

AI Risk Management Assurance Network (AIRMAN) - White Paper

Aidan Kierans
University of Connecticut

The AI Risk Management Assurance Network (AIRMAN) addresses a critical gap in AI safety: the disconnect between existing AI assurance technologies and standardized safety documentation practices. While the market shows high demand for both quality/conformity tools and observability/monitoring systems, currently used solutions operate in silos, offsetting risks of intellectual property leaks and antitrust action at the expense of risk management robustness and transparency. This fragmentation not only weakens safety practices but also exposes organizations to significant liability risks when operating without clear documentation standards and evidence of reasonable duty of care.

Our solution creates an open-source standards framework that enables collaboration and knowledge-sharing between frontier AI safety teams while protecting intellectual property and addressing antitrust concerns. By operating as an OASIS Open Project, we can provide legal protection for industry cooperation on developing integrated standards for risk management and monitoring.

The AIRMAN is unique in three ways: First, it creates a neutral, dedicated platform where competitors can collaborate on safety standards. Second, it provides technical integration layers that enable interoperability between different types of assurance tools. Third, it offers practical implementation support through templates, training programs, and mentorship systems.

The commercial viability of our solution is evidenced by strong willingness-to-pay across all major stakeholder groups for quality and conformity tools. By reducing duplication of effort in standards development and enabling economies of scale in implementation, we create clear value for participants while advancing the critical goal of AI safety.

1. Problem overview

As artificial intelligence systems grow more capable and autonomous, frontier AI development poses increasing systemic risks that could affect society at a massive scale. A critical gap exists in the current AI safety landscape: while individual companies are developing sophisticated tools for AI quality management and operational monitoring, these efforts remain dangerously siloed and fragmented.

The liability implications of this fragmentation are significant. Courts assess duty of care based not only on regulatory compliance but also on broader standards including industry best practices, responses to known threats, and organizational capacity for continuous improvement. Without comprehensive documentation linking development practices to safety outcomes, AI companies face increased liability exposure - especially since established legal precedent suggests courts may

hold companies to "reasonable" safety standards rather than just common industry practices (Kierans, Ritticher, and Sonsayar).

This division creates two distinct but disconnected domains. quality, conformity & audit management platforms focus on compliance and best practices through standardized review procedures, but often lack integration with operational monitoring that could provide crucial real-time risk data. Observability, monitoring and incident response solutions excel at tracking deployments and managing incidents but typically operate independently from quality management systems, creating potential blind spots in comprehensive risk assessment.

This fragmentation is particularly problematic for frontier AI development, where complex development pipelines require seamless integration between quality assurance and operational monitoring, novel capabilities may emerge during training demanding real-time risk assessment, and systemic risks require coordinated response across organizational boundaries. Without a focused initiative to develop integrated standards for safety documentation and verification, the industry risks perpetuating dangerous gaps in risk management while duplicating effort through siloed solutions.

2. Your solution

The AI Risk Management Assurance Network (AIRMAN) creates an open-source standards framework by enabling collaboration between frontier AI safety teams while protecting intellectual property. By 2027, our initiative will establish the definitive framework for documenting and verifying AI safety practices across the development lifecycle.

The solution consists of three core components: 1) A "Documentation Integration Layer" that standardizes information sharing between quality management and monitoring tools, 2) A "Safety Verification Framework" establishing methods for documenting and validating safety practices, and 3) A set of "Risk Assessment Protocols" for structuring internal documentation of risks and mitigation strategies.

This structure addresses the fragmentation in current AI safety documentation by bridging the silos between conformity tools and operational monitoring. It enables verification that documented practices are being followed and creates traceable links between safety decisions and outcomes. The framework provides comprehensive coverage across the entire AI development lifecycle while maintaining compatibility with major AI development workflows. The modular architecture is also highly scalable; as new capabilities emerge, we can quickly develop and integrate corresponding safety documentation standards. The open-source nature means that as more resources become available, the community can expand the framework's coverage while maintaining consistency through our established governance structure.

Our competitive advantage comes from three key factors. First, we leverage OASIS Open's established framework to provide legal protection for industry cooperation while developing integrated standards. Second, our focused approach allows faster development of technical specifications compared to broader initiatives. Third, we

engage a wider range of stakeholders including civil society organizations and academic researchers, enabling us to develop more robust and practical standards.

3. Pilot experiment or demo

The key output was this document's analysis of existing AI assurance technologies and practices, identifying specific technical requirements for bridging current gaps. The main challenge was determining how to enable meaningful collaboration on safety standards while protecting intellectual property and addressing antitrust concerns. I addressed this through careful study of successful precedents in other industries and existing frameworks like OASIS Open.

This research phase revealed valuable insights about requirements for technical feasibility, though there are significant limitations. As a solo project completed within a short sprint, I was not able to develop concrete technical artifacts or integration examples. Future work should focus on expanding the team to develop these technical components and launching pilot programs with industry partners.

4. Process

Our implementation strategy focuses on rapid development of core standards while building sustainable structures for industry adoption and governance. Operating through OASIS Open's established framework enables us to quickly create technical specifications while engaging key stakeholders from industry, civil society, and academia.

Timeframe	What will you do?
Next 3 months	- Begin OASIS Open Project by drafting a charter, identifying sponsors, soliciting feedback, and establishing a core group - Create detailed technical specifications for Documentation Integration Layer APIs and data schemas - Launch outreach program to engage civil society organizations, legal experts, and academic researchers
2025	- Develop reference implementations and validation tools - Release v1.0 of core standards including API specifications, data models, and validation frameworks - Establish working relationships with key regulatory bodies - Launch pilot programs with early adopter organizations
2026	- Scale adoption through integration with major AI development platforms and assurance tools - Release enhanced standards (v2.0) incorporating lessons learned from early implementations - Expand certification program to include auditor training and qualification - Establish formal liaison relationships with regulatory bodies
2027	- Achieve recognition of standards in key regulatory

	frameworks (EU AI Act, US guidelines) - Establish comprehensive training and support ecosystem - Create knowledge sharing networks for implementers - Develop advanced features for emerging AI capabilities
--	--

The timeline emphasizes early delivery of practical tools and templates that organizations can begin implementing immediately, while systematically developing the broader ecosystem needed for widespread adoption.

5. Impact on AI safety & key risks

The AI Risk Management Assurance Network delivers significant safety impact by addressing critical gaps in current documentation and verification practices. By establishing comprehensive standards across the AI development lifecycle, we create clear chains of responsibility through integrated quality management and operational monitoring. This directly improves safety by facilitating early detection of emerging risks and reducing the likelihood of "strategic overlooking" through robust documentation requirements. Our solution strengthens the overall safety culture by reducing duplication of effort in standards development, creating economies of scale in safety tool development, and building shared understanding of best practices across organizations, governments, and civil society.

The initiative faces several key risks requiring careful mitigation. First, insufficient participation from major AI developers could limit effectiveness - we'll address this through early engagement via OASIS Open relationships and demonstrating a clear value proposition. Second, competing standards could emerge with contradictory recommendations, potentially reducing impact. We plan to mitigate this by quickly establishing our framework as the de facto standard while actively collaborating with initiatives like CoSAI to ensure complementary rather than competing approaches. By 2026, we aim to take a leadership role in guiding industry-wide standards development.

Our framework aligns closely with existing industry standards such as the NIST AI Risk Management Framework and EU AI Act requirements, maintained through regular engagement with standards bodies and regulatory agencies. Progress will be monitored through adoption rates across different organization sizes, documentation completeness scores, and risk assessment coverage. Key performance indicators focus on both technical implementation (integration rates, support resolution times) and safety outcomes (incident response times, cross-organization learning rates). Through this focused approach combining robust standards with practical implementation support, AIRMAN can deliver meaningful improvements to AI safety documentation practices at scale while actively managing key risks to successful deployment.

6. Appendix

A. Market Analysis & Industry Context

The current AI assurance technology market shows significant demand across all major stakeholders for tools that can help manage AI risk. Through our research, we've identified two primary categories of solutions that organizations are willing to invest in:

1. Quality and conformity management tools that focus on compliance and best practices
2. Observability and monitoring systems that track deployments and manage incidents

However, these solutions currently operate in isolation from each other, primarily due to concerns around intellectual property protection and potential antitrust issues. This fragmentation creates several critical inefficiencies:

- Organizations must duplicate effort in developing safety standards
- Monitoring systems lack integration with quality management frameworks
- Cross-organizational learning is limited
- Response times to emerging risks are slower than necessary
- Safety documentation practices vary widely between organizations

B. Nuclear Precedent

The nuclear sector provides a sobering example of how fragmented safety documentation can mask fundamental issues. Research by Juraku & Sugawara (2021) examining Japan's nuclear industry revealed that:

- Documentation focused on superficial compliance rather than active safety engagement
- "Strategic overlooking" of uncomfortable knowledge became common practice
- Formal governance procedures failed to prevent the Fukushima disaster
- Integration between safety systems and operational monitoring was insufficient

We believe that AIRMAN can prevent similar failures from occurring in the AI sector.

C. Regulatory Framework Analysis

Our solution is designed to align with and support compliance with emerging regulatory requirements:

European Union AI Act

The EU AI Act establishes specific documentation requirements for both high-risk AI systems and general-purpose AI models with systemic risk. Key requirements include:

- Detailed documentation of accuracy metrics and robustness measures
- Comprehensive cybersecurity protocols
- Specifications for input data requirements
- Performance data across different user groups

- Documentation of testing procedures and results

The process of implementing these requirements remains to be specified, and open-source efforts could lead to stronger engagement that more effectively incorporates stakeholder feedback.

US Standards & Guidelines

While currently voluntary, US standards through the NIST AI Risk Management Framework suggest similar documentation requirements. Our framework will use these guidelines as a starting point while maintaining flexibility for future regulatory changes — an especially important point given the changing US political landscape.

D. Technical Implementation Details

Our solution's architecture consists of four main technical components:

1. Documentation Integration Layer

This component enables standardized information sharing between different tools and systems through:

- Common APIs for safety-critical information exchange
- Standardized data formats and schemas
- Integration hooks for existing platforms
- Secure data exchange protocols

2. Safety Verification Framework

This framework provides:

- Standardized documentation methods across the AI development lifecycle
- Clear mapping between procedures and monitoring data
- Validation protocols for safety practices
- Traceability requirements linking decisions to outcomes

3. Risk Assessment Protocols

These protocols establish:

- Structured templates for risk identification and documentation
- Standardized formats for mitigation strategies
- Real-time capability monitoring during AI training
- Integration requirements with existing risk management systems

4. Interoperability Framework

This framework ensures system-wide compatibility through:

- Modular architecture specifications
- Comprehensive API documentation
- Common data schema definitions
- Integration patterns for various tool types

- Security and access control requirements
- Flexible deployment configuration options

E. References

AIAT Report. "Risk & Reward: AI Assurance Technology Market Report 2024." AI Assurance Technology Market Report, (May 2024). <https://aiat.report>

Juraku, K., & Sugawara, S. E. (2021). Structural Ignorance of Expertise in Nuclear Safety Controversies: Case Analysis of Post-Fukushima Japan. *Nuclear Technology*, 207(9), 1423–1441.
<https://doi.org/10.1080/00295450.2021.1908075>

Kierans, Aidan, Kaley Rittichier, and Utku Sonsayar. "Safety Documentation and Liability: Managing Systemic Risks in Frontier AI Development." University of Connecticut, 2025 [preprint]. Non-public link:
<https://docs.google.com/document/d/16UUo9JdcK2zXCx1lmkyHsFIhb8a2x6KIpDMFW9xruRs/edit?usp=sharing>