JUDITH C. WINN

Proposal and Comment Information

Title: Anti-Money Laundering and Countering the Financing of Terrorism Programs, R-1835

Comment ID: FR-0000-0133-02-C26

Subject

Docket No. R-1835 and RIN 7100-AG78

Submitter Information

Name: Judith C. Winn

Submitted Date: 09/01/2026

Dear Sir or Madam,

Please find attached my comment on the proposed rule on Anti-Money Laundering and Countering the Financing of Terrorism Programs, Docket No. R-1835, RIN 7100-AG78, 91 FR 42363 (July 9, 2026).

The comment is attached as a PDF and is submitted in my personal capacity.

Kind regards
Judith C. Winn

JUDITH C. WINN, CISSP, CCISO
Information Security and AI Governance, Knysna, Western Cape, South Africa

August 31, 2026

Benjamin W. McDonough
Secretary, Board of Governors of the Federal Reserve System
20th Street and Constitution Avenue NW
Washington, DC 20551

Re: Anti-Money Laundering and Countering the Financing of Terrorism Programs
Docket No. R-1835; RIN 7100-AG78; 91 FR 42363 (July 9, 2026)

Dear Mr. McDonough:

I write in response to the Board's request for comment on the above proposal. I lead the information security and AI governance function at a payment service provider, and my work covers payment security and financial-crime controls in multiple regulated jurisdictions. I hold the CISSP and CCISO credentials, ISO/IEC 27001 Lead Implementer certification, and a Master of Science in Cybersecurity. I submit this comment in my personal capacity; my professional affiliation is given for identification only, and nothing here represents the position of my employer. I also maintain an open, freely licensed threat model in this subject area, cited at the end of this comment; I have no commercial interest in it.

My comment responds to the Board's requests for comment on the risk assessment process, on when banks must review or update those processes, and on the treatment of innovative technology.

Summary

Autonomous AI agents that initiate and authorize payments on a customer's instruction are in production in the United States market. I am not going to argue that they create new money laundering typologies. I went looking for those and largely did not find them. I set that out below, because the contrary claim is being made freely at the moment and I do not think it holds.

The narrower point is this. Agent-mediated payments generate a class of evidence that the Bank Secrecy Act reporting taxonomy has no structured place to record. They also discard authorization evidence that is richer than what the regulated record currently carries. These are reporting and data problems. Neither is a new criminal method, and both can be addressed by extending instruments the Board and FinCEN already maintain.

I. What this comment does not claim

One clarification before the substance, because the current discussion around agentic AI and financial crime contains a good deal of assertion.

I examined ten candidate money laundering typologies specific to agent-initiated and machine-to-machine payments, and checked each against existing coverage. Nine reduce cleanly to typologies that are already named, documented and instrumented. Delegated agent authority is nominee control with a remote controller, described in FATF's 2018 report on professional money laundering. Agent-driven transaction fragmentation is structuring, reached by 31 U.S.C. § 5324 and by the funnel account advisory at FIN-2014-A005. Machine-to-machine settlement layering is structuring combined with service-based trade laundering. Agent-operated intermediate wallets are peel chains, and deterministic wallet generation has made those holding nodes free and disposable since 2012. Agents change the cost and the speed of these schemes. They do not change how the schemes work.

I set this out because if I claimed a novel threat and the claim did not hold, the recommendations resting on it would not hold either. So far as I can establish, no enforcement action, indictment or financial intelligence unit typology report attributes laundering to an autonomous agent. FATF published Horizon Scan: AI and Deepfakes on December 22, 2025, which it describes as part of its staged approach to emerging technologies and which sets out hypothetical scenarios in which AI agents automate laundering. I am not aware of any AML standard-setter that has yet published a typology, grounded in observed cases, treating autonomous agents as a distinct risk category.[1]

There is a second point that follows from the first. What agent adoption changes is not the method. It is whether the method remains visible.

In my experience, automated transaction behavior has until now been an anomaly signal in its own right. Machine timing and machine consistency were themselves grounds for scrutiny, because a human customer did not produce them. As a material share of legitimate volume becomes agent-initiated, that signal stops discriminating. An agent operated by a criminal, whether through stolen credentials or in concert with an insider, produces activity that is authenticated, within policy, and hard to distinguish from the same customer's legitimate agent activity.

I raise it because the gaps set out in Sections III through V are what would ordinarily replace a lost signal, and none of them is instrumented today. My concern is that a detection signal is being retired by commercial adoption while its replacement does not yet exist.

II. Signature validity is not intent integrity

The protocols now carrying agent-initiated payments secure them through cryptographic proof of instruction: the signed Intent and Cart Mandates of the Agent Payments Protocol, HTTP Message Signatures under RFC 9421 in Visa's Trusted Agent Protocol, Stripe's scoped Shared Payment Tokens, Mastercard's Agentic Tokens. Each establishes that a particular instruction was issued by a particular agent and has not since been altered. That is a real and necessary security property.

None of them establishes that the agent was operating under its principal's control when it formed the instruction. That is a different property. Indirect prompt injection defeats it. An attacker who places instructions in content the agent processes, whether a web page, a product listing, a message or a tool response, can redirect the agent's behavior while every credential, signature and audit record remains valid.

In March 2026 Palo Alto Networks Unit 42 published two attack scenarios specific to agentic commerce: concealed instructions injecting unauthorized line items into the Cart Mandate of Google's Universal Commerce Protocol, which is compatible with the Agent Payments Protocol, and malicious metadata driving a commerce protocol state machine to issue refunds without verification.[2] The first attacks the mandate object these protocols offer as their principal security control; the second attacks the verification logic around it. The underlying mechanism has been demonstrated in shipping products. Brave's security research team concluded in October 2025 that indirect prompt injection is *"not an isolated issue, but a systemic challenge facing the entire category"* of AI-powered browsers, noting that such a system executes with its principal's authenticated privileges.[3]

For an anti-money-laundering program, the resulting transactions are well-formed, authenticated and correctly attributed to a real customer who did grant standing authority. The customer is an unwitting participant, a category the Bank Secrecy Act already recognizes through FinCEN's 2020 advisory on money mules. The category of person is not new. What is new is that the compromise happens at the reasoning layer rather than the credential layer, so no authentication signal is disturbed. Much of this will present first as fraud rather than laundering. It reaches the Bank Secrecy Act regime because the resulting flows, and the reports filed on them, do.

III. The instruction channel has no place in the reporting taxonomy

FinCEN's advisory on cyber-events directs that suspicious activity reports involving cyber-events should include, to the extent available, *"relevant IP addresses and their timestamps," "device identifiers" and "methodologies used."*[4] That vocabulary was built for compromise at the network layer, and it works well there.

There is no key term, structured field or indicator convention for the instruction channel in an agent-mediated payment: the address and content hash of the ingested artifact that carried the injected instruction, the identity of the tool or model-context server whose description delivered it, or the agent session in which it executed.

A single injection campaign reaches many victims through one artifact, and that artifact is often the only element common across all of them. It is what would allow a financial intelligence unit to cluster a large number of otherwise unrelated reports into a single case. Under current practice it is recorded, if at all, as free text in a narrative, where it is neither searchable nor shareable as an indicator.

IV. No record distinguishes agent-initiated activity

No currency transaction report or suspicious activity report field records that a transaction was initiated by an autonomous agent acting on a customer's standing authority rather than by the customer directly.

This is not an attribution failure. The account holder is identifiable, the transmittal order is retained, and the existing provisions for recording a person on whose behalf a transaction is effected remain available. The customer can be named, and should be.

What it means is that no dataset exists in which agent-initiated activity can be measured. No detection model can be trained on the distinction, and no supervisor can assess prevalence. In my view the absence of documented cases discussed in Section I should be read partly as an absence of collection capability rather than as evidence of absence.

V. The best available authorization evidence is discarded at the regulated boundary

I would ask the Board to give this section particular attention. What it describes is a chance to improve the regulated record.

The agentic payment protocols now in production generate signed, non-repudiable records identifying the authorizing principal, the scope of the authority granted, and the specific action taken under it. In several respects this is richer authorization evidence than a conventional wire message carries. Section II establishes that a signed mandate does not prove the instruction reflected the principal's intent. It does prove who granted the authority, what scope was granted, and what was executed under it, and preserving that record is what allows an investigator to compare what was authorized with what was actually done.

Neither the recordkeeping and travel rule field set at 31 CFR 1010.410(f) nor the revised FATF Recommendation 16 field set has anywhere to carry it. The payment rail generates the data and it is discarded at the point where the regulated record begins. A receiving institution performing screening or a supervisor conducting an examination sees none of it.

Extending the field set to carry an authorizing-principal identifier would improve the quality of the regulated record. It would also give supervisors a way to distinguish agent-initiated flows, which addresses Section IV.

VI. Implications for the risk assessment process and the update requirement

The proposal requires a bank to update its risk assessment process *"promptly upon any change that the bank knows or has reason to know significantly changes"* its money laundering and terrorist financing risks. In my professional view, material customer adoption of agent-initiated payment channels is such a change. It alters what evidence of the customer's intent exists, which monitoring signals remain reliable, and what a filed report can usefully record.

The adoption figures are already published. A Cloud Security Alliance survey of financial-sector professionals, reported in June 2026, found that 62 percent of respondents say their organization is using AI agents. Of those, 55 percent operate under limited autonomy, 33 percent under conditional autonomy, and 5 percent have already granted high autonomy. Eighty-five percent of respondents expect AI agents to initiate and execute payment transactions.[5] Mastercard enabled agent-initiated payments for Citi and U.S. Bank cardholders in September 2025.[6]

Behavioral transaction monitoring is affected as well. It infers intent from patterns characteristic of human decision-making. As agent-originated activity enters shared customer segments, I would expect the reference population to shift, pooled variance to widen, and per-customer deviation scores to fall

across the segment, including for customers who have adopted no agent at all. Existing model-tuning obligations cover this in principle. The question is whether tuning cadence and segmentation practice keep pace with a shift driven by commercial adoption that no individual bank controls.

VII. Recommendations

1. Clarify that a bank's risk assessment process should account for non-human transaction initiation, and indicate whether the Board regards this as a characteristic of a distribution channel, of the customer relationship, or as a distinct dimension of risk.

2. Clarify whether material customer adoption of agent-initiated payment channels constitutes a change that significantly changes a bank's money laundering and terrorist financing risks for the purposes of the update requirement.

3. In coordination with FinCEN, establish a structured indicator set and key term for the instruction channel in agent-mediated payments, as an extension of the existing cyber-event indicator regime, so that a common artifact across many reports becomes searchable and shareable.

4. In coordination with FinCEN, introduce a means of recording that a reported transaction was agent-initiated, so that prevalence can be measured and supervisory expectations can be set on measured data.

5. Support extension of the recordkeeping and travel rule field set to carry an authorizing-principal identifier, so that authorization evidence generated by agentic payment rails is preserved in the regulated record rather than discarded at the boundary.

These recommendations are sequenced by burden. Recommendation 3 requires an advisory rather than a rulemaking, Recommendation 4 yields the prevalence data on which any further step should depend, and Recommendation 5 need proceed only if that measured data justifies the implementation cost. On the distinction the proposal draws between establishing and maintaining a program, I would add one observation. Controls for agent-initiated transactions are likely to require revision more often than most anti-money-laundering controls, because the underlying protocols are themselves changing on a timescale of months. A program reasonably designed in January may not be reasonably designed in September, through no failure of maintenance. Guidance acknowledging that reasonable design is assessed against the state of knowledge at the time of design would help institutions act before the position settles. On the proposal's treatment of innovative technology, the protocols that create the monitoring problem described above are the same protocols that generate the evidence described in Section V. Guidance should neither treat agent initiation as inherently elevated risk nor allow the authorization evidence it produces to be discarded.

The intent-integrity analysis in Section II is developed at greater length, and mapped to existing control frameworks, in an open threat model I maintain for autonomous agents in payment workflows: J. C. Winn, Agentic Finance Threat Model (2026), DOI 10.5281/zenodo.22115281, licensed CC BY 4.0. It is offered as the basis for the analysis above rather than as a proposal for the Board. I would be glad to expand on any part of this comment if it would assist the Board's staff.

Thank you for the opportunity to comment.

Respectfully submitted,

Judith C. Winn, CISSP, CCISO

Submitted in a personal capacity

NOTES AND SOURCES

1. FATF, Horizon Scan: AI and Deepfakes, December 22, 2025.
2. Palo Alto Networks Unit 42, "Who's Really Shopping? Retail Fraud in the Age of Agentic AI," March 20, 2026.
3. Brave Software security research, "Unseeable prompt injections in screenshots: more vulnerabilities in Comet and other AI browsers," October 21, 2025 (updated October 31, 2025).
4. FinCEN Advisory FIN-2016-A005, Advisory to Financial Institutions on Cyber-Events and Cyber-Enabled Crime, October 25, 2016. See also FIN-2020-A003 on money mules for the unwitting-participant category referenced in Section II.
5. Cloud Security Alliance, "Is Financial Services Ready for Agentic Payments?", June 23, 2026, reporting findings from State of Cloud and AI for Financial Services 2026.
6. Mastercard, "Making agentic AI commerce a reality with embedded, trusted payments," September 30, 2025.