

Process vs. Behavioral Attestation

Why Development Process Certification Is Insufficient for Autonomous Systems

February 2026 — Public Document

Every major AI governance framework in use today—ISO/IEC 42001, NIST AI RMF, the EU AI Act, SOC 2—verifies that an organization has the right processes in place. None of them verify that the AI system actually behaves as those processes require. ODDC closes this gap by certifying behavioral conformance: not what you say your system does, but what your system demonstrably does under sustained autonomous operation.

Field	Details
Document	Process vs. Behavioral Attestation
Classification	Public Document — Informative
Effective Date	February 2026
Category	White Paper
Owner	Sentinel Authority
Contact	info@sentinelauthority.org

This document references publicly available government reports, court filings, regulatory publications, and peer-reviewed research. This document does not constitute legal advice.

This white paper provides a detailed comparative analysis of the world's leading AI governance and safety frameworks, examining each one's specific provisions for verifying autonomous system behavior at runtime. It demonstrates, framework by framework, that the current standards ecosystem addresses organizational governance comprehensively while leaving a critical gap in behavioral verification—the gap between documented intent and operational reality.

The consequences of this gap are not theoretical. ECRI, the independent healthcare research organization, ranked AI systems deployed without proper oversight as the number one health technology hazard for 2025. The Department of Justice has subpoenaed pharma and digital health companies over AI deployed in electronic medical record systems. A 2024 academic analysis identified 51 court cases involving software-related patient injuries from clinical decision support, drug management, and surgical robotics. In autonomous transportation, repeated software recalls affecting thousands of vehicles demonstrate that governance processes did not prevent systematic behavioral failures.

This paper argues that the current standards are not wrong—they are incomplete. ODDC is not a replacement for ISO 42001 or the NIST AI RMF. It is the missing behavioral layer that makes these governance frameworks enforceable at the system level.

1. The Governance Achievement

Before examining what is missing, it is important to acknowledge what has been accomplished. The AI governance ecosystem that has emerged since 2023 represents a significant advance in how organizations manage AI risk. NIST's AI Risk Management Framework provides a structured approach to identifying, assessing, and mitigating AI risks. ISO/IEC 42001 provides the world's first certifiable AI management system standard. The EU AI Act establishes the world's first comprehensive AI regulation with binding obligations. These frameworks, individually and collectively, have established essential organizational infrastructure for AI governance.

But there is a consistent pattern across all of them: they verify what an organization does to manage AI systems. They do not verify what the AI systems do. This is not an oversight—it reflects a deliberate design choice. Management system standards have always focused on organizational processes because these are what auditors can evaluate using established conformity assessment methodologies. The problem is that for autonomous systems operating in safety-critical domains, organizational processes are necessary but not sufficient. A hospital can have world-class AI governance policies and still deploy a clinical decision support system that makes dangerous recommendations. A trucking company can maintain comprehensive safety management documentation and still operate autonomous vehicles with software defects that affect the entire fleet.

2. Framework-by-Framework Analysis

The following analysis examines each major governance framework's specific provisions for verifying autonomous system behavior at runtime. For each framework, we identify what it verifies, what it does not verify, and where the behavioral gap exists.

2.1 ISO/IEC 42001: AI Management Systems

ISO/IEC 42001, published in December 2023, is the world's first international standard specifically designed for AI management systems. It follows the Plan-Do-Check-Act methodology familiar from ISO 9001 and ISO 27001, providing a structured framework for establishing policies, assessing risks, implementing controls, and conducting management reviews. As of 2025, a CSA benchmark report indicates that 76% of organizations plan to pursue frameworks like ISO 42001.

What ISO 42001 verifies: organizational AI management system with defined policies, roles, and responsibilities; risk assessments covering the AI system lifecycle; documentation for system design, development, and deployment; management reviews on a regular cadence; and processes for monitoring, measurement, and continual improvement.

What ISO 42001 does not verify: that the AI system actually operates within the boundaries defined by those policies; that the risk controls documented in the management system are enforced at runtime; or that the system's behavior matches its documented specifications under sustained autonomous operation. A UNIDO analysis noted that there is currently no ecosystem of conformity assessment for digital services equivalent to that of tangible or manufactured products. ISO/IEC 42001 certifies that your organization governs AI responsibly. It does not certify that your AI system behaves responsibly.

2.2 NIST AI Risk Management Framework (AI RMF 1.0)

The NIST AI RMF, released in January 2023 with expanded companion playbooks through 2024–2025, provides a voluntary, technology-agnostic framework organized around four core functions: Govern, Map, Measure, and Manage. It verifies that governance structures, risk identification, measurement approaches, and risk management processes exist. It does not verify that measurement approaches are applied to runtime behavior, that governance processes translate into enforceable operational constraints, or that an autonomous system's behavior remains within the risk tolerances defined during the Map and Measure phases. NIST's own standardization plan, NIST AI 100-5e2025, explicitly acknowledges that conformity assessment with other standards is a Tier 2 priority item requiring more scientific work before standardization.

2.3 EU AI Act (Regulation 2024/1689)

The EU AI Act entered into force on August 1, 2024, establishing the world's first comprehensive, binding AI regulation. High-risk system obligations become applicable on August 2, 2026. The Act requires conformity assessments including quality management systems, technical documentation, and accuracy, robustness, and cybersecurity requirements. However, the standards needed to perform these assessments are not ready. CEN and CENELEC were unable to meet the August 2025 deadline. The first

harmonized AI standard—prEN 18286 for AI quality management systems—only entered public enquiry on October 30, 2025. The EU AI Act creates a regulatory mandate for conformity evidence that does not yet exist.

2.4 SOC 2 and Related Assurance Frameworks

SOC 2, developed by the AICPA, provides assurance over five trust service criteria: security, availability, processing integrity, confidentiality, and privacy. It verifies that security controls are in place and operating effectively. It does not verify that an AI system's outputs are correct or safe, that the system operates within its designed operational domain, or that autonomous decision-making adheres to specified constraints. SOC 2 was designed for service organizations processing data, not for autonomous systems making consequential decisions in physical environments.

2.5 Sector-Specific Frameworks

Several sector-specific frameworks have attempted to address AI safety within their domains, each with the same structural limitation. The FDA's SaMD framework addresses pre-market review but post-market surveillance depends on manufacturer-reported adverse events. NHTSA's framework relies on voluntary safety self-assessments and post-hoc recall authority. ISO 26262 and ISO/PAS 21448 (SOTIF) are applied during development to identify and mitigate hazards but provide no mechanism for continuous verification that the deployed system operates within the boundaries established during development. The gap between development-phase safety analysis and deployment-phase behavioral conformance is precisely where autonomous system failures occur.

3. The Comprehensive Comparison

Verification	ISO 42001	NIST AI RMF	EU AI Act	SOC 2	ODDC
Organizational governance	✓ Full	✓ Full	✓ Full	✓ Full	— Not in scope
Risk assessment process	✓ Full	✓ Full	✓ Required	✓ Partial	— Not in scope
Technical documentation	✓ Required	✓ Guidance	✓ Required	✓ Required	✓ ODD specification
Operational boundary specification	— Not required	— Guidance only	— Implicit	— Not addressed	✓ Machine-readable ODD
Runtime boundary enforcement	— Not addressed	— Not addressed	— Not specified	— Not addressed	✓ ENVELO Interlock
Continuous behavioral verification	— Not addressed	— Not addressed	— Intent without mechanism	— Not addressed	✓ CAT-72 testing
Tamper-evident conformance records	— Not addressed	— Not addressed	— Not specified	— Audit logs only	✓ Cryptographic attestation
Independent third-party verification	✓ Accredited CABs	— Voluntary	✓ Required for high-risk	✓ CPA firms	✓ Sentinel Authority

The table reveals a clear pattern: existing frameworks thoroughly address organizational governance while leaving behavioral verification unaddressed. This is not a criticism—these frameworks were designed for organizational assurance. The gap exists because no framework was designed for autonomous system behavioral assurance. ODDC fills exactly this gap.

4. Real-World Consequences of the Gap

4.1 Healthcare: Governance Without Behavioral Guardrails

ECRI ranked AI systems deployed without proper oversight as the number one health technology hazard for 2025, and insufficient AI governance as the number two patient safety threat. The Department of Justice's subpoenas of pharmaceutical and digital health companies targeted situations where governance documentation existed but the systems themselves produced harmful outputs. The 2024 analysis of 51 court cases involving software-related patient injuries spans drug management, clinical decision support, and surgical robotics. In each category, the deploying organization could point to governance policies, risk assessments, and vendor documentation. None of these process artifacts prevented the behavioral failure.

4.2 Autonomous Vehicles: Testing That Missed Deployment Behavior

Waymo's autonomous vehicles are among the most extensively tested in the industry, with over 100 million miles and a published safety framework. The Swiss Re study documented an 88% reduction in property damage claims. By any process-attestation standard, Waymo's governance would receive high marks. Yet in deployment, Waymo's vehicles systematically violated school bus traffic safety laws across multiple cities. The Austin Independent School District documented 20 separate incidents during the 2025–2026 school year. NHTSA's recall ultimately affected 3,067 vehicles. A system verified to operate within its declared ODD through continuous conformance testing would have flagged the school bus recognition failure before it produced 20 incidents involving children.

4.3 Industrial Robotics: Documentation That Didn't Prevent Injury

In September 2025, a \$51 million lawsuit was filed against Tesla and Fanuc after a robotic arm struck a worker with approximately 8,000 pounds of force at Tesla's Giga Texas facility. OSHA data documents 77 robot-related workplace accidents (2015–2022) resulting in 93 injuries. NIOSH recorded 61 robot-related fatalities (1992–2015). Industrial robotics installations operate under extensive process documentation: ISO 12100, ISO 13849, ANSI/RIA R15.06. Complete process documentation does not prevent a robot from operating outside its programmed parameters. Runtime behavioral enforcement is the missing layer.

5. How ODDC Completes the Stack

5.1 The Complementary Architecture

ISO 42001 + ODDC: ISO 42001 certifies organizational governance. ODDC certifies that the deployed system’s behavior conforms to the specifications documented in that management system. Together, they provide end-to-end assurance from organizational governance to runtime behavior.

NIST AI RMF + ODDC: The AI RMF’s Govern, Map, Measure, and Manage functions establish policy and risk infrastructure. ODDC provides the verification layer confirming risk responses are actually effective at constraining system behavior. This directly addresses NIST AI 100-5e2025’s identified need for verification and validation tools.

EU AI Act + ODDC: The AI Act requires providers to demonstrate that high-risk systems meet accuracy, robustness, and cybersecurity requirements. ODDC’s formal ODD specification, ENVELO Interlock enforcement, and CAT-72 testing provide a concrete mechanism for generating the conformity evidence the Act demands.

SOC 2 + ODDC: SOC 2 verifies infrastructure controls around the AI system. ODDC verifies behavioral constraints within the AI system. The combination provides assurance that both the environment and the system itself are operating as intended.

5.2 The Three Verification Layers

Layer	Component	Function
Layer 1	Formal ODD Specification	Machine-readable operational boundaries defining exactly where, when, and how the system is designed to operate. Not marketing language—a formal contract between the system developer and the certification authority.
Layer 2	ENVELO Interlock	Runtime enforcement mechanism that prevents the autonomous system from operating outside its declared ODD. Cannot be disabled or overridden by the system it governs. Non-bypassable enforcement, not policy.
Layer 3	CAT-72 Conformance Assurance Test	72 cumulative hours of system operation under varied conditions verifying sustained behavioral conformance. Results are cryptographically signed, independently stored, and tamper-evident.

6. The Standards Development Opportunity

NIST AI 100-5e2025 categorizes conformity assessment with AI standards as a Tier 2 priority—needed but requiring more scientific work before standardization. The EU AI Act’s harmonized standards gap

creates an immediate market opportunity: organizations need conformity evidence before the standards defining how to produce it are finalized. CEN/CENELEC's missed deadline demonstrates that institutional standards development is proceeding more slowly than the regulatory calendar demands.

ODDC represents a pragmatic bridge: a deployable conformance verification methodology that can be applied now, refined as standards mature, and ultimately incorporated into harmonized standards frameworks. For organizations facing imminent EU AI Act compliance deadlines, ODDC provides auditable behavioral evidence while the formal standards ecosystem catches up.

7. Conclusion

NIST AI RMF tells you how to govern. ISO 42001 certifies that governance is in place. The EU AI Act requires conformity evidence. ODDC certifies that the system actually behaves as governance requires. This is not a competing standard—it is the missing layer that makes every other framework enforceable at the system level.

The gap between process attestation and behavioral attestation is the defining challenge of autonomous system safety. It is the gap that ECRI identifies when it warns about AI without proper oversight. It is the gap that NIST acknowledges when it identifies conformity assessment as a Tier 2 priority. It is the gap that the EU AI Act creates when it mandates conformity evidence that harmonized standards have not yet defined how to produce.

References

- [1] ISO/IEC 42001:2023. Artificial Intelligence — Management System. December 2023.
- [2] NIST. AI Risk Management Framework (AI RMF 1.0). NIST AI 100-1. January 2023.
- [3] NIST. A Plan for Global Engagement on AI Standards. NIST AI 100-5e2025. April 2025.
- [4] European Union. Regulation (EU) 2024/1689 (Artificial Intelligence Act). August 1, 2024.
- [5] European Commission. Digital Omnibus Proposal, November 19, 2025.
- [6] CEN/CENELEC. prEN 18286: AI Quality Management System. October 30, 2025.
- [7] European Commission. General-Purpose AI Code of Practice. July 10, 2025.
- [8] CSA. 2025 Compliance Benchmark Report: 76% plan to pursue ISO 42001.
- [9] UNIDO. Overview of ISO/IEC 42001 and AI System Conformity Assessment. 2025.
- [10] ISACA. ISO 42001: Balancing AI Speed & Safety. October 2025.
- [11] ECRI. Top 10 Health Technology Hazards for 2025. AI without oversight ranked #1.
- [12] Stanford HAI. Assessment of Trustworthy AI in the Context of Healthcare.
- [13] U.S. Department of Justice. Subpoenas to pharma/digital health companies over AI in EMR. 2024.
- [14] Bates, D. et al. Software-Related Patient Injuries: Analysis of 51 Court Cases. 2024.
- [15] Waymo Recall of 3,067 Robotaxis. School bus safety violations. December 2025.
- [16] Austin ISD. Letter to Waymo documenting 20 illegal school bus passes. 2025–2026.

- [17] Tesla-Fanuc \$51M Lawsuit. Robotic arm incident, Giga Texas. September 2025.
- [18] OSHA. Robot-Related Workplace Accidents: 77 incidents, 93 injuries (2015–2022).
- [19] NIOSH. Robot-Related Fatalities: 61 deaths (1992–2015).
- [20] Swiss Re and Waymo. Comparative Safety Performance. 25.3M miles. December 2024.
- [21] NHTSA. Standing General Order on ADS Incident Reporting.
- [22] Frontier Model Forum. Comments on NIST TEVV Standard Outline. September 2025.
- [23] Sentinel Authority. ODDC Overview v4.0, ENVELO Requirements v3.0, CAT-72 Procedure v5.0.

— End of Document —

© 2026 Sentinel Authority. Distribution permitted with attribution.