

Governance, risk, and compliance frameworks for AI Security: A review of emerging standards and challenges

Abimbola Filani ¹ and Jochebed Akoto Opoku ^{2,*}

¹ Department of Information Technology (DIT), Trine University, IN, USA.

² Department of Telecommunication Engineering, Kwame Nkrumah University of Science and Technology, Ghana.

Magna Scientia Advanced Research and Reviews, 2026, 17(02), 30–040

Publication history: Received on 26 May 2026; revised on 30 June 2026; accepted on 02 July 2026

Article DOI: <https://doi.org/10.30574/msarr.2026.17.2.0120>

Abstract

Artificial intelligence (AI) adoption is accelerating across industries, introducing novel governance, risk, and compliance (GRC) challenges that traditional cybersecurity frameworks cannot fully address. Standards such as ISO/IEC 27001 and the NIST Risk Management Framework safeguard IT assets but do not comprehensively mitigate AI-specific risks like adversarial attacks, model drift, and ethical concerns such as fairness and accountability. This gap raises critical questions about how organizations can govern AI responsibly while maintaining security and compliance. This paper reviews emerging AI-GRC frameworks and regulations, including the OECD AI Principles, ISO/IEC 42001, the NIST AI Risk Management Framework, and the EU AI Act, alongside industry standards from Microsoft, Google, and IBM. Through comparative analysis, we examine how these frameworks address governance structures, risk assessment methodologies, compliance mechanisms, and transparency requirements. We also explore integration strategies with existing cybersecurity and enterprise risk models. Our findings reveal a fragmented ecosystem with overlapping principles but inconsistent enforcement and technical depth. While global standards emphasize values such as transparency and accountability, operational guidance on adversarial robustness and lifecycle risk monitoring remains limited. The review identifies best practices for embedding GRC into AI development pipelines, including continuous monitoring, documentation artifacts, and risk-tiering strategies. It also highlights gaps in interoperability, audit tooling, and liability regimes. This research speaks to policymakers, compliance officers, cybersecurity professionals, and AI developers seeking harmonized governance approaches. Future work should prioritize unified audit standards, empirical evaluation of governance effectiveness, and integration of AI risk metrics into ESG reporting to ensure trustworthy and sustainable AI deployment.

Keywords: AI Governance; Adversarial attacks; Risk Management; Compliance; Transparency

1. Introduction

The growing adoption of artificial intelligence (AI) across industries is drastically changing business operations, public services, and societal infrastructure. AI promises to drive innovation in sectors ranging from healthcare to telecommunications by providing benefits such as increased automation, predictive insights, and scalability (Hamdouni, 2025). However, this move introduces several new and different security, privacy, and ethical problems. AI-enabled systems, unlike traditional information systems, are vulnerable to model poisoning, adversarial assaults, data integrity manipulation, and the transmission of unintended bias. These concerns reflect AI models' dynamic, data-driven, and often opaque character, which calls into question the efficacy of established governance, risk, and compliance frameworks (Rezaei, 2025).

* Corresponding author: Jochebed Akoto Opoku

Traditional cybersecurity GRC frameworks, such as ISO/IEC 27001 for information security management and the NIST Risk Management Framework (RMF) for federal information systems, offer systematic techniques to safeguard IT assets, manage threats, and ensure regulatory compliance (Nastoska et al., 2025). However, they are not entirely tailored to the specific threat vectors of AI systems, which involve not only technological vulnerabilities but also ethical and societal elements. For instance, consider algorithmic fairness, explainability, and accountability. As a result, there is an increasing need to create structured opportunities for governance and compliance measures customized specifically to AI systems (Ahanger et al., 2025; Panful et al., 2025).

Clarifying vital concepts can assist in addressing these issues. AI governance refers to the regulations, monitoring processes, and accountability structures that regulate how AI systems are created, implemented, and used. It guarantees that AI is consistent with the organization's goals, beliefs, and legal and ethical requirements (Hamdouni, 2025). According to one definition, AI governance at the organizational level is "a system of rules, practices, processes, and technological tools that ensure an organization's use of AI technologies aligns with its strategies, objectives, and values; fulfils legal requirements; and meets ethical AI principles" (Batoool et al., 2025). The term "AI risk" refers to the wide variety of potential hazards caused by AI. These include technical, ethical, and social dangers. Adversarial attacks, data governance problems, biased decisions, privacy violations, and broader negative societal consequences are all examples.

Compliance means adhering to laws, regulations, standards, and organizational policies. These standards impose requirements on how AI systems are built, implemented, and managed. Transparency, fairness, accountability, robustness, and human-centeredness are all examples of trustworthy AI principles. They serve as recommendations for making AI systems more dependable, ethical, and consistent with social norms. For example, research on trustworthiness examines fairness, transparency, privacy, security, and responsibility throughout the AI lifecycle (Nastoska et al., 2025). It views trust as something that must be assessed in multiple dimensions, not just one.

Despite increased regulatory and standards activity in AI, there is still no clear agreement on standardized frameworks that combine governance, risk, and compliance for AI-enabled systems. This review presents a complete overview of global AI-GRC frameworks. It identifies gaps, overlaps, and potential for harmonization. It also makes suggestions to academics, governments, and industry. Overall, we want to deepen our understanding of how to control AI responsibly and productively (Pelc, P. 2024).

2. Overview of Emerging Frameworks

2.1. International and Cross-Sector Frameworks

As AI systems become more complex, new global frameworks for responsible AI governance have emerged. These frameworks provide normative, policy, and technological direction. The OECD AI Principles were the first intergovernmental standard for trustworthy AI. More than 40 governments have embraced them (Mitchell et al., 2019). The principles emphasize inclusive growth, human-centered values, transparency, robustness, and responsibility. They serve as the foundation for subsequent national and regional initiatives, while also contributing to the work of the Global Partnership on AI (Marwala, 2024).

Another significant milestone is ISO/IEC 42001:2023, the Artificial Intelligence Management System Standard. It specifies the prerequisites for an organizational AI management system. The standard connects policy and operations, establishing a framework for continuous improvement, risk assessment, and governance of AI lifecycle activities. It is among the first standards to translate broad ideas into auditable management practices (Shetty et al., 2025). The UNESCO Recommendation on the Ethics of AI provides a worldwide, values-based perspective. It establishes ethical standards such as respect for human rights, equality, environmental sustainability, and responsibility. It outlines the requirements for an organizational AI management system. The standard integrates policy and operations, providing a framework for continuous improvement, risk assessment, and governance of AI lifecycle activities. It is one of the first standards to convert broad concepts into auditable management practices. The UNESCO Recommendation on the Ethics of AI offers a global, value-based perspective. It establishes ethical principles such as human rights, equality, environmental sustainability, and accountability (Marwala, 2024).

2.2. United States Frameworks

The NIST AI Risk Management Framework (AI RMF), which was launched in 2023, is the first US effort to formalize AI-specific risk governance. It is organized around four main functions: governance, mapping, measurement, and management. The framework provides practical tools to assist organizations in identifying and eliminating risks

associated with bias, transparency, robustness, and privacy (Mökander et al., 2023). Although voluntary, the AI RMF has quickly become a valuable resource for both industry and government. It also works with current cybersecurity standards, such as the NIST Cybersecurity Framework 2.0. The Executive Order on Safe, Secure, and Trustworthy AI supplements this foundation. It directs government entities to put AI risk management into effect. It also mandates that the deployment of AI in vital industries adhere to strict safety and privacy standards (Torres & Ali-Vehmas, 2025; Filani et al., 2026).

Along with general frameworks, there are sector-specific approaches. For example, the Food and Drug Administration (FDA) provides advice for medical AI devices. The Federal Financial Institutions Examination Council (FFIEC) provides AI risk advisories to the financial sector. Taken collectively, the US AI framework ecosystem combines technological standards and policy approaches. It develops a hybrid model that blends voluntary compliance with regulatory oversight (Alsaigh et al., 2024).

2.3. European Union

The European Union AI Act is the world's first legally binding, risk-based AI regulatory framework. It categorizes AI systems under four danger levels: unacceptable, high, limited, and minimum. Each level has its own compliance requirements. The Act has received appreciation for its broad breadth. However, other researchers contend that its prescriptive approach may stifle innovation or generate regulatory ambiguity (Aboy & Vayena, 2024). To complement the Act, the European Union Agency for Cybersecurity (ENISA) released the AI Threat Landscape. They emphasize transparency, human agency, and accountability in AI design and development. Taken together, these instruments form a sophisticated yet complex regulatory ecology. This ecosystem strives to strike a balance between innovation and protection (Madhavan et al., 2025).

2.4. Industry and Consortium Frameworks

Industry actors have played an important role in putting AI GRC into action. The ISO/IEC JTC 1/SC 42 technical committee is drafting worldwide AI standards for nomenclature, bias mitigation, and lifecycle management. Inside enterprises, frameworks such as the Microsoft Responsible AI Standard, the Google AI Principles, and the IBM AI Governance Framework establish AI governance policies and procedures. These frameworks serve to assure accountability and compliance throughout the product lifecycle. They frequently follow public-sector standards such as NIST AI RMF and ISO 42001, but place a greater emphasis on internal audits, documentation, and cross-functional oversight committees. This type of industry self-regulation promotes global alignment of trustworthy AI processes. It also demonstrates a shift toward implementing AI ethics on a large scale (Nastoska et al., 2025).

Table 1 Summary of Major AI Governance, Risk, and Compliance Frameworks

Framework	Scope	Domain	Focus	Maturity
OECD AI Principles	International, cross-sector	All AI systems	Values-based governance	High
ISO/IEC 42001: AIMS Standard	International, organizational	All AI systems	Governance & risk management	Emerging
UNESCO Recommendation on AI Ethics	International	Ethical AI	Human rights and sustainability principles	Medium
G7 Hiroshima Process on AI	Multilateral	Policy coordination	Global cooperation	Early
NIST AI Risk Management Framework	US, cross-sector	AI systems and services	Risk management lifecycle	High
EU AI Act (Reg 2024/1689)	EU, cross-sector	AI systems in the market	Risk-based compliance regulation	Medium
ENISA AI Threat Landscape	EU, cybersecurity	AI security risks	Threat taxonomy and risk mapping	Medium
ISO/IEC JTC 1/SC 42 Technical Standards	International	AI lifecycle	Technical standardization	Emerging

Microsoft Responsible AI Standard v2	Corporate	Product development	Operational governance	Medium
Google AI Principles	Corporate	R&D and deployment	Ethical principles and accountability	Medium
IBM AI Governance Framework	Corporate	AI solutions and services	Governance controls and compliance	Medium

(European Commission High-Level Expert Group on Artificial Intelligence. (2019).Oluoch, I. (2024).)

Table 1 shows frameworks for a maturing yet fragmented ecology. Global bodies such as the OECD and UNESCO establish high-level ethical standards. Technical organizations such as ISO and NIST translate these standards into actionable management systems. Regional regulators, such as those overseeing the EU AI Act, then enforce compliance. There is growing agreement on fundamental concepts such as transparency, fairness, and accountability. However, there are still variances in enforcement mechanisms, maturity levels, and how well these frameworks align with current cybersecurity requirements. For example, ISO 42001 and NIST AI RMF provide useful operational tools. However, just a few frameworks explicitly address adversarial robustness and model poisoning, two important risks to AI security (Madhavan et al., 2025).

3. Comparative Framework Analysis

3.1. Thematic Dimensions of AI GRC

Comparing AI governance, risk, and compliance frameworks across topics reveals how different tools prioritize roles, procedures, and controls. Governance models vary from high-level value statements to comprehensive regulation schemes. Normative mechanisms such as the OECD Principles and the UNESCO Recommendation emphasize values and organizational accountability (Shetty et al., 2025). They leave the specific assignment of roles to member states and particular organizations. In contrast, the EU AI Act and other state measures establish clearer responsibilities for system owners, providers, and deployers. They specify documentation requirements, compliance testing, and, in some situations, the appointment of human supervisors for high-risk systems. Internal governance patterns are added by industry frameworks such as Microsoft and IBM. These include responsibility matrices, review boards, and permission gates for operationalizing responsibilities across the product lifecycle (Orhan et al., 2025).

Another significant distinction is the process of managing risks. The NIST AI RMF is clearly risk-focused. It provides lifecycle-based approaches to mapping, measuring, and managing AI risk. This strongly supports both quantitative and process-based risk engineering. ISO/IEC 42001 applies the familiar management system approach to an AI scenario. It promotes constant improvement and risk assessment within an “AI management system”. Other frameworks, such as ENISA's AI Threat Landscape, offer technical taxonomies and attacker-specific mappings. These materials complement risk assessments, but they do not prescribe how businesses should structure their overall risk governance (Radanliev, P. 2025).

However, global adoption of these standards is still underway. Transparency and explainability are present in most frameworks; however, they are used in various ways. High-level principles view transparency and explainability as moral and practical necessities. NIST and ENISA provide technical recommendations to help put these concepts into action. They emphasize documentation, provenance tracking, and tools like model cards as practical procedures. Human oversight and larger ethical standards are important to UNESCO and OECD guidelines. Industry frameworks also help by emphasizing human-in-the-loop controls and internal review boards (Radanliev, P. 2025).

Security and privacy constraints receive varying degrees of emphasis among frameworks. Cybersecurity agencies and threat-focused frameworks, such as ENISA, prioritize security. They explain adversarial attack surfaces in depth and propose mitigation measures. The NIST AI RMF also addresses security, but only as part of a broader risk profile. Many ethical and value-based systems emphasize privacy as a human right. However, they frequently provide less data on specific technological controls. This results in an implementation gap between high-level ethical purpose and day-to-day engineering practice (Madhavan et al., 2025).

3.2. Mapping to Cybersecurity and Risk Models

AI-GRC frameworks complement existing cybersecurity and enterprise-risk models, but they also introduce new tensions. The NIST Cybersecurity Framework (CSF) and ISO 31000 complement the NIST AI RMF and ISO/IEC 42001's focus on lifecycle and risk management (Ispas et al., 2025). All of these promote systematic processes for identifying,

assessing, and treating risks. FAIR (Factor Analysis of Information Risk) provides quantitative risk models that can be used for AI. It is useful for estimating the frequency and amount of loss events. However, AI introduces novel failure mechanisms, including model inversion and data poisoning. These require updated threat modeling assumptions and new assessment terms. This is consistent with governance components in the EU AI Act and organizational AI standards (Aboy & Vayena, 2024).

There are some important integration points between AI-GRC and other frameworks. Organizations can utilize the NIST CSF or ISO 27001 as a baseline for security controls. They can then layer AI-specific risk processes from the NIST AI RMF and ISO 42001 on top. However, disagreements do arise. Prescriptive regulation, such as the EU AI Act, might impose compliance requirements that do not align with voluntary business practices. These responsibilities may also conflict with quantitative risk approaches, which understate sociotechnical harm. They may also contrast with data-sharing policies advocated by worldwide scientific groups (Orhan et al., 2025; Aboy & Vayena, 2024).

4. Risk and Compliance in AI Systems

4.1. AI-Specific Risk Categories

AI systems present distinct risk domains that governance and compliance programs must explicitly address:

Technical risks (model, data, infrastructure).

Models can fail in several ways, including hallucination, drift, and overfitting, which compromise reliability and accuracy. Data-related issues such as poor quality, biased sampling, and labeling errors further exacerbate these vulnerabilities, introducing systemic bias and reducing model robustness. Infrastructure adds another layer of complexity; weaknesses in compute isolation, insecure model-serving environments, and supply-chain dependencies create exploitable attack surfaces. Collectively, these risks can lead to operational failures and safety hazards once AI systems are deployed, underscoring the need for rigorous testing, independent validation, and hardened deployment pipelines to mitigate these threats.

These risks can cause operational failures and safety harms once models are deployed in real settings. They require technical controls such as robust testing, independent validation datasets, and hardened deployment pipelines (Suyono et al., 2025).

Ethical and societal risks.: Beyond malfunction, AI systems can incorporate or exacerbate social harm. Examples include discriminatory outcomes, invasions of privacy, loss of human autonomy, and the dissemination of misinformation. These risks are frequently context-specific. They advocate for cross-disciplinary mitigation, such as ethics reviews, stakeholder engagement, and fairness testing (Ratti & Morrison, 2025).

Regulatory and legal risks (liability, IP, privacy):. AI also poses legal and regulatory problems. These result from privacy violations such as the misuse of training data, intellectual property disputes involving training data or outputs, and uncertain accountability when autonomous systems cause harm. Deployers or "AI operators" are increasingly being held accountable under new legislation and court judgments. This necessitates documentation, access controls, and rigorous risk allocation through contracts (Ahmed et al., 2025).

4.2. Risk Assessment Methodologies

Model Risk Management (MRM):. MRM applies proven financial and engineering risk-management methodologies to AI models by inventorying models, categorizing criticality, defining acceptable performance criteria, and establishing governance for model creation, testing, and deployment. Recent guidance (both regulatory and industrial) stresses lifecycle MRM for generative and large-scale models, such as independent validation and periodic reassessment following retraining or distributional shifts (Mökander et al., 2023).

Threat modeling for AI systems.: Beyond traditional IT dangers, threat modeling includes adversarial ML (poisoning, evasion), model-extraction assaults, and misuse scenarios (wrongful automation). Effective threat modeling combines security engineering and domain knowledge to identify plausible adversary goals, system attack surfaces, and mitigations (robust training, anomaly detection, rate limitations). Recent research suggests incorporating ethical misuse scenarios into threat models to capture non-technical damages as part of a comprehensive risk assessment (Madhavan et al., 2025).

Continuous risk monitoring: Static pre-deployment checks are insufficient for systems that learn and interact in dynamic contexts. Continuous monitoring of inputs, model outputs, performance indicators, and security telemetry allows for early discovery of conceptual drift, fairness regressions, and operational anomalies. Organizations should provide pipelines with automatic warning, frequent revalidation, and human-in-the-loop escalation methods (Paidy, 2023).

4.3. Compliance Mechanisms

Auditing and transparency tools.: Automated audit tools like logging, provenance capture, and explainability modules collaborate with manual audit processes like red-team exercises and ethics reviews. Together, they provide the evidence base that regulators and certifiers require. Model cards, datasheets for datasets, and data assertions are now widely recognized as best practices. They contribute to accountability, transparency, and repeatability in AI systems (Mökander et al., 2023; Filani & Opoku, 2025).

Certification schemes and standards.: Standards and certification channels for AI are rapidly expanding. ISO/IEC 42001 specifies a formal management-system approach to trustworthy AI. Other programs and vendor-led verification schemes, such as early industry AI assurance efforts and sector-specific seals, offer further audits and certifications. Certification can improve assurance, but organizations should not treat it as a checkbox exercise. They should also verify that certification scopes match their actual operating risks (Torres & Ali-Vehmas, 2025).

Regulatory compliance and documentation requirements: Regulations are progressively requiring auditable records of dataset provenance, development decisions, validation outcomes, and incident responses. Maintaining such documents benefits not only legal defense, but also continual improvement and external audits. Tools that automate portions of this paperwork decrease burden while improving fidelity (Pelc, P., 2024).

5. Integrating Risk and Compliance into the AI Lifecycle

A practical governance approach is to map controls across the AI lifecycle (Data → Model → Deployment → Monitoring).

Table 2 Summary of common risks and representative controls for each stage of the AI lifecycle, synthesized from guidance in NIST, ISO, and industry MRM publications (AI, N. 2024; Mitchell et al., 2019; Ispas et al., 2025).

Lifecycle Stage	Typical Risks	Representative Controls
Data	Bias, privacy violations, poor provenance	Data inventories, datasheets, access controls, consent management.
Model Development	Overfitting, lack of robustness, opaque behavior	Model cards, validation suites, adversarial testing, and versioning.
Deployment	Unsecure serving, misuse, and scalability failures	Hardened serving, RBAC, rate limits, contract & SLA controls.
Monitoring & Maintenance	Concept drift, fairness regressions, and incidents	Continuous monitoring, performance/regression tests, and incident playbooks.

AI risk and compliance are both transdisciplinary and dynamic. Standards (ISO 42001), government frameworks (NIST), and industry MRM guidelines all rely on lifecycle thinking, transparency artifacts (model cards, datasheets), and continuous assurance as fundamental building elements. However, major gaps exist, particularly in terms of enforceable liability regimes, auditability of black-box models, and scalable independent verification; thus, businesses must combine standards adoption with pragmatic technical controls and legal/ethical oversight (Rezaei, M., 2025; Birkstedt et al., 2023).

5.1. Challenges in Implementing AI GRC Frameworks

Implementing governance, risk, and compliance (GRC) frameworks for artificial intelligence is a multidimensional challenge involving technical, organizational, legal, ethical, and interoperability issues. These challenges often interact and amplify each other, making effective AI governance difficult in practice. The following subsections examine these challenges, supported by tables that summarize key insights from existing literature.

5.2. Technical Challenges

Despite an increase in tools intended for AI audits, the landscape remains immature, fragmented, and lacking cohesive standards. A study by Ojewale et al. (2024) assessing 390 AI auditing tools revealed that most tools fail to support critical stages such as harm discovery, advocacy, and continuous evaluation. This widespread tooling gap weakens organizations' ability to ensure accountability and compliance across an AI system's lifecycle. Unlike traditional software systems, AI models continue to evolve after deployment, influenced by new data or environmental drift. This makes static governance models ineffective because model behavior at deployment may not reflect behavior during later operation. Rezaei, M. (2025) emphasizes that unless governance frameworks account for drift, silent degradation may occur without detection.

5.3. Organizational Challenges

Organizational barriers like skills gaps, siloed teams, and audit costs significantly weaken AI governance structures (Dodds et al., 2025). These factors can undermine even the most sophisticated technical frameworks. Most organizations lack personnel with hybrid expertise in machine learning, ethics, risk assessment, and regulatory compliance. Kokina and Davenport (as cited in Alsaif et al., 2023) emphasize that traditional auditors and compliance officers lack foundational knowledge in AI interpretability, making effective oversight challenging. AI governance requires alignment between data science, operations, compliance, cybersecurity, and legal teams. However, many organizations still operate in silos, creating fragmentation. A study in *AI and Ethics* found that most governance teams prioritize operational risk but fail to incorporate ethical or social impact considerations (Batoool et al., 2023; Papagiannidis et al., 2025). Comprehensive AI governance, such as regular audits, fairness testing, documentation, and human-in-the-loop reviews, demands significant financial resources. OECD (2024) reports that cost is among the top barriers to AI oversight in governments and small enterprises.

5.4. Legal and Policy Uncertainty

Legal ambiguity creates structural challenges for AI GRC implementation. Rapidly evolving regulations, cross-border data constraints, and inconsistent definitions of "AI system" complicate compliance. Organizations that operate internationally confront inconsistent rules, for example, differing definitions of high-risk AI between the EU AI Act and frameworks in the United States. Alsaigh et al. (2025) highlight that such inconsistencies create regulatory arbitrage opportunities and uncertainty in compliance strategies. AI models often depend on large, diverse datasets, yet many jurisdictions impose data localization requirements that constrain data movement. This creates friction for model training, inference, and governance monitoring. Oluoch, I. (2024) notes that without a unified definition of AI, regulation varies across sectors and jurisdictions, making harmonization extremely difficult.

5.5. Ethical and Societal Challenges

Even where technical and legal controls exist, ethical AI governance requires deliberate, continuous engagement with fairness, human oversight, and accountability principles. Bias in AI emerges from data, algorithms, or broader systemic inequalities. A review by Batoool et al. (2023) found that fairness is often sidelined in governance, with organizations focusing more on risk or performance metrics. Opaque AI decision-making obscures responsibility. In high-risk settings such as credit scoring, healthcare, and policing, unclear accountability can directly lead to harm with limited pathways for redress. Ensuring meaningful human oversight requires more than a human "in the loop"; it requires well-trained experts empowered to override automated systems. Many organizations have not yet institutionalized this structure (Ratti & Morrison, 2025; Karimian et al., 2022).

5.6. Interoperability Issues

Lack of interoperability is one of the most persistent systemic challenges for AI GRC implementation. Governance frameworks (e.g., NIST AI RMF, ISO/IEC 42001, EU AI Act, OECD guidelines) often use different terminology, risk taxonomies, lifecycle definitions, and control requirements (Aboy & Vayena, 2024). Alsaigh et al. (2024) emphasize that these incompatibilities create inefficiencies and increase compliance burdens. No global consensus exists on metrics for fairness, robustness, accuracy, explainability, or trustworthiness. As a result, organizations often build ad-hoc evaluation frameworks that are not comparable across industries.

6. Best practices and integration strategies

Academic research and early adopters' experiences are all contributing to the development of best practices for AI governance, risk, and compliance.

Instead of developing entirely new governance structures, the most effective approaches build upon existing GRC architectures. Examples include ISO 27001–based information security management systems and enterprise risk frameworks. On top of these, organizations implement AI-specific processes and controls (Paidy, 2023). According to recent reviews, responsible AI cannot be realized solely through ethical ideals. It also necessitates firm organizational structures, defined lifecycle processes, and quantifiable metrics. These components must be integrated into regular development and operating workflows (Shetty et al., 2025).

6.1. Governance design principles

At the governance level, organizations are increasingly formalizing accountability for AI systems. Empirical research on responsible AI governance demonstrates a transition from informal, project-level ethics debates to organized arrangements that include board monitoring, designated leadership roles, and cross-functional committees (Pelc, P., 2024). According to Papagiannidis et al., leading organizations form AI or data ethics committees that include technical, legal, risk, and business stakeholders and have explicit mandates to examine high-risk use cases, grant exceptions, and monitor rehabilitation. According to Radanliev, P. (2025), trustworthy AI principles are only effective when they are linked to concrete governance needs like role descriptions, escalation procedures, and documentation tasks.

Alignment with recognized AI governance standards is another developing trend.

Organizations with established risk and compliance functions frequently use the NIST AI Risk Management Framework or ISO/IEC 42001 as reference models. They then integrate these into their current GRC structures. In practice, this entails applying AI-specific controls and processes to existing risk registers, policy frameworks, and assurance activities (Mitchell et al., 2019). It does not imply creating a distinct AI silo (Suyono et al., 2025). At the same time, many organizations accept high-level, trustworthy AI guidelines developed by organizations such as the OECD, UNESCO, and the EU High-Level Expert Group. They apply these concepts as normative anchors. They then convert these into internal conduct and policy statements that inform AI project decisions (Marwala, 2024).

6.2. Embedding GRC in AI development

The second set of best practices focuses on integrating GRC into AI development pipelines. The goal is to avoid viewing governance as something that is only reviewed at the end. Technical and organizational guidance now agree on "security by design" and "ethics by design" as essential principles. In reality, this entails including security needs, adversarial risk assessments, privacy measures, and bias evaluation at each level (Paidy, 2023, Papagiannidis et al., 2025). These stages are: data gathering, feature engineering, model selection, training, and deployment. Ahangar et al.(2025) suggest that trustworthy AI in industrial contexts necessitates robustness and explainability assessments throughout the pipeline, rather than only upon deployment. They also suggest that governance approaches should be consistent with common technical practices like version control and continuous integration (Egele et al., 2024).

AI system risk grading and tiering are becoming more prevalent. According to Papagiannidis et al.,(2025), some organizations categorize AI systems into risk levels based on domain, individual impact, and business process dependency, and then tie governance requirements such as mandatory impact assessments, independent review, or post-deployment monitoring to these levels. This rationale is consistent with the EU AI Act's risk-based frameworks and financial model risk management, which require stricter regulations for higher-risk systems (Aboy & Vayena, 2024). Documented development patterns and artifacts contribute to AI-GRC evidence being available by default. Work on model cards, datasheets for datasets, and algorithmic impact evaluations demonstrates that common templates can help teams. These templates allow them to consistently communicate the model's purpose, training data, limits, and potential harms (Alsaigh et al., 2024; Egele et al., 2024).

When these artifacts are integrated into MLOps workflows, they are generated as part of routine operations. They may be generated during code review, deployment, or monitoring.

Over time, they accumulate a reusable compliance record for audits and regulatory inquiries. Microsoft and IBM's governance white papers describe comparable responsible AI documentation standards. Developers complete standard forms. Central governing bodies then assess these forms as part of their oversight procedure (Egele et al., 2024; Papagiannidis et al., 2025).

6.3. Audit and assurance practices

Best assurance practices are shifting away from periodic, static audits and toward more continuous, risk-sensitive procedures. Studies on AI deployment in regulated industries demonstrate that periodic, checklist-based inspections are insufficient for systems that evolve and function in complicated contexts. Instead, organizations are starting to track crucial indications in near real time (Wijaya et al., 2025). These include performance degradation, distribution shifts, abnormality rates, and user complaints. They then incorporate these indicators into risk registers and incident management procedures. Ahangar et al.(2025) state that in industrial settings, continual monitoring of model drift and resilience is critical for safety and quality assurance. They propose tool-supported pipelines that initiate retraining or mitigation when predetermined thresholds are met (Mökander et al., 2023).

Transparency or responsibility reports are another emerging assurance practice, especially among large technology firms and public-sector bodies. These reports describe where and how AI systems are used, summarize risk assessments, and outline mitigation and oversight mechanisms. Radanliev (2025) highlights such reporting as a way to turn abstract ethical commitments into observable institutional behavior, but also warns that it must be linked to verifiable processes or it risks becoming just a public show of ethics with no real substance

Independent or third-party assurance is becoming increasingly popular for high-risk AI systems. ISO/IEC 42001 specifies a standard for a certifiable AI management system. It enables businesses to demonstrate that they have organized AI governance processes in place. At the same time, organizations like AI Verify offer test-and-report platforms. These can serve as external benchmarks for claims of robustness, justice, and accountability (Suyono et al., 2025). Empirical examination of these mechanisms remains restricted. However, the current trend is to combine strong internal control with some type of external evaluation or benchmarking. This is especially true in situations when regulation or market pressure necessitates increased confidence (Torres & Ali-Vehmas, 2025; Moiseev et al., 2023).

7. Conclusion

This review has shown that AI governance, risk, and compliance are evolving quickly but remain fragmented. Global principles from bodies such as the OECD and UNESCO, technical standards from ISO and NIST, regional regulations such as the EU AI Act, and industry frameworks from major vendors all contribute important pieces. Yet they differ in scope, enforceability, and depth of technical guidance, especially around adversarial robustness, socio-technical harms, and liability. A cross-cutting finding is that effective AI-GRC builds on existing cybersecurity and enterprise-risk frameworks rather than replacing them. Mapping AI-specific controls onto known structures such as ISO 27001, NIST CSF, and FAIR enables more coherent governance. Best practices also converge on “by design” and “by default” approaches: integrating security, ethics, and documentation into AI pipelines, using standard artefacts in MLOps, and shifting assurance toward continuous, risk-sensitive monitoring with selective third-party validation. Future work should focus on harmonized audit standards, shared benchmarks for AI risk metrics, and integration of AI risk into ESG and board-level reporting. Without such consolidation, organizations will struggle to govern AI systems in a way that is both responsible and operationally practical.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

- [1] Aboy, M., Minssen, T., & Vayena, E. (2024). Navigating the EU AI Act: implications for regulated digital medical products. *NPJ Digital Medicine*, 7(1), 237.
- [2] Ahangar, M. N., Farhat, Z. A., & Sivanathan, A. (2025). AI trustworthiness in manufacturing: challenges, toolkits, and the path to industry 5.0. *Sensors*, 25(14), 4357.
- [3] Ahmed, Z., Filani, A., Osifowokan, A. S., & Hutchful, N. (2025). The Impact of Data Breaches in US Healthcare: A Cost-Benefit Analysis of Prevention vs. Recovery. <https://doi.org/10.30574/wjarr.2025.27.3.3274>
- [4] AI, N. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile. *NIST Trustworthy and Responsible AI Gaithersburg, MD, USA*.

- [5] Alsaigh, R., Mehmood, R., Katib, I., Liang, X., Alshanqiti, A., Corchado, J. M., & See, S. (2024). Harmonizing AI governance regulations and neuroinformatics: perspectives on privacy and data sharing. *Frontiers in Neuroinformatics*, 18, 1472653.
- [6] Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: a systematic literature review. *AI and Ethics*, 1-15.
- [7] Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: themes, knowledge gaps and future agendas. *Internet Research*, 33(7), 133-167.
- [8] Dodds, T., Vandendaele, A., Simon, F. M., Helberger, N., Resendez, V., & Yeung, W. N. (2025). Knowledge silos as a barrier to responsible AI practices in journalism? Exploratory evidence from four Dutch news organisations. *Journalism Studies*, 26(6), 740-758.
- [9] Egele, R., Junior, J., CS, J., van Rijn, J. N., Guyon, I., Baró, X., ... & Wan, J. (2024). Ai competitions and benchmarks: Dataset development. *arXiv preprint arXiv:2404.09703*.
- [10] European Commission High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. Publications Office of the European Union.
- [11] Filani, A. and Opoku, J. A. "Protecting Electronic Health Records (EHRs): Advances and Challenges in Data Security and Privacy." *Sarcouncil Journal of Engineering and Computer Sciences* 4.12 (2025): pp 1-9.
- [12] Filani, A., Addotey, N., & Opoku, J. A. Cybersecurity Threat Landscape in the US Healthcare Sector: Trends, Risks, and National Implications. <https://doi.org/hbpwnq>
- [13] Hamdouni, A. (2025). The role of artificial intelligence in enhancing ESG performance: Evidence from Saudi listed firms. *Journal of Risk and Financial Management*, 18(10), 572. <https://doi.org/10.3390/jrfm18100572>
- [14] Ispas, L., Mironeasa, C., Severin, T. L., Cerlincă, D. A., & Mironeasa, S. (2025). Artificial Intelligence Applications in Risk Management Within Integrated Management Systems: A Review. *Systems*, 13(11), 967.
- [15] Karimian, G., Petelos, E., & Evers, S. M. (2022). The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. *AI and Ethics*, 2(4), 539-551.
- Madhavan, K., Yazdinejad, A., Zarrinkalam, F., & Dehghantanha, A. (2025). Quantifying Security Vulnerabilities: A Metric-Driven Security Analysis of Gaps in Current AI Standards. *arXiv preprint arXiv:2502.08610*.
- Marwala, T. (2024). Framework for the Governance of Artificial Intelligence.
- [16] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. Model cards for model reporting, Proceedings of the conference on fairness, accountability, and transparency, 2019, pp. 220–229.
- [17] Moiseev, I., Balabaeva, K., & Kovalchuk, S. (2025). Open and Extensible Benchmark for Explainable Artificial Intelligence Methods. *Algorithms*, 18(2), 85.
- [18] Mökander, J., Schuett, J., Kirk, H. R., & Floridi, L. (2023). Auditing large language models: a three-layered approach. AI and Ethics. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (Vol. 1, pp. 4275-4293).
- Nastoska, A., Jancheska, B., Riziniski, M., & Trajanov, D. (2025). Evaluating trustworthiness in AI: Risks, metrics, and applications across industries. *Electronics*, 14(13), 2717.
- [19] Ojewale, V., Steed, R., Vecchione, B., Birhane, A., & Raji, I. D. (2025, April). Towards AI accountability infrastructure: Gaps and opportunities in AI audit tooling. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-29).
- [20] Oluoch, I. (2024). Crossing boundaries: The ethics of AI and geographic information technologies. *ISPRS International Journal of Geo-Information*, 13(3), 87.
- [21] Orhan, Z., Orhan, M., Lund, B. D., Mannuru, N. R., Bevara, R. V. K., & Porter, B. (2025). Artificial Intelligence Standards in Conflict: Local Challenges and Global Ambitions. *Standards*, 5(4), 27.
- [22] Paidy, P. (2023). Leveraging AI in Threat Modeling for Enhanced Application Security. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 4(2), 57-66.
- [23] Panful, B., Apafllo, B., Filani, A., Nnadi, K., & Hutchful, N. (2025). Human factor vulnerabilities in energy industry cybersecurity: Assessing employee awareness and behavior in breach prevention. *International Journal for Multidisciplinary Research (IJFMR)*, 7(6), 1-18. <https://doi.org/10.36948/ijfmr.2025.v07i06.63932>
- [24] Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, 34(2), 101885.

- Pelc, P. (2024). Cybersecurity issue in the executive order on the safe, secure, and trustworthy development and use of artificial intelligence from october 30, 2023. *Cybersecurity and Law*, 11(1), 47-57.
- Radanliev, P. (2025). AI ethics: Integrating transparency, fairness, and privacy in AI development. *Applied Artificial Intelligence*, 39(1), 2463722.
- [25] Ratti, E., Morrison, M., & Jakab, I. (2025). Ethical and social considerations of applying artificial intelligence in healthcare—a two-pronged scoping review. *BMC Medical Ethics*, 26(1), 68.
- [26] Rezaei, M. (2025). Artificial intelligence in knowledge management: Identifying and addressing the key implementation challenges. *Technological Forecasting and Social Change*, 217, 124183.
- [27] Shetty, D. K., Arjunan, R. V., Cenitta, D., Makkithaya, K., Hegde, N. V., Salu, S., ... & Pallela, P. K. (2025). Analyzing AI Regulation through Literature and Current Trends. *Journal of Open Innovation: Technology, Market, and Complexity*, 100508.
- [28] Suyono, W. P., Puspa, E. S., Anugrah, S., & Firnanda, R. (2025). Artificial Intelligence in Auditing: A Systematic Review of Tools, Applications, and Challenges. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 3393-3401.
- [29] Torres, A. P. G., & Ali-Vehmas, T. (2025). AI regulation: maintaining interoperability through value-sensitive standardisation. *Ethics and Information Technology*, 27, 26.
- [30] Wijaya, J. R. T., Prasetyo, I., Rahmatika, D. N., & Indriasih, D. (2025). ARTIFICIAL INTELLIGENCE AND AUDIT QUALITY: AN EMPIRICAL LITERATURE REVIEW FROM SCOPUS DATABASE. *Fokus Ekonomi: Jurnal Ilmiah Ekonomi*, 20(1), 61-76.