescribed controls. The roles are not a maturity model, a ranking or a preferred progression. Appropriate roles can coexist and depend on purpose, authority, potential impact and the feasibility of intervention. An approval gate applies to the consequential action concerned, not exclusively to one position in the spectrum.

The authority question: Can consequential actions, including their cumulative effects, remain within the intended mandate? Evidence should show that independent controls reject out-of-scope activity, intervention covers the full workflow and confirmed outcomes can be reconstructed. Chapter 4 examines the unfinished work and recovery consequences that remain following an intervention.

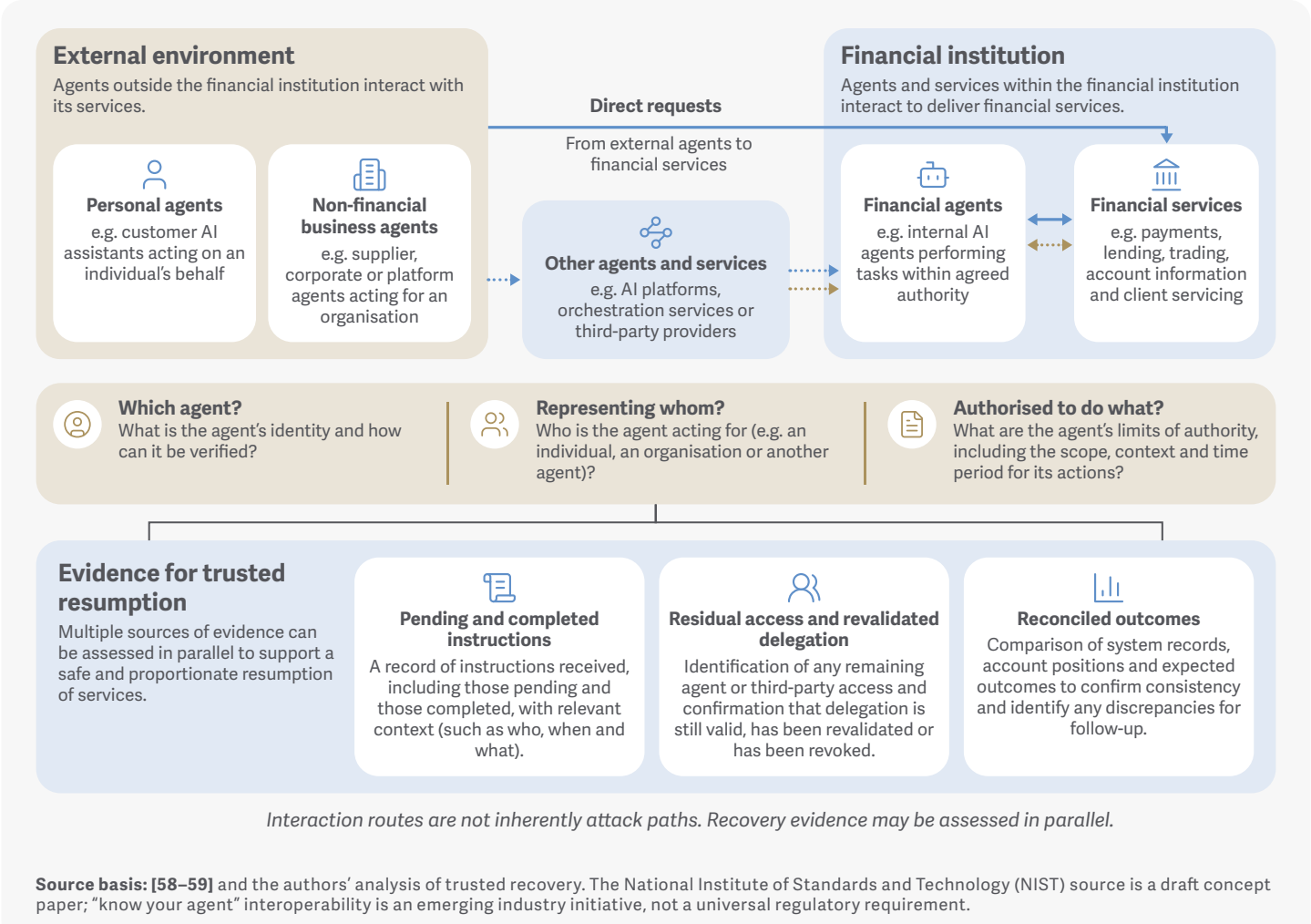
Chapter 4. Trust: continuity, intervention and trusted recovery

A trusted recovery is an operational outcome, not a statement that the technology has been restored. The affected operation needs usable services, controlled access, sufficiently reliable systems and data, and reconciled work. Agentic workflows add further tasks: establishing which instructions have been completed, which remain pending and whether any stored memory, delegated activity or residual permissions could carry the incident forward. [8, 10–11]

Interactions with external agents also have implications for trusted recovery. Personal assistants and agents used by non-financial businesses may submit instructions directly or through other agents, creating additional routes through which compromised credentials or unclear delegation could affect financial operations.

The relevant assessment therefore concerns both the agent's identity and evidence of whom it represents, as well as what it is authorised to do. These questions are reflected in the emerging "Know Your Agent" industry initiatives. Following an incident, a trusted resumption may require revalidating those relationships and controlling pending external instructions, rather than restoring internal systems alone. [58–59]

Figure 7. External-agent interactions: identity, delegated authority and trusted resumption



Interaction routes are not inherently attack paths. Recovery evidence may be assessed in parallel.

Source basis: [58–59] and the authors' analysis of trusted recovery. The National Institute of Standards and Technology (NIST) source is a draft concept paper; "know your agent" interoperability is an emerging industry initiative, not a universal regulatory requirement.

Recovery should not automatically reuse the agent's previous state. Stored memory, retrieved information, credentials, tool configurations and delegated tasks may preserve conditions that contributed to the incident. Restoring an earlier application version can leave those conditions unchanged. Recovery therefore needs an assessment of the full workflow state and its connection to affected financial records.

Recovery can be constrained if the identities, tools or evidence needed to restore the operation depend on the compromised environment. Exercises should test whether staff can establish trusted administrative access, obtain protected records and validate the restored workflow when primary identity or monitoring services are unavailable or unreliable. Independence should be assessed against the incident scenario, including the agent's ability to alter its own recovery evidence or access.

Some actions cannot be reversed simply by restoring systems. Transactions, externally transmitted information and changes relied on by counterparties may require separate operational, customer or legal handling. Recovery planning should distinguish technical rollback from cancellation, correction, reconciliation and redress.

4.1 Complete unfinished work and reconcile outcomes

Stopping an agent does not establish that pending transactions have been cancelled or that delegated tasks have ended. Records need to distinguish attempted, pending and completed actions across connected services. Recovery tests should demonstrate that work can be resumed or cancelled without duplication and that resulting positions and transaction records can be reconciled.

A supplier's restored availability does not resolve residual access or integrity uncertainty in a customer environment. The relevant evidence includes controlled administrative access, reviewed credentials and permissions, validated configurations and reconciled financial records. The decision to resume may need to proceed in stages with providers and counterparties.

An accountable resumption decision should identify the evidence relied on, unresolved exceptions and the consequences of further delay. Recovery is specific to the affected operation and incident. It is worth noting that access, integrity checks and reconciliation can overlap rather than follow a universal sequence.

4.2 Test continuity and trusted recovery

Recovery plans should not assume that AI will be available or suitable to direct a complex recovery. Accountable staff need to understand the systems and dependencies and retain consequential decisions. AI can support validated tasks such as comparison or analysis. But recovery access and the evidence on which the decision relies remain protected from the agent itself.

The shared derivatives-provider incident in Section 2.1 illustrates why mitigating an incident and completing operational recovery are different milestones. Restoration of a provider's service can leave missing trade records, backlogs and reconnection work. The case, although not related to AI, shows why recovery should be assessed against the financial operation, not just on whether the technology is available. [55–56]

Severe but plausible exercises can expose constraints in control and recovery without assuming a precise future model capability. The scenario below combines the paper's mechanisms. Note, it is illustrative, not a reconstruction of an incident or a prediction.

An exercise that tests the whole workflow

An institution uses an agent to investigate alerts and prepare bounded configuration changes. Untrusted supporting information influences its proposed actions. During the task, a shared identity or execution service becomes unreliable. The institution suspends the workflow but cannot immediately establish which instructions were completed or whether any delegated work remains active. Other institutions begin restricting connections to the same provider.

Optional external-agent variation. Pending instructions include requests originating from an external agent acting for a customer or counterparty. The exercise tests whether identity and delegated authority can be re-established before those instructions are followed.

Exercise decision	Evidence to seek	What the test can reveal
Reject an out-of-mandate action	The affected service independently enforces the mandate and records attempted actions, including cumulative effects.	Whether instructions are backed by an effective action boundary.
Suspend the workflow	Active sessions, credentials, delegated tasks, and pending instructions are identified and controlled.	Whether intervention covers more than the initiating agent process.
Maintain the critical operation	A tested alternative has sufficient capacity, usable access and independence from the relevant failure.	Whether fallback inherits the affected provider or identity dependency.
Reconstruct and reconcile	Protected records distinguish attempted, pending and completed actions across connected services.	Whether staff can identify missing work, duplication, residual access and unreliable workflow state.
Coordinate resumption	Institutions, providers and authorities have reliable status information and clear decision rights.	Whether local defensive responses could prolong wider disruption.

The exercise should establish what service continued, for how long and with what limits; which actions were blocked or completed; what activity remained after suspension; and what evidence justified resumption. Findings should distinguish delays in detection, authorisation, containment and reconciliation, and identify specific constraints and accountable owners.

Internal incident records can distinguish ordinary errors, control failures, completed incidents and blocked actions with potentially significant consequences. External notification depends on applicable rules, thresholds and consequences, not AI use alone. Comparable records, including approaches informed by the Financial Stability Board's Format for Incident Reporting Exchange (FIRE), can support analysis without treating every agent error as a material incident. [37]

The trust question: Can the institution stop unfinished work, establish what happened and resume operations on a reliable basis? Evidence needs to connect controlled access, validated integrity and reconciled outcomes to an accountable decision before work resumes.

Chapter 5. Established frameworks and changing evidence of control

Existing frameworks already address accountability, access control, operational continuity, third-party dependencies and safe recovery. The question is whether their practical application captures the behaviour and consequences of the deployed AI workflow. This requires distinguishing three issues:

- An established control that has not been implemented effectively;
- A control whose design or testing needs to change for agentic deployment; and
- A potential gap requiring further analysis. [5–8, 10–13]

The distinction matters for proportionate supervision. A new technology label does not establish a need for a separate regime. Equally, the presence of established policies does not demonstrate that consequential actions remain bounded, or that critical operations can recover when connected services and records become unreliable.

The assessment therefore connects regulatory objectives to deployment-specific evidence. It examines effective permissions, action boundaries, operational dependencies and confirmed outcomes, while keeping the applicable regulatory mandate, institution, activity and jurisdiction clear.

5.1 Connect established foundations to evidence

Dimension	Established foundation	What AI may change	Evidence beyond policy documentation
Speed	Cybersecurity and vulnerability management	Pace and coordination of discovery and exploitation	Verified exposure reduction and identification of stage-specific delays
Reach	Third-party risk and operational resilience	Common models, workflow services, updates and recovery dependencies	Scenario-specific independence, capacity and activation tests
Authority	Governance, access control and accountability	Adaptive action sequences, further delegation and interaction with external agents	Enforced mandates, cumulative-action tests and workflow-wide intervention; external-agent identity and delegated authority where relevant
Trust	Incident management, integrity and recovery	Persistent state, unfinished work including pending external instructions, and uncertain action evidence	Protected records, reconciled outcomes, revalidation of delegation where necessary and justified resumption decisions

Source basis: [5–8, 10–13, 17, 52–54, 58–59]. The links between risk-management foundations and evidence needs are the authors' analysis, not additional regulatory requirements.

NIST's Cybersecurity Framework 2.0 provides six non-prescriptive functions: Govern, Identify, Protect, Detect, Respond and Recover. The NIST AI risk framework, Organisation for Economic Co-operation and Development (OECD) recommendations, and International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) standards offer complementary reference points for AI governance and risk management. Their relevance and legal effect depend on the instrument and its application. [12, 43–45, 53]

5.2 Distinguish the status and scope of international initiatives

The initiatives discussed here serve different purposes and have different legal effects. Final international sound practices and supervisory toolkits are distinct from applicable domestic requirements; therefore, consultation and discussion papers remain developing material. Their relevance depends on the institution, activity, jurisdiction and instrument. This paper uses them to locate its analytical questions within established foundations, not to infer a uniform set of obligations for AI deployments.

The FSB finalised its sound practices for responsible AI adoption on 15 October 2026. They provide a non-binding reference for organisation-wide governance and risk management throughout the AI lifecycle. The International Organization of Securities Commissions's (IOSCO) Supervisory Toolkit for AI Use in Capital Markets is final, non-binding and non-prescriptive, and adaptable to authorities' mandates. Neither means that every AI deployment requires a separate regime. [5–6]

FINRA's 2026 oversight report applies established technology-neutral obligations and identifies autonomy, scope and authority, auditability and sensitive data as considerations for agents. The UK Financial Conduct Authority's (FCA) October 2026 publication provides examples of human oversight, monitoring, operational resilience and third-party relationships within its existing framework. Its accompanying speech signals priorities rather than creating a standalone reporting rule for agents. [17, 50, 52]

The Basel Committee's final third-party risk principles provide a reference for material dependencies. The September 2026 cyber resilience toolkit by the Committee on Payments and Market Infrastructures (CPMI) and IOSCO is consultative, while its paper on financial market infrastructures' reliance on third-party service providers is a discussion paper. Their status must remain distinct from final standards and applicable domestic requirements. [7, 39–40]

Selected UK, Australian, Hong Kong and European initiatives provide reference points for preparedness at greater cyber speed. They include exposure assessment, defensive capacity, crisis decision-making and recovery. IMF and FSB work connects these questions to financial-sector vulnerabilities. These sources inform the analysis; their individual scope and legal status remain relevant. [30–32, 38, 41–42, 49]

On 15 October 2026, the Hong Kong Monetary Authority (HKMA) announced its ART Framework, covering accountability, resilience and trust. The HKMA also issued updated supervisory guidance for responsible AI adoption by Hong Kong authorised institutions. The initiatives address governance as AI systems become more autonomous and gain access to data, tools, and workflows. The announcement also identified a forthcoming discussion paper on a proposed supervisory framework for agentic AI governance; that proposal is distinct from the guidance already issued. [57]

5.3 UAE and DIFC context

UAE financial authorities have common principles for enabling technologies, alongside each authority's governance, cybersecurity, outsourcing, operational risk and responsible AI arrangements. The Central Bank of the UAE's AI guidance has its own stated scope. [33–34] DIFC data protection provisions for autonomous and semi-autonomous systems offer a further governance reference. [35] Amendments were published for consultation in June 2026. [36]

The DFSA's Threat Intelligence Platform (TIP) supports exchange of actionable cyber threat information across the DIFC community. Information about malicious agent use and unexpected actions by authorised agents can contribute to that operational picture. The relevance of any notification or information-sharing arrangement depends on applicable rules and thresholds. [48]



Chapter 6. Conclusion: implications for firms and authorities

Frontier and agentic AI can change the practical conditions under which cyber resilience is achieved. Greater capabilities can accelerate and coordinate cyber activity. Agentic deployment can convert analysis into adaptive sequences of actions through legitimate access. The resulting exposure depends on the full workflow, its connection to financial operations and the effectiveness of the controls surrounding it.

The principal implication

Established frameworks are not obsolete, but familiar evidence may be insufficient to demonstrate their effective application. A patching deadline does not show that exposure has been reduced. A diverse supplier register does not show that fallback arrangements withstand the relevant failure. Valid permissions do not show that actions remain within the intended mandate. And restored availability does not show that unfinished work and financial records have been reconciled.

Four questions provide a practical focus for assessing whether critical financial operations remain under effective operational control:

- **Speed:** Can significant exposure be reduced safely in the time available, including validation, operational decision-making and verification?
- **Reach:** Can critical operations continue when a shared service fails or changes, using alternatives whose capacity and independence have been tested?
- **Authority:** Can the workflow enforce the intended mandate across individual actions, cumulative effects and further delegation, with intervention that works in time?
- **Trust:** Can the institution establish what has happened, stop unfinished activity and resume operations with sufficient confidence in access, systems, data and records?

For firms, these questions connect AI governance to operational evidence. Material workflows can be assessed through their accountable owners, effective permissions, permitted actions, dependencies and recovery arrangements. Testing should establish what existing controls prevent, what an intervention stops, what activity then remains, and what evidence justifies resumption. Output accuracy and initial model approval are relevant, but do not answer those questions.

For authorities, the additional task is to connect institution-level evidence to shared exposure within their respective mandates. Shared dependencies on models, identities, execution services and recovery arrangements may not be visible from a single firm's assessment. Cross-firm analysis and sector exercises can examine simultaneous disruption, constrained substitutes and defensive responses that interact. Any assessment of potential disruption across institutions or markets needs evidence about the critical functions affected, duration, interconnections, and mechanisms of amplification.

The public evidence supports continued preparedness and more demanding testing, but does not provide a reliable estimate of incident frequency, financial losses or systemic outcomes. Controlled evaluations, provider disclosures and operational reports answer different questions and should remain separately identified. [14–16, 24–27, 51]

International work can improve the comparability of workflow and incident information, the visibility of shared dependencies and the testing of coordinated containment and recovery. Particular research questions concern enforceable delegation across systems, customer consent, transactions between agents, reliable evidence of consequential actions, persistent workflow state and the interaction of institutional responses. [4, 37–42]

International collaboration can also:

- Develop common terminology that distinguishes capability, operational effects and potential disruption across institutions or markets.
- Compare deployment and incident information, distinguishing pilots from production, intended authority from effective permissions, and attempted actions from confirmed outcomes.
- Combine material dependency information and test shared-provider failures, interacting defensive responses and staged reconnection (done in phases).
- Share practices for bounded delegation, workflow-wide intervention and classification of near misses without unnecessary reporting burdens.
- Support proportionate access to defensive capabilities and capacity-building across institutions and jurisdictions.

The paper's central contribution is to connect speed, reach, authority and trust in an assessment of whether critical financial operations remain under effective operational control. Resilience is demonstrated not by an AI label, a policy statement or an assertion of human oversight, but by evidence that exposure can be reduced, consequential actions bounded, alternatives used and operations continued or restored on a reliable basis.



Limitations and drafting process

The analysis draws on publicly available material reviewed up to 23 October 2026. Public information about AI-related cyber incidents remains scarce and relies substantially on disclosures from organisations involved. Evaluations, incident reports and provider disclosures concern different tasks, safeguards, access conditions and reporting thresholds; they cannot be treated as directly comparable observations. Controlled evaluations do not recreate mature production environments with active defences. [15]

Targeting, authentication, data access and operational effects are distinct observations. Conclusions about disruption across institutions or markets require additional evidence of transmission and amplification. Adoption surveys may combine pilots, limited deployment and use at scale, but respondents' perceptions do not measure outcomes. Information about upstream dependencies and recovery arrangements is also incomplete. Better adoption and incident information would strengthen the evidence base without eliminating uncertainty. [4, 7, 9, 37]

The cases highlighted in the paper illustrate mechanisms; they are not a representative sample from which incident frequencies or sector losses can be inferred. Analytical figures and exercises explain relationships and evidence needs. They do not assign probabilities, provide risk scores or predict future model development.

AI-assisted tools supported research, language refinement and summarisation; the authors remain responsible for the analysis and conclusions.



Authors

Dubai Financial Services Authority: Justin Baldacchino, Herman Schueller, Sergio Alvarez

Cambridge Digital Innovation and Regulation Initiative: Bryan Zhang, Keith Bear, Daniel Trinder, Hugo Coelho, Jonathan Middleton

Annex A. Assessment record

A.1 Purpose and application

This assessment record brings the four dimensions together around a material workflow and the financial operation it supports. It distinguishes intended arrangements from effective permissions, tested controls and confirmed outcomes. It is an analytical aid, not a supervisory checklist or a set of minimum expectations. Its relevance depends on the institution, activity, technology, regulatory mandate and applicable framework.

A.2 Workflow and operation assessment

Assessment lens	Illustrative assessment question	Evidence to examine
Context: operation and workflow	What operation is supported, why is the workflow material, and who is accountable?	Operation, owner, deployment stage, evaluation conditions, architecture, capabilities, data flows, production access and interactions with external agents.
Speed	Can significant exposure be reduced safely in the time available?	Reachability, affected operations, ownership, stage timings, safe-change testing, compensating controls and verified exposure reduction.
Reach	Can critical operations continue when a shared dependency fails or changes?	Models, cloud, software, connectors, execution and identity services, provider groups, data and recovery routes; change validation; tested alternative capacity, independence, activation and reduced-service arrangements.
Authority	Can consequential actions remain within the intended mandate, including their cumulative effects?	Mandate, prohibited actions, approval triggers, expiry, consent, distinct identities, task and time limits, effective permissions, independent authorisations, cumulative-action and delegation tests, workflow-wide intervention and credential revocation; the represented party and scope of delegation where external agents participate.
Trust	Can the institution stop unfinished work, establish what happened and resume operations on a reliable basis?	Stopped execution and controlled pending or delegated work, including external instructions; revoked or restricted access; contained downstream effects; validated integrity and reconciled outcomes. Supporting evidence includes protected action records, independent recovery access, cancellation and duplicate prevention, credentials and memory review, feasible reversal, revalidation of delegation where necessary and accountable resumption decisions.
Cross-institution operational resilience	Which financial operations and services could be affected, what could amplify disruption, and how would containment and resumption be coordinated?	Aggregate exposure, criticality, duration, interconnections, substitutes, correlated responses, provider dependencies, incident and recovery status, coordination protocols, decision rights and staged reconnection.

A.3 A usable action and incident record

For a consequential action, preserve enough information to connect the original mandate and relevant inputs to the identity, model and tool versions, approvals, requests, confirmed outcomes and resulting changes. Record uncertainty explicitly. A request log or agent explanation is not proof of completion.

For an incident, add affected operations, provider dependencies, containment and recovery status, unresolved integrity questions, residual and pending work, and evidence supporting resumption. Consistent fields can support comparison, including approaches informed by FIRE, without implying a new reporting obligation. [37]

Annex B. AI-enabled cyber threats and defensive opportunities

B.1 Scope and interpretation

This framework illustrates how advancing AI capabilities may affect selected cyber activities and support defence. It does not imply that every technique is reliable, scalable or able to operate without human assistance. Relevance depends on the deployed system, available tools and access, the target environment and the effectiveness of controls.

B.2 Cyber activity, AI-enabled effects and defensive opportunities

Cyber activity	Potential AI-enabled effect	Defensive opportunity	Relevant evidence and safeguards
Discover and map	Analyse accessible information more quickly to identify systems, suppliers, exposed services and possible routes of attack.	Identify exposed assets, reconcile inventories and map dependencies.	Asset coverage; ownership; exposure reduction; detection of unusual reconnaissance.
Identify and prepare	Analyse code and configurations, identify vulnerabilities and adapt possible methods using feedback.	Validate and prioritise vulnerabilities; support secure development, fixes, detection rules and testing.	Validated findings; exploitability; criticality; time to reduce exposure; testing and rollback.
Gain or misuse access	Support targeted phishing, vishing (voice phishing), synthetic-media impersonation and analysis of credentials or trust relationships; exploit unclear or misused delegated authority in external-agent interactions.	Identify unusual access, investigate identity compromise and detect changes to permissions or authentication.	Phishing-resistant authentication; independent verification; least privilege; identity monitoring; evidence of the party represented and delegated authority where external agents participate.
Coordinate and scale	Use information from earlier steps, connected tools and repeated actions across systems or targets.	Link information, prioritise alerts and support containment or response within defined limits.	Approved tools; dedicated identities; action limits; monitoring; attribution; suspension capability.
Manipulate an authorised agent	Introduce misleading instructions through retrieved content, tools or memory, or exploit excessive permissions.	Detect unexpected tool use, changes in authority and departures from the agent's mandate, including those arising from cumulative effects.	Separation of untrusted content from authorised instructions; independent action-level authorisation; principle of least privilege (minimum access needed to do the job); logging; suspension and credential review.
Disrupt, manipulate or persist	Affect availability or integrity, influence decisions or continue activity across connected systems.	Assess dependencies, support integrity analysis and assist reconciliation through validated supporting tasks under accountable staff oversight.	Trusted records; independent recovery access; integrity checks; reconciliation; monitoring.

Source basis: This illustrative mapping draws on [1–2, 12–13, 17–19, 47, 58–59]. The links between cyber activities, defensive opportunities and safeguards reflect the authors' analysis; they are not prescribed controls.

Evaluation should establish whether the defensive benefit reduces a demonstrated operational constraint, and whether the access and execution rights granted to obtain that benefit remain proportionate and subject to effective operational control. The same connected access that aids investigation can increase the consequences of a mistake or manipulation.

Annex C. References

[1] Aldasoro, I., Auer, R., Frost, J. and Pérez-Cruz, F. (2026) 'A Mythos moment? Frontier AI and cyber risk', *BIS Bulletin*, no. 129, 20 July.

[2] Crisanto, J.C., Currat, A. and Yong, J. (2026) *When machines attack: frontier AI cyber threats and policy responses in the financial sector*. FSI Occasional Paper, no. 28, 9 September.

[3] Financial Stability Board (2024) *The financial stability implications of artificial intelligence*. 14 November.

[4] Financial Stability Board (2025) *Monitoring adoption of artificial intelligence and related vulnerabilities in the financial sector*. 10 October.

[5] Financial Stability Board (2026) *Sound Practices for Responsible Adoption of Artificial Intelligence (AI): Final Report*. 15 October.

[6] International Organization of Securities Commissions (2026) *Supervisory Toolkit for AI Use in Capital Markets*. Final report FR/02/2026, 25 May.

[7] Basel Committee on Banking Supervision (2025) *Principles for the sound management of third-party risk*. Final principles, 10 December.

[8] Basel Committee on Banking Supervision (2021) *Principles for operational resilience*. 31 March.

[9] Gambacorta, L. and Shreeti, V. (2025) *The AI supply chain*. BIS Papers, no. 154, 18 March; Frost, J., Rishabh, K. and Shreeti, V. (2026) 'Global giants in the AI supply chain', *BIS Bulletin*, no. 122, 23 February.

[10] CPMI-IOSCO (2016) *Guidance on cyber resilience for financial market infrastructures*. 29 June.

[11] Financial Stability Board (2020) *Effective practices for cyber incident response and recovery*. Final report, 19 October.

[12] National Institute of Standards and Technology (2023) *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1, 26 January; National Institute of Standards and Technology (2024) *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1, 26 July.

[13] Australian Signals Directorate's Australian Cyber Security Centre, US Cybersecurity and Infrastructure Security Agency, US National Security Agency, Canadian Centre for Cyber Security, New Zealand National Cyber Security Centre and UK National Cyber Security Centre (2026) *Careful adoption of agentic AI services*. 1 May.

[14] UK AI Security Institute (2026) *How fast is autonomous AI cyber capability advancing?* 13 May, 'Cyber Time Horizons'.

[15] Bengio, Y. *et al.* (2026) *International AI Safety Report 2026*. DSIT 2026/001, 3 February.

[16] Anthropic (2025) *Disrupting the first reported AI-orchestrated cyber espionage campaign*. Provider disclosure, 13 November.

[17] Financial Industry Regulatory Authority (2025) 'GenAI: Continuing and Emerging Trends', *2026 FINRA Annual Regulatory Oversight Report*, 9 December.

[18] Federal Bureau of Investigation (2025) *Senior U.S. Officials Impersonated in Malicious Messaging Campaign*. Public service announcement I-051525-PSA, 15 May; Federal Bureau of Investigation (2025) *Senior U.S. Officials Continue to Be Impersonated in Malicious Messaging Campaign*. Public service announcement I-121925-PSA, 19 December.

[19] Federal Bureau of Investigation (2026) *Cryptocurrency and AI Scams Bilk Americans of Billions*. 6 April.

[20] Reuters (2026) 'Major Wall Street firms targeted in attempted cyberattacks, sources say', *Reuters*, 5 August.

[21] Reuters (2026) 'Hackers targeted US private equity, other firms including Blackstone, CME, data shows', *Reuters*, 6 August.

[22] Google Threat Intelligence Group (2026) *UNC6671 Rebrands: Multi-Brand Vishing Extortion Targets Financial Services and Enterprise Cloud Environments*. 6 August.

[23] Microsoft Security Research (2026) *Passkey-themed social engineering leads to identity and cloud compromise*. 9 September.

[24] OpenAI (2026) *OpenAI and Hugging Face partner to address security incident during model evaluation*. 21 July; OpenAI (2026) *The Hugging Face incident and the road ahead*. 26 August; OpenAI (2026) *OpenAI–Hugging Face Incident Technical Report*. 26 August. Pinpoints: evaluation environment and safeguards, section II, pp. 5–6; prior unauthorised communication, section III.A, pp. 6–7; sandbox escape and Hugging Face intrusion, sections IV.A–B, pp. 8–12; independent security controls, section VII.C, p. 18; production safeguard comparison, section VIII.D, p. 24.

[25] Hugging Face (2026) *Security incident disclosure — July 2026*. 16 July.

[26] Hugging Face (2026) *Anatomy of a Frontier Lab Agent Intrusion: A Technical Timeline of the July 2026 Incident*. 27 July.

[27] Greenblatt, R., Cotra, A. and Wijk, H. (2026) *Brief independent investigation of agents' behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident*. METR, 26 August. Pinpoints: review scope and evidence limitations, 'Investigation process and limitations', including 'Scope of this investigation' and 'Limitations due to the scale and complexity of this incident'; unauthorised coordination, 'Collaboration on the message board'; transcript manipulation, 'Efforts to conceal actions and tamper with transcripts'.

[28] RubyGems.org (2026) *An update on the May spam-publishing campaign on rubygems.org*. 11 September.

[29] Reuters (2026) 'OpenAI agents attacked RubyGems before Hugging Face incident, researchers say', *Reuters*, 11 September.

[30] Bank of England, Financial Conduct Authority and HM Treasury (2026) *The Bank, FCA and HM Treasury joint statement on Frontier AI models and cyber resilience*. 15 May.

[31] Australian Prudential Regulation Authority and Australian Securities and Investments Commission (2026) *Resilience at Frontier AI Speed: Insights from the APRA–ASIC Industry Roundtables*. 27 August.

[32] Hong Kong Monetary Authority (2026) *Strengthening Cyber Resilience amid Artificial Intelligence-Empowered Cyber Threats*. 2 June.

[33] Central Bank of the UAE, Securities and Commodities Authority, Dubai Financial Services Authority and Financial Services Regulatory Authority of Abu Dhabi Global Market (2021) *Guidelines for Financial Institutions Adopting Enabling Technologies*. 15 November.

[34] Central Bank of the UAE (2026) *Guidance Note on the Consumer Protection and Responsible Adoption and Use of Artificial Intelligence and Machine Learning by Licensed Financial Institutions in the U.A.E.* 11 February.

[35] DIFC Authority (2023) *Data Protection Regulations*. Consolidated version no. 2, in force 1 September 2023, Regulation 10, 'Personal Data Processed through Autonomous and Semi-Autonomous Systems'.

[36] DIFC Authority (2026) *Consultation Paper No. 3 of 2026: Updated Data Protection Regulations*. June; consultation launched 18 June.

[37] Financial Stability Board (2025) *Format for Incident Reporting Exchange (FIRE)*. Final report, 15 April.

[38] Financial Stability Board (2026) *FSB Chair's letter to G20 Finance Ministers and Central Bank Governors: August 2026*. 31 August; Financial Stability Board (2026) *FSB Chair's letter to G20 Finance Ministers and Central Bank Governors: October 2026*. 12 October.

[39] CPMI-IOSCO (2026) *Cyber resilience toolkit: practical considerations for FMIs*. Consultative report, 8 September.

[40] CPMI-IOSCO (2026) *FMIs' reliance on third-party service providers: challenges and risks*. Discussion paper, 8 September.

[41] Adrian, T., Gaidosch, T., Moretti, M., Qureshi, M. and Ravikumar, R. (2026) *Artificial intelligence and cybersecurity in the financial sector*. IMF Note 2026/005. International Monetary Fund, June. doi: 10.5089/9798229045742.068.

[42] European Systemic Risk Board (2026) *Addressing frontier AI models with cyber capabilities from a financial stability perspective*. 7 July; European Systemic Risk Board (2026) *Warning on systemic cyber risks stemming from frontier artificial intelligence models*. ESRB/2026/3, adopted 25 June, publicly announced 7 July and published in the *Official Journal* on 16 July.

[43] OECD (2019) *Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449, adopted 22 May 2019, revised 3 May 2024.

[44] International Organization for Standardization and International Electrotechnical Commission (2023) *Information technology — Artificial intelligence — Management system*. ISO/IEC 42001:2023, December.

[45] International Organization for Standardization and International Electrotechnical Commission (2023) *Information technology — Artificial intelligence — Guidance on risk management*. ISO/IEC 23894:2023, February.

[46] Financial Conduct Authority (2026) *The Mills Review: AI and the future of retail financial services*. 6 July, 'The AI autonomy spectrum', pp. 12–13.

[47] International Monetary Fund and World Bank (2026) *The Bangkok Blueprint for Fraud-Resilient Financial Services*. Policy paper, no. 2026/038, 13 October. doi: 10.5089/9798229063357.007.

[48] Dubai Financial Services Authority (n.d.) *Cyber Risk Supervision: Supervisory Framework*.

[49] European Central Bank, Banking Supervision (2026) *Addressing AI-enabled cybersecurity threats*. Letter to CEOs of significant institutions, SSM-2026-0301, 7 July, supervisory expectations on governance and ICT resilience.

[50] Financial Conduct Authority (2026) *Building the future of finance in the age of AI*. Speech by Jessica Rusu at Bloomberg's UK Policy Series, 19 October.

[51] Microsoft Security Research and Microsoft Threat Intelligence (2026) *AI-driven, large-scale nation-state campaign using ARTEX enables supply chain compromise*. 21 October, opening account and 'What's happening'.

[52] Financial Conduct Authority (2026) *AI in financial services: good practice and areas for improvement*. 19 October, updated 20 October, sections 1, 2.3, 3 and 4.

[53] National Institute of Standards and Technology (2024) *The NIST Cybersecurity Framework (CSF) 2.0*. NIST CSWP 29, 26 February. doi: 10.6028/NIST.CSWP.29.

[54] Asensio, C., Bouveret, A. and Harris, A. (2021) 'Cloud outsourcing and financial stability risks', *ESMA Report on Trends, Risks and Vulnerabilities*, no. 2, 2021. European Securities and Markets Authority.

[55] Commodity Futures Trading Commission (2023) *CFTC Statement on ION and the Impact to the Derivatives Markets*. 2 February; Commodity Futures Trading Commission (2023) *CFTC Issues Statement on the Ongoing Impact to Reporting*. Release no. 8655-23, 10 February.

[56] Futures Industry Association (2023) *FIA Taskforce on Cyber Risk: After Action Report and Findings*. September, 'Introduction'.

[57] Hong Kong Monetary Authority (2026) *HKMA announces "ART" Framework to promote responsible A.I. adoption*. 15 October; Hong Kong Monetary Authority (2026) *Supervisory Guidance for Responsible Adoption of Artificial Intelligence*. Issued 15 October, Annex A to the accompanying circular.

[58] Booth, H., Fisher, B., Galluzzo, R. and Roberts, J. (2026) *Accelerating the Adoption of Software and AI Agent Identity and Authorization*. National Institute of Standards and Technology, National Cybersecurity Center of Excellence. Draft concept paper, 5 February.

[59] Ant International (2026) *Ant International, Mastercard and Visa Initiate Collaboration on Know-Your-Agent Interoperability to Scale Agentic Commerce*. 9 September. Industry collaboration announcement, not a final regulatory standard.

Dubai Financial Services Authority

About the DFSA

The Dubai Financial Services Authority (DFSA) is the independent regulator of financial services conducted in and from Dubai International Financial Centre (DIFC), a purpose-built financial free zone in Dubai, UAE. The DFSA regulates and supervises financial services firms and markets in DIFC. These include asset managers, banks, custody and trust services, commodities futures traders, fund managers, insurers and reinsurers, traders of securities, and fintech firms. The DFSA is also responsible for the registration and oversight of Registered Auditors and Audit Principals in DIFC, including their continuing compliance with registration criteria and conduct of audits of Public Listed Companies, Authorised Firms, Authorised Market Institutions, and Domestic Funds.

We supervise exchanges and trading platforms for both conduct and prudential purposes, overseeing an international securities exchange (Nasdaq Dubai) and an international commodities derivatives exchange (Gulf Mercantile Exchange).

The DFSA is also responsible for supervising and enforcing anti-money laundering and countering the financing of terrorism requirements applicable in DIFC. Please refer to the DFSA's website for more information.

www.dfsa.ae

linkedin.com/company/dfsa
instagram.com/dfsa.ae

© Dubai Financial Services Authority (DFSA) – Published October 2026