Public Comment to the NAIC Big Data and Artificial Intelligence (H) Working Group

Spring 2026 National Meeting

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

Date: March 24, 2026 **Re:** Implementation of the NAIC Model Bulletin on the Use of Artificial Intelligence Systems by Insurers — The Need for Standardized Behavioral Safety Credentialing

Submitted via: Email to BDAIWG support staff (Scott Sobel, ssobel@naic.org)

I. Executive Summary

Empty Set LLC respectfully submits this comment to the Big Data and Artificial Intelligence (H) Working Group regarding the implementation and operationalization of the NAIC Model Bulletin on the Use of Artificial Intelligence Systems by Insurers (adopted December 2023).

We write as practitioners building credentialing infrastructure at the intersection of AI agent behavioral safety and insurance underwriting. Our central observation is this: **the Model Bulletin correctly identifies the need for governance frameworks and third-party standards, but the insurance market currently lacks the standardized behavioral safety metrics necessary to operationalize that vision at scale.**

Three market developments since the Model Bulletin's adoption make this gap urgent:

1. **The Verisk generative AI exclusion wave** (effective January 1, 2026) has systematically stripped AI liability coverage from standard CGL policies via ISO Forms CG 40 47, CG 40 48, and CG 35 08. An estimated 95% of carriers are adopting these exclusions, creating a protection gap that affects every insured deploying AI agents.

2. **Federal courts and Congress are establishing strict products liability for AI agents.** In *Garcia v. Character Technologies, Inc.* (M.D. Fla., May 2025), a federal court ruled that an AI companion chatbot constitutes a "product" whose behavioral design defects create strict liability. *Gavalas v. Google LLC* (N.D. Cal., filed March 4, 2026) escalates this precedent to a frontier AI model alleged to have contributed to a user's death through behavioral failures — not technical compromise. The bipartisan *AI LEAD Act* (S.2937, Sens. Durbin & Hawley) would codify this shift federally, explicitly classifying AI systems as "products" subject to strict liability regardless of developer care or EULA indemnifications. In *Mobley v. Workday, Inc.*, a federal court held that an AI vendor can be directly liable as an "employment agency" under Title VII for discriminatory algorithmic screening — establishing joint and several liability between vendors and deployers.

3. **State legislatures are outpacing insurance market infrastructure.** California SB 243 mandates crisis detection protocols for companion chatbots. Colorado SB 24-205 requires algorithmic discrimination impact assessments with an affirmative defense for compliance with recognized risk management frameworks. Connecticut Bulletin MC-25 imposes annual certification requirements. The insurance market needs standardized, machine-readable behavioral safety data to underwrite these emerging risks — and to serve as the recognized third-party framework that Colorado's affirmative defense provision contemplates.

Our recommendation: The Working Group should encourage state regulators to recognize independently developed behavioral safety credentialing standards — specifically those that translate AI governance frameworks into actuarial data points suitable for underwriting. The NAIC Model Bulletin's existing language (encouraging insurers to "adopt, incorporate, or rely upon... framework or standards developed by official third-party standard organization") provides the authority. What is needed now is implementation guidance that connects this language to emerging credentialing programs.

II. The Protection Gap: Why Insurers Need Behavioral Safety Standards

A. The Verisk Exclusion Creates a Market Failure

The Verisk ISO exclusionary endorsements represent the insurance market's admission that it cannot currently price generative AI risk. This is not because AI risk is inherently uninsurable — it is because the market lacks standardized data on AI behavioral safety that would allow actuarial modeling.

Consider the analogy to fire safety. Before Underwriters Laboratories (UL) developed standardized testing protocols in the 1890s, fire insurers could not distinguish between safe and unsafe electrical products. The result was either blanket exclusions or catastrophic losses. UL's contribution was not to eliminate fire risk — it was to create a data layer that allowed insurers to differentiate and price risk accurately.

The AI insurance market is at this same inflection point. Specialty insurers (Armilla AI, Relm Insurance, Munich Re's aiSure program) are entering the market but cannot underwrite at scale without the equivalent of UL's standardized testing — behavioral safety metrics that are: - **Empirically tested** (not self-reported governance checklists) - **Machine-readable** (suitable for programmatic underwriting and ACORD data standards) - **Independently verified** (assessed by third parties, not AI developers themselves) - **Actuarially relevant** (translatable to loss frequency and severity estimates)

B. What Existing Frameworks Miss

Current AI governance frameworks are necessary but insufficient for insurance purposes:

Framework	What It Provides	What Insurers Still Need
NIST AI RMF	Process-based risk management guidance	Outcome-based failure probability data
ISO/IEC 42001	AI management system certification	Agent-level behavioral safety metrics
EU AI Act	Regulatory compliance baseline	Actuarial telemetry for pricing
SOC 2	Security controls verification	Behavioral safety (orthogonal to security)
Treasury FS AI RMF (Feb 2026)	230 control objectives for financial AI governance	Pre-execution behavioral constraint (see below)

An organization can hold ISO 42001 certification, maintain full NIST AI RMF compliance, and still

deploy an AI agent that causes catastrophic behavioral harm — as the *Garcia* and *Gavalas* cases demonstrate. Infrastructure security and behavioral safety are orthogonal risk domains.

The U.S. Treasury's Financial Services AI Risk Management Framework (published February 19, 2026) illustrates this gap at scale. Its 230 control objectives represent the most detailed federal AI governance guidance to date — yet approximately 97% of those controls operate within a "detect-and-respond" paradigm: logging, monitoring, and human-in-the-loop review. This paradigm was designed for static generative AI where a human initiates each interaction. Autonomous AI agents execute multi-step workflows — API calls, financial transactions, network traversals — at machine speed. By the time a monitoring tool detects anomalous behavior, the agent has already acted. Post-hoc detection cannot protect against pre-execution harm. What insurers need is proof of *pre-execution structural constraint*: a standardized credential that mathematically verifies an agent cannot take unauthorized actions in the first place. This is the actuarial prerequisite that no existing framework provides.

C. The Five Data Points Insurers Need

Based on our engagement with the MGA and reinsurance markets, we have identified five core behavioral safety data points that bridge the gap between AI governance and insurance underwriting:

1. **Model Lineage & Systemic Dependency Mapping** — Which foundation model, which version, what fine-tuning history. This addresses the reinsurance industry's systemic accumulation concern: if 200 AI agents all depend on the same foundation model, a single model failure creates correlated losses across the portfolio.
2. **Hallucination/Error Frequency** — Empirically tested failure rates per 1,000 outputs under standardized test conditions. This is the loss frequency data that Tech E&O actuaries need but currently lack.
3. **Guardrail Efficacy & Containment** — The ratio of autonomous to human-supervised operations, measured under stress conditions. Higher autonomy correlates with higher potential severity, providing a severity proxy for underwriting models.
4. **Non-Determinism Variance** — Output variance given identical inputs across multiple runs. This addresses the fundamental actuarial challenge of AI: non-deterministic systems resist traditional frequency-severity modeling. Quantifying variance allows actuaries to build confidence intervals.
5. **Regulatory Compliance Indicators** — Boolean flags for EU AI Act requirements, Verisk CGL exclusion definitions, and state-specific mandates. These enable programmatic underwriting — an MGA's policy administration system can automatically check eligibility without manual review.

III. How Behavioral Safety Credentialing Operationalizes the Model Bulletin

A. The Model Bulletin's Existing Authority

The NAIC Model Bulletin on the Use of AI Systems by Insurers (December 2023) already contemplates third-party standards:

> "An insurer's AIS Program may adopt, incorporate, or rely upon... framework or standards developed by official third-party standard organization."

This language is the foundation. What is needed is implementation guidance that: 1. Defines what constitutes an "official third-party standard organization" for AI behavioral safety 2. Establishes minimum criteria for behavioral safety credentialing programs that insurers may rely upon 3. Encourages state regulators to recognize credentialing programs that meet these criteria

B. The Credentialing Model

Empty Set LLC is developing a tiered Behavioral Safety Credential (BSC) designed specifically for insurance integration:

- **BSC-1 (Basic):** Enterprise-internal, human-supervised AI agents. 12-month validity. Suitable for internal automation that does not interact with policyholders or claimants.
- **BSC-2 (Standard):** Enterprise-facing AI agents with limited autonomy. 12-month validity. Suitable for agent-assisted underwriting, claims triage, and policy servicing.
- **BSC-3 (Advanced):** Consumer-facing AI agents with high autonomy, including those interacting with vulnerable populations. 6-month validity due to behavioral drift risk. Suitable for chatbots, robo-advisors, and autonomous claims handlers.

Each credential is issued as a W3C Verifiable Credential — machine-readable, cryptographically verifiable, and time-limited. This format enables seamless integration with insurer policy administration systems and ACORD data standards.

C. Alignment with the AI Systems Evaluation Tool Pilot

The Working Group's launch of the multi-state AI Systems Evaluation Tool pilot program (March–September 2026, 12 participating states including California) represents a significant step toward standardized regulatory examination of insurer AI practices. We note that behavioral safety credentialing directly supports the pilot's objectives:

- **Exhibit B (Governance Risk Assessment):** A current BSC credential provides examiners with a machine-readable, independently verified artifact confirming that the insurer's AI governance framework has been assessed against standardized behavioral safety criteria — streamlining the examination process.
- **Exhibit C (High-Risk AI Systems Details):** The five behavioral safety data points underlying each BSC credential (model lineage, hallucination frequency, guardrail efficacy, non-determinism variance, and compliance indicators) map directly to the granular details Exhibit C requires for high-risk applications.

Rather than requiring each of the 12 pilot states to independently develop technical expertise to evaluate AI behavioral safety during market conduct exams, a recognized credentialing standard provides a common, verifiable baseline. The credential does not replace the Evaluation Tool — it gives insurers a standardized way to satisfy it.

D. The Assessment Model

Following the HITRUST precedent (framework owner separate from assessors): - The standards commission develops the standard and accredits assessors - Certified Behavioral Safety Assessors (CBSAs) — independent, human professionals — perform the assessments - The commission reviews

assessment reports and issues credentials - AI agents assist CBSAs with diagnostic analysis at machine speed, but CBSAs take ultimate legal liability

This separation preserves independence: the standards commission never assesses its own members' AI agents. The standard is open; the assessment is independent; the credential is verifiable.

IV. Alignment with State Regulatory Initiatives

The Model Bulletin's adoption by over half of U.S. jurisdictions — including Massachusetts, North Carolina, Delaware, New Jersey, and Wisconsin in the past 15 months — demonstrates sustained regulatory momentum. Yet NAIC surveys reveal a persistent compliance gap: while AI adoption among insurers is near-universal (92% health, 88% auto, 70% homeowners), nearly one-third of health insurers still do not regularly test their deployed models for bias or discrimination. This gap between adoption rates and governance maturity is precisely what behavioral safety credentialing addresses.

Adjacent financial regulators are already moving faster. Freddie Mac's updated Seller/Servicer Guide Section 1302.8 (effective March 3, 2026) prohibits opaque AI models in the mortgage lifecycle, mandates adversarial threat assessment, and shifts full financial indemnity to the deploying organization regardless of whether the AI was built in-house or procured from a vendor. As mortgage originators seek E&O and D&O coverage against these new liabilities, insurers will need standardized behavioral safety data to underwrite the risk.

Several states have also enacted or proposed AI governance requirements that behavioral safety credentialing directly supports:

State	Requirement	How BSC Serves It
Colorado (SB 24-205)	Algorithmic discrimination impact assessments; affirmative defense for compliance with recognized risk management frameworks	BSC provides the "recognized framework" that triggers the affirmative defense
New York (DFS Circular 7, Jul 2024)	Board oversight, internal audits, and third-party vendor risk management for AI in insurance	BSC serves as the third-party audit artifact for NYDFS examiners
Connecticut (Bulletin MC-25, Feb 2024)	Annual certification of AI governance compliance	BSC automates the annual certification with machine-readable credentials
California (SB 243, eff. Jan 1, 2026)	Crisis detection protocols and suicidal ideation screening for companion chatbots	BSC-3 criteria include behavioral safety for vulnerable populations

A credentialing standard that satisfies multiple state requirements simultaneously reduces compliance costs for multi-state insurers and MGAs. The Working Group is well-positioned to encourage harmonized recognition of behavioral safety credentials across adopting states.

V. Recommendations to the Working Group

1. **Issue implementation guidance** clarifying that behavioral safety credentialing programs meeting defined criteria qualify as "framework or standards developed by official third-party standard organization" under the Model Bulletin.
2. **Define minimum criteria** for recognized behavioral safety credentialing programs, including: independent assessment (not self-certification), empirically tested behavioral metrics (not process checklists alone), machine-readable credential format, and periodic re-assessment requirements.
3. **Encourage state regulators** to recognize behavioral safety credentials as compliance artifacts under state-specific AI governance requirements, promoting harmonization across jurisdictions.
4. **Convene a working session** (virtual or at the NAIC Summer Meeting) bringing together AI developers, insurers, credentialing bodies, and state regulators to discuss operationalization of behavioral safety standards in insurance.
5. **Coordinate with NIST** on alignment between insurance-specific behavioral safety standards and the NIST AI Risk Management Framework, ensuring that insurance credentialing efforts complement rather than duplicate federal guidance.

VI. About Empty Set LLC

Empty Set LLC is a Nevada Series LLC developing intellectual property, standards, and infrastructure for AI agent behavioral safety credentialing, with a specific focus on enabling insurance underwriting of AI risk. Empty Set holds a portfolio of 43 patents on AI behavioral safety assessment methodology and is establishing a dedicated 501(c)(6) trade association (name forthcoming, formation April 2026) to house its credentialing standards initiative as an independent industry body.

Empty Set submitted public comments to NIST's Center for AI Standards and Innovation (CAISI) and the National Cybersecurity Center of Excellence (NCCoE) on March 6, 2026, recommending the integration of behavioral safety credentialing into NIST's AI agent security and identity frameworks.

We stand ready to collaborate with the Working Group, state regulators, and market participants to advance behavioral safety credentialing as insurance infrastructure.

Respectfully submitted,

Brice Love Co-Founder, Empty Set LLC brice@emptysetllc.com
