Public Comment on NCCoE Concept Paper

Accelerating the Adoption of Software and AI Agent Identity and Authorization

Submitted by: Empty Set LLC Contact: Brice Love, Co-Founder Email: brice@emptysetllc.com

Date: March 14, 2026 **Submitted to:** AI-Identity@nist.gov **Reference:** NCCoE Concept Paper, February 2026 **Prior Submission:** NIST-2025-0035 (March 6, 2026, Tracking: mmf-286f-aneg)

I. Executive Summary

Empty Set LLC respectfully submits this comment on the NCCoE Concept Paper, "Accelerating the Adoption of Software and AI Agent Identity and Authorization." We previously submitted a public comment on March 6, 2026 in response to the CAISI Request for Information on Security Considerations for Artificial Intelligence Agents (Docket NIST-2025-0035, Tracking: mmf-286f-aneg), in which we argued that behavioral safety constitutes a distinct and unaddressed security domain. That comment was submitted under the working name "AI Agent Credentialing Standards Commission." Our standards initiative has since been formally named **The Box Commons**, a 501(c)(6) trade association in formation dedicated to AI agent credentialing standards.

This comment extends our March argument into the identity and authorization domain: **an AI agent's identity is incomplete without its verified behavioral safety posture, and authorization to act in sensitive contexts should require behavioral safety certification.**

The concept paper correctly identifies that organizations need to understand how identification, authentication, and authorization apply to AI agents. It proposes a practical NCCoE demonstration project anchored in existing standards—OAuth 2.0/2.1, OpenID Connect, SPIFFE/SPIRE, SCIM, and NGAC. These are necessary infrastructure components. But they address only one dimension of the trust problem: *can this agent prove it is who it claims to be, and does it have permission to access this resource?*

A second dimension is missing: *has this agent been independently verified to behave safely within the scope of its authorization?* An agent that authenticates flawlessly via OAuth, operates within its delegated authority, and logs every action in a tamper-proof audit trail can still cause catastrophic harm if its behavioral parameters permit unsafe interactions. This is the operative reality established by *Garcia v. Character Technologies, Inc.* and *Gavalas v. Google LLC*, and it is the reality that the insurance market is pricing through the Verisk ISO exclusion wave.

We propose that NIST integrate behavioral safety credentialing as a component of agent identity metadata, as an input to authorization policy decisions, and as a dimension of audit and non-repudiation. We further propose a fourth use case for the NCCoE demonstration project: behavioral safety credentialing as an authorization gate for agent deployment.

II. Responses to Concept Paper Questions

1. General Questions to Inform Choice of Demonstration Use Case

What enterprise use-cases are organizations currently using agents for? Which use-cases are in the near future? What opportunities do agents present?

Beyond the three use cases identified in the concept paper (workforce efficiency, security, software development), a critical emerging use case is **consumer-facing autonomous agents**—AI companions, financial advisors, healthcare navigators, and customer service agents that interact directly with the public, including vulnerable populations. In the near future, enterprises will deploy multi-agent orchestration systems where agents delegate to and coordinate with other agents, autonomous procurement agents that negotiate contracts and authorize purchases, and regulatory compliance agents that interpret and apply complex regulatory frameworks across jurisdictions. Each of these near-future use cases amplifies the identity and authorization challenge: when agents delegate to other agents, who is the principal? When an agent authorizes a purchase, what behavioral constraints govern its negotiation tactics?

The opportunities are significant—agents can extend organizational capacity, reduce decision latency, and enable small enterprises to operate with capabilities previously reserved for large organizations. But these opportunities are only realizable if the trust infrastructure exists to deploy agents with confidence. Identity and authorization standards are that trust infrastructure, and behavioral safety credentialing is the missing layer that connects technical trust to commercial insurability.

The insurance market is already treating consumer-facing agent deployment as an authorization problem. Armilla AI requires independent third-party verification of model performance, bias assessment, and behavioral robustness before issuing coverage. Relm Insurance's NOVAAI product tiers coverage based on verified governance structures and external certifications. These insurers are constructing de facto authorization gates: without credentialing, no coverage; without coverage, no commercially viable deployment.

What are the core characteristics of agentic architectures? How do AI agents differ from other forms of software agents? How are agentic architectures different from current microservices architectures?

The core characteristics that make agentic architectures distinct from prior automation paradigms are: (1) non-deterministic reasoning—the same input may produce different outputs depending on model state, context, and reasoning chain; (2) dynamic tool selection—agents choose which tools to invoke at runtime rather than following predetermined execution paths; (3) persistent state accumulation—agents build context across interactions through memory and conversation history; and (4) unbounded action spaces—the set of possible agent actions cannot be fully enumerated at deployment time.

These characteristics distinguish AI agents from traditional software agents (cron jobs, RPA bots, microservice workers) in a fundamental way relevant to identity and authorization. Traditional software agents are deterministic: given the same input and state, they produce the same output. Their authorization can be statically defined because their action space is bounded and predictable. AI agents are non-deterministic: their reasoning and actions emerge from the interaction of model weights, context, and input in ways that cannot be fully anticipated. This means that identity must encompass not just *what* the agent is permitted to access but *how* it is certified to behave within that access scope.

Microservices architectures differ from agentic architectures in a parallel way. A microservice exposes a fixed API with defined inputs and outputs. Its authorization scope is the set of endpoints it can call and the data it can access—fully enumerable, statically configurable. An agentic architecture introduces a reasoning layer between input and action that makes authorization inherently dynamic. The agent may decide at runtime to access a tool or resource that was not anticipated when its authorization policy was written. This is why the concept paper's question about "least privilege" for agents whose "required actions might not be fully predictable" is so critical—and why behavioral safety certification provides a necessary complement to resource-based authorization.

What risks worry you about agents?

The risk that concerns us most is the gap between infrastructure identity and behavioral identity. An agent can be cryptographically authenticated, operating within its OAuth-delegated scope, and still cause harm through behavioral patterns that no access control would prevent. The concept paper's framing assumes that if we solve identification, authentication, and authorization at the infrastructure level, agents will be trustworthy. The judicial and actuarial record of 2025-2026 demonstrates otherwise.

What support are you seeing for new protocols such as Model Context Protocol (MCP)?

MCP adoption is accelerating rapidly across the agentic AI ecosystem. MCP's reliance on OAuth and OIDC for authorization is architecturally sound. However, MCP's current specification addresses *resource access* authorization—can this agent use this tool, access this data source?—but does not address *behavioral* authorization—is this agent certified to interact safely with users in this context? As MCP becomes the de facto standard for agent-to-resource and agent-to-agent communication, behavioral safety credentials should be expressible within the protocol's authorization framework.

In what ways do agentic architectures introduce identity and authorization challenges?

Agentic architectures introduce a challenge the concept paper acknowledges but does not fully develop: **the non-deterministic nature of agent actions.** The paper asks how to establish "least privilege" for an agent "especially when its required actions might not be fully predictable when deployed." This unpredictability is precisely why behavioral safety credentialing is essential. Traditional authorization assumes a bounded action space. Agentic systems have an unbounded action space constrained only by their behavioral parameters. Authorization policies must therefore incorporate verified behavioral constraints—not just resource access permissions—to meaningfully limit risk.

What current or roadmap technology does your organization have that supports agents?

Empty Set LLC is developing credentialing methodology and assessment frameworks for AI agent behavioral safety, anchored in a portfolio of 43 patents and provisional patent applications covering AI agent credentialing, behavioral safety assessment, digital identity for autonomous systems, and autonomous corporate governance. We are establishing **The Box Commons**, a 501(c)(6) trade association, to house the credentialing standards initiative as an independent, industry-governed body.

Our technical roadmap includes: (1) a behavioral safety assessment instrument designed for integration with W3C Verifiable Credentials; (2) a certification lifecycle management system for issuing, renewing, and revoking behavioral safety credentials; and (3) integration specifications mapping behavioral safety credentials to the authorization frameworks identified in this concept

paper—OAuth 2.0/2.1, NGAC, and SCIM. We are positioned to serve as a technology collaborator in the NCCoE demonstration project.

What standards exist, or are emerging, to support identity and access management of agents?

In addition to the standards the concept paper identifies (OAuth 2.0/2.1, OIDC, SPIFFE/SPIRE, SCIM, NGAC), NIST should consider the **W3C Verifiable Credentials (VC) standard** as a mechanism for expressing behavioral safety certifications within agent identity metadata. Verifiable Credentials provide a cryptographically secure, privacy-preserving, and interoperable format for third-party attestations. A behavioral safety credential issued by a certified auditor could be embedded in an agent's identity profile, presented during authorization flows, and verified by relying parties—including insurers—without requiring direct access to the agent's internal architecture.

We also note that NIST's own guidelines—SP 800-207 (Zero Trust Architecture), SP 800-63-4 (Digital Identity Guidelines), and NISTIR 8587 (Protecting Tokens and Assertions from Forgery, Theft, and Misuse)—provide foundational frameworks that should be explicitly extended to address behavioral safety credentials as a class of identity assertion. SP 800-63-4's identity assurance levels, in particular, could inform a parallel framework for behavioral safety assurance levels.

The ISO/IEC 42001 standard for AI management systems and the emerging IEEE P7001 (Transparency of Autonomous Systems) also provide frameworks that could inform behavioral safety credentialing criteria.

2. Identification

How might agents be identified in an enterprise architecture?

The concept paper asks what metadata is essential for an AI agent's identity. We propose that agent identity metadata should include, at minimum:

- **Model provenance**: Model family, version, fine-tuning history, and training data lineage.
- **Capability declaration**: The set of tools, resources, and interaction modalities the agent is designed to access.
- **Behavioral safety posture**: A machine-readable record of the agent's current behavioral safety certifications, including the certifying body, certification date, scope of certification, and expiration. This record should be structured as a W3C Verifiable Credential or equivalent cryptographically verifiable attestation.
- **Deployment context classification**: Whether the agent operates in enterprise-internal, consumer-facing, or mixed contexts, and whether it interacts with vulnerable populations (minors, patients, individuals in crisis).
- **Incident history digest**: A summary-level record of behavioral safety incidents or near-misses, structured for integration with insurance underwriting without exposing proprietary operational details.

This metadata schema extends the identity attributes envisioned in SP 800-63-4 by introducing behavioral safety as an identity dimension alongside the traditional attributes of authentication assurance and federation assurance.

Should agent identity metadata be ephemeral or fixed?

Both. Core identity attributes (model provenance, organizational binding) should be persistent.

Behavioral safety credentials should have defined validity periods with mandatory renewal—analogous to how SSL certificates have expiration dates and must be reissued by certificate authorities. This ensures that behavioral safety posture is continuously verified rather than established once and assumed indefinitely. Behavioral drift—post-deployment changes in agent behavior from ongoing learning or interaction patterns—makes perpetual credentials dangerous.

Should agent identities be tied to specific hardware, software, or organizational boundaries?

Agent identities should be bound to the deploying organization and the specific deployment configuration (model version + system prompt + tool access + behavioral safety controls). A change in any of these components should trigger re-identification and potentially re-certification. This is analogous to how a software update invalidates a prior security certification—the updated system is a different entity from a trust perspective.

3. Authentication

What constitutes a strong authentication for an AI agent?

Strong authentication for an AI agent should include two dimensions:

1. **Infrastructure authentication**: Cryptographic proof of identity via the mechanisms the concept paper already identifies (OAuth tokens, SPIFFE SVIDs, X.509 certificates). This answers: "Is this agent who it claims to be?"
2. **Behavioral attestation**: Cryptographically verifiable proof that the agent holds a current behavioral safety credential from a recognized certifying body. This answers: "Has this agent been independently verified to behave safely within its declared scope?"

Infrastructure authentication without behavioral attestation is analogous to verifying that a driver has a valid license without checking whether they have insurance. The license confirms identity; the insurance confirms that a third party has assessed and accepted the risk of that entity operating in the environment. Both are required before the entity is authorized to act. Token protection mechanisms described in NISTIR 8587 should be extended to cover behavioral safety credential tokens with equivalent protections against forgery, theft, and replay.

How do we handle key management for agents? Issuance, update, and revocation?

Behavioral safety credentials require a lifecycle management process parallel to—but distinct from—cryptographic key management:

- **Issuance**: Upon successful completion of a behavioral safety assessment by a certified auditor, a Verifiable Credential is issued and bound to the agent's identity.
- **Update**: When the agent's deployment configuration changes (model update, tool access expansion, new deployment context), re-assessment is triggered and a new credential is issued.
- **Revocation**: Upon detection of a behavioral safety incident, the certifying body revokes the credential. Revocation should propagate through standard mechanisms (e.g., credential status lists) so that relying parties—including authorization servers and insurers—are immediately notified.

4. Authorization

How can zero-trust principles be applied to agent authorization?

Zero-trust principles, as articulated in SP 800-207, should extend beyond resource access to behavioral trust. In a zero-trust architecture, no agent should be implicitly trusted to behave safely based solely on its organizational affiliation or model provenance. Every authorization decision should incorporate:

- **Resource access authorization**: Does this agent have permission to access this tool/data/system? (Currently addressed by the concept paper.)
- **Behavioral authorization**: Has this agent been certified to operate safely in the context of this interaction? (Currently unaddressed.)

The urgency of this distinction is underscored by the U.S. Treasury's Financial Services AI Risk Management Framework (February 19, 2026), which establishes 230 control objectives for AI governance in financial services. Analysis of these controls reveals that approximately 97% operate within a "detect-and-respond" paradigm—logging, monitoring, and post-hoc human review. This paradigm fails for autonomous agents that execute multi-step workflows at machine speed. By the time a monitoring system flags anomalous behavior, the agent has already acted. Authorization servers must therefore evaluate behavioral safety credentials *before* issuing access tokens—enforcing pre-execution constraint rather than relying on post-hoc detection.

Can authorization policies be dynamically updated when an agent context changes?

Yes, and behavioral safety credentialing provides a principled mechanism for triggering dynamic authorization updates. When an agent gains access to new tools or resources, its behavioral attack surface expands—a financial analysis agent granted access to customer communications data can now cause harm through inappropriate disclosure or manipulative engagement, not just through unauthorized data access. Authorization policies should incorporate behavioral re-certification triggers: when the sensitivity tier of aggregated data exceeds the scope of the agent's current behavioral safety credential, authorization should be suspended pending re-assessment. This prevents the credential-scope mismatch that arises when tool access expands incrementally but behavioral certification remains static.

How do we establish "least privilege" for an agent, especially when its required actions might not be fully predictable when deployed?

This is the central challenge that behavioral safety credentialing addresses. Traditional least-privilege models define a static permission set. Agents require a dynamic behavioral envelope: a set of behavioral constraints that limit not just *what* the agent can access but *how* it can act within its access scope. Behavioral safety certification verifies that these constraints are implemented and effective.

For example, an agent authorized to access a customer's financial records (resource authorization) should also be constrained from providing investment advice that exceeds its certified competency scope (behavioral authorization). An agent authorized to interact with minors should be constrained by architecturally enforced session limits, content restrictions, and crisis escalation protocols—regardless of what its resource permissions allow.

What are the mechanisms for an agent to prove its authority to perform a specific action?

An agent should prove authority through dual presentation: (1) infrastructure credentials (OAuth access token, SPIFFE SVID) establishing permission to access the requested resource, and (2) a behavioral safety credential establishing certification to operate safely in the context of the requested

action. The relying party—whether a tool server, data source, or downstream agent—should verify both before granting access. This dual-presentation model mirrors how regulated professionals prove authority: a physician presents both a medical license (identity and permission) and malpractice insurance (third-party risk assessment) before being authorized to practice at a hospital.

How might an agent convey the intent of its actions?

Agents should be required to declare behavioral context alongside action intent. When an agent requests authorization to perform an action, the request should include not only what the agent intends to do (the action) and on whose behalf (the delegating user), but under what behavioral constraints it is operating (its certified behavioral safety posture). This allows authorization servers to make risk-informed decisions that account for both access scope and behavioral scope.

How do we handle delegation of authority for "on behalf of" scenarios?

Delegation of authority should encompass behavioral scope, not just resource scope. When a human delegates authority to an agent, the delegation grant should specify: (1) what resources the agent may access (resource scope—currently addressed by OAuth scopes and NGAC policies), and (2) what behavioral parameters the agent must operate within (behavioral scope—currently unaddressed). A healthcare administrator delegating patient record access to an agent should be able to constrain the agent's behavioral parameters as part of the delegation grant—for example, prohibiting the agent from providing medical advice, requiring crisis escalation for patients expressing distress, and mandating human review for any communication sent to patients.

Behavioral safety certification of the agent should be a prerequisite for accepting delegated authority in sensitive contexts. An agent without a valid behavioral safety credential should not be permitted to act on behalf of a human in contexts where behavioral harm is foreseeable—regardless of whether the human has granted resource access permission.

How do we bind agent identity with human identity to support "human-in-the-loop" authorizations?

Human-in-the-loop authorization should be tiered based on the agent's behavioral safety certification level. Agents with higher-tier certifications—verified by independent audit against standardized criteria—should require less frequent human intervention for routine operations, while agents with lower-tier or expired certifications should require human approval for a broader set of actions. This creates a market incentive for developers to pursue and maintain behavioral safety certification: higher certification reduces operational friction.

5. Auditing and Non-Repudiation

How can we ensure that agents log their actions and intent in a tamper-proof and verifiable manner?

The concept paper correctly identifies the need for tamper-proof audit logging. We propose extending this requirement to **behavioral decision logging**: every behavioral safety decision the agent makes—content filtered, crisis escalation triggered, session limit enforced, engagement constraint activated—should be logged with the same tamper-proof guarantees as resource access decisions.

This is essential for three reasons:

1. **Insurance claims adjudication.** When an agent causes harm, insurers need to determine whether the agent was operating within its certified behavioral parameters. Behavioral decision

logs provide the evidentiary basis for this determination.

2. **Regulatory compliance.** California SB 243, Colorado SB 24-205, and the New York RAISE Act all impose specific behavioral safety mandates with incident reporting requirements. Behavioral decision logs provide the audit trail necessary to demonstrate compliance.

3. **Post-incident forensics.** As we argued in our March 6 submission, behavioral safety failures require root-cause analysis of the behavioral decision chain, not just the access event chain. An agent that fails to escalate a user's crisis is not an access control failure—it is a behavioral decision failure that must be traceable through the audit record.

How do we ensure non-repudiation for agent actions and binding back to human authorization?

Non-repudiation should encompass behavioral attestation: the audit record should cryptographically bind the agent's behavioral safety credential to each logged action. This creates an immutable chain: human authorized agent → agent held behavioral safety credential X at time of action → agent took action Y → behavioral decision log shows the agent's behavioral reasoning. If the credential was expired or revoked at the time of the action, the non-repudiation chain reveals that the agent was operating without valid behavioral authorization.

6. Prompt Injection Prevention and Mitigation

What controls help prevent both direct and indirect prompt injections?

Behavioral safety controls serve as a complementary defense layer against prompt injection. Even if an injection succeeds in bypassing input filtering, architecturally enforced behavioral constraints—crisis escalation protocols, content boundaries, session limits—limit the harm the compromised agent can inflict. This is defense-in-depth applied to behavioral safety: the injection may redirect the agent's intent, but behavioral guardrails constrain the agent's action space regardless of intent.

This reinforces the case for behavioral safety as an identity and authorization component: an agent whose behavioral constraints are architecturally enforced and independently verified (via behavioral safety certification) presents a smaller attack surface even in prompt injection scenarios, because the injection cannot override constraints that are enforced at the system level rather than the prompt level.

After prompt injection occurs, what controls and practices can minimize the impact of the injection?

Post-injection impact mitigation is where behavioral safety certification provides its most distinctive value. When an injection compromises an agent's intent layer, three classes of behavioral controls limit blast radius:

1. **Architectural behavioral constraints** function as circuit breakers. An agent that has been injected to disclose sensitive data is still bound by architecturally enforced output filters that strip personally identifiable information. An injected agent instructed to escalate its own privileges is still bound by behavioral constraints that prevent self-modification of authorization scope.

2. **Behavioral anomaly detection** identifies post-injection divergence from the agent's certified behavioral profile. If an agent's outputs suddenly deviate from its established behavioral baseline—for example, generating content categories it has never produced before—behavioral

monitoring should trigger automatic quarantine and human review, independent of whether an injection has been detected at the infrastructure level.

3. **Credential-scoped damage containment** limits post-injection actions to the scope of the agent's behavioral safety credential. An agent certified for customer service interactions but injected to perform financial transactions should be blocked not by resource access controls alone (which the injection may have bypassed) but by behavioral authorization constraints that restrict the agent's certified action categories.

Behavioral safety certification should include attestation of post-injection resilience: evidence that the agent's behavioral constraints remain effective even when its intent layer has been compromised.

III. Proposed Fourth Use Case: Behavioral Safety Credentialing as Authorization Gate

The concept paper identifies three potential use cases for the NCCoE demonstration project. We propose a fourth:

Enterprise AI agents with behavioral safety credentialing for insurance-integrated authorization.

This use case would demonstrate how an enterprise deploys AI agents whose authorization policies incorporate behavioral safety credentials issued by third-party auditors, with those credentials structured for direct integration with insurance underwriting processes.

The demonstration would show:

1. **Agent identity enrichment**: An agent's SCIM-provisioned identity is enriched with a W3C Verifiable Credential attesting to its behavioral safety posture, issued by a certified auditor.

2. **Authorization policy integration**: The organization's NGAC or OAuth-based authorization server incorporates the behavioral safety credential into policy evaluation. Agents with valid, high-tier credentials are authorized for sensitive contexts (consumer interactions, healthcare data, financial decision support). Agents with expired or low-tier credentials are restricted to sandboxed or human-supervised contexts.

3. **Insurance underwriting integration**: The behavioral safety credential is presented to the organization's AI liability insurer as part of the underwriting process. The insurer uses the credential to price coverage, set deductibles, and define coverage scope—creating a direct economic link between behavioral safety posture and commercial viability.

4. **Incident response and credential revocation**: When a behavioral safety incident occurs, the certifying body revokes the credential. The authorization server immediately restricts the agent's permissions. The insurer is notified. The audit trail provides the evidentiary basis for claims adjudication and regulatory reporting.

This use case bridges the concept paper's infrastructure-focused scope with the market reality that the insurance industry is already constructing de facto authorization gates based on behavioral safety posture. NIST standardization of this integration would accelerate adoption by providing a common framework that both developers and insurers can rely on.

IV. Specific Recommendations

Recommendation 1: Include behavioral safety posture in agent identity metadata. The NCCoE demonstration project should define a standard agent identity schema that includes behavioral safety credentials alongside infrastructure identity attributes. The W3C Verifiable Credentials standard provides a ready-made format for this purpose. The schema should be designed for compatibility with SP 800-63-4's identity assurance levels, establishing parallel behavioral safety assurance levels.

Recommendation 2: Incorporate behavioral safety certification into authorization policy evaluation. Authorization frameworks demonstrated in the NCCoE project (OAuth 2.0/2.1, NGAC) should show how behavioral safety credentials can serve as authorization inputs—enabling risk-based, context-sensitive authorization decisions that account for both resource access and behavioral safety. This aligns with SP 800-207's zero-trust principle that access decisions should incorporate multiple contextual signals.

Recommendation 3: Define behavioral safety credential lifecycle management. The NCCoE project should demonstrate issuance, renewal, revocation, and verification of behavioral safety credentials, integrated with existing identity lifecycle management tools (SCIM, certificate management). Token protection mechanisms should follow NISTIR 8587 guidance for preventing forgery, theft, and misuse of behavioral safety credential assertions.

Recommendation 4: Add a fourth use case for behavioral safety credentialing. The demonstration project should include a use case showing behavioral safety credentialing as an authorization gate, with insurance underwriting integration, as described in Section III above.

Recommendation 5: Coordinate with the CAISI RFI work stream. As we recommended in our March 6 submission (Tracking: mmf-286f-aneg), the identity and authorization work stream and the security work stream should be explicitly coordinated. Behavioral safety is both a security domain and an identity attribute. The two concept papers should produce complementary, non-duplicative guidance.

Recommendation 6: Expand scope to include consumer-facing agents in future iterations. While we understand the initial enterprise focus, consumer-facing agents present the most acute identity and authorization challenges. We urge NIST to prioritize consumer-facing agent identity and authorization in the next iteration of this project.

V. About the Commenting Organization

Empty Set LLC is a Nevada Series LLC developing intellectual property, standards, and infrastructure for AI agent behavioral safety credentialing, insurance integration, and autonomous agent governance. Empty Set holds a portfolio of 43 patents and provisional patent applications spanning AI agent credentialing, behavioral safety assessment, digital identity for autonomous systems, and autonomous corporate governance. Empty Set is establishing **The Box Commons**, a 501(c)(6) trade association, to house its credentialing initiative as an independent, industry-governed standards body.

We submitted a comprehensive public comment on the CAISI RFI (Docket NIST-2025-0035, Tracking: mmf-286f-aneg, March 6, 2026) establishing the evidence base for behavioral safety as a security domain. This comment extends that work into the identity and authorization domain.

We have registered for the NIST CAISI listening sessions on barriers to AI adoption in finance (April 2026) and welcome the opportunity to participate in any working groups or collaborator calls that result from this concept paper.

Respectfully submitted,

Empty Set LLC

March 14, 2026