

Leading into the AI Revolution

A proposed federal framework
for agentic finance

Prepared by Lead



Introduction

Agentic finance is the use of autonomous or semi-autonomous AI agents to carry out commerce and other financial transactions on behalf of a consumer or business. It covers both agentic commerce (an agent making a purchase) and agentic payments (an agent moving money). Instead of a person browsing, approving each step, and clicking to pay, an agent takes a stated goal and executes it: discovering options, negotiating, and settling payment within the authority it has been given. For banks, agentic finance is not a wholly new product. Rather, it is a controlled extension of existing bank services, sitting within the current regulatory framework but breaking two assumptions at the core of U.S. banking law: that humans initiate transactions and that institutions control decision-making. AI agents invert both. The legal and risk issue is not whether an AI agent can call an API or transact; it is whether banks can clearly identify who is acting, on whose authority, under what limits, with what monitoring, and with a clean audit trail when something goes wrong.

Lead has built an infrastructure and governance layer for agentic finance, translating existing rails into a controlled operating model that regulators, partners, and enterprise customers can trust. We are excited by many of the recent governmental initiatives moving in this direction: the Unleashing AI Innovation in Financial Services Act ([H.R. 4801](#)), which establishes supervised innovation labs and risk-based “rules of the road” for AI in financial services; the AI PLAN Act ([H.R. 2152](#)), directing a whole-of-government assessment of AI-driven financial-crime risk; the FUTURES Act ([H.R. 8278](#)), mandating a review of supervisory technology; the Bank Fraud Technology Advancement Act ([H.R. 8671](#)), expanding community-bank access to AI-powered fraud detection; the GUARD Act ([H.R. 2978](#)), supporting state and local law enforcement against sophisticated scams; and Sen. Mark Warner’s AI Agent Act, which would direct NIST to identify technical standards and open protocols for secure agent access.

Agentic finance is not a net new product or regulatory free-for-all. Rather, it is a controlled extension of existing bank services within the existing framework of regulation but with a need for change across all domains. In practice, that means narrow use cases, explicit permissions, transaction-level controls, human override paths, logging, clear contractual allocation of responsibility, and conservative partner governance before broader rollout. As a responsible and innovative bank, Lead’s priority



is to ensure that innovation can flourish under a clear, transparent, and modernized regulatory framework that protects consumers and promotes financial health.

The main risk categories are:

- **Authority and agency:** The ability to distinguish between the end user, the agent, the developer/platform, and the merchant, and define whose instruction legally controls a payment or account action.
- **Fraud and unauthorized activity:** Agent-initiated payment flows raise questions around impersonation, delegated authority, transaction manipulation, and whether existing fraud controls are adequate for autonomous or semi-autonomous actions. “Agentic cascades,” i.e. scenarios where multiple autonomous agents interact (A2A), creating systemic feedback loops, amplified liquidity, or market volatility that could lead to instant de-leveraging or flash crashes, are a distinct concern. Standard incident-response runbooks assume human or single-system failures and may not address multi-agent, machine-speed market contagion.
- **BSA/AML and sanctions accountability:** If agents initiate or route transactions, Lead still owns the core compliance posture and needs clear responsibility for screening, monitoring, escalation, and suspicious-activity handling.
- **Consumer protection and error resolution:** If an agent acts incorrectly, the allocation of responsibility among bank, platform, merchant, and end user becomes a core legal issue, especially where the experience feels consumer-facing, even if the underlying stack is B2B.
- **Model and explainability risk:** If underwriting, approvals, or transaction decisions are influenced by agentic systems, Lead needs enough visibility into logic, constraints, and overrides to defend the control environment.
- **Third-party risk:** If the agent layer sits with Stripe, OpenAI, Google, or another platform, Lead’s practical exposure is classic third-party and fourth-party risk wrapped in a new UX.
- **Regulatory narrative risk:** The market may move faster than the legal framework, so overclaiming “agentic banking” before controls are mature could challenge the long-term viability of the model. Each regulated institution must adopt and maintain a robust risk-appetite and management program for such innovations.
- **Technical security and adversarial risk:** Prompt injection, tool hijacking, credential compromise, and agent-to-agent trust exploitation are distinct from traditional fraud and third-party risk, and require information-security-specific controls, metrics, and incident classes to detect and contain them.



There are several domains where the existing framework must be refined for responsible innovation and continued U.S. excellence.

- On identity, AI agents break the assumption that a “customer” is a natural person or legal entity acting directly, so there must be legal recognition of AI agents as authorized transacting delegates and a move toward cryptographic identity (wallet-based or PKI-backed agent IDs), with financial-crimes modernization focused on continuous behavioral monitoring rather than static onboarding and SAR requirements reinterpreted to define suspicious AI behavior.
- On liability, the allocation of liability for agent-assisted transactions requires a framework for legally valid “AI consent” (API signatures, MPC, and smart contracts would all satisfy this), so that a human who authorizes an agent retains liability for its actions, clearly and conspicuously disclosed and consented to, with a kill switch to remove authorization.
- On model risk, traditional oversight paradigms must be reimaged for LLMs, with explainability paramount even as model evolution and interpretation stay nimble.
- On cybersecurity and privacy, autonomous systems expand the attack surface, so we support updating third-party risk-management frameworks for AI-specific risks, with auditable decision trails and consent-scoped data use so that AI cannot reuse financial data beyond its consent scope. Where Lead or its clients operate globally, the framework must also align with international regimes (EU AI Act, GDPR, Basel III/IV) to avoid regulatory arbitrage when agents move funds across borders.

Much of what follows can begin now, through supervisory guidance and examination. The roadmap at the end sequences the rest. Appendix A sets out the internal control model related to “Know your Agent” – an example of the governance layer needed for robust oversight in an agentic finance world.



Recommendation: Create a Federal “Authorized Financial Agent” Regime

Banking law assumes a human or legal entity directly authorizes transactions, while agentic finance will evolve into software agents negotiating, instructing, initiating, and settling transactions and making financial decisions on behalf of a principal. Regulators should create a recognized category of “Authorized Financial Agent,” meaning an AI system that can act only within a verifiable delegation from a consumer or business. **To enable it, regulators should redefine the core primitives of banking: identity, authorization, liability, and supervision, not just amend individual rules.**

The recommendations below are a single numbered list. Each is tagged by the primitive it serves, Identity, Authorization, Liability, or Supervision.

1

Identity, KYC, AML, and sanctions

Identity

(see also [Appendix A](#)
for a Model KYA
Framework)

ISSUE

Existing BSA/AML and CIP rules identify customers, beneficial owners, and account users, but not autonomous agents acting on their behalf.

RECOMMENDED POLICY

FinCEN should require banks to maintain a principal-agent identity record for any AI agent authorized to transact.

WHAT SHOULD CHANGE

- Add agent identity to CIP/CDD records.
- Require each agent to hold a unique, non-transferable cryptographic credential tied to a verified human or legal entity (e.g., an mTLS certificate or agent-bound token), issued only on completion of the identity and verification record, so



shared credentials cannot break the identity linkage and an unregistered agent cannot authenticate.

- Require banks to know the agent’s scope of authority: spending limits, merchant categories, jurisdictions, duration, and revocation terms.
- Require monitoring for agent-specific suspicious behavior (e.g. rapid account switching, abnormal transaction velocity, sanctions-evasion patterns, prompt-injection-driven payments).
- Update SAR guidance to include AI-agent compromise, unauthorized delegation, and model manipulation.
- Reference emerging interoperable credential standards (FIDO Alliance Agentic Authentication Technical Working Group, Visa Trusted Agent Protocol, Mastercard Agent Pay) so the rule does not lock in a single vendor scheme.

PROPOSED POLICY TEXT

“Financial institutions must identify and verify any autonomous digital agent authorized to initiate or materially influence a covered financial transaction, must issue that agent a unique, non-transferable, cryptographically verifiable credential technically bound to the identity record, and must maintain a verifiable linkage between the agent, the principal, and the scope of delegated authority. Credential issuance must be conditioned on completion of the identity and verification record.”

FinCEN’s role is directly relevant because it administers AML obligations under the BSA framework, and recent FinCEN activity continues to focus on illicit-finance modernization and cyber-enabled fraud.

2

Agent licensing and certification

Identity

ISSUE

Any AI agent connected to banking rails could become a payment, fraud, or consumer-protection risk.

RECOMMENDED POLICY

Establish a certification regime for high-risk financial agents.



WHAT SHOULD CHANGE

- Tier by capability: low-risk (budgeting, alerts, information retrieval); medium-risk (bill pay, merchant selection, recurring purchases); high-risk (payments, credit applications, investments, account opening, funds movement).
- Require certification for high-risk agents before bank integration.
- Require a defined adversarial-testing standard as a certification prerequisite for medium- and high-risk agents, covering prompt injection, mandate escape, credential exfiltration, tool misuse, data-scope overreach, and multi-agent confused-deputy scenarios.
- Require recertification after any material change to the underlying model, framework, or tool access.

PROPOSED POLICY TEXT

“No autonomous agent may initiate covered financial transactions through a regulated bank unless it satisfies technical, security, identity, auditability, and consumer-protection certification requirements, including a defined adversarial-testing standard covering prompt injection, mandate escape, credential exfiltration, tool misuse, data-scope overreach, and multi-agent confused-deputy scenarios with documented pass/fail criteria. Recertification is required upon any material change to the underlying model, framework, or tool access.”

3

Authorization and consent for payments

Authorization

ISSUE

Regulation E and payment law distinguish authorized from unauthorized transfers, but do not define what valid authorization looks like when an AI agent initiates the transaction.

RECOMMENDED POLICY

CFPB, Federal Reserve, and banking agencies should define “machine-executed authorization.”

WHAT SHOULD CHANGE

- Treat an AI transaction as authorized only if it falls within a recorded delegation,



evidenced by a verifiable signed mandate as opposed to an unsigned instruction. Thus, acting within delegation can be checked cryptographically rather than asserted after the fact.

- Require granular consent: amount caps, merchant categories, recurring permissions, geographic restrictions, and expiration.
- Require real-time revocation by the principal, defined against a maximum-latency standard so the right is measurable rather than aspirational.
- Distinguish the authorization requirements for one-time transactions from those for recurring transactions (e.g., same amount, same merchant).
- Require transaction receipts showing whether a human, API, or AI agent initiated the transfer, and require banks to preserve authorization logs for disputes.

PROPOSED POLICY TEXT

“An electronic fund transfer initiated by an autonomous agent is authorized only where the agent acted within a consumer’s prior, specific, revocable, and auditable delegation, evidenced by a verifiable signed mandate, and where the institution can demonstrate that revocation of that delegation takes full technical effect within a defined maximum latency.”

This matters because Regulation E limits consumer liability for unauthorized electronic fund transfers and has specific error-resolution and liability rules.

4

Payments infrastructure and settlement

Authorization

ISSUE

Agentic finance and transactions may involve high-frequency microtransactions, dynamic negotiation, conditional payment instructions, and machine-to-machine settlement.

RECOMMENDED POLICY

Create standards for **programmable payment authorization**.

WHAT SHOULD CHANGE

- Allow payment instructions to carry machine-readable conditions.



- Require standardized APIs for delegated payment authority.
- Require transaction metadata identifying agent, principal, merchant, purpose, and authorization token.
- Require payment networks to support revocation, dispute evidence, and agent traceability.
- Create risk-based throttles for autonomous transaction bursts.
- Require that mandate limits (spend ceilings, counterparty lists, geography) be enforced at the payment-instrument level, not only in application logic.
- Require that in agent-to-agent transaction chains, effective authority is the intersection, not the union, of the delegations in the chain, so authority cannot expand as it passes through multiple agents.

PROPOSED POLICY TEXT

“Payment systems should support verifiable delegated authority, machine-readable transaction conditions, and auditable agent-originated payment metadata, and must ensure that in multi-agent transaction chains, effective delegated authority is the intersection of all delegations in the chain, enforced where possible at the payment-instrument level.”

5

Data privacy and GLBA

Authorization

ISSUE

AI agents need access to sensitive financial data, but existing privacy notices and data-sharing consents were not designed for autonomous data use.

RECOMMENDED POLICY

Congress should pass legislation amending the GLBA to require **agent-specific data permissions**.

WHAT SHOULD CHANGE

- Separate consent for human use, bank use, third-party use, and AI-agent use.
- Require data minimization for agent tasks and purpose limitation: agents can use data only for the delegated task.
- Prohibit hidden reuse of customer financial data for model training without



opt-in consent.

- Require data-lineage records showing what data the agent accessed and why.
- Enforce purpose and scope limitation technically, at the data-access layer, with monitoring to detect an agent accessing data beyond its authorized scope.

PROPOSED POLICY TEXT

“A financial institution may permit an autonomous agent to access customer financial information only under a specific, revocable, purpose-bound authorization, and must enforce purpose and scope limitation technically at the data-access layer, with monitoring to detect and remediate agent data access exceeding the delegated scope.”

6

Consumer disclosures

Authorization

ISSUE

Consumers may not know whether they are interacting with a bank employee, chatbot, third-party AI, or autonomous agent.

RECOMMENDED POLICY

Congress should pass legislation to require clear AI interaction and authority disclosures.

WHAT SHOULD CHANGE

- Disclose, reasonably and prominently, when a consumer is interacting with AI.
- Disclose whether the AI can bind the bank or merely provide information.
- Disclose when a consumer’s own agent is initiating an action.
- Provide a plain-language explanation of risks, limits, and revocation methods.
- Require confirmation for high-risk transactions.

PROPOSED POLICY TEXT

“Consumers must be clearly informed when an AI system is used in a financial interaction and whether that system has authority to provide advice, make decisions, initiate transactions, or bind the institution.”



7

Merchant and platform obligations

Authorization

ISSUE

Merchants may manipulate AI agents through hidden instructions, dark patterns, misleading metadata, or adversarial prompts.

RECOMMENDED POLICY

Prohibit agent-targeted manipulation in financial commerce.

WHAT SHOULD CHANGE

- Ban hidden prompts or machine-readable instructions designed to override user intent.
- Require merchants to present accurate price, terms, refund, and subscription data in machine-readable form.
- Require platforms to disclose when offers are personalized to exploit agent behavior.
- Create liability for merchants that knowingly cause an agent to exceed authority.

PROPOSED POLICY TEXT

“A merchant or platform may not knowingly design content, code, metadata, or instructions to cause an autonomous financial agent to act outside the principal’s delegated authority or contrary to disclosed transaction terms.”

8

Design and interaction standards for delegated authority

Authorization

ISSUE

Consumer-protection regimes assume the consumer is present at the transaction. Agentic finance breaks that assumption because the principal who authorized the agent is absent when it executes, meaning comprehension, consent, and friction must be handled at the delegation moment. Emerging protocols (ACP, AP2, Verifiable Intent, Trusted Agent Protocol, under the FIDO Alliance) solve machine trust, but no framework addresses the human experience layer.



RECOMMENDED POLICY

Treat interaction design as regulatory infrastructure for delegated authority, with risk-tiered standards. Support a graded, externally-maintained conformance standard, **TRACE (Transparency, Revocability, Auditability, Comprehension, and Equivalence)**, governed by a multi-stakeholder body and referenceable in supervisory guidance, procurement, and contract.

TRACE CONFORMANCE LEVELS

- **TRACE-1**, the floor for any agent operating on banking infrastructure: machine-readable agent identity, principal-facing receipts, functional revocation, plain-language scope disclosure, and an accessible kill switch.
- **TRACE-2**, the baseline for any institution enabling consumer-facing agentic transactions or business agents with material transaction authority: adds standardized scope representation, provenance signaling, comprehension-proportional confirmation, the full revocation spectrum, and accessibility and inclusive-access validation.
- **TRACE-3**, required for high-stakes use cases (treasury agents, lending agents, agents serving structurally limited principals): adds independently-audited comprehension testing, real-time principal notification with intervention windows, and plain-language decision rationale in audit trails.

WHAT SHOULD CHANGE

- Give third parties with legitimate oversight rights (co-account holders, treasury supervisors, parents, compliance officers, auditors) visibility into agent activity and a way to intervene.
- Prohibit interfaces that induce over-delegation (over-broad defaults, auto-renewing permissions, granting/revoking asymmetry), the principal-side inverse of merchant manipulation.
- Establish provenance signaling that makes an agent's identity and principal relationship unambiguous to counterparties, the analog to the HTTPS lock icon.
- Meet established accessibility standards (ADA, Section 508, WCAG) and validate across language, literacy, age, and financial circumstance, and, for institutional principals, governance maturity.
- Require a standardized, principal-facing representation of agent scope at delegation (a Schumer box for agents), with real comprehension for high-scope grants rather than clickthrough.
- Support the full revocation spectrum (pause, narrow, suspend, revoke), making



it no harder than granting, and give the principal a plain-language record to reconstruct what the agent did and dispute it.

PROPOSED POLICY TEXT

“Banking organizations and platforms enabling autonomous agent transactions must maintain interaction and disclosure standards sufficient for the principal to comprehend the scope of authority delegated, audit actions taken under it, and revoke it across the full spectrum of pause, scope-reduction, suspension, and full revocation with friction no greater than that required to grant it. Such standards shall be risk-tiered, accessible, functionally equivalent across principal capacity, applicable to third-party platforms mediating the principal experience, and may be evidenced through a graded, independently-auditable conformance standard layered above existing infrastructure-layer protocols.”

9

Liability for agent errors, fraud, and misuse

Liability

ISSUE

Current rules do not clearly allocate liability when an AI agent makes a wrong purchase, is manipulated, exceeds instructions, or is compromised.

RECOMMENDED POLICY

Create a tiered liability framework, effectuated via policy guidance and potential legislation.

WHAT SHOULD CHANGE

- Consumer is not liable where the agent exceeded delegated authority.
- Bank is liable where it had control over the delegation and failed to enforce it.
- AI provider is liable where failure arose from model defect, insecure tool use, or inadequate safeguards, including absent mandate enforcement, missing or untested revocation capability, or failure to mitigate known prompt-injection or tool-hijacking techniques.
- Merchant is liable where it knowingly induced the agent to exceed authority.
- Any other participant in the transaction chain, such as an issuer-processor, program manager, payment network, or orchestration platform, is liable for



failures within its own span of control.

- Shared liability where multiple control failures contributed.

PROPOSED POLICY TEXT

“No consumer shall bear liability for an agent-initiated transaction outside the documented delegation of authority, except where the consumer knowingly enabled or ratified the transaction after disclosure. An AI provider is liable where failure arose from model defect, insecure tool use, or inadequate safeguards, including failure to implement mandate enforcement, tested revocation capability, or mitigations for known adversarial-manipulation techniques.”

10

AI lending and underwriting

Liability

ISSUE

AI models may make or influence credit decisions, but ECOA and Regulation B require specific reasons for adverse action. Black-box AI can make this difficult.

RECOMMENDED POLICY

Congress should pass legislation, or the CFPB should update existing guidance, to require explainable credit decisions for any AI-influenced lending action.

WHAT SHOULD CHANGE

- Require model-level and decision-level explainability.
- Require adverse-action notices to describe the actual factors used.
- Prohibit generic explanations such as “algorithmic risk score too low.”
- Require bias testing before deployment and continuously after deployment.
- Require audit trails showing inputs, model version, and decision rationale.

PROPOSED POLICY TEXT

“A creditor using an AI or complex model must be able to generate specific, accurate, consumer-comprehensible reasons for any adverse action, including the material variables that actually affected the decision.”



The CFPB has already stated that lenders using AI or complex models must provide specific and accurate reasons for adverse actions, and that complex algorithms do not excuse ECOA/Regulation B obligations.

11

Fair lending and discrimination

Liability

ISSUE

AI systems can create disparate impact through proxies, alternative data, behavioral signals, or optimization targets. Agents will also make thousands of non-credit decisions (merchant recommendations, personalized pricing, dynamic risk profiling) that could trigger UDAAP or fair-lending issues (e.g., digital redlining) if they perpetuate systemic biases.

RECOMMENDED POLICY

Make AI fair-lending audits mandatory for covered banking use cases, and add a “Non-Credit Algorithmic Fairness” audit requirement for any agent that affects a customer’s financial experience or product eligibility.

WHAT SHOULD CHANGE

- Require pre-deployment disparate-impact testing.
- Require periodic protected-class proxy analysis.
- Require challenger models to test whether less discriminatory alternatives exist.
- Require banks to document why high-risk variables are necessary.
- Require remediation plans when disparities appear.

PROPOSED POLICY TEXT

“Banks must demonstrate that AI systems used in credit, pricing, servicing, collections, or account access do not create unlawful disparate treatment or unjustified disparate impact.”



12

Human override and dispute resolution

Liability

ISSUE

Consumers need a practical remedy when an AI agent makes a harmful or unexpected financial decision.

RECOMMENDED POLICY

Create a mandatory **human escalation right**.

WHAT SHOULD CHANGE

- Consumers can suspend any agent immediately, within a measurable maximum latency that the institution tests, and can dispute agent-initiated transactions.
- Banks must provide human review for high-impact outcomes.
- Banks must explain whether the agent, bank system, merchant, or third party caused the issue.
- Provisional-credit rules should apply where the transaction is plausibly unauthorized or outside delegation.

PROPOSED POLICY TEXT

“Consumers must have the right to suspend, revoke, dispute, and obtain human review of any financial transaction or decision materially caused by an autonomous agent, with suspension taking effect within a defined maximum latency that the institution tests and can demonstrate.”

13

Third-party AI vendors

Supervision

ISSUE

Banks will rely on model providers, cloud providers, agent platforms, orchestration tools, and data vendors. Existing third-party risk guidance covers vendors generally but does not impose AI-specific controls.



RECOMMENDED POLICY

Designate critical AI providers as **critical banking technology service providers** when their models materially affect banking operations.

WHAT SHOULD CHANGE

- Require model documentation, testing rights, audit rights, incident notification, and termination rights.
- Require banks to assess training data, model governance, cybersecurity, and subcontractors.
- Require fallback plans if a model provider fails, changes model behavior, or withdraws access.
- Require contractual restrictions on using bank/customer data for model training without consent.
- Extend diligence to the model and framework supply chain: hosting-environment security attestations, vendor vulnerability disclosure and patch commitments, and notification when the underlying model version changes materially.

PROPOSED POLICY TEXT

“A banking organization using a third-party AI system for a material banking function must retain sufficient access, auditability, contractual control, and operational substitutability to manage the risk, and must extend diligence to the hosting environment, orchestration framework, and underlying model version, with notification obligations for material model changes.”

The Fed, FDIC, and OCC already issued interagency guidance stating that banks should identify, assess, monitor, and control third-party relationship risks across the relationship lifecycle.

14

Model risk management

Supervision

ISSUE

Existing model risk management frameworks were built for statistical models, not autonomous agents that plan, reason, call tools, and adapt to context. Current supervisory guidance has left generative and agentic AI outside model-risk scope, leaving a



gap around exactly the systems posing the greatest model risk.

RECOMMENDED POLICY

The prudential regulators should expand the scope of model risk management guidance to cover agentic systems.

WHAT SHOULD CHANGE

- Classify AI agents as high-risk models when they can move money, extend credit, deny access, or communicate binding financial terms.
- Require model inventory, versioning, validation, red-teaming, and drift monitoring.
- Require “tool-use controls” for any agent connected to payment rails, accounts, credit systems, or customer data.
- Require independent validation before production use, and kill-switches and human escalation for high-risk actions.
- Weight ongoing oversight at least as heavily as pre-deployment validation, applying the mandate-enforcement, drift-monitoring, revocation, and adversarial-testing controls.

PROPOSED POLICY TEXT

“Any AI system capable of autonomous financial action, customer-impacting decisioning, or regulatory-compliance execution shall be treated as a material model subject to independent validation, monitoring, and governance, with ongoing oversight, including mandate enforcement, drift monitoring, tested revocation, and adversarial testing, treated as a supervisory focus equal to pre-deployment validation.”

15

Operational risk and cybersecurity

Supervision

ISSUE

AI agents introduce new threat vectors: prompt injection, tool hijacking, data poisoning, identity spoofing, autonomous fraud, and model compromise.

RECOMMENDED POLICY

Add AI-specific operational incident categories to bank cyber rules.



WHAT SHOULD CHANGE

- Require agents to be cryptographically declared and verified at the network/API layer, so the bank can confirm an authorized agent is connecting. Treat undeclared automation against customer accounts as presumptively hostile, rather than routing it through fraud models tuned for human behavior.
- Require least-privilege architecture. Keep credentials in a central vault with automated rotation, and detect agent credentials leaking into prompts, model inputs or outputs, or logs.
- Define AI-specific incident classes (model compromise, unauthorized agent action, data poisoning, prompt-injection exploitation, transaction malfunction), each with a dedicated response runbook (for compromised credentials, hijacked mandates, model or tool-chain poisoning, and rogue unregistered agents). Require tamper-evident logging, and notification when an agent causes material customer or payment harm.
- Require continuous behavioral-drift monitoring: track each agent's normal pattern (transaction velocity, counterparty set, data-access and tool-call profile) and flag deviations, with thresholds that can trigger automatic suspension, not just human-initiated review.
- Before deployment, require agents to pass a defined set of adversarial ("red team") tests: prompt injection, mandate escape, credential theft, tool misuse, data-scope overreach, and multi-agent confused-deputy scenarios, with documented pass/fail criteria and containment measures (throttling, credential revocation, agent suspension).
- Require a measured, board-reportable kill-switch: test and report the actual time from the decision to suspend an agent to full technical shut-off (credentials, sessions, payment instruments, tool access), rather than simply attesting that a kill-switch exists.

PROPOSED POLICY TEXT

"Banking organizations must maintain controls to detect, contain, and report AI-system compromise or autonomous behavior that may materially affect customers, payments, compliance, safety, or soundness, and must demonstrate, through periodic testing, the time between an incident determination and full technical containment of the affected agent."



16

Deposit insurance and AI-driven runs

Supervision

ISSUE

AI agents could move deposits at machine speed in response to rumors, rate changes, social signals, or liquidity warnings.

RECOMMENDED POLICY

Require banks to model and control **agentic liquidity-flight risk** in living-will or other contingency-planning scenarios.

WHAT SHOULD CHANGE

- Include AI-triggered withdrawal scenarios in liquidity stress testing.
- Require circuit breakers for mass agent-driven transfers.
- Require banks to monitor abnormal automated outflow patterns.
- Require regulator notification for significant AI-driven deposit movement.
- Require controls for business customers using treasury agents.

PROPOSED POLICY TEXT

“Liquidity risk management must account for automated and agent-coordinated deposit movement, including machine-speed withdrawal scenarios.”

17

Prudential supervision and safety-and-soundness

Supervision

ISSUE

AI can affect credit, liquidity, operational, compliance, market, and reputational risk simultaneously.

RECOMMENDED POLICY

Supervisors should require an **AI risk management program** for banks using material AI, the enterprise umbrella that ties together the component controls in third-party risk, model risk, and operational risk.



WHAT SHOULD CHANGE

- Require board-level AI risk oversight and senior-management accountability.
- Maintain an enterprise AI inventory with risk-tiering of models and agents.
- Require independent testing and internal-audit review, drawing on the third-party, model-risk, and operational controls rather than duplicating them.
- Require periodic regulatory reporting for high-risk use cases.

PROPOSED POLICY TEXT

“Banking organizations using material AI systems must maintain an enterprise AI governance framework commensurate with the system’s impact on customers, financial condition, compliance, and safety and soundness.”

18

Records, auditability, and examination

Supervision

ISSUE

Regulators cannot supervise agentic commerce unless they can reconstruct what happened.

RECOMMENDED POLICY

Require **examinable AI audit trails**.

WHAT SHOULD CHANGE

- Preserve prompts, instructions, tool calls, model versions, data sources, decisions, overrides, and transaction outputs.
- Require tamper-evident logs, specified technically through hash-chained or write-once storage.
- Require audit trails to be explainable to examiners.
- Require a standardized evidence format usable in dispute resolution, SAR filings, and examination.
- Require retention periods aligned with underlying banking laws.

PROPOSED POLICY TEXT



“Banks must maintain AI-system records sufficient to reconstruct any customer- or transaction-impacting AI action for examination, dispute resolution, and enforcement, stored in hash-chained or write-once form that demonstrates integrity, in a standardized format usable in SAR filings and examination.”

19

Machine-readable compliance

Supervision

ISSUE

Agentic finance will operate too quickly for manual compliance workflows.

RECOMMENDED POLICY

Regulators should publish machine-readable supervisory rules and safe-harbor APIs.

WHAT SHOULD CHANGE

- Standardize schemas for consent, delegation, revocation, transaction metadata, audit logs, and dispute evidence, aligned with the interoperable agent-identity and mandate standards so compliance and payment-authorization schemas converge instead of duplicate.
- Create regulatory APIs for sanctions screening, licensing checks, and reporting.
- Permit safe harbor for banks that use certified compliance controls.
- Require explainable policy engines for agent actions.

PROPOSED POLICY TEXT

“Regulators should define machine-readable compliance standards for autonomous financial transactions, including standardized delegation, monitoring, reporting, and revocation protocols.”



20

Agent threat-intelligence information-sharing safe harbor

Supervision

ISSUE

Agent-specific attacks (e.g. a prompt-injection payload, a compromised agent framework, or a credential-theft technique) will replay across institutions faster than traditional fraud. Current safe harbors for information sharing were not written with agent behavioral signatures or compromise indicators in mind, creating uncertainty that slows cross-institution sharing when speed matters most.

RECOMMENDED POLICY

Extend safe-harbor protections under Section 314(b) to the sharing of agent compromise indicators, behavioral signatures of malicious or hijacked agents, and known adversarial manipulation techniques among financial institutions and designated information-sharing organizations.

WHAT SHOULD CHANGE

- Clarify that sharing agent-specific indicators of compromise is protected activity under safe harbor.
- Support a designated sharing mechanism for agent-specific threat intelligence, with FinCEN and the banking agencies receiving aggregated trend reporting.
- Require standardized indicator formats so shared signatures are usable across institutions' detection systems.

PROPOSED POLICY TEXT

“Financial institutions may share indicators of AI-agent compromise, malicious agent behavioral signatures, and known adversarial-manipulation techniques with other financial institutions and designated information-sharing organizations under the safe-harbor protections applicable to suspicious-activity information sharing.”



Priority Implementation Roadmap

This roadmap sorts the recommendations by the authority each requires. It presents how soon each recommendation can move, e.g. what supervisors can do now through guidance and examination, what needs formal rulemaking, and what ultimately requires new statutes. Some items span phases by design, starting as guidance and hardening into rule or law as the framework matures.

Phase 1: Immediate supervisory guidance

- Define AI agents as third-party or internal systems subject to bank risk management.
- Require inventory, governance, audit logs, and consumer disclosures.
- Clarify that existing consumer protection, AML, fair lending, and payment laws apply.

Phase 2: Formal rulemaking

- Amend Regulation E for agent-initiated transfers.
- Amend AML/CIP rules for agent identity.
- Issue CFPB rules on AI disclosures and adverse action explanations.
- Expand third-party risk management to critical AI providers.

Phase 3: New statutory framework

- Create Authorized Financial Agent legal status.
- Create federal certification for high-risk financial agents.
- Amend GLBA for agent-specific data permissions.
- Establish liability allocation across consumers, banks, AI providers, merchants, and payment networks.
- Create real-time supervisory reporting for systemic agentic finance risk.



Appendix A:

Lead Bank Model KYA Program

1. Purpose and Scope

This document defines the discrete oversight responsibilities of five control functions: **Financial Crimes, Legal, Enterprise Risk Management (ERM), Regulatory Compliance, and Information Security (InfoSec)**, across the full lifecycle of agentic AI systems that initiate or facilitate transactions, customer interactions, or regulated decisions at or through a bank.

It applies to:

- Agents deployed internally by financial institutions.
- Agents deployed by partners, fintech operators, or merchants that transact through a bank.
- Agents acting on behalf of bank customers (consumer or business) against bank-issued credentials, accounts, or credit.
- Multi-agent (agent-to-agent) interactions where at least one principal is a bank customer, partner, or the bank itself.

It does not replace any existing program (BSA/AML/CFT/OFAC, Compliance Management System, Model Risk & Governance, Information Security, TPRM, Complaints, Fair Lending). It defines how each program's existing authorities extend to a new class of actor, the agent, and how the functions coordinate.

2. Operating Principles

- **No function owns Know Your Agent ("KYA") alone.** FinCrime is the Program DRI, but every oversight function below has a non-delegable role.
- **The agent is treated as a distinct actor with its own risk profile**, not as an extension of the principal's risk profile, and not as a "tool" of the operator.
- **Each function retains its existing escalation, veto, and reporting rights** under its source program. KYA adds responsibilities; it does not subtract



authority.

- **Pre-deployment review is mandatory for Tier 1 and Tier 2 agents** from every function listed below before risk committee or other governance presentation.
- **Post-deployment monitoring is continuous**, with quarterly attestations into the relevant internal governance body and exception-based escalation at any time.
- **All five functions have unilateral kill-switch authority** for agents within their domain (e.g., InfoSec for credentialing compromise; FinCrime for BSA/sanctions; Legal for legal-privilege or litigation-hold conflict; Regulatory Compliance for consumer-harm patterns; ERM for enterprise-risk threshold breach). Exercise is reported to the relevant governance committee within 24 hours.

3. Financial Crimes: Program DRI

3.1 Mandate

Financial Crimes owns the KYA AML Program end-to-end. Enterprise Risk and Financial Crimes is accountable to the relevant governance committee for the design, operation, and effectiveness of the Program.

3.2 Pre-deployment oversight

- Sets and maintains the **KYA risk-tiering methodology** (Tier 1-4) and approves the tier of every in-scope agent.
- Maintains the **KYA Standards Document** (operational standards) and the **KYA File template** (per-agent record of identity, operator, mandate, controls, monitoring plan).
- Reviews and approves the **mandate scope** of every Tier 1 and Tier 2 agent before production deployment, including spend ceilings, counterparty universe, geography, product scope, and revocation pathway.
- Performs **agent-specific BSA/AML/CFT/OFAC risk assessment** for each Tier 1/Tier 2 agent, documented in the KYA File and reconciled to the partner or enterprise risk assessment.
- Approves the **transaction monitoring plan** for each Tier 1/Tier 2 agent, including third party provider rule mapping and agent-tagging requirements.
- Reviews sanctions screening posture for any agent that can initiate cross-border payments, counterparty selection, or KYC/KYB-equivalent decisions.



3.3 Ongoing oversight

- Operates **agent-aware transaction monitoring** through the AML vendor, with rule packs or tags that distinguish agent-initiated activity from human-initiated activity.
- Investigates and dispositions alerts on agent-initiated activity; files SARs where warranted, including the agent's identifier and mandate excerpt in the narrative.
- Tracks **agent-driven typologies** (e.g., agent-mediated structuring, mule recruitment via agent intermediaries, sanctions evasion through agent layering, ATO amplification) and publishes a quarterly typology update to the Working Group.
- Performs **annual KYA refresh** for every in-scope agent, mirroring the CDD periodic-review cadence.
- Owns **agent offboarding**: formal revocation of mandate, credential disablement coordination with InfoSec, and retention of the KYA File for the regulatory record-retention period.

3.4 Reporting

- To relevant governance committees.

3.5 Non-delegable authorities

- Tier assignment.
- Mandate approval and modification for Tier 1/Tier 2 agents.
- SAR filing decisions on agent-initiated activity.
- OFAC blocking and rejection decisions where an agent is the initiator.
- Kill-switch exercise for BSA/sanctions cause.

4. Legal: Mandate, Liability, and Contractual Architecture

4.1 Mandate

Legal owns the legal-enforceability layer of the KYA Program: the contractual instruments that bind principals, operators, and Lead; the disclosure and consent architecture; the liability allocation; and the regulatory interpretation of agent authority. CLRO has program-level co-approval rights with the BSA Officer.



4.2 Pre-deployment oversight

- Drafts and maintains **mandate template language** for inclusion in:
 - Program Management Agreements (PMAs) for partner-deployed agents.
 - Consumer and business deposit and credit account agreements for customer-deployed agents acting on Lead-issued accounts.
 - Internal terms of use for Lead-deployed agents.
- Reviews and approves the **legal characterization** of each Tier 1/Tier 2 agent, specifically whether the agent is acting as (a) an agent of the customer, (b) an agent of the operator, (c) an agent of Lead, or (d) an independent third-party service provider. This characterization drives Reg E/Z liability allocation and is recorded in the KYA File.
- Reviews and approves **authorization and consent flows** for any agent that can initiate EFTs, originate credit, or take adverse action against a consumer.
- Reviews any **A2A (agent-to-agent) interaction patterns** where Lead's agent or a partner's agent transacts with a counterparty agent, including pass-through liability for the counterparty's actions.
- Identifies and tracks **regulatory interpretation gaps** (Reg E "authorized vs. unauthorized" doctrine applied to agents, Reg Z disclosure timing when an agent originates credit, ECOA adverse action when an agent makes the decision, UDAAP risk in agent-mediated marketing).

4.3 Ongoing oversight

- Maintains the **regulatory interpretation log**: written positions Lead has taken on novel agentic-commerce legal questions, with rationale and source authority.
- Reviews **dispute and complaint patterns** with Regulatory Compliance to identify legal-liability themes.
- Manages **regulator engagement** on KYA matters, including any examination requests, supervisory letters, or proactive briefings (coordinated with Compliance and the proposed regulator engagement track Erica has initiated).
- Manages **litigation, subpoenas, and discovery** involving agent activity, including litigation-hold scope over agent logs, KYA Files, and mandate evidence.
- Owns **intellectual property and licensing** review of third-party agentic protocols (ACP, AP2, KYAPay, etc.) before adoption.



4.4 Reporting

- To relevant governance committees.

4.5 Non-delegable authorities

- Approval of mandate template language and any deviation.
- Legal characterization of an agent's principal relationship.
- Approval to engage with a regulator on a KYA matter.
- Litigation hold scope over agent evidence.
- Kill-switch exercise for legal-privilege, litigation-hold, or enforcement cause.

5. Enterprise Risk Management: Aggregate Exposure and Risk Appetite

5.1 Mandate

ERM owns the enterprise-level view of agentic-commerce risk: aggregation across partners, concentration limits, risk-appetite calibration, and integration into the Enterprise Risk Assessment (ERA) and the Enterprise Risk Profile. ERM ensures KYA does not become a siloed FinCrime artifact and that agentic risk is reflected in the Bank's overall risk posture.

5.2 Pre-deployment oversight

- Establishes and maintains **agentic-commerce risk appetite statements** for Board approval, including:
 - Maximum aggregate agent-initiated transaction volume as a percent of total transaction volume.
 - Concentration limits by operator, by protocol, and by agent type.
 - Maximum tolerated residual risk by tier.
 - Tolerance for novel-protocol pilots vs. mature-protocol production deployments.
- Reviews each Tier 1 agent's **inherent and residual risk rating** before deployment and confirms it fits within the approved risk appetite.
- Maintains the **KYA risk taxonomy** within the broader Enterprise Risk Taxonomy, ensuring agent-specific risks (mandate drift, principal repudiation, agent impersonation, multi-agent collusion, protocol exploitation) are explicit and not absorbed into generic third-party or operational risk



categories.

- Coordinates with TPRM to ensure operator-level due diligence for agent operators is consistent with third-party risk standards.

5.3 Ongoing oversight

- Aggregates KYA File data quarterly to produce an **enterprise agentic-commerce risk dashboard**: count of in-scope agents by tier, by operator, by protocol; aggregate transaction volume and dollar value; aggregate loss experience; aggregate alert and SAR volume; aggregate complaint volume; concentration metrics against limits.
- Updates the **Enterprise Risk Assessment (ERA)** annually to include agentic-commerce risk as a named risk domain.
- Reviews and challenges Tier-3/Tier-4 agent self-attestations on a risk-based sampling basis.
- Coordinates **scenario analysis and stress testing** for agentic-commerce risk (e.g., a coordinated agent failure event, a protocol-level exploit, a regulator-driven pause).
- Owns **issue management** for KYA-derived issues into the enterprise Issue Management process, with tracking to closure.

5.4 Reporting

- To relevant governance committees.

5.5 Non-delegable authorities

- Risk-appetite statement design and breach escalation.
- Enterprise risk assessment integration of agentic-commerce risk.
- Concentration-limit design and monitoring.
- Kill-switch exercise for risk-appetite or concentration-limit breach.

6. Regulatory Compliance: Consumer Protection, Disclosures, and Complaints

6.1 Mandate

Regulatory Compliance owns the consumer-protection layer of agentic commerce and payments: disclosures, consent, fair-lending applicability, UDAP, complaints, and the Compliance Management System (CMS) extension to agent-mediated activity. The CCO holds program-level concurrence



rights with the BSA Officer and CLRO for any consumer-facing agentic-commerce deployment.

6.2 Pre-deployment oversight

- Reviews and approves **all consumer-facing disclosures** for agents that interact with consumers, originate credit, initiate EFTs, take adverse action, or market products. Includes Reg E initial and change-in-terms disclosures, Reg Z disclosures, ECOA adverse-action notices, true lender and UDAAP-relevant marketing claims.
- Reviews **consent capture flows** to ensure they meet legal authorization standards (works with Legal) and consumer-protection standards (clarity, conspicuousness, revocability).
- Reviews **fair-lending applicability** for any agent that influences a credit decision or its terms, applying the Fair Lending risk assessment methodology to the agent itself (not just the underlying model) and confirming with FinCrime/MRM whether ECOA / Reg B testing is required.
- Reviews **marketing and product representations** made by or about agents (including “AI-assisted” or “agent-recommended” framing) for UDAAP and advertising-rule compliance.
- Confirms each Tier 1/Tier 2 agent has a complete **complaint-handling pathway**, i.e., the consumer knows how to complain about agent behavior, the complaint is routed correctly, and the agent’s role is captured in the complaint record.

6.3 Ongoing oversight

- Operates **agent-tagged complaint intake** so any complaint touching agent-initiated activity is flagged, routed, and trended separately.
- Performs **quarterly complaint trend analysis** for agent-initiated activity, with root-cause and consumer-harm assessment; escalates patterns to FinCrime, Legal, and the Working Group.
- Performs **periodic disclosure and consent testing** for each Tier 1/Tier 2 agent (sample-based) to confirm what the consumer actually sees matches the approved language.
- Coordinates **Reg E error-resolution** for unauthorized-EFT claims involving agents, specifically the disputed question of “was this transaction authorized by the consumer” when an agent executed it. Documents the dispositional standard Lead applies and tracks consistency.



- Maintains **Fair Lending monitoring** for agent-influenced credit decisions, including disparate-impact testing on a defined cadence.
- Owns **regulatory change management** for new or amended consumer-protection rules relevant to agentic commerce and payments (CFPB guidance, prudential-regulator AI guidance, state UDAP).

6.4 Reporting

- To relevant governance committees.

6.5 Non-delegable authorities

- Approval of consumer-facing disclosures and consent flows for agent activity.
- Reg E error-resolution disposition standard for agent-initiated EFTs.
- Fair-lending testing scope for agent-influenced credit decisions.
- Kill-switch exercise for consumer-harm pattern, fair-lending disparity, or UDAAP cause.

7. Information Security: Identity, Credentialing, and Technical Control

7.1 Mandate

InfoSec owns the technical identity and access-control layer for agents: how an agent is provisioned with credentials, how those credentials are scoped and revoked, how the agent's actions are logged and made tamper-evident, and how the agent's technical environment is hardened. The CISO holds program-level co-approval rights with the BSA Officer for any agent that is provisioned with Lead credentials or accesses Lead systems.

7.2 Pre-deployment oversight

- Operates the **Agent Identity Registry**: the system of record for every in-scope agent's cryptographic identity, principal binding, credential set, and entitlement scope.
- Issues **agent credentials** under a documented standard (preferred: short-lived, scoped, revocable, cryptographically bound to a principal mandate). Coordinates with Legal on mandate-credential binding.
- Reviews and approves **technical scope** for each Tier 1/Tier 2 agent: API entitlements, rate limits, data-access scope, environment isolation, secret management.



- Performs **security architecture review** of the agent's deployment environment, including the operator's environment for partner-deployed agents (coordinated with TPRM).
- Reviews **third-party protocol adoption** (ACP, AP2, KYAPay, MCP, A2A) for cryptographic soundness, identity-binding rigor, and revocation pathway integrity before Lead adopts.
- Reviews **prompt-injection, jailbreak, and adversarial-input** posture for any agent that consumes external instructions or untrusted context, coordinated with MRM and Product/Data Science.

7.3 Ongoing oversight

- Operates **continuous credential lifecycle management** (provisioning, rotation, revocation, emergency disablement) with documented SLAs.
- Operates **agent activity logging and integrity**: tamper-evident logs of every agent action, mandate evidence, principal authorization, and credential use; retention aligned to the record-retention policy.
- Operates **anomaly and security-incident detection** specific to agent behavior: credential misuse, mandate-scope violation attempts, lateral-movement attempts, exfiltration patterns, impersonation attempts.
- Manages **agent security incidents** under the existing Incident Response plan with KYA-specific runbooks (credential compromise of an agent; operator compromise resulting in agent takeover; protocol-level exploit).
- Coordinates with FinCrime on **agent-driven ATO and impersonation typologies**, contributing technical indicators to monitoring rules.
- Performs **annual penetration testing** that includes agent-specific attack scenarios for Tier 1 agents.

7.4 Reporting

- To relevant governance committees.

7.5 Non-delegable authorities

- Agent credential issuance, rotation, and revocation.
- Technical scope approval for Tier 1/Tier 2 agents.
- Protocol cryptographic and identity-binding acceptance.
- Kill-switch exercise for credentialing, identity-binding, or security cause.



8. Coordination, Handoffs, and Decision Rights

8.1 Required pre-deployment concurrence (Tier 1 and Tier 2 agents)

No Tier 1 or Tier 2 agent may be presented to MRCC for deployment approval without written concurrence from all five functions on the items listed in Sections 3-7 above. Absent any single concurrence, the agent is not eligible for deployment. The KYA Working Group is the forum for resolving concurrence disagreements; unresolved disagreements escalate to the relevant governance committee.

8.2 Standing handoff matrix

Trigger	Originating function	Notify within	Receiving functions
New agent intake (any tier)	Sponsoring business + Product	2 business days	FinCrime (lead), all four other functions
Tier 1/Tier 2 mandate change request	Sponsoring business	Pre-change	FinCrime (lead), Legal, InfoSec
Credential compromise	InfoSec	Same day	FinCrime, Legal, ERM, Reg Compliance
SAR pattern on agent activity	FinCrime	Per SAR cadence	Legal, ERM
Reg E unauthorized-EFT claim involving agent	Reg Compliance	2 business days	Legal, FinCrime
Fair-lending disparity signal	Reg Compliance	5 business days	Legal, FinCrime, ERM
Risk-appetite breach	ERM	Same day	All four other functions + MRCC
Litigation hold or regulator inquiry	Legal	Same day	All four other functions
Kill-switch exercise (any function)	Exercising function	24 hours	All four other functions + MRCC



9. Three-Lines-of-Defense Mapping

Line	Roles
1LOD	Product, Data Science, Engineering, BaaS Partnerships, sponsoring business units. Build and operate agents; perform first-line controls; own KYA File completeness for their agents.
2LOD	FinCrime, Legal, ERM, Regulatory Compliance, InfoSec, TPRM, MRM. Independent oversight, challenge, approval, monitoring, and reporting as defined in Sections 3-7.
3LOD	Internal Audit. Independent assurance over the KYA Program design and operating effectiveness on the audit plan cadence.

