US and International AI Governance

A Structured Practitioner's Independent White Paper

Author: Tylor J. G. Miranda

Affiliation: USAF Veteran; Enterprise GIS, Secure Federal Systems, THORIUM AI Holdings

Correspondence: www.linkedin.com/in/tylormiranda

Location: Alaska, United States

Author Note

This article is a structured practitioner review, not a causal empirical study or formal compliance assessment. It reflects an independent review of major AI governance frameworks, selected AI incident and implementation literature, and documented AI risk patterns from approximately 2018 through 2026.

The framework scoring is based on single-reviewer structured expert judgment and should be read as an analytical aid rather than a validated empirical ranking, legal opinion, audit result, or certification assessment.

No external funding was received for this review. The author has no financial conflict of interest related to the frameworks, standards, or sources discussed.

Abstract

AI governance frameworks are broadly correct about the risks created by modern AI systems. The major documents reviewed in this article consistently identify accountability, transparency, fairness, privacy, safety, security, human oversight, documentation, lifecycle monitoring, and remediation as core governance concerns. However, evidence from 2018 through 2026 suggests that identifying these risks is not enough. Organizations continue to face both real-world AI harms and implementation problems when governance expectations are not translated into repeatable controls.

This article presents a structured practitioner review of major AI governance frameworks, including the NIST AI Risk Management Framework, ISO/IEC 42001, OECD AI Principles, UNESCO Recommendation on the Ethics of Artificial Intelligence, EU AI Act, Council of Europe Framework Convention on AI, U.S. federal AI guidance, U.S. enforcement agency materials, and IEEE standards. The article then compares these frameworks against documented evidence from AI incident research, peer-reviewed studies, public-sector adoption research, enterprise survey data, and AI safety reporting from approximately 2018 through 2026.

The review finds that the major frameworks are directionally aligned with observed AI risks. Real-world evidence shows recurring concerns involving discrimination, biometric error, healthcare bias, privacy exposure, weak accountability, unclear human oversight, synthetic media harms, cybersecurity misuse, and disorganized adoption. These evidence patterns broadly align with the risks identified in major governance documents.

The central gap is not risk identification. The central gap is operationalization.

This article does not claim to prove that operationalization is the sole or primary cause of AI governance failure. Rather, it argues that an operationalization gap is a plausible and important explanation supported by converging evidence: framework documents often leave implementation decisions to organizations; observed AI harms align with framework-identified risks; incident research shows accountability, reporting, and corrective-action gaps; responsible-AI tooling under-supports deployers,

institutional leaders, end users, and impacted communities; and adoption evidence suggests AI use can expand faster than guidance, training, and oversight.

The article concludes that responsible AI requires more than principles, more than compliance, and more than technical testing. It requires a practical operating model: use-case intake, AI inventory, risk tiering, evidence management, vendor review, role-based training, human oversight design, incident escalation, lifecycle monitoring, and executive reporting.

1. Introduction

AI governance has moved from a specialized policy issue into a core enterprise concern. Organizations are adopting generative AI tools, machine learning systems, automated decision systems, AI-enabled analytics, vendor AI platforms, and increasingly agentic systems at a pace that often exceeds traditional governance cycles.

At the same time, governments, standards bodies, intergovernmental organizations, enforcement agencies, and research institutions are clarifying expectations for responsible AI. A multinational organization may need to consider the NIST AI Risk Management Framework, ISO/IEC 42001, OECD AI Principles, UNESCO Recommendation on the Ethics of Artificial Intelligence, EU AI Act, Council of Europe Framework Convention on AI, U.S. federal AI guidance, FTC enforcement materials, EEOC guidance on algorithmic employment discrimination, DOJ civil rights materials, and IEEE standards related to ethical system design, transparency, well-being, and organizational governance.

The result is not a lack of guidance. The result is a dense and expanding guidance environment.

At a high level, these sources increasingly agree. They repeatedly emphasize accountability, transparency, risk management, documentation, fairness, privacy, security, human oversight, monitoring, and lifecycle governance. This convergence is useful. It gives organizations a shared vocabulary for responsible AI.

But shared vocabulary is not the same as operational capability.

This article presents a structured practitioner review of major AI governance frameworks and compares those frameworks against documented AI harms and implementation problems from approximately 2018 through 2026. It does not claim to prove that operationalization is the sole or primary cause of AI governance failure. Rather, it argues that the operationalization gap is a plausible and important explanation supported by converging evidence: frameworks often leave implementation decisions to organizations; observed AI harms broadly align with framework-identified risks; incident research shows accountability, reporting, and corrective-action gaps; responsible-AI tooling under-supports deployers, institutional leaders, end users, and impacted communities; and adoption evidence suggests AI use can expand faster than guidance, training, and oversight.

The practical question is no longer whether AI governance frameworks recognize risk. Most do. The more important question is whether organizations can turn those recognized risks into ownership, workflow, evidence, training, oversight, monitoring, and executive decision-making.

2. Research Questions

This review is organized around one central question:

Are organizations struggling with responsible AI because they lack principles, or because existing principles have not been translated into practical enterprise operating processes?

The review also addresses six supporting questions:

1.	**Which AI governance frameworks are primarily values-based, which are**

	process-oriented, and which are legally binding or enforcement-oriented?
2.	Which frameworks provide the clearest operational guidance?
3.	Which frameworks are strongest on ethics, human rights, and public values?
4.	Which frameworks are strongest on documentation, lifecycle governance, risk management, and auditability?
5.	Where do frameworks leave gaps in role assignment, training, adoption support, executive reporting, evidence management, and cross-functional implementation?
6.	Do documented AI harms and implementation problems from 2018 through 2026 align with the risk categories identified by major frameworks?

3. Methodology

This article uses a structured practitioner review supported by two methods.

The first method is a comparative documentation review of major AI governance frameworks and official supporting materials.

The second method is an evidence-based framework-to-reality comparison. This comparison evaluates whether the risk categories identified in those frameworks align with documented AI harms, AI incident research, adoption studies, public-sector use studies, enforcement evidence, and peer-reviewed or academically published studies from approximately 2018 through 2026.

The purpose of the evidence comparison is not to prove that any single framework succeeds or fails. The purpose is to test whether the risks described in major AI governance frameworks align with observed harms and implementation patterns. The question is practical: if the frameworks identify the right risks, where does enterprise implementation still break down?

3.1 Framework Corpus

The primary framework corpus includes:

1.	**NIST AI Risk Management Framework 1.0.**
2.	NIST AI RMF Playbook and NIST AI Resource Center materials.
3.	ISO/IEC 42001:2023, Artificial intelligence management systems.
4.	OECD AI Principles.
5.	OECD Due Diligence Guidance for Responsible AI.
6.	UNESCO Recommendation on the Ethics of Artificial Intelligence.
7.	Regulation (EU) 2024/1689, the EU Artificial Intelligence Act.
8.	Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law.
9.	U.S. federal AI governance and procurement guidance.
10.	U.S. enforcement guidance and agency materials from the FTC, EEOC, and DOJ.
11.	IEEE standards and official standards summaries related to ethical system design, transparency, well-being impact assessment, and organizational AI governance.

3.2 Framework Coding Method

Framework scoring was conducted as single-reviewer structured expert judgment. Scores are not presented as validated empirical measurements, compliance determinations, or formal audit findings. They are comparative indicators of how directly each framework supports enterprise implementation across ten operational dimensions.

The scoring scale uses three levels only:

High indicates that the framework directly and meaningfully addresses the dimension through explicit requirements, structured guidance, or implementation-relevant detail.

Medium indicates that the framework addresses the dimension, but does so broadly, partially, or in a way that requires substantial organizational interpretation.

Low indicates that the framework mentions the dimension minimally, addresses it only indirectly, or leaves most implementation decisions to the organization.

No inter-rater reliability calculation was performed. The matrix should therefore be read as an analytical aid rather than a formal coding study.

One important limitation of the matrix is that some dimensions show low variance across the corpus. Scope and Risk Model score High or Medium for most frameworks because nearly all reviewed sources define an AI governance scope and address risk in some form. These dimensions help describe the corpus, but they do less analytic work in distinguishing frameworks. The stronger comparative signal appears in Role Clarity, Training/Adoption, Reporting, Enforcement, and Operational Specificity.

Each framework was assessed across ten implementation dimensions:

1.	**Scope.**
2.	Legal status.
3.	Risk model.
4.	Lifecycle coverage.
5.	Role clarity.
6.	Documentation requirements.
7.	Training and adoption guidance.
8.	Reporting expectations.
9.	Enforcement mechanism.
10.	Operational specificity.

Operational specificity means the extent to which a framework can be translated directly into organizational workflows, templates, approval gates, evidence packages, training paths, and reporting routines without substantial additional interpretation.

3.3 Evidence Corpus

The evidence comparison covers approximately 2018 through 2026. This period captures several overlapping developments: algorithmic bias research, biometric performance concerns, healthcare algorithmic harm, public-sector AI use, AI incident database growth, generative AI adoption, synthetic media risk, enterprise AI implementation gaps, AI agent transparency concerns, and the rapid expansion of AI governance activity.

Evidence was grouped into six categories:

1.	**Longitudinal AI index and governance trend reporting.**
2.	AI incident databases and incident-taxonomy research.
3.	Peer-reviewed or academically published studies of AI harm.
4.	Government and enforcement evidence.
5.	Enterprise adoption and implementation studies.
6.	AI agent and general-purpose AI safety research.

3.4 Evidence Grading Method

The evidence base was graded by source type and evidentiary role.

Grade A evidence includes peer-reviewed or academically published empirical studies with specific measured outcomes. These sources were used to document measurable harms or performance differences.

Grade B evidence includes official government, standards-body, intergovernmental, or regulatory materials. These sources were used to establish governance expectations, legal obligations, enforcement positions, and formal policy direction.

Grade C evidence includes longitudinal index reporting, incident database research, systematic reviews, and preprint research where formal peer-review status is not confirmed. These sources were used to identify broader patterns, trend indicators, incident categories, and governance research gaps.

Grade D evidence includes consulting surveys, enterprise adoption surveys, practitioner reports, and business reporting. These sources were used only as indicators of adoption pressure, organizational readiness, or implementation challenges. They were not treated as proof of governance effectiveness or failure.

This distinction matters because the evidence classes do different work. A peer-reviewed algorithmic-bias study can show measurable harm. An incident database can reveal reporting and accountability patterns, but not a complete census of all AI harms. A policy index can show increased regulatory attention, but not governance quality. A consulting survey can show adoption pressure, but not whether responsible AI is working.

4. Framework Classification

The reviewed frameworks fall into three broad families: values and ethics frameworks, management and process frameworks, and binding or enforcement-oriented frameworks. These categories overlap, but the distinction helps explain why different documents are useful for different organizational purposes.

4.1 Values and Ethics Frameworks

Values and ethics frameworks define what AI should protect, promote, and avoid. They are strongest when the question is: what should responsible AI mean in a democratic, human-centered, rights-respecting society?

This family includes the OECD AI Principles, UNESCO Recommendation on the Ethics of Artificial Intelligence, and Council of Europe Framework Convention on AI.

These frameworks emphasize human rights, democratic values, transparency, accountability, fairness, safety, sustainability, privacy, human agency, non-discrimination, and protection of the rule of law.

They are valuable because they establish legitimacy. They prevent AI governance from becoming only a technical, legal, or compliance exercise. They remind organizations that AI risk is not limited to cybersecurity, model accuracy, or legal exposure. AI can affect human dignity, autonomy, employment, public services, access to opportunity, democratic participation, social trust, institutional accountability, and procedural fairness.

However, values frameworks often leave implementation gaps. They rarely provide detailed internal workflow designs. They do not usually specify the exact intake form, approval gate, training curriculum, dashboard structure, risk register, incident taxonomy, evidence package, or escalation path an enterprise should use. Their strength is normative direction. Their weakness is operational incompleteness.

4.2 Management and Process Frameworks

Management and process frameworks define how an organization should structure governance, controls, monitoring, documentation, and improvement cycles. They are strongest when the question is: how should an organization manage AI as an enterprise capability?

This family includes NIST AI RMF, the NIST AI RMF Playbook, ISO/IEC 42001, and IEEE governance-oriented standards.

These sources move closer to practical implementation. NIST provides the four-function structure of Govern, Map, Measure, and Manage. ISO/IEC 42001 provides a management-system approach for establishing, implementing, maintaining, and continually improving AI governance. IEEE standards address topics such as ethical system design, measurable transparency, well-being impact assessment, and organizational governance.

These frameworks are valuable because they begin to define repeatable enterprise functions. They address inventories, role assignment, documentation, measurement, monitoring, governance structures, risk management, and continuous improvement.

However, even these frameworks do not fully solve implementation. NIST provides outcomes and suggested actions, but organizations must still design their own templates, workflows, governance cadence, dashboards, evidence standards, approval authorities, and handoffs. ISO/IEC 42001 supports a certifiable management system, but implementation still requires organizational design and day-to-day procedures. IEEE standards are useful but modular; organizations must assemble them into a coherent operating model.

4.3 Binding and Enforcement-Oriented Frameworks

Binding and enforcement-oriented frameworks define obligations, penalties, compliance duties, and regulatory consequences. They are strongest when the question is: what must an organization do to comply with law, procurement rules, or enforcement expectations?

This family includes the EU AI Act, U.S. federal AI governance and procurement guidance, and enforcement activity or guidance from agencies such as the FTC, EEOC, and DOJ.

The EU AI Act is the most comprehensive binding legal framework in the reviewed corpus. It establishes risk categories, prohibitions, requirements for high-risk systems, transparency duties, provider and deployer obligations, documentation duties, human oversight expectations, post-market monitoring, serious incident reporting, and administrative fines.

U.S. federal AI guidance is binding mainly within the federal agency context, but it is highly instructive for enterprise governance because it addresses inventories, Chief AI Officers, governance boards, high-impact AI, procurement controls, testing, documentation, monitoring, and risk management. U.S. private-sector enforcement remains distributed across existing authorities. The FTC may address deceptive or unfair AI claims. The EEOC may address discriminatory employment practices involving AI. The DOJ may address civil rights and related enforcement concerns.

These frameworks are valuable because they create consequences. They turn abstract governance into compliance risk.

However, binding law and enforcement guidance still do not automatically create internal operating capability. A legal obligation to maintain documentation does not tell a product team how to gather the documentation. A requirement for human oversight does not define the reviewer's training, authority, independence, and documentation duties. A serious incident reporting obligation does not automatically create an internal escalation matrix. Compliance requirements still must be translated into enterprise process.

5. Framework Scoring Matrix

Framework	Scope	Legal Status	Risk Model	Lifecycle Coverage	Role Clarity	Documentation	Training / Adoption	Reporting	Enforcement	Operational Specificity
NIST AI RMF 1.0	High	Low	High	High	Medium	Medium	Medium	Medium	Low	Medium
NIST AI RMF Playbook	High	Low	High	High	Medium	High	Medium	Medium	Low	High
ISO/IEC 42001	High	Medium	High	High	High	High	Medium	Medium	Medium	High
OECD AI Principles	High	Low	Medium	Medium	Medium	Low	Low	Low	Low	Low
OECD Due Diligence Guidance	High	Low	High	High	Medium	Medium	Medium	Medium	Low	Medium
UNESCO AI Ethics Recommendation	High	Low	Medium	Medium	Medium	Medium	High	Medium	Low	Medium
EU AI Act	High	High	High	High	High	High	Medium	High	High	High
Council of Europe AI Convention	High	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium
U.S. Federal AI Guidance / OMB	Medium	High	High	High	High	High	High	High	Medium	High
FTC / EEOC / DOJ Enforcement Materials	Medium	High	Medium	Low	Medium	Low	Low	Medium	High	Medium
IEEE Standards Ecosystem	Medium	Medium	High	High	Medium	High	Medium	Medium	Medium	High

The matrix should not be read as a ranking of overall quality. Several frameworks are designed for different purposes. A values framework may score lower on operational specificity while still being essential for democratic legitimacy and human-rights alignment. A binding regulation may score higher on enforcement while still requiring substantial internal translation before an organization can implement it. A voluntary risk framework may score lower on enforcement but still provide a strong operating structure.

The matrix shows that most reviewed sources recognize AI risk in some form. The more important distinction is whether the framework helps organizations assign responsibility, collect evidence, train users, monitor deployment, report to executives, and respond to incidents.

6. Evidence Review, 2018–2026

The evidence review is not intended to serve as a complete empirical assessment of AI harms. Instead, it identifies several documented evidence patterns that are relevant to the article's central thesis.

The strongest evidence pattern is not that frameworks are detached from reality. It is the opposite. The evidence broadly aligns with the risks the frameworks identify.

AI systems have produced measurable disparities in facial analysis, biometric systems, healthcare risk scoring, public-sector AI use, employment-related automation, synthetic media, and enterprise adoption. Several of these harms map directly onto framework categories such as fairness, non-discrimination, transparency, human oversight, accountability, data governance, privacy, safety, and lifecycle monitoring.

However, the evidence also shows why framework-level alignment is not enough. Even when a framework correctly identifies a risk, organizations still need procedures, controls, training, evidence, monitoring, and decision rights to manage that risk.

This section therefore does not claim that documented AI harms prove operationalization is the only or primary governance failure. It uses documented harms and adoption patterns to show that the risks identified by governance frameworks are material, while incident-accountability research and responsible-AI tooling research suggest that organizational implementation remains uneven.

6.1 Numerical Evidence Summary

Evidence Area	Finding	Evidence Grade	Governance Implication
Global AI policy activity	Stanford AI Index reporting indicates that legislative attention to AI has grown substantially across countries over the past decade.	C	Governments are reacting quickly, but policy volume does not automatically create organizational implementation capacity.
U.S. AI regulatory activity	Stanford AI Index reporting indicates that U.S. federal and state AI-related policy activity increased substantially from 2023 to 2024.	C	Organizations face a more complex compliance environment and need internal obligation-mapping capability.
AI incident tracking	Paeth et al. reviewed an AI Incident Database corpus of more than 750 AI incidents.	C	Incident reporting is becoming a governance discipline, but taxonomies, thresholds, and evidence standards remain immature.
AI incident accountability	Richards, Benn, and Zilka analyzed 962 incidents and 4,743 related reports from the AI Incident Database and found that identifiable responsible parties do not necessarily lead to increased accountability.	C	Accountability cannot be assumed from role labels alone; organizations need defined response obligations and corrective-action ownership.
AI privacy and ethical incidents	Hadan et al. analyzed 202 real-world AI privacy and ethical incidents and found gaps in reporting, corrective measures, and legal action.	C	Accountability gaps persist after incidents occur, especially where reporting duties and corrective ownership are unclear.
Responsible AI tools	Kuehnert, Kim, Forlizzi, and Heidari categorized more than 220 responsible-AI tools and found that most support designers and developers during data-centric or modeling stages, while institutional leaders, deployers, end users, and impacted communities are underserved.	C	The responsible-AI tool ecosystem reflects a governance-to-operations gap: too much support for technical development, too little for deployment, oversight, governance, and adoption.
Facial analysis bias	Buolamwini and Gebru reported that commercial gender-classification systems had error rates as high as 34.7% for darker-skinned women, compared with 0.8% for lighter-skinned men.	A	Fairness and non-discrimination are measurable; demographic performance differences can be substantial.
Facial recognition demographic effects	NIST demographic-effects work found that many facial recognition systems produced different false positive rates across demographic groups, with substantially higher false positive rates for some groups in some systems.	B	Accuracy, demographic testing, and deployment-context validation are essential before high-stakes use.
Healthcare algorithmic bias	Obermeyer et al. found that a commercial health-risk algorithm using cost as a proxy for health need caused Black patients to be assigned lower risk than equally sick white patients.	A	Even race-neutral variables can encode structural inequality. Risk assessments must test proxies, outcome definitions, and data assumptions.
Public-sector generative AI use	Bright et al. surveyed 938 UK public service professionals and found that generative AI use was already present in public-sector work, while clear workplace guidance lagged behind use and trust.	C	Adoption can move faster than policy, training, and oversight, creating shadow-use and human-oversight risk.
Enterprise AI adoption	Enterprise survey reporting indicates rapid growth in AI use across business functions, while measurable impact and governance maturity remain uneven.	D	Enterprise use is widespread, but adoption does not equal responsible or effective implementation.
General-purpose AI safety	The International AI Safety Report synthesizes expert assessment across governments and international bodies.	B/C	General-purpose AI risk is now treated as a global governance and scientific issue, but mitigation still requires institutional controls.

7. Framework-to-Reality Comparison

The framework review shows what guidance documents say. The evidence review shows whether those risks are material in practice.

The evidence broadly aligns with the frameworks. AI governance documents repeatedly warn about discrimination, opacity, privacy harm, weak accountability, unsafe deployment, insufficient human oversight, and lifecycle monitoring failures. Documented AI incidents and empirical studies show that these categories are not hypothetical.

The problem is that identifying risk is not the same as controlling risk.

Real-World Risk Category	Evidence Pattern	Frameworks That Address It	What the Frameworks Generally Say	Remaining Operating Gap
Bias and discrimination	Gender Shades showed large demographic error differences in commercial facial analysis. Obermeyer et al. showed that cost-based proxies can understate health need for Black patients.	NIST AI RMF, EU AI Act, EEOC, DOJ, UNESCO, OECD, Council of Europe.	AI should be fair, non-discriminatory, tested, documented, and subject to oversight.	Organizations need concrete bias-testing ownership, proxy-variable review, demographic performance thresholds, corrective-action triggers, and business-user training.
Human rights, democracy, and rule of law	Biometric misidentification, opaque automated decisions, synthetic media, and high-stakes public-sector AI can affect liberty, due process, access to services, identity, and public trust.	Council of Europe, UNESCO, OECD, EU AI Act, DOJ.	AI lifecycle activities should respect human rights, democracy, rule of law, remedies, safeguards, transparency, and accountability.	Rights-based principles need deployment controls: notice, contestability, appeal channels, evidentiary documentation, review authority, and clear responsibility for remedies.
Transparency and explainability	AI systems, including emerging general-purpose and agentic systems, may provide limited information about safety, evaluation, data use, and societal impacts.	NIST AI RMF, IEEE transparency standards, EU AI Act, OECD, UNESCO, ISO/IEC 42001.	Users and affected parties should receive meaningful information; systems should support transparency, traceability, and documentation.	Transparency must become artifacts: user notices, model cards, system cards, agent permission disclosures, evaluation summaries, and approval records.
Human oversight	Public-sector generative AI evidence suggests that users may trust AI outputs even where workplace guidance is limited.	EU AI Act, NIST AI RMF, UNESCO, OECD, Council of Europe.	AI systems should include human oversight and meaningful human agency.	Human oversight must be operationalized through reviewer training, override authority, escalation rules, documentation duties, and oversight-quality monitoring.
Incident reporting and corrective action	AI Incident Database research shows that AI incidents are difficult to classify and that public accountability after AI harms can be inconsistent.	EU AI Act, NIST AI RMF, ISO/IEC 42001, OMB guidance, OECD due diligence.	Organizations should monitor systems, report serious incidents where required, communicate harms, and provide remediation.	Incident taxonomies, near-miss reporting, internal escalation paths, root-cause analysis, and corrective-action tracking are still immature.
Privacy and data governance	AI privacy and ethical incident research identifies reporting and corrective-action gaps after privacy-related harms.	FTC, NIST AI RMF, ISO/IEC 42001, EU AI Act, OECD, UNESCO.	Data should be governed, privacy protected, and AI use documented.	Organizations need approved-tool lists, data-classification rules, vendor data clauses, employee-use controls, and monitoring for shadow AI.
Enterprise adoption and change management	Enterprise and public-sector evidence suggests adoption can move faster than clear guidance, role-specific training, and workflow integration.	NIST AI RMF, ISO/IEC 42001, OMB guidance, UNESCO, OECD.	Governance culture, literacy, management systems, training, and adoption support are needed.	Adoption is outpacing enablement. Organizations need role-based learning paths, workflow integration, communications, metrics, and executive reporting.
Responsible AI tooling imbalance	Research on more than 220 responsible-AI tools found that institutional leaders, deployers, end users, and impacted communities are underserved.	NIST AI RMF, ISO/IEC 42001, OECD due diligence, IEEE, UNESCO.	Responsible AI should apply across roles and lifecycle stages.	Tooling over-supports technical development and under-supports deployment, business ownership, governance, and affected communities.
AI agents and autonomous execution	Agentic systems increase the risk of opacity, unauthorized action, weak monitoring, and difficult rollback.	NIST AI RMF, ISO/IEC 42001, EU AI Act, IEEE governance standards, emerging AI safety frameworks.	Higher-capability systems require risk management, evaluation, transparency, safeguards, and monitoring.	Agentic systems need permission tiers, audit logs, sandboxing, rollback, action thresholds, and continuous monitoring.

8. Findings

Finding 1: The Frameworks Are Directionally Aligned With Observed Risks

The evidence does not suggest that major AI governance frameworks are detached from reality. In many cases, the evidence supports the relevance of the risks they identify.

The frameworks warn about fairness; real-world systems have produced measurable demographic disparities.

The frameworks warn about transparency; real-world AI systems may provide limited information about safety, evaluation, and societal impacts.

The frameworks warn about accountability; incident studies show unclear responsibility, inconsistent response, and limited corrective action.

The frameworks warn about human oversight; adoption studies suggest that users may trust AI outputs while lacking clear workplace guidance.

The frameworks warn about documentation and lifecycle monitoring; incident evidence shows that failures can be difficult to classify, trace, and correct after deployment.

The stronger conclusion is that AI governance frameworks have broadly identified many of the right risks, but organizations still lack reliable operating models for managing those risks and the implementation problems that follow.

Finding 2: Scope and Risk Are No Longer the Main Differentiators

The framework matrix shows that most reviewed sources define broad scope and address risk in some form. That matters, but it is no longer the most important distinction among leading AI governance sources.

The real differences appear in operational dimensions.

Some frameworks are stronger on role clarity. Some are stronger on reporting. Some are stronger on enforcement. Some are stronger on training and adoption. Some are stronger on documentation and management-system design. These differences matter because implementation failures usually occur at the handoff between broad risk recognition and practical organizational execution.

In other words, the question is no longer only whether a framework recognizes AI risk. Most do. The more important question is whether the framework helps an organization assign owners, collect evidence, train users, review systems, monitor deployment, escalate incidents, and report to leadership.

Finding 3: Governance Activity Is Increasing Faster Than Governance Maturity

The period from 2018 through 2026 shows a sharp increase in AI policy activity, standards development, enforcement attention, AI safety reporting, and enterprise AI adoption.

This creates pressure on organizations. Regulation is expanding. Public expectations are increasing. Employee use is spreading. Vendor AI is entering procurement channels. General-purpose and agentic systems are becoming more capable. Incident reporting is becoming more visible.

The mismatch is the problem.

Governance activity is increasing.

AI adoption is increasing.

Incident evidence is accumulating.

But operational maturity remains uneven.

That produces a central enterprise risk: organizations may be surrounded by guidance while still lacking the internal machinery to comply with it.

Finding 4: Human Oversight Is Named More Often Than It Is Operationalized

Human oversight is one of the most repeated concepts in AI governance. It appears across values-based frameworks, risk-management frameworks, legal obligations, and enforcement-oriented guidance. The concept is sound: AI systems should not remove meaningful human agency from high-impact decisions.

The problem is that human oversight is often named more clearly than it is operationalized.

A human who lacks training, context, authority, independence, time, or documentation duties is not necessarily providing meaningful oversight. In some workflows, the human reviewer may become a procedural formality rather than an effective control. This is especially likely when the AI output appears technical, authoritative, efficient, or difficult to challenge.

A mature oversight model should define eight elements:

1.	What the human reviewer must check.
2.	What information the reviewer receives.
3.	When the reviewer may override the AI.
4.	When the reviewer must escalate.
5.	How disagreement with the AI is documented.
6.	What training the reviewer must complete.
7.	How oversight quality is monitored.
8.	Who audits the oversight process.

This is one of the clearest examples of the missing middle. The frameworks correctly identify human oversight as necessary, but organizations still need to design the workflow that makes oversight meaningful.

Finding 5: Responsible AI Tools Do Not Fully Support the People Who Need to Operate Governance

The evidence on responsible AI tooling is especially important. A systematic review of more than 220 responsible-AI tools found that existing tools tend to support designers and developers during data-centric or modeling stages, while institutional leaders, deployers, end users, and impacted communities are underserved.

That finding directly supports the operationalization-gap thesis.

The problem is not only that organizations lack policy. It is that the tools and artifacts needed to carry governance into deployment, adoption, oversight, training, and impact response are underdeveloped.

Responsible AI cannot be limited to model teams. It must be usable by procurement officers, HR managers, product owners, cybersecurity teams, legal reviewers, executives, frontline employees, and affected communities.

Finding 6: Documentation Requirements Are Growing, but Evidence Management Remains Weak

Documentation is central to NIST AI RMF, ISO/IEC 42001, the EU AI Act, OMB guidance, IEEE standards, and due diligence practices. However, organizations still need to decide what documentation is sufficient, where it is stored, who validates it, and how it is used.

A documentation requirement without an evidence-management process can produce fragmented records. A legal team may hold one document, a vendor another, a product team another, and a security team another. No one may have a complete record of the use case, risk classification, model purpose, dataset assumptions, testing results, human oversight design, vendor commitments, launch approval, and post-deployment monitoring.

Responsible AI requires evidence discipline.

9. Counterargument: Flexibility Is Necessary, but Structure Is Still Required

A reasonable counterargument is that AI governance frameworks should not prescribe a single operating model. Organizations differ by sector, size, use case, jurisdiction, data environment, technical maturity, and risk tolerance. A healthcare triage tool, an employment screening system, a customer-service chatbot, a military decision-support tool, and a logistics optimization model should not all be governed through identical procedures.

That objection is correct.

Over-prescription could create brittle compliance programs, encourage checklist behavior, and make frameworks obsolete as AI capabilities change. It could also prevent organizations from adapting controls to context-specific risks.

The practitioner operating model proposed in this article should therefore not be read as a universal compliance checklist. It is better understood as a reusable implementation architecture. The specific artifacts, metrics, and workflows should be adapted by sector, system type, legal environment, and risk tier.

The key point is not that every organization needs the exact same process. The key point is that every organization needs some controlled process for intake, inventory, risk classification, evidence management, vendor review, training, oversight, monitoring, incident escalation, and executive reporting.

Flexibility is necessary. Absence of operating structure is not.

10. Practitioner Implementation Appendix: Operating Artifacts, Metrics, and Roadmap

The operating artifacts, metrics, and roadmap in this section are authored practitioner recommendations derived from the framework comparison and evidence synthesis. They are not presented as direct requirements from any single framework, nor as empirically validated best practices. Their purpose is to translate the review's findings into an implementation architecture that organizations can adapt.

A. Minimum Viable AI Governance Operating Model

A minimum viable AI governance operating model should include:

An executive sponsor.

A named AI governance lead.

A cross-functional AI governance board.

A single AI intake process.

An enterprise AI inventory.

A risk-tiering rubric.

A legal and policy obligation mapping process.

A required evidence package for higher-risk use cases.

A vendor AI due diligence process.

A human oversight design standard.

A training and communication plan.

A monitoring and incident escalation process.

A recurring executive reporting cadence.

The operating model should not begin with a large policy library. It should begin with visibility and control.

Organizations need to know what AI is being used, by whom, for what purpose, with what data, in what jurisdiction, affecting which people, with which vendor dependencies, and under which risk tier.

Once visibility exists, policy becomes enforceable. Without visibility, policy is performative.

B. Core Governance-to-Operations Workflow

A mature enterprise AI governance process should include the following handoff chain:

1.	**Policy and standards interpretation.**
2.	Use-case intake.
3.	AI inventory entry.
4.	Initial risk classification.
5.	Jurisdiction and obligation mapping.
6.	Data, privacy, security, legal, and business review.
7.	Vendor due diligence where applicable.
8.	Impact assessment or risk assessment.
9.	Human oversight and transparency design.
10.	Evidence package completion.
11.	Approval or exception decision.
12.	User training and launch communication.
13.	Deployment monitoring.
14.	Incident escalation.
15.	Executive reporting.
16.	Periodic reassessment.
17.	Retirement, replacement, or modification control.

Each step requires an owner, template, decision rule, record, and handoff. Without those elements, governance remains descriptive rather than operational.

C. Persona-Specific Implementation Needs

AI governance is cross-functional. The same framework means different things to different teams.

Legal and compliance teams need obligation registers, jurisdiction mapping, provider/deployer responsibility matrices, regulatory watch logs, policy interpretation briefs, contract clause libraries, fundamental rights impact assessment templates, data protection impact assessment alignment guides, records retention schedules, and exception logs.

Security teams need AI threat model templates, model/system security baselines, abuse-case libraries, red-team plans, logging standards, AI incident taxonomies, containment and rollback playbooks, security review gates, vendor security questionnaires, and monitoring requirements.

Data governance and analytics teams need dataset provenance sheets, data suitability checklists, lineage registers, data rights matrices, training data documentation standards, evaluation data documentation standards, bias and representativeness assessments, feedback-loop control plans, model or system cards, and retention plans.

Product and user-experience teams need use-case intake forms, risk-tier rubrics, human oversight design checklists, user notice templates, explanation pattern libraries, appeal or contestability workflows, launch-readiness checklists, user testing plans for AI disclosures, and post-launch feedback processes.

HR and workforce teams need employee AI acceptable-use policies, algorithmic employment assessment templates, accommodation review checklists, manager training decks, role-based AI literacy curricula, employee FAQs, human review standards for employment-related AI, vendor assessments for HR AI tools, and workforce impact assessments.

Procurement and vendor management teams need AI vendor intake questionnaires, contract clause libraries, data rights checklists, model transparency requirements, audit and documentation requirements, service-level monitoring terms, subprocessor and subcontractor AI disclosures, vendor incident notification requirements, exit and data return plans, and third-party risk scoring matrices.

Operations and business units need plain-language process guides, use-case decision trees, role-specific checklists, approved AI tool lists, prohibited use lists, escalation contact lists, human review procedures, operational runbooks, monitoring checklists, and change request processes.

Executives and boards need AI portfolio dashboards, quarterly board packets, AI risk appetite statements, KPI/KRI scorecards, exception and waiver summaries, high-risk system inventories, incident and trend reports, investment and adoption roadmaps, governance maturity assessments, and annual AI governance effectiveness reviews.

These artifacts are not equally necessary for every organization. They should be selected based on use case, risk tier, sector, legal environment, and organizational maturity.

D. First-Generation AI Governance Metrics

AI governance should be measured. Otherwise, it becomes documentation without management.

Recommended first-generation metrics include:

Percentage of AI use cases inventoried.

Percentage of AI use cases risk-tiered.

Number of high-risk or high-impact systems.

Percentage of high-risk systems with completed assessments.

Percentage of systems with assigned business owners.

Percentage of vendor AI tools with completed due diligence.

Percentage of systems with completed security review.

Percentage of systems with completed privacy review.

Percentage of systems with documented human oversight.

Training completion by role.

Average time from intake to decision.

Number of open exceptions.

Number of unresolved monitoring findings.

Number of AI incidents or near misses.

Time from incident detection to escalation.

Number of systems pending reassessment.

Number of retired or decommissioned AI systems.

Executive action closure rate.

These metrics are not merely administrative. They make governance visible. They allow leaders to see whether responsible AI is functioning as an operating capability.

E. Prioritized Implementation Roadmap

An enterprise trying to operationalize AI governance should move in phases.

Phase 1: Establish ownership.

Name the executive sponsor, identify the AI governance lead, stand up the cross-functional governance board, define the governance charter, define risk appetite, and assign business ownership rules.

Phase 2: Build visibility.

Launch the AI inventory, publish the use-case intake process, classify known systems, identify high-risk or high-impact use cases, map jurisdictions and actor roles, and identify existing vendor AI exposure.

Phase 3: Create controls.

Publish the risk-tiering rubric, create required evidence package templates, build legal, privacy, security, and data review gates, create vendor due diligence procedures, define human oversight standards, and define transparency and notice requirements.

Phase 4: Train the organization.

Build role-based learning paths, train executives on portfolio risk and decision rights, train legal, security, HR, product, procurement, and operations teams, publish general employee AI acceptable-use guidance, and create communication materials explaining why the governance model exists.

Phase 5: Monitor and report.

Build the executive dashboard, create incident and near-miss reporting, define escalation thresholds, schedule periodic reassessments, track open exceptions and corrective actions, and report AI portfolio status to executives.

Phase 6: Improve and institutionalize.

Audit evidence quality, update templates based on incidents and lessons learned, retire or modify systems that fail governance standards, update training based on new risks, and reassess governance maturity annually.

11. Discussion

The reviewed frameworks show that AI governance is maturing in three directions at once.

First, values are becoming clearer. OECD, UNESCO, and the Council of Europe have helped define AI governance in terms of human rights, fairness, dignity, democracy, accountability, transparency, and social responsibility.

Second, management systems are becoming stronger. NIST, ISO/IEC 42001, and IEEE standards are helping organizations move from values into process, documentation, lifecycle thinking, design practices, and governance functions.

Third, enforcement is becoming more real. The EU AI Act creates binding obligations. U.S. agencies are applying existing law to AI-related harms. Federal procurement guidance is creating operational expectations for government AI acquisition and use.

However, these developments do not remove the implementation gap. In some ways, they make the gap more urgent. As expectations multiply, organizations need internal translation capacity.

That translation capacity is not purely legal, technical, educational, or compliance-based. It is a cross-functional governance and change management capability.

Organizations must translate principles into policies, policies into procedures, procedures into workflows, workflows into training, training into behavior, behavior into evidence, evidence into reporting, and reporting into executive decisions.

That translation layer is the governance-to-operations handoff.

12. Limitations

This review has several limitations.

First, it is a structured practitioner review, not a causal empirical study of organizational AI governance programs. It does not directly observe internal governance operations, employee behavior, executive decision-making, or implementation outcomes inside specific organizations.

Second, the framework scoring is single-reviewer structured expert judgment. It is intended to support comparison, not to certify compliance, assign legal risk, or produce a validated empirical ranking.

Third, some standards are not fully accessible without purchase or subscription. ISO/IEC 42001 and several IEEE standards are therefore assessed using official public summaries and explanatory materials rather than full-text reproduction.

Fourth, AI governance is changing quickly. U.S. federal guidance, EU implementing acts, codes of practice, standards, enforcement priorities, and enterprise practices may evolve.

Fifth, operational specificity may vary by sector. Healthcare, defense, finance, employment, education, critical infrastructure, consumer technology, and public administration may require different controls.

Sixth, AI incident databases and media-reported incident datasets are useful but incomplete. They help show patterns, but they do not provide a complete census of AI harms.

Seventh, adoption and financial-impact statistics should be interpreted carefully. AI use, AI adoption, AI maturity, and AI value realization are not the same thing.

Eighth, the practitioner appendix is authored recommendation. The artifacts, metrics, and roadmap are proposed implementation aids derived from the review, not direct requirements from any one framework.

13. Conclusion

The framework review and evidence comparison support the same conclusion from two directions.

The framework review shows that major AI governance documents converge around common principles: accountability, transparency, fairness, privacy, safety, security, human oversight, documentation, lifecycle risk management, and remediation.

The evidence review shows that real-world AI harms and implementation problems occur in many of those same categories. AI systems have produced or enabled discrimination, privacy exposure, misinformation, synthetic media harm, unsafe recommendations, cybersecurity misuse, automation bias, weak accountability, and organizational confusion over responsibility.

Together, these findings suggest that the central governance challenge is not simply a lack of risk identification. The risks have been identified. The harder challenge is operationalization.

AI governance is right about the risks. It is weaker on the operating model.

That does not mean the frameworks are inadequate. It means they are incomplete unless organizations build the operational layer beneath them.

Responsible AI requires more than principles, more than compliance, and more than technical testing. It requires organizational translation.

An organization does not become responsible by naming fairness. It becomes responsible by testing for disparate performance, reviewing proxy variables, assigning ownership, documenting results, and correcting failures.

An organization does not create meaningful human oversight by stating that a human is involved. It creates oversight by training the reviewer, defining override authority, requiring documentation, and monitoring whether human review is meaningful.

An organization does not manage incidents by acknowledging that incidents occur. It manages incidents through reporting channels, incident taxonomies, escalation paths, root-cause analysis, and corrective-action tracking.

The next stage of AI governance must therefore be evidence-based and operational. Organizations should not only ask whether they have principles. They should ask whether those principles are visible in workflow, training, documentation, monitoring, and executive decisions.

The organizations that succeed will not be the ones with the longest AI policy. They will be the ones that can prove, through process and evidence, that responsible AI is being practiced by the people who actually build, buy, deploy, manage, and use AI systems.

LinkedIn Introduction

AI governance frameworks are not wrong about the risks.

The major frameworks correctly identify the categories of harm now appearing in real-world AI incidents: discrimination, opacity, privacy exposure, deepfakes, unsafe recommendations, weak human oversight, cybersecurity misuse, and accountability gaps.

The issue is not that responsible AI lacks principles.

The issue is that principles must be translated into operating procedures: intake, risk tiering, training, vendor review, documentation, monitoring, incident response, and executive reporting.

That operating layer is where AI governance either becomes real or remains performative.

References

Bengio, Y., Mindermann, S., Privitera, D., Besiroglu, T., Bommasani, R., Casper, S., Choi, Y., Fox, P., Garfinkel, B., Goldfarb, D., Heidari, H., Ho, A., Kapoor, S., Khalatbari, L., Longpre, S., Manning, S., Mavroudis, V., Mazeika, M., Michael, J., Newman, J., Ng, K. Y., Okolo, C. T., Raji, D., Sastry, G., Seger, E., Skeadas, T., South, T., Strubell, E., Tramèr, F., Velasco, L., Wheeler, N., and others. (2025). International AI Safety Report. arXiv:2501.17805.

Bright, J., Enock, F. E., Esnaashari, S., Francis, J., Hashem, Y., & Morgan, D. (2024). Generative AI is already widespread in the public sector. arXiv:2401.01291.

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research, 81, 77–91.

Council of Europe. (2024). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law.

European Union. (2024). Regulation (EU) 2024/1689, Artificial Intelligence Act.

Federal Trade Commission. AI-related consumer protection and deceptive claims enforcement materials.

Grother, P., Ngan, M., & Hanaoka, K. (2019). Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects. National Institute of Standards and Technology.

Hadan, H., Mogavi, R. H., Zhang-Kennedy, L., & Nacke, L. E. (2025). Who is Responsible When AI Fails? Mapping Causes, Entities, and Consequences of AI Privacy and Ethical Incidents. arXiv:2504.01029.

IEEE. IEEE 7000, IEEE 7001, IEEE 7010, IEEE 2863, and related autonomous and intelligent systems standards materials.

International Organization for Standardization. (2023). ISO/IEC 42001:2023, Information technology — Artificial intelligence — Management system.

Kuehnert, B., Kim, R. M., Forlizzi, J., & Heidari, H. (2025). The "Who," "What," and "How" of Responsible AI Governance: A Systematic Review and Meta-Analysis of Actor- and Stage-Specific Tools. arXiv:2502.13294.

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework 1.0.

National Institute of Standards and Technology. AI RMF Playbook and AI Resource Center materials.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

OECD. AI Principles.

OECD. Due Diligence Guidance for Responsible AI.

Paeth, K., Atherton, D., Pittaras, N., Frase, H., & McGregor, S. (2024). Lessons for Editors of AI Incidents from the AI Incident Database. arXiv:2409.16425.

Richards, I., Benn, C., & Zilka, M. (2025). From Incidents to Insights: Patterns of Responsibility following AI Harms. arXiv:2505.04291.

Stanford Institute for Human-Centered Artificial Intelligence. AI Index reports.

UNESCO. Recommendation on the Ethics of Artificial Intelligence.

U.S. Department of Justice. AI, automated systems, and civil rights materials.

U.S. Equal Employment Opportunity Commission. AI and algorithmic fairness materials.

U.S. Office of Management and Budget. Federal AI governance and AI acquisition memoranda.