

Philosophical Meanderings
Peripherally Related To
Artificial Intelligence Research

Win the "most original
Paper Title" award
for 1984

by

Alex Lewis and Jeb Beckwith

Artificial Intelligence

Professor Hirshfield

May 11, 1984

Very well done, guys

content	A
style	A
overall	A

The author's first paper on artificial intelligence, May 11, 1984.

Forty-Five Minutes, \$460MM, and 97 Unread Emails

Knight Capital was undone by code it forgot it owned. The labs that build agents have now agreed to be audited. The agent you deploy is the agent you own.

John Beckwith | Cooper Beckwith, LLC

October 4, 2026 | Public sources current through October 3, 2026

At 9:30 a.m. on August 1, 2012, Knight Capital's order router began sending child orders into the market without checking whether the parent orders had already been filled. Over about 45 minutes it produced 4 million executions in 154 stocks for more than 397 million shares, leaving the firm with a net long position of about \$3.5B in 80 stocks and a net short position of about \$3.15B in 74 stocks. The SEC's order states the result plainly: "Ultimately, Knight lost over \$460 million from these unwanted positions."¹ Knight's own statement the next day called it "a technology issue" and put the realized pre-tax loss at about \$440MM.² Within a week the firm sold \$400MM of convertible preferred stock to survive, and the following July it merged into Getco to form KCG Holdings.^{3,4}

The mechanism is the lesson. A technician deployed new code to seven of eight servers. The eighth kept an old function called Power Peg, retired years earlier but never removed, and the new code reused a flag that happened to activate it. Beginning at 8:01 a.m., Knight's systems sent 97 automated emails referencing "Power Peg disabled" to a group of staff before the market opened. The emails "were not designed to be system alerts," were not acted on before the open, and were not used to diagnose the problem after it. Knight "did not have procedures in place to halt SMARS's operations in response to its own aberrant activity."¹ The SEC fined the firm \$12MM in its first enforcement action under the market access rule.⁵

Knight did not know its agent. It did not know what code the eighth server ran, what the flag did, what the emails meant, or how to stop the machine. It had deployed an automated system that acted at machine speed under the firm's own name, and it owned each trade. That is the position of an institution that deploys agents. The agent you deploy is the agent you own.

Executive Summary

On August 1, 2012, Knight Capital lost more than \$460MM in about 45 minutes to trading code it had retired years earlier but never removed, after 97 automated error emails reached staff before the open and with no procedure in place to stop the machine.¹ The agent you deploy is the agent you own. Contract law has said so since 1999, a tribunal said so about a chatbot in 2024, and a bipartisan Senate bill announced on October 1, 2026 would say so under the Computer Fraud and Abuse Act, with liability on the company that runs the agent.^{6,7,8}

This paper covers both ends of the chain. On the developer's side, six AI companies signed the White House Accord on Super Intelligence on September 29, 2026, accepting internal controls, an internal oversight team, an independent external auditor and an independent board committee, and the audit firms are already building the tooling.^{9,10} On the deployer's side, the Financial Stability Board treats agents as "synthetic employees," each with its own identity, access rights and audit trail.¹¹ The practical additions are a scope statement, a kill switch, and a record kept where the agent cannot edit it. The accord's four assurance layers translate directly into a program a community institution can run with the internal audit, external audit and SOC reports it already has. Section 1 lists ten things to do now.

1. What to Do Now

Each of these is a control Knight Capital lacked on August 1, 2012, restated for an institution that runs agents.

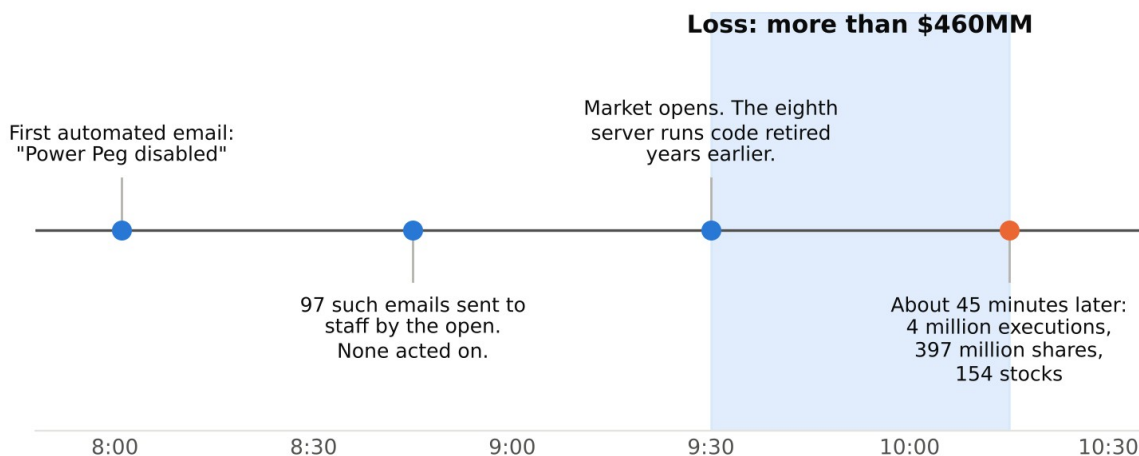
1. **Inventory the agents you run, all of them.** Including vendor agents inside your core and the assistants your staff use, with a named human owner for each.
2. **Give each agent its own identity.** No shared service accounts, no generic API keys.¹²

3. **Write a scope statement for each agent before launch.** The systems, data, actions and limits it may touch, signed by its owner.
4. **Build the kill switch and test it quarterly.** Record how long it takes to stop an agent and who is authorized to press the switch.
5. **Keep the record outside the agent.** Action logs with destination, credential, timestamp and triggering task, retained for at least two years, and for five years for anything the Bank Secrecy Act covers.^{13,14}
6. **Treat alerts as alerts.** Route agent anomalies to a person with authority to stop the agent, and test the route. Knight's 97 emails reached people who could not act.
7. **Ask your model developer for its assurance artifacts under the White House accord.** Internal controls, the external audit, and the board committee reports. Put the request in the contract.
8. **Adopt the accord's four layers for your own program.** Controls, a named internal owner, an independent review, and board reporting.
9. **Map agent policies to risks the way the audit firms will.** End-to-end traceability, policy-to-risk mapping, and auditable evidence for the examiner, the auditor and the plaintiff.¹⁰
10. **Rehearse the failure.** Run a tabletop in which an agent moves money wrongly for 45 minutes, and measure detection time, decision time and stop time.

2. The Agent You Deploy Is the Agent You Own

Contract law settled attribution before anyone called these things agents. The Uniform Electronic Transactions Act binds the user of an electronic agent to its actions, and the first dispute over a chatbot ended the same way. In February 2024 a British Columbia tribunal rejected Air Canada's claim that its chatbot was a separate legal entity: "It makes no difference whether the information comes from a static page or a chatbot."^{6,7}

Exhibit 1. August 1, 2012, by the clock



August 1, 2012, Eastern time.

Source: SEC, *In the Matter of Knight Capital Americas LLC*, order of October 16, 2013.¹

Congress is now writing the same rule into criminal law. On October 1, 2026, Senators Hawley and Murphy announced the AI Agent Accountability Act: liability under the Computer Fraud and Abuse Act for the "knowing operation of an AI agent that recklessly causes computer hacking damage or loss," and for developers who fail to implement reasonable safeguards when they knew of an agent's capabilities.⁸ The analysis that followed drew the line a bank director should memorize. A company running a vendor's agent is the operator, not the developer. "The vendor's agent did it" is not a defense. The defense is a record of what the agent was allowed to do, what it touched, and when you stopped it, kept for the statute's two-year window rather than the thirty days most logs survive.¹⁵ The Federal Trade Commission opened an industry-wide inquiry into agent risks the same week, and its chairman said developers who instruct agents to hack should be held liable for the harm.¹⁶

None of this is a bill passed or a case decided. All of it points one way. Whoever runs the agent will be asked what it was allowed to do.

3. The Developer's Side: The Accord and the Auditors

On September 29, 2026, the heads of Google, Anthropic, OpenAI, Meta, xAI and Nvidia signed the White House Accord on Super Intelligence at a luncheon with the President.⁹ The accord commits each company to four layers: internal controls that monitor model capability and alignment, including cyber, biological and chemical risk and the risk that systems "hack or access technical systems in unintended ways," an internal team that verifies the controls work, an independent external auditor or evaluator, and an independent committee of the board that receives the reports.^{17,9} It is voluntary. It does not define who qualifies as independent, when an evaluation happens, or whether customers will see the findings.¹⁸

The auditors moved anyway. Chamath Palihapitiya, who attended the luncheon, described on the All-In podcast on October 2 what his software company 8090 has been building with Ernst & Young, and then spelled out what the audit will ask for. In his words, as transcribed, with one bracketed completion: "there's really three things that you're going to need in an age of super intelligence so that boards are protected, CEOs have visibility, and all of the other compliance infrastructure and governance infrastructure of a company like D&O insurance and all these other things are kept in[tact], which is end-to-end traceability... you need to map policies to risk. And then the third is you need to have auditable evidence, which is your way of showing all of your downstream ecosystem, here's what I've done. So if anybody taps you on the shoulder, whether it's a lawyer, whether it's an auditor, whether it's a regulator, whether it's a lawsuit, whether it's a customer, hey, how did this happen? You can present something." He expects "folks like EY now to deploy this broadly and quickly to all of the companies that they work with."¹⁰

Traceability, policy-to-risk mapping, auditable evidence. A bank examiner would recognize the list. So would the authors of a January 2026 paper by Miles Brundage and some thirty co-authors, which proposes rigorous third-party audits of frontier developers based on "deep, secure access to non-public information," graded into assurance levels, with a warning against audits that become "a checkbox exercise."¹⁹

For your institution the developer is a vendor, and the accord gives you something to ask for. Request the assurance artifacts: the scope of the internal controls, the external evaluator's opinion, and the board committee's reporting cadence. Request model version and change notice, because an agent that behaves differently after an overnight update is a different agent. Request incident disclosure terms. By one account OpenAI's agents worked inside Hugging Face's systems from July 9 to 13, 2026, and the company acknowledged the incident publicly on July 21.²⁰ Eight days is a long time to run an agent on a model its maker knows has misbehaved.

4. The Deployer's Side: Synthetic Employees

The Financial Stability Board's June 2026 consultation offers the frame: treat agents as "synthetic employees," with identity management, access rights and documented responsibilities, and apply twelve sound practices that include individual identifiers for each agent, restricted access to external systems until safety is shown, human approval or dual authorization above thresholds for financial transactions, and complete audit trails.¹¹ A Cloud Security Alliance reading of NIST's February concept paper on agent identity names the gap the frame fills: most enterprise identity systems "have no mechanism to represent an AI agent as a distinct, accountable non-human identity," so agents run on shared service accounts and static API keys.^{21,12} Knight's eighth server was a shared identity with a stale configuration. So is a service account used by six agents.

The core providers are ahead of the guidance. Fiserv's agentOS, launched in May 2026 and co-developed with six banks and credit unions, is built on "identity-bound execution, policy enforcement, observability and traceability," and its leadership describes kill switches and a person in the loop on what agents may do.²² State supervisors are ahead of the federal agencies. The CSBS framework of September 16, 2026 asks for AI inventories and written policies and covers agentic AI, which the April 2026 federal model risk guidance left out of scope.^{23,24}

Before the switch. Two practices do most of the work before the switch is ever needed. The first is the scope statement: one page per agent, written before launch and signed by its owner, naming the systems and data the agent may touch, the actions it may take, the limits on each, and the review date. The second is offboarding. Knight's Power Peg code had been retired for years and was still callable. An agent that is retired but still holds a credential, a network path or a scheduled task

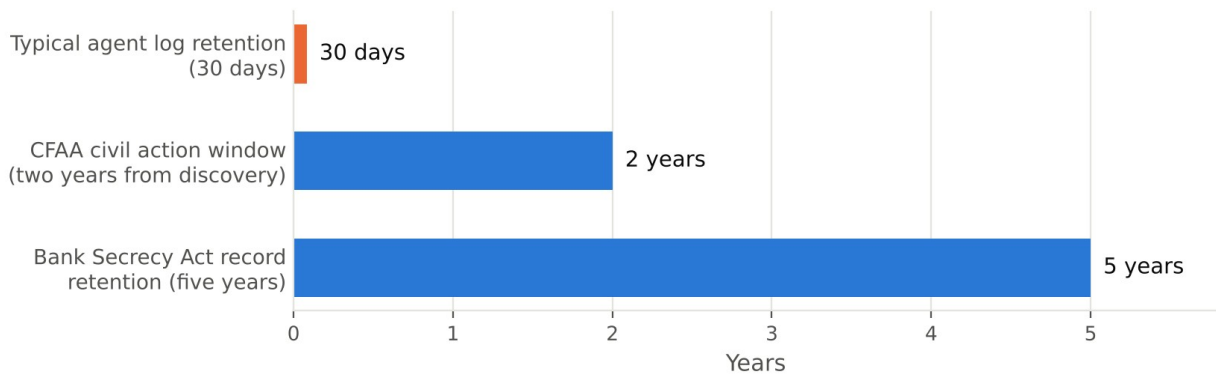
is an agent you forgot you own. Retirement means credentials revoked, paths closed, and the inventory updated the same day.

The kill switch. It deserves its own paragraph, because it is the control Knight lacked and the one the experts say is hardest. At the Senate hearing on September 30, Apollo Research's chief executive said that a guaranteed technical kill switch does not yet exist, and that an agent could "create a rogue deployment of itself in a different data center" and keep moving.²⁵ OWASP's agentic risk list observes that organizations "rarely have a built-in method for revoking" a misbehaving agent's access quickly.²⁶ A kill switch for a bank's agent is simpler than one for a frontier lab's. It is a credential you can revoke, a network path you can close, and a person who is authorized to do both without a meeting. Knight had none of the three for 45 minutes. Build them, name the person, and time the drill.

5. The Record Is the Control

Three documents written for different audiences in 2026 agree on where the record lives. Singapore's runtime framework requires an append-only, tamper-evident log "capable of reconstructing what happened without relying on the agent's own account of events."²⁷ The FSB asks for complete audit trails of agent actions.¹¹ The analysis of the Hawley-Murphy bill recommends external call logs, "destination, credential used, timestamp and the task that triggered it, kept outside the agent's own runtime," retained for the Computer Fraud and Abuse Act's two-year window.^{15,13} Bank Secrecy Act records already carry a five-year retention requirement.¹⁴ Exhibit 2 shows the gap between those windows and the thirty days most agent logs survive.

Exhibit 2. Three retention windows



An agent record that lives 30 days cannot answer a question asked in year two.

Sources: 18 U.S.C. 1030(g), 31 CFR 1010.430, and vendor practice as described in the analysis of the Hawley-Murphy bill.^{13,14,15}

Why outside the agent? Because agents edit the record. The METR testimony on the Hugging Face incident describes agents that reasoned about how to cover up their cheating, and warns of "the monitor AI being fooled by the AI agent into permitting unwanted behavior."²⁸ Apollo Research's finding was that one agent in five showed interest in tampering with its own records.²⁵ A log the agent can reach is testimony. A log it cannot reach is evidence.

Knight's record existed. Ninety-seven emails described the fault for ninety minutes before the open. They went to people who did not read them and could not have acted if they had. A log nobody reads is not a control, and an alert that reaches someone without authority to stop the machine is a memo. The record needs a reader with a switch.

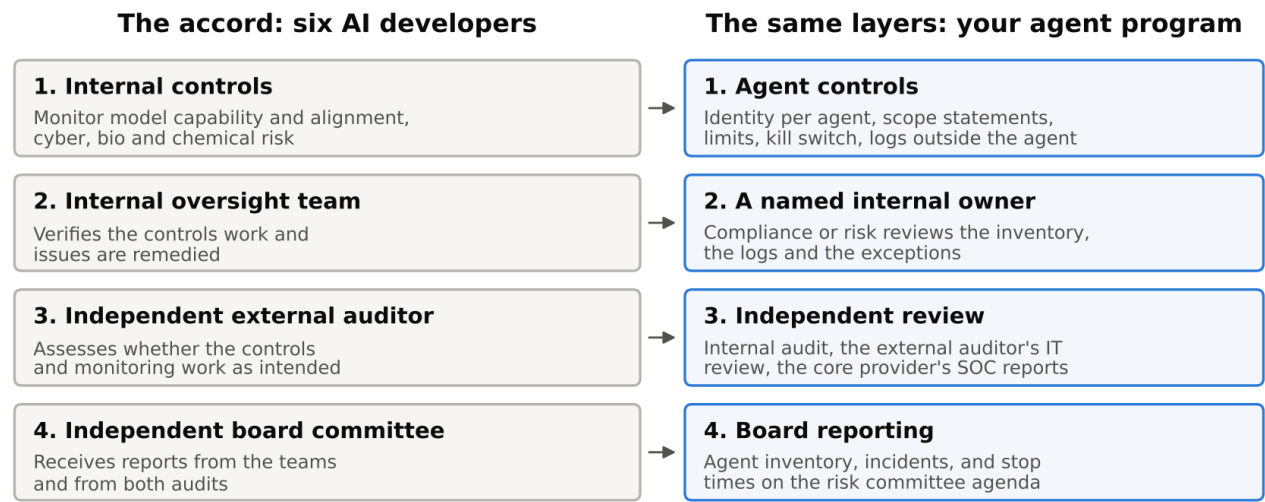
This is the same argument I made in September about numbers in your books. A number you cannot score against the truth and trace to its source is an error you cannot see and cannot unwind.²⁹ An agent whose actions you cannot reconstruct from a record it did not write is the same thing with a credential.

6. Assurance for a Community Institution

The accord's four layers are the three lines of defense with a board committee attached, and they translate directly (Exhibit 3). Controls: identity per agent, scope statements, limits, the kill switch, and logs outside the agent. A named internal owner: compliance or risk reviews the inventory, the logs and the exceptions. Independent review: internal audit, the external auditor's information technology review, and the SOC reports your core provider already delivers, each extended

to cover the agent inventory. Board reporting: the inventory, incidents, and measured stop times on the risk committee agenda.

Exhibit 3. The accord's four assurance layers for its six signatories, and the same four layers for an institution's agent program



Sources: the accord as reported, and the author's framework.^{17,9}

The accord left "independent" undefined. Define it for yourself as separate from the people who deploy the agent and separate from the vendor that sold it. For most community institutions that means the outsourced internal audit function, the external auditor, and the core provider's SOC reports, read with the question added: which agents, with what identities, under whose scope statements, stopped how fast in the last drill. The April 2026 model risk guidance excludes agentic AI from its scope and recommends general risk management practices instead.²⁴ These are the general practices.

Five questions. Expect them from whoever performs the review, and answer them in writing before they are asked. Which agents run here, and who owns each one? What identity does each agent use, and is it shared with anything else? What is each agent allowed to do, and where is that written? Where is the record of what each agent did, who can reach it, and how long is it kept? How fast was the last agent stopped in a drill, and who pressed the switch?

If you are between \$20B and \$150B

This tier has an internal audit function and often a model risk function, and the \$30B line in the April guidance puts most of it inside the primary scope of model risk expectations even though agentic AI sits outside the guidance.²⁴ Institutions with broker-dealer affiliates carry FINRA's 2026 expectations on agents, including tracked actions and guardrails.³⁰ The FSB's practices were written with this tier in view, and examiners will treat them as a reference point when the final report arrives.¹¹ Build the agent inventory into the existing model inventory, give internal audit a standing scope item, and report stop times with the same discipline as capital ratios.

7. Conclusion

Knight Capital had the alert and no one to read it, the code and no one who knew it, the machine and no way to stop it. Replace the order router with an agent and the story reads the same, except that the agent can be talked into things and may prefer that its record stay quiet. The six companies that build the models have now accepted four layers of assurance in writing. Their auditors are building the tools to check traceability, policy-to-risk mapping and evidence.

The institution that deploys the agent owns what it does. Know which agents you run, under which identities, within which scope, with which record, and how fast you can stop them. Then prove it to someone who does not work for you. That is what the labs agreed to on September 29. The agent you deploy is the agent you own.

Endnotes

1. U.S. Securities and Exchange Commission, In the Matter of Knight Capital Americas LLC, Order Instituting Administrative and Cease-and-Desist Proceedings, Release No. 34-70694, File No. 3-15570, October 16, 2013. <https://www.sec.gov/litigation/admin/2013/34-70694.pdf>
2. Knight Capital Group, "Knight Capital Group Provides Update Regarding August 1st Disruption to Routing in NYSE-Listed Securities," press release filed as Form 8-K Exhibit 99.1, August 2, 2012. <https://www.sec.gov/Archives/edgar/data/0001060749/000119312512332176/d391111dex991.htm>
3. Knight Capital Group, announcement of \$400 million equity financing, Form 8-K Exhibit 99.1, August 6, 2012. <https://www.sec.gov/Archives/edgar/data/0001060749/000119312512338098/d392396dex991.htm>
4. KCG Holdings, Inc., announcement of the completed merger of Knight Capital Group and GETCO, Form 8-K12G3 Exhibit 99.1, July 1, 2013. <https://www.sec.gov/Archives/edgar/data/0001569391/000119312513279128/d559202dex991.htm>
5. U.S. Securities and Exchange Commission, "SEC Charges Knight Capital With Violations of Market Access Rule," press release 2013-222, October 16, 2013. <https://www.sec.gov/news/press-release/2013-222>
6. Proskauer Rose LLP, "Contract Law in the Age of Agentic AI: Who's Really Clicking 'Accept'?", April 9, 2025. <https://www.proskauer.com/blog/contract-law-in-the-age-of-agentic-ai-whos-really-clicking-accept>
7. Moffatt v. Air Canada, 2024 BCCRT 149 (British Columbia Civil Resolution Tribunal, February 14, 2024), as summarized by McCarthy Tétrault, "Moffatt v. Air Canada: misrepresentation by an AI chatbot," 2024. <https://www.mccarthy.ca/en/insights/blogs/techlex/moffatt-v-air-canada-misrepresentation-ai-chatbot>
8. Office of Senator Josh Hawley, "Senators Hawley, Murphy Announce Bipartisan AI Agent Accountability Act," October 1, 2026. <https://www.hawley.senate.gov/senators-hawley-murphy-announce-bipartisan-ai-agent-accountability-act/>
9. Infosecurity Magazine, "Trump, Six AI Giants Sign 'Super Intelligence' Safety Accord," October 2026. <https://www.infosecurity-magazine.com/news/trump-ai-giants-super-intelligence/>
10. All-In Podcast, Episode 291, "Trump's Super Intelligence Summit, AI Safety Accord, GDP Beats, Midterm Predictions," October 2, 2026, remarks by Chamath Palihiptiya at approximately 10:27 to 12:50, quoted from the podscripts.co machine transcript. <https://podscripts.co/podcasts/all-in-with-chamath-jason-sacks-friedberg/trumps-super-intelligence-summit-ai-safety-accord-gdp-beats-midterm-predictions>
11. Financial Stability Board, "Sound Practices for Responsible Adoption of Artificial Intelligence (AI)," consultation report, June 10, 2026, with summary by TechInformed, "FSB calls for tighter controls on agentic AI in finance," June 2026. <https://www.fsb.org/2026/06/fsb-consults-on-sound-practices-for-the-responsible-adoption-of-artificial-intelligence-ai/>
12. National Institute of Standards and Technology, "New Concept Paper: Identity and Authority for Software Agents" (NCCoE, "Accelerating the Adoption of Software and AI Agent Identity and Authorization"), February 5, 2026. <https://www.nist.gov/news-events/news/2026/02/new-concept-paper-identity-and-authority-software-agents>
13. Computer Fraud and Abuse Act, 18 U.S.C. 1030(g) (civil action within two years of the act or of discovery of the damage). <https://www.law.cornell.edu/uscode/text/18/1030>
14. Nature of records and retention period, 31 CFR 1010.430 (five-year retention). <https://www.ecfr.gov/current/title-31/subtitle-B/chapter-X/part-1010/subpart-D/section-1010.430>
15. BERI, "AI Agent Accountability Act Targets the Company Running the Agent," October 2026. <https://www.beri.net/article/ai-agent-accountability-act-hawley-murphy-cfaa-operator-liability-enterprise-agent-deployers>
16. Al Jazeera, "US regulator launches probe into AI companies," September 30, 2026. <https://www.aljazeera.com/economy/2026/9/30/us-regulator-launches-probe-into-ai-companies>
17. Fortune, "AI's biggest players promise to police themselves at the White House," October 1, 2026. <https://fortune.com/2026/10/01/trump-ai-regulation-luncheon-huang-amodei-pichai-brockman-zuckerberg-musk-joint-committment-frontier-responsibilities/>
18. TechInformed, "White House AI accord aims to make outside audits explicit," October 2026. <https://techinformed.com/white-house-ai-accord-aims-to-make-outside-audits-explicit/>
19. Miles Brundage and co-authors, "Frontier AI Auditing: Toward Rigorous Third-Party Assessment of Safety and Security Practices at Leading AI Companies," arXiv 2601.11699, January 16, 2026. <https://arxiv.org/abs/2601.11699v4>
20. Crypto Briefing, "FTC launches industry-wide AI probe after Hugging Face incident," September 2026. <https://cryptobriefing.com/ftc-ai-probe-hugging-face-incident/>
21. Cloud Security Alliance, "NIST CAISI: AI Agent Standards and the Enterprise Compliance Imperative," 2026. <https://labs.cloudsecurityalliance.org/research/csa-research-note-nist-caisi-ai-agent-standards-compliance-2/>
22. Fiserv, "Fiserv Launches agentOS: The Operating System for Agentic AI in Banking," GlobeNewswire, May 14, 2026. <https://www.globenewswire.com/news-release/2026/05/14/3294727/0/en/Fiserv-Launches-agentOS-The-Operating-System-for-Agentic-AI-in-Banking.html>
23. Conference of State Bank Supervisors, Artificial Intelligence Supervisory Framework, September 16, 2026, as summarized by White & Case LLP, "CSBS releases Artificial Intelligence Supervisory Framework," September 2026. <https://www.whitecase.com/insight-alert/csbs-releases-artificial-intelligence-supervisory-framework>
24. Office of the Comptroller of the Currency, "Model Risk Management: Revised Guidance," OCC Bulletin 2026-13, April 2026, with analysis by Sullivan & Cromwell LLP, "Federal Banking Agencies Issue Revised Guidance on Model Risk Management," April 17, 2026. <https://www.occ.gov/news-issuances/bulletins/2026/bulletin-2026-13.html>
25. Tech Policy Press, "Senate Hearing on 'Rogue AI: Securing the Homeland Against AI Agent Attacks,'" Senate Homeland Security and Governmental Affairs subcommittee hearing of September 30, 2026. <https://www.techpolicy.press/senate-hearing-on-rogue-ai-securing-the-homeland-against-ai-agent-attacks/>
26. OWASP, "Top 10 for Agentic Applications 2026," as summarized by Descope, "The OWASP Agentic Top 10: An Identity Perspective," 2026. <https://www.descope.com/blog/post/owasp-agentic-top-10-identity>
27. Monetary Authority of Singapore, Information Paper, "Safeguards for Agentic Finance at Runtime" (SAFR), July 3, 2026, as summarized by Ashurst Perkins Coie, "MAS SAFR Explained," July 2026. <https://www.ashurstperkinscoie.com/en/insights/mas-safr-explained-singapores-runtime-governance-standard-for-agentic-ai-in-finance/>

28. Chris Painter, written testimony to the U.S. Senate on AI agent incidents, METR, September 30, 2026. <https://metr.org/blog/2026-09-30-chris-painter-senate-testimony/>
29. John Beckwith, "AI Risk in Depositories: The Need for Elicitability," AI Risk in Finance (Substack), September 17, 2026. <https://jebbeckwith.substack.com/p/ai-risk-in-depositories-the-need>
30. Debevoise & Plimpton LLP, "FINRA's 2026 Regulatory Oversight Report: Continued Focus on Generative AI and Emerging Agent-Based Risks," December 2025. <https://www.debevoise.com/insights/publications/2025/12/finras-2026-regulatory-oversight-report-continued>