

Addressed to: Sen. Mark R. Warner (D-VA), Co-Chair, Senate Cybersecurity Caucus

The AI Safety Investigation Act: A Federal AI Incident Investigation and Reporting Authority

Logan Bradley and Ishan Ajwani

Executive Summary

This brief proposes the AI Safety Investigation Authority (AISIA), an independent federal agency modeled on the National Transportation Safety Board (NTSB), mandated to investigate significant AI incidents, maintain a centralized incident database, and issue non-binding safety recommendations. Unlike prescriptive regulatory approaches that risk rapid obsolescence, an investigative body generates the empirical foundation for evidence-based AI governance while resolving growing fragmentation across state-level AI safety regimes.

Problem Statement

The U.S. faces three converging AI governance failures. First, AI incidents are proliferating without systematic documentation. The OECD's 2026 analysis found that media-reported AI incidents rose from 92 per month in 2022 to 324 per month in 2025 [1]. Yet no federal mechanism exists to collect or learn from these failures, as existing repositories like the Partnership on AI's Incident Database rely on crowdsourced media reports and lack investigative authority [2].

Second, the regulatory landscape is fragmenting. California's SB 53 requires frontier developers to report incidents within 15 days [3]. New York's RAISE Act requires 72-hour reporting to the Department of Financial Services, with penalties up to \$3 million [4]. Colorado and Texas have enacted further frameworks [5]. These laws define key terms differently, creating compliance burdens that major developers have called unsustainable [6].

Third, the federal response has been inadequate. The Administration's March 2026 National AI Legislative Framework proposes preempting state laws in favor of a "minimally burdensome national standard," but offers no incident investigation mechanism [7]. The bipartisan Secure AI Act (Warner-Tillis, 2024) proposed voluntary databases but was never enacted [8]. No federal entity has authority to investigate an AI incident after it occurs.

Policy Recommendation: The AI Safety Investigation Act

Congress should establish AISIA as an independent agency with three core functions:

1. Mandatory incident reporting and investigation. Developers and deployers must report safety incidents—where an AI system contributes to death, serious injury, critical infrastructure disruption, or damages exceeding \$1 million—within 72 hours. Reporting is triggered by outcomes, not model size, ensuring coverage adapts as capabilities evolve. For mandatory-threshold investigations, AISIA holds subpoena authority over model logs, training records, and decision traces, ensuring investigations are not limited to what developers voluntarily disclose [9]. AISIA also retains discretionary authority to investigate below-threshold incidents revealing systemic risk patterns.

2. Centralized incident database. AISIA maintains a public, standardized repository with taxonomies aligned to the OECD's international framework [10] and Georgetown CSET's reporting schema [11], subsuming fragmented state channels into a single federal system and providing the evidence base both the Administration and state legislatures have

identified as essential for future rulemaking.

3. Non-binding safety recommendations. Like the NTSB, AISIA issues recommendations without direct enforcement power. This preserves industry flexibility while creating public accountability: the NTSB has issued over 15,000 recommendations since 1967, roughly 82% of which have been closed in “acceptable” status, driven by reputational incentives and independent investigative weight [12].

Implementation

AISIA would be governed by a five-member bipartisan board confirmed by the Senate for staggered five-year terms, independent of the Commerce Department and OSTP. Initial authorization of \$75 million annually supports approximately 150 staff in a phased three-year ramp-up, scaling toward the NTSB’s FY2025 level of \$145–150 million [13]. AISIA’s preemption is narrowly scoped: the Act satisfies state incident *reporting* obligations (under SB 53, the RAISE Act, and similar statutes) through federal compliance, but explicitly preserves all state enforcement authority, substantive safety standards, whistleblower protections, and consumer protection powers [14].

Impacts and Trade-offs

The primary benefit is information infrastructure. Aviation safety transformed after the NTSB’s creation in 1967; U.S. commercial aviation fatality rates have since declined by over 90% [15]. An equivalent evidence base for AI would enable calibrated rather than reactive regulation. A key challenge is that AI incidents differ from physical accidents: failures are probabilistic, opaque, and contested, and investigations require access to proprietary model internals in ways that physical wreckage analysis does not. AISIA addresses this through a tiered disclosure model: full technical findings—including model architecture, training data, and decision traces—are preserved in restricted annexes accessible only to AISIA and the investigated entity, while public reports focus on systemic factors and recommendations [16]. An evidentiary privilege (paralleling 49 U.S.C. § 1154(b)) prevents AISIA findings from use in civil litigation [17].

To ensure report integrity without chilling disclosure, the Act employs a dual-channel approach. For mandatory reports, developers who knowingly submit false or materially misleading statements face independent civil penalties—modeled on § 1421(5) of New York’s RAISE Act [4]—enforceable by the Attorney General. For voluntary and below-threshold reports, full immunity is preserved to encourage near-miss reporting, mirroring the FAA’s Aviation Safety Reporting System [18]. This incentivizes honest mandatory disclosure while maintaining the high-trust voluntary channel that makes incident databases comprehensive.

Resilience to Technological Change

This proposal is mechanism-agnostic. Because reporting is triggered by outcomes rather than model architecture or compute thresholds, the framework adapts automatically. The NTSB has maintained relevance across seven decades—from propeller aircraft to autonomous vehicles—because investigation follows failure rather than attempting to anticipate it [9]. An AI investigative body exhibits the same adaptive resilience, providing durable governance regardless of whether frontier risk emerges from large language models, autonomous agents, biological design tools, or capabilities not yet imagined.

References

- [1] OECD, "Trends in AI Incidents and Hazards Reported by the Media," OECD Artificial Intelligence Papers (Feb. 2026). <https://doi.org/10.1787/4f5ff43c-en>
- [2] Partnership on AI, "AI Incident Database" (2024). <https://incidentdatabase.ai>
- [3] Cal. SB 53, Transparency in Frontier Artificial Intelligence Act, signed Sept. 29, 2025. See Brookings Institution, "What Is California's AI Safety Law?" (Dec. 23, 2025). <https://www.brookings.edu/articles/what-is-californias-ai-safety-law/>
- [4] N.Y. RAISE Act (S6953B/A6453B), signed Dec. 19, 2025. See Jones Walker LLP, "What Frontier Model Developers Need to Know" (Jan. 2, 2026). <https://www.joneswalker.com/en/insights/blogs/ai-law-blog/new-yorks-raise-act-what-frontier-model-developers-need-to-know.html>
- [5] Colo. Rev. Stat. §§ 6-1-1701–1707 (Colorado AI Act); Tex. Responsible AI Governance Act (June 2025).
- [6] C. Lehane, OpenAI Chief Global Affairs Officer, quoted in IAPP, "Hochul Enacts New York's AI Safety and Transparency Bill" (Dec. 2025). <https://iapp.org/news/a/hochul-enacts-new-yorks-ai-safety-and-transparency-bill>
- [7] White House, "President Donald J. Trump Unveils National AI Legislative Framework" (March 20, 2026). <https://www.whitehouse.gov/releases/2026/03/president-donald-j-trump-unveils-national-ai-legislative-framework/>
- [8] S.4230, Secure A.I. Act of 2024, 118th Congress, introduced by Sens. Warner (D-VA) and Tillis (R-NC), May 1, 2024. <https://www.congress.gov/bill/118th-congress/senate-bill/4230/text>
- [9] 49 U.S.C. Ch. 11; Independent Safety Board Act of 1974, Pub. L. 93-633.
- [10] OECD, "Defining AI Incidents and Related Terms," OECD Artificial Intelligence Papers No. 16 (May 2024). https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/defining-ai-incidents-and-related-terms_88d089ec/d1a8d965-en.pdf
- [11] R.B.L. Dixon and H. Frase, "AI Incidents: Key Components for a Mandatory Reporting Regime," Center for Security and Emerging Technology, Georgetown Univ. (Jan. 2025). <https://cset.georgetown.edu/publication/ai-incidents-key-components-for-a-mandatory-reporting-regime/>
- [12] NTSB, FY2024 Performance and Accountability Report. <https://www.nts.gov/about/reports/Documents/FY2024%20Performance%20and%20Accountability%20Report.pdf>
- [13] NTSB, Fiscal Year 2025 Budget Request. <https://www.nts.gov/about/reports/Documents/Fiscal%20Year%202025%20Budget%20Request.pdf>
- [14] Exec. Order, "Ensuring a National Policy Framework for Artificial Intelligence" (Dec. 11, 2025). <https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/>
- [15] Bureau of Transportation Statistics, "U.S. General Aviation Safety Data." <https://www.bts.gov/content/us-general-aviation-safety-data>
- [16] A. Ortega, "AI Threats to National Security Can Be Countered Through an Incident Regime," arXiv:2503.19887 (Apr. 2025). <https://arxiv.org/html/2503.19887>
- [17] 49 U.S.C. § 1154(b).
- [18] FAA Aviation Safety Reporting System (ASRS), administered by NASA. <https://asrs.arc.nasa.gov>