RAIS Journal for Social Sciences | Vol. 10, No. 1, 2026
ISSN 2574-0245 (Print) | ISSN 2574-1179 (Online) | DOI:10.5281/zenodo.19657276

Designing an AI Governance Program: A Control-Based Model for Risk and Compliance

Miranda Stanfield

Capitol Technology University, Laurel, MD, USA
miranda.stanfield@gmail.com

Abstract: As organizations increasingly use artificial intelligence (AI) for organizational decision-making, cybersecurity, and compliance, the limits of principle-based AI governance have become clear. Frameworks have revealed limitations in principle-oriented AI governance approaches. While frameworks such as the NIST AI Risk Management Framework provide broad, high-level guidance, many organizations still lack practical, auditable mechanisms to operationalize AI governance within their enterprise governance, risk, and compliance (GRC) programs. This research introduces a control-based AI governance model that embeds AI oversight into existing internal controls and risk management structures. The model organizes governance through administrative, technical, and operational controls, including integrating AI risk assessment, compliance mapping, and continuous monitoring throughout the AI lifecycle. Governance controls are mapped to the NIST AI Risk Management Framework and NIST Special Publication 800-53 to demonstrate compatibility and traceability, without requiring demonstration of operational compatibility or mandating specific technologies. This study presents a practical, technology-neutral approach to help organizations implement AI governance and align it with their GRC efforts, promoting an agnostic approach that advances the operationalization of AI governance and its integration with enterprise GRC practices.

Keywords: AI Governance, Control-Based Governance, Artificial Intelligence Risk Management, Governance, Risk, and Compliance (GRC), NIST AI Risk Management Framework, NIST SP 800-53, AI Lifecycle Governance, Auditability and Assurance, Enterprise Risk Management

Introduction

The rapid deployment of artificial intelligence (AI) and automated decision systems across organizations has expanded existing risks and introduced new vulnerabilities with operational, financial, legal, and societal consequences. Cybersecurity incidents, including data breaches and cyberattacks, compromise sensitive information, disrupt critical services, and impose substantial direct and indirect costs, including long-term reputational damage (Liu & Babar, 2024; Romanosky, 2016). Empirical evidence indicates that persistent weaknesses in data security and repeated breaches reduce stakeholder trust and diminish brand value, even when short-term market responses are inconsistent (Kamiya et al., 2021). Concurrently, organizations deploying AI for analytics, recruitment, and decision support face governance challenges related to algorithmic bias, limited transparency, and regulatory exposure. Studies have demonstrated that biased or opaque AI systems can produce discriminatory outcomes, particularly under data protection and anti-discrimination law regimes (Barocas & Selbst, 2016; Chen, 2023). Such systems have been shown to intensify existing inequities and threaten equitable access to services, underscoring the need for

explicit, auditable requirements for fairness, transparency, and accountability throughout the AI development and deployment lifecycle (Raji et al., 2020).

Beyond individual organizations, deficiencies in AI and cybersecurity governance reduce public trust, worsen structural inequalities, and raise concerns about rights and accountability, particularly in the public sector and high-stakes decision-making domains (Busuioc, 2021; Kuziemski & Misuraca, 2020; Manias et al., 2023). Despite growing recognition of these risks, existing governance approaches remain insufficient. Principles-based frameworks articulate values without specifying enforceable controls (Mittelstadt, 2019), while audit-oriented proposals focus on accountability mechanisms without providing operational guidance for integration into enterprise risk management (Mökander et al., 2021; Raji et al., 2020). Critically, no established model maps AI governance controls to existing organizational compliance infrastructure—such as Governance, Risk, and Compliance (GRC) frameworks—in a form that is both auditable and technology-agnostic. This gap leaves organizations without actionable guidance for managing AI-related risk within the structures they already operate.

This study addresses that gap by proposing a control-based AI governance model that is, to the authors' knowledge, the first to structure administrative, technical, and operational controls across the full AI lifecycle in direct alignment with both the NIST AI Risk Management Framework and NIST Special Publication 800-53. Unlike prior governance proposals that operate as parallel or advisory structures, the model is designed for integration within enterprise GRC frameworks, enabling AI-related risks to be managed through established organizational mechanisms and subject to existing audit and assurance functions. Governance controls are mapped to NIST standards to support traceability and auditability across implementation contexts without prescribing specific technologies. The remainder of this paper is organized as follows. Section 2 reviews existing literature on AI governance, cybersecurity risk management, and GRC frameworks to establish the theoretical foundation for the proposed model. Section 3 presents the control-based governance framework, including control taxonomy, lifecycle mapping, and GRC integration architecture. Section 4 discusses implementation considerations, practical implications for compliance and risk functions, and limitations of the proposed approach. Section 5 concludes with directions for future research.

Problem Statement

Cybercrime is estimated to cost the global economy USD 1 trillion in 2020, representing an increase of more than 50% since 2018 (Cremer et al., 2022), highlighting the growing consequences of cybersecurity risks that outpace existing governance and control mechanisms. Recent research indicates that, despite the growing sophistication of cyber threats, traditional cybersecurity governance, risk, and compliance (GRC) approaches remain heavily manual, siloed, and reactive, creating persistent gaps in risk identification and mitigation across modern digital infrastructure (Avianti et al., 2025; Birkstedt et al., 2023). The general problem is that many cybersecurity governance, risk, and compliance practices remain siloed and retrospective, relying on periodic reviews rather than continuous oversight, which limits an organization's ability to keep pace with AI-driven cyber risks and evolving regulatory demands. The specific problem is that organizations integrating AI and large language models into security and compliance operations often lack a unified, control-focused cybersecurity GRC framework that systematically governs AI-related risks and emerging legal obligations across the AI lifecycle. Although recent studies have examined AI-enhanced GRC, AI-driven compliance, and cybersecurity risk analysis for AI systems, the literature offers limited practical guidance on continuously monitored control-based models that unify the cybersecurity GRC and AI governance. This study addresses this gap by

proposing and evaluating an integrated, control-based cybersecurity GRC model tailored to govern the adoption of large language models.

Significance of the Study

This study is important because it addresses the major governance gap caused by the rapid adoption of AI and large language models in organizational cybersecurity, risk, and compliance. Although monitoring, threat detection, and control testing are becoming more automated, many organizations still rely on governance methods built for manual, periodic, and non-AI-driven environments. This mismatch increases the risk of security incidents, regulatory issues, and damage to reputation, as misalignment heightens exposure to security threats, regulatory noncompliance, and reputational damage. AI-specific risks such as model misuse, data leaks, unclear decisions, and bias can exceed the safeguards designed to control them, resulting in unintended outcomes like opaque decision-making and bias. From a research and scholarly standpoint, this study advances applied research at the intersection of AI governance, cybersecurity, risk, and compliance (GRC). Most existing literature often treats AI ethics, cybersecurity risk management, and regulatory compliance as separate topics and rarely emphasizes integrated, control-level governance.

This study fills that gap by offering a practical, scalable model that embeds AI governance within existing cybersecurity and GRC systems. It provides a framework for professionals to use in governance, risk, compliance, and assurance. The model helps organizations systematically identify AI-related risks and offers an implementation-ready framework for governance, risk, compliance, and assurance in practice. The proposed model supports the systematic identification of AI-related risks, the integration of emerging regulatory requirements into existing GRC programs, and the ongoing use of continuous monitoring and assurance mechanisms aligned with how AI systems operate. More broadly, this study provides guidance for regulators and standards bodies by illustrating how governance principles such as transparency, accountability, and fairness can be translated into tangible, auditable controls and enforceable practices. It responds to the call for unified, practical AI governance models that scale across different organizations and support various organizational contexts, thereby strengthening responsible AI deployment. More broadly, this study provides regulators and standards bodies with practical guidance by demonstrating how governance principles such as transparency, accountability, and fairness can be operationalized as real, auditable controls and enforceable practices. The research addresses the need for unified, practical, and scalable AI governance models that can adapt across organizations and promote responsible AI deployment.

Theoretical Framework

This study draws on multiple complementary theoretical perspectives to examine the implementation of organizational governance, risk management, and control within complex environments. Systems, institutional, and internal control theories are employed as conceptual lenses to analyze how governance mechanisms are structured, coordinated, and operationalized within organizations deploying artificial intelligence. These theories are discussed at a conceptual level to explain organizational behavior, cross-functional integration, and control alignment, without extending the technical system design. Collectively, they establish a coherent theoretical basis for analyzing AI governance as an organizational and control-based phenomenon.

Systems Theory

Systems theory provides a foundational lens for examining organizations as complex, interdependent structures rather than as collections of isolated components. Originating in

Bertalanffy's (1968) General System Theory, this approach emphasizes that system behavior and outcomes arise from interactions among constituent elements, and therefore requires holistic analysis rather than reductionist decomposition (Mulej et al., 2004; Ramage & Shipp, 2009). The core principles commonly associated with systems thinking include openness to environmental influences, dynamic interactions among subsystems, and regulation through feedback mechanisms rather than linear causality (von Bertalanffy, 1968; Mulej et al., 2004). This perspective emphasizes structural interdependence, adaptation, and equilibrium as the defining features of complex organizational systems.

In organizational research, systems theory conceptualizes organizations as sociotechnical systems composed of interconnected social and technical subsystems that must be jointly aligned to function effectively (Emery & Trist, 1960; Trist & Bamforth, 1951). From this standpoint, governance outcomes and risk exposure emerge from interactions among human actors, institutional arrangements, technologies, and operational processes rather than from discrete technical or procedural failures. Governance thus operates as a system-level function shaped by cross-functional dependencies and feedback cycles, which helps explain why formally defined policies or controls may fail to achieve the intended outcomes when systemic alignment is weak (von Bertalanffy, 1968; Ramage & Shipp, 2009).

In governance and enterprise risk management (ERM) scholarship, systems theory positions risk management as an integrated component of organizational control systems rather than as an isolated compliance activity. Prior research indicates that effective risk governance depends on coordination among governance structures, decision rights, internal controls, and monitoring mechanisms distributed across organizational levels (Jankensgård, 2019). System-oriented analyses further associate ERM maturity with enterprise-wide integration and continuous feedback processes, in contrast to compartmentalized or functionally siloed risk practices (Mikes & Kaplan, 2012, 2013).

Systems theory also informs analyses of cross-functional risk ownership by emphasizing that organizational risks arise through interactions across sociotechnical subsystems, rather than within individual functional domains. Sociotechnical frameworks describe how accountability for risk is distributed among interconnected roles, processes, and technologies, necessitating governance mechanisms that support coordination, communication, and shared oversight (Emery & Trist, 1960). This distributed conception of risk ownership is particularly relevant in complex organizations, where operational, compliance, legal, and technical functions are tightly coupled (Jankensgård, 2019; Mikes & Kaplan, 2013). In the context of artificial intelligence, systems theory supports conceptualizing AI governance as a system-level control function embedded within broader organizational governance and risk structures. AI-related risks arise from interactions among data governance practices, automated decision-making processes, human oversight mechanisms, and regulatory constraints, indicating the need for integrated, adaptive governance approaches (Taeihagh, 2021; Wirtz et al., 2022; NIST, 2023). System-oriented scholarship further frames AI oversight as a continuous governance process integrating design-time controls, operational monitoring, and institutional accountability, rather than as a discrete technical intervention (Taeihagh, 2021; NIST, 2023).

Overall, systems theory provides a coherent framework for examining AI governance as an organizational capability formed by systemic integration, control alignment, and adaptive oversight. By emphasizing interdependence, feedback mechanisms, and coordination across sociotechnical subsystems, this perspective explains why governance effectiveness depends on aligning structures, processes, and monitoring practices rather than on isolated technical solutions. As such, systems theory provides a rigorous foundation for analyzing AI-related risks as system-level phenomena that require integrated governance responses within enterprise risk and control architectures.

Institutional Theory

Institutional Theory explains the adoption of organizational governance structures primarily as a response to external legitimacy pressures rather than as an outcome of internal efficiency improvement. Contemporary institutionalist scholarship characterizes organizational governance practices as shaped by socially constructed rules, norms, and expectations that define acceptable conduct within institutional environments (Meyer & Rowan, 1977; Scott, 2014). Governance frameworks are often adopted to demonstrate conformity with prevailing institutional expectations and to secure legitimacy, even when their operational effectiveness has not been empirically established (Meyer & Rowan, 1977; DiMaggio & Powell, 1983).

Neo-institutional theory is commonly described as encompassing historical, sociological, and rational-choice strands, each offering complementary explanations of how governance structures emerge, persist, and diffuse across organizations (Scott, 2014). Historical institutionalism emphasizes path dependence, suggesting that governance arrangements tend to evolve incrementally and remain anchored to prior decisions rather than undergoing comprehensive redesign (Mahoney & Thelen, 2010). Sociological institutionalism focuses on the role of norms, professional standards, and shared meanings in forming acceptable governance practices, whereas rational choice institutionalism frames governance adoption as a strategic response to the incentives and constraints embedded within institutional environments (DiMaggio & Powell, 1983; Scott, 2014).

A core explanatory concept within Institutional Theory is institutional isomorphism, which accounts for convergence in organizational governance practices. The literature distinguishes between coercive isomorphism arising from regulatory mandates and stakeholder pressures, normative isomorphism associated with professional and industry standards, and mimetic isomorphism, which occurs when organizations emulate peers under conditions of uncertainty (DiMaggio & Powell, 1983). Prior empirical studies indicate that such pressures frequently result in symbolic adoption, in which formal governance structures are introduced to signal conformity, while substantive implementation remains limited or selectively applied (Meyer & Rowan, 1977; Okhmatovskiy & David, 2012).

The extent of the decoupling between formal governance structures and operational practices varies across sectors and national contexts. Cross-national research suggests that similar institutional expectations can produce divergent implementation outcomes depending on how governance structures mediate political, social, and professional pressures (Scott, 2014; Mahoney & Thelen, 2010). When external standards are voluntary, ambiguous, or costly to operationalize, organizations may rely on internally defined governance arrangements that satisfy visibility and signaling requirements while only partially addressing underlying institutional demands (Okhmatovskiy & David, 2012; Bromley & Powell, 2012). These institutional dynamics are especially relevant in enterprise governance, risk, and compliance environments, as well as in the governance of artificial intelligence. Institutional research indicates that emerging technologies introduced amid regulatory uncertainty and evolving professional norms generate heightened legitimacy pressures, encouraging organizations to demonstrate responsible and ethical deployment (DiMaggio & Powell, 1983; Taeihagh, 2021). Recent analyses of AI adoption further suggest that governance initiatives regularly emphasize principle-based commitments, such as ethics statements and transparency pledges, which function as legitimacy signals in the absence of embedded lifecycle controls, accountability mechanisms, and enforcement structures (Bromley & Powell, 2012; Reis & Pinheiro Junior, 2025).

Overall, Institutional Theory frames AI governance adoption as a response to external legitimacy pressures that influence organizational behavior independent of operational efficiency considerations. By distinguishing symbolic conformity from embedded governance, this theory helps to explain why principle-based AI commitments may satisfy institutional expectations without producing effective oversight or control integration. This

perspective reinforces the need to translate legitimacy-driven governance signals into enforceable control mechanisms inside existing governance, risk, and compliance structures.

Internal Control Theory

Internal control theory is commonly described as originating from early internal check practices designed to reduce errors and fraud through the separation of duties and cross-verification mechanisms. Early control systems emphasized distributing authority across individuals or functions to prevent any single actor from exercising unchecked control over business processes, thereby supporting basic accountability and operational reliability (Zhang, 2014). These foundational practices were later formalized as organizational processes intended to structure operations and limit improper or unauthorized activities.

Throughout the twentieth century, the internal control literature gradually expanded in scope as organizations confronted increasing operational complexity and regulatory expectations. Internal control frameworks evolved from narrow mechanisms focused on asset safeguarding and financial accuracy to broader systems encompassing operational effectiveness, risk management, and regulatory compliance. In the United States, legislative responses to major financial scandals accelerated this evolution, particularly through reforms that emphasized formalized documentation, management accountability, and independent evaluation of control effectiveness. The Sarbanes–Oxley Act of 2002 is widely regarded as a major milestone in this transition, strengthening the role of internal control as a governance mechanism rather than a purely accounting function (Heier et al., 2005). The Act introduced mandatory management assessments of internal control effectiveness and required independent auditor attestation, institutionalizing internal control as a core component of corporate governance and financial reporting accountability (Heier et al., 2005).

Subsequent contributions from standard-setting bodies further shaped modern internal control theory by integrating governance and risk management principles into comprehensive frameworks. The Committee of Sponsoring Organizations of the Treadway Commission (COSO) articulated internal control as a multidimensional system composed of interrelated components, including the control environment, risk assessment, control activities, information and communication, and monitoring (COSO, 2013). This framework represents a broader conceptual shift toward aligning internal controls with organizational governance objectives and evolving accountability expectations, and has since become the most widely adopted internal control framework globally (COSO, 2013; Ghyati & Kasbaoui, 2023). Scholarly reviews of the internal control literature further confirm that the concept has progressively shifted from a narrow financial reporting function toward a holistic governance instrument that addresses strategic, operational, and compliance objectives (Ghyati & Kasbaoui, 2023).

Overall, the literature characterizes internal control theory as progressing from discrete internal checks to integrated frameworks that support organizational objectives, risk management, compliance, and oversight. This evolution positions internal control as a core construct in contemporary governance systems, particularly in environments characterized by increasing regulatory scrutiny and operational complexity (Heier et al., 2005; COSO, 2013).

Methodology for the Systematic Literature Review

A systematic literature review was conducted to examine the research on AI governance, organizational risk, and compliance integration. Literature was retrieved using Capitol Technology University's virtual library resources on EBSCOhost, IEEE Xplore, ScienceDirect, ProQuest, and Google Scholar, ensuring interdisciplinary coverage of cybersecurity, governance, risk management, and information systems research. The review followed a structured evidence synthesis approach, emphasizing transparent retrieval,

screening, and thematic analysis. The searches employed targeted keywords and Boolean operators (AND/OR), including "AI governance," "artificial intelligence risk management," "AI compliance," "algorithmic accountability," "AI ethics frameworks, and" "governance, risk, and compliance (GRC)." The inclusion criteria prioritized peer-reviewed journal articles, conference proceedings, and recognized publications on governance or standards relevant to AI risk oversight, while non-scholarly or tangential sources were excluded. The selected studies were analyzed using thematic synthesis to identify dominant governance approaches, implementation challenges, and gaps related to operationalization. To support a control-based orientation, the findings were coded using a governance and control lens aligned with the NIST AI Risk Management Framework (AI RMF), NIST SP 800-53 control families, and NIST Cybersecurity Framework (CSF) 2.0. This mapping enabled differentiation between principle-based governance discussions and literature addressing concrete controls, monitoring mechanisms, and accountability structures across the AI lifecycle. This methodology is designed to reduce selection bias, increase reproducibility, and provide an empirical foundation for developing a control-focused AI governance model integrated into enterprise Governance, Risk, and Compliance programs.

Review of Literature

Existing scholarship on AI governance has expanded rapidly in response to the widespread adoption of automated and algorithmic systems in organizational operations. Early research primarily emphasized ethical principles, such as fairness, transparency, accountability, and explainability, framing AI governance as a normative or value-based concern. While this literature establishes important conceptual foundations, it often lacks guidance on how ethical principles translate into enforceable organizational practices (Mittelstadt, 2019). Subsequent studies have focused on regulatory guidance, standards, and policy frameworks developed by governments and standards bodies, emphasizing compliance obligations and oversight expectations (Birkstedt et al., 2023). However, governance is frequently addressed at a high level, with limited attention paid to internal control design, risk ownership, and integration with established GRC functions. Consequently, many of the proposed models remain descriptive rather than operational (Papagiannidis et al., 2025).

There is broad agreement in the literature that principle-based approaches alone are insufficient for managing AI-related risks (Mittelstadt, 2019; Birkstedt et al., 2023). Persistent challenges include fragmented governance structures, limited auditability, and the absence of control mechanisms embedded across the AI life cycle (Raji et al., 2020; Mökander et al., 2022). These gaps demonstrate the need for governance approaches that integrate AI oversight into existing risk and compliance infrastructure, which informs the rationale for a control-based AI governance model.

Proposed Control-Based AI Governance Model

Design Principles

The proposed control-based AI governance model is informed by recurring limitations identified in the AI governance literature, particularly the persistent gap between principle-oriented guidance and operationally enforceable governance mechanisms. To address this gap, the model was structured around three design principles: alignment with enterprise GRC structures, scalability and auditability, and technology-agnostic governance.

Alignment With Existing Enterprise GRC Structures

The first design principle emphasizes integration with established enterprise GRC and ERM structures. Prior research indicates that AI governance initiatives are more effective when

embedded within existing organizational risk and compliance architectures, rather than being implemented as standalone ethical or technical programs (Papagiannidis et al., 2025; Birkstedt et al., 2023). Embedding AI governance within enterprise GRC enables organizations to leverage mature control environments, define risk ownership, and establish assurance processes that are already used for cybersecurity, privacy, and regulatory compliance (Schneider et al., 2023; Mäntymäki et al., 2022). The literature further highlights the importance of explicitly assigning ownership to AI-related risks within enterprise risk frameworks to support accountability and continuous oversight (COSO, 2017; Mökander et al., 2022). However, empirical studies have cautioned that structural alignment alone does not ensure effective governance. Governance effectiveness depends on the extent to which AI-specific risks are translated into enforceable controls and incorporated into routine GRC activities such as risk assessments, control testing, and compliance reporting (Papagiannidis et al., 2025; Mökander et al., 2022).

Consistent with this principle, the proposed model is designed to be incorporated within established control-based frameworks, including the NIST AI Risk Management Framework and NIST Special Publication 800-53, without constraining the governance design to any single regulatory or technical standard.

Scalability and Auditability

The second design principle addresses scalability and auditability as the foundational requirements for operational AI governance. Because AI systems are deployed across multiple use cases, business units, and jurisdictions, governance mechanisms must support consistent oversight without relying on manual or ad hoc reviews. Empirical research has identified significant challenges in scaling principle-based AI governance frameworks, particularly in decentralized and regulated organizational contexts (Papagiannidis et al., 2025; Mökander et al., 2022; Birkstedt et al., 2023).

Research on AI assurance and continuous auditing demonstrates that embedding auditability into AI system lifecycles improves governance effectiveness by promoting systematic evidence collection and ongoing control verification (Minkkinen et al., 2022; Waltersdorfer et al., 2024; Raji et al., 2020). Control-by-design approaches that include standardized documentation artifacts, audit trails, and continuous monitoring support scalable assurance and reduce reliance on retrospective assessments (Stettinger et al., 2024; Schneider et al., 2023). Lifecycle-oriented governance further supports sustained oversight by consistently applying controls across the development, deployment, monitoring, retraining, and retirement stages (Lu et al., 2022; Mäntymäki et al., 2022).

Technology-Agnostic Governance

The third design principle highlights technology-agnostic governance to ensure its relevance across various AI models, vendors, and deployment scenarios. Scoping reviews and systematic analyses demonstrate convergence around governance approaches that are intentionally risk-based and technology-neutral, including frameworks such as the EU AI Act and NIST AI Risk Management Framework, which are designed to accommodate heterogeneous AI techniques and evolving use contexts (Papagiannidis et al., 2025; Birkstedt et al., 2023). The literature indicates that technology-agnostic governance is primarily achieved through lifecycle processes, organizational oversight structures, and standardized documentation practices rather than model-specific technical prescriptions (Lu et al., 2022; Schneider et al., 2023). These mechanisms enable governance structures to remain stable as underlying technologies evolve while highlighting the need for complementary control architectures that translate abstract principles into enforceable governance mechanisms (Papagiannidis et al., 2025; Mäntymäki et al., 2022).

Novel Contribution Relative to Existing NIST-Aligned Work

Although the NIST AI Risk Management Framework and NIST Special Publication 800-53 provide authoritative guidance on AI risk domains and control catalogs, they do not specify how AI governance controls should be officially organized within enterprise GRC programs. Consequently, organizations are left to operationalize NIST-aligned guidance independently across fragmented governance, risk, and assurance activities. The proposed model advances the existing NIST-aligned work by offering a control-based governance architecture that structures AI oversight around explicit design principles, rather than individual controls or compliance obligations. In doing so, the model functions as an organizing layer that enables consistent, auditable implementation of NIST-aligned controls without redefining or duplicating existing standards.

Governance Control Categories

The proposed control-based AI governance model operationalizes oversight through three complementary categories of governance control: administrative, technical, and operational. Drawing on internal control theory and organizational governance research, this categorization reflects how enterprises structure authority, assurance, and risk mitigation across complex socio-technical systems (COSO, 2017). As noted by Birkstedt et al. (2023), distinguishing control types is essential for moving AI governance from abstract principles to enforceable organizational practices.

Administrative Controls

Administrative controls establish a formal governance foundation for AI oversight by defining policies, accountability structures, and decision-making authorities. According to Schneider et al. (2023), organizational AI governance typically begins with policy instruments that specify acceptable use, documentation requirements, and escalation thresholds. These controls are reinforced through oversight bodies, such as governance committees or designated risk owners, which Papagiannidis et al. (2025) identify as necessary for managing cross-functional review and approval activities. Prior research has consistently emphasized that accountability mechanisms must be explicitly defined to prevent AI governance from remaining advisory. Mittelstadt (2019) argues that the absence of enforceable accountability leads to governance frameworks that lack institutional authority, a finding echoed by Birkstedt et al. (2023) in their analysis of symbolic AI ethics initiatives. Administrative controls, therefore, function as the primary mechanism for institutionalizing AI governance expectations and anchoring them within organizational authority structures (Mäntymäki et al., 2022).

Technical Controls

Technical controls operationalize governance requirements within data and model pipelines by enforcing constraints, monitoring system activity, and supporting validation activities. Raji et al. (2020) described technical governance mechanisms as essential for closing the accountability gap between high-level oversight and system-level implementation. These controls include data governance practices, such as access restrictions, lineage tracking, and dataset documentation, as well as model monitoring and testing processes designed to identify performance degradation and deviations from approved specifications (Lu et al., 2022). Unlike administrative controls, technical controls are embedded directly within system architectures and workflows. According to Minkkinen et al. (2022), embedded instrumentation is a prerequisite for continuous audit and evidence-based assurance. Studies on AI assurance further demonstrate that auditability depends on the availability of technical

artifacts that can support independent verification of control effectiveness, particularly in regulated environments (Waltersdorfer et al., 2024; Stettinger et al., 2024).

Operational Controls

Operational controls govern how AI systems are supervised and managed within day-to-day organizational processes. As described by Mäntymäki et al. (2022), operational governance mechanisms translate formal policies and technical constraints into routine practices such as human-in-the-loop reviews, incident response, and change management. These controls ensure that AI-enabled processes remain subject to organizational oversight throughout system use. Research on corporate AI governance highlights that operational controls are particularly important for managing emergent risks that arise after deployment. Mökander et al. (2022) noted that many governance failures occur during system operations, where insufficient procedures for escalation or intervention limit effective risk mitigation. By embedding governance requirements into operational workflows, these controls support timely incident responses and the evolution of control systems over time (Lu et al., 2022).

Control Category Coherence

Collectively, administrative, technical, and operational controls form an integrated governance system that supports enforceable and auditable AI oversight. Consistent with COSO's internal control framework, administrative controls establish governance intent, technical controls generate verifiable evidence, and operational controls enable active risk management during system use (COSO, 2017). This structured categorization provides a coherent foundation for subsequent control mapping and enterprise integration, as discussed in the following section.

Integration with Enterprise GRC

Effective AI governance requires integration with enterprise GRC functions to ensure that AI-related risks are identified, owned, monitored, and assured using established organizational mechanisms. Prior research emphasizes that governance frameworks operating outside enterprise GRC structures often lack enforceability, auditability, and sustained oversight, particularly in regulated environments (Papagiannidis et al., 2025; Birkstedt et al., 2023). Building on previously defined control categories, this section describes how the proposed model embeds AI governance into enterprise risk assessment, compliance mapping, and continuous monitoring and assurance processes.

Risk Assessment and Risk Ownership

Integration with enterprise GRC begins with incorporating AI-related risks into formal risk assessments and risk ownership processes. According to COSO's enterprise risk management framework, risks must be systematically identified, assessed, and assigned to accountable owners to support effective governance and decision making (COSO, 2017). Consistent with this approach, the proposed model treats AI risks as part of the enterprise risk portfolio, rather than as a separate or exceptional category. The literature highlights that unclear ownership of AI risks is a recurring governance failure, often resulting in fragmented oversight across technical, legal, and business functions (Mökander et al., 2022; Birkstedt et al., 2023). By integrating AI risk assessments into enterprise GRC workflows, organizations can assign ownership to designated risk owners responsible for oversight throughout the AI lifecycle. This approach aligns AI governance with existing risk escalation, review, and reporting mechanisms, enabling consistent treatment of AI risks alongside cybersecurity, privacy, and operational risks (Schneider et al., 2023; COSO, 2017).

Compliance Mapping

Compliance mapping is an essential part of the connection between AI governance and enterprise GRC functions. As regulatory and standards-based expectations for AI governance expand, organizations face increasing challenges in translating high-level requirements into operational controls (Birkstedt et al., 2023). Prior studies note that many organizations rely on informal or ad hoc compliance interpretations, which limit transparency and assurance (Papagiannidis et al., 2025). The proposed model addresses this challenge by supporting structured compliance mapping that links AI governance controls to applicable regulatory, legal, and standards-based requirements. According to Schneider et al. (2023), effective AI governance requires traceability among the governance objectives, implemented controls, and external obligations. By embedding compliance mapping into enterprise GRC tooling and processes, organizations can maintain a documented relationship between AI use cases, control implementations, and compliance obligations without duplicating governance structures. This integration supports consistent reporting and reduces the risk of misalignment between governance intention and regulatory expectations.

Continuous Monitoring and Assurance

Continuous monitoring and assurance form the final integration layer between AI governance and enterprise GRC. Research on AI assurance has consistently emphasized that periodic or static reviews are insufficient for systems that evolve over time through retraining, data changes, or operational drift (Minkkinen et al., 2022; Waltersdorfer et al., 2024). Consequently, the effectiveness of governance depends on the ability to generate ongoing evidence of control performance. Within the proposed model, continuous monitoring mechanisms are aligned with enterprise GRC assurance processes, including internal audits, compliance testing, and control assessments. Raji et al. (2020) argued that auditability requires traceable artifacts across the AI lifecycle, whereas Stettinger et al. (2024) highlighted the importance of structured assurance processes for high-risk AI systems. Embedding these mechanisms into enterprise GRC enables organizations to evaluate AI governance controls using established assurance methodologies rather than creating parallel audit processes.

The table introduced in this section illustrates how AI governance controls map to enterprise GRC processes across the risk assessment, compliance, and assurance functions. These artifacts provide a structured representation of integration points and support consistent implementation across organizational contexts.

Control Mapping to NIST Frameworks

Purpose of the Control Mapping

The purpose of control mapping is to demonstrate how the proposed control-based AI governance model can be operationalized within established, authoritative governance and risk management frameworks without redefining or duplicating existing standards. While prior sections establish the model's design principles, control categories, and integration with enterprise GRC, this subsection delivers a structured mechanism for translating these elements into implementable governance artifacts. This mapping aligns the governance control categories with the NIST AI Risk Management Framework and NIST Special Publication 800-53 to support traceability, auditability, and consistent interpretation throughout organizational contexts. This alignment validates that the proposed model is compatible with widely adopted risk-based and control-based frameworks, while remaining independent of any single regulatory regime.

Method for Mapping Controls

Control mapping was developed using a structured, deductive approach established in the internal control theory and risk-based governance principles. Rather than mapping specific technical implementations, the analysis focuses on controlling intent and governance functions. Each governance control category defined in the previous sections was operationalized into control objectives associated with oversight, accountability, and assurance across the AI lifecycle. These control objectives were mapped to the functional domains of the NIST AI Risk Management Framework, focusing on governance, risk identification, risk analysis, and risk management activities. In parallel, control objectives were mapped to the relevant NIST SP 800-53 control families based on their primary governance functions, including planning, risk assessment, audit and accountability, system protection, and continuous monitoring. Mapping reflects functional alignment rather than one-to-one correspondence, recognizing that individual governance controls may support multiple framework functions, depending on the organizational context.

Introduction to the Control Mapping Table

Table 1 presents the consolidated control mapping across four dimensions: governance control category, control objective, alignment with NIST AI Risk Management Framework functions, and alignment with NIST SP 800-53 control families. The fifth dimension identifies the primary AI lifecycle phase associated with each control objective to support lifecycle-based governance and assurance. The table was designed as an operational reference rather than as a prescriptive checklist. This illustrates how governance intent can be translated into recognizable control constructs, while allowing flexibility in implementation. By presenting mapping at the control objective level, the table avoids dependence on specific technologies, vendors, or system architectures.

How to Read and Use the Table

The table should be read from left to right, beginning with the governance control category that establishes the nature of control. The control objective column clarifies the governance purpose, independent of the implementation detail. The subsequent columns indicate alignment with the NIST AI Risk Management Framework functions and NIST SP 800-53 control families, enabling organizations to situate AI governance controls within existing risk management and assurance structures. This table can support governance design, gap analysis, and assurance activities. Governance and compliance teams may use mapping to assess AI-related risk coverage, whereas audit and assurance functions may use it to support control testing and evidence alignment. From a research perspective, the table functions as an instantiation artifact that demonstrates how the proposed model bridges conceptual AI governance principles with established control frameworks and provides a foundation for future empirical examination.

Table 1 presents the mapping of the proposed governance control categories to the NIST AI Risk Management Framework and NIST Special Publication 800-53 control families. This illustrates how administrative, technical, and operational control objectives align with the established risk management and assurance structures throughout the AI life cycle. The mapping demonstrates the operational compatibility of the proposed model with authoritative governance frameworks, while preserving implementation flexibility and avoiding dependence on specific technologies or deployment architectures.

Table 1. Stanfield's Control-Based AI Lifecycle Mapping of Governance Control Categories to NIST AI Risk Management Framework Functions and NIST SP 800-53 Control Families

Control Category	Control Objective	NIST AI RMF Function	NIST SP 800-53 Control Family	AI Lifecycle Phase	Example Governance Artifacts
Administrative	Establish a formal enterprise AI governance policy and scope	GOVERN	PM (Program Management), PL (Planning)	Strategy & Design	AI Governance Charter; Enterprise AI Policy
Administrative	Define AI accountability, decision rights, and risk ownership	GOVERN	PM (Program Management)	Strategy & Design	AI RACI Matrix; Governance Role Definitions
Administrative	Establish third-party AI risk governance policies	GOVERN	SR (Supply Chain Risk Management), SA (System & Services Acquisition)	Procurement	Vendor AI Risk Review Checklist; Third-Party Risk Policy
Administrative	Define enterprise AI audit and governance review cadence	GOVERN	CA (Assessment, Authorization & Monitoring), PM (Program Management)	Operations	AI Governance Review Plan; Audit Schedule
Administrative	Require AI risk assessments prior to deployment	MAP	RA (Risk Assessment), CA (Assessment, Authorization & Monitoring)	Development	AI Risk Assessment Report; Risk Register Entry
Administrative	Categorize AI systems by risk level, use context, and affected population	MAP	RA (Risk Assessment), PL (Planning)	Strategy & Design	AI System Inventory; Risk Categorization Register
Technical	Identify and assess potential for discriminatory or disparate impact outcomes	MAP	RA (Risk Assessment), SI (System & Information Integrity)	Development	Bias Risk Assessment; Fairness Evaluation Report
Technical	Map data provenance, lineage, and quality risks for training datasets	MAP	SA (System & Services Acquisition), SR (Supply Chain Risk Management)	Data Preparation	Data Lineage Map; Training Data Quality Report
Technical	Validate AI models prior to production release	MEASURE	CA (Assessment, Authorization & Monitoring), SA (System & Services	Development	Model Validation Report; Testing Summary

Control Category	Control Objective	NIST AI RMF Function	NIST SP 800-53 Control Family	AI Lifecycle Phase	Example Governance Artifacts
			Acquisition)		
Technical	Monitor model performance and detect drift	MEASURE	SI (System & Information Integrity), CA (Assessment, Authorization & Monitoring)	Operations	Drift Monitoring Dashboard; Performance Metrics Report
Technical	Log AI system activity for traceability and auditability	MEASURE	AU (Audit & Accountability)	Operations	Audit Logs; Monitoring Dashboard
Operational	Conduct periodic AI governance and risk reviews	MEASURE	CA (Assessment, Authorization & Monitoring), RA (Risk Assessment)	Operations	Governance Review Report; Risk Reassessment Summary
Operational	Enforce access controls on AI training and operational data	MANAGE	AC (Access Control), IA (Identification & Authentication)	Data Preparation / Operations	Access Control Matrix; Privileged Access Review
Operational	Protect AI models from unauthorized modification	MANAGE	CM (Configuration Management), SI (System & Information Integrity)	Operations	Configuration Baseline; Change Control Log
Operational	Execute AI incident response playbooks	MANAGE	IR (Incident Response), SI (System & Information Integrity)	Operations	AI Incident Playbook; After-Action Report
Operational	Retire AI systems exceeding defined risk tolerance	MANAGE	SA (System & Services Acquisition), CM (Configuration Management), MP (Media Protection)	Retirement	AI Retirement Plan; Risk Closure Documentation

Implications for Practice and Policy

The proposed control-based AI governance model has practical and policy implications for organizations and regulators seeking to operationalize AI oversight within existing governance and risk management structures. For practitioners, the model delivers a structured pathway for integrating AI governance into enterprise GRC programs by aligning governance controls with established risk assessment, compliance, and assurance processes, thereby reducing the reliance on ad hoc and principle-only approaches. The control categorization and mapping framework enables organizations to assign clear accountability,

support auditability, and maintain consistency across the AI lifecycle, while remaining flexible to organizational size, sector, and regulatory context. From a policy perspective, the model offers a mechanism for translating high-level regulatory expectations and standards-based guidance into enforceable, auditable governance practices, without prescribing specific technologies or implementation methods. By demonstrating compatibility with widely adopted frameworks, such as the NIST AI Risk Management Framework and NIST SP 800-53, the model supports regulatory harmonization and reduces fragmentation across jurisdictions and compliance regimes. Collectively, these implications indicate that control-based AI governance can serve as a practical intermediary between evolving policy objectives and organizational implementation, strengthening oversight while preserving the capacity for adaptation in rapidly changing AI-enabled environments.

Recommendations for Future Research

Future research should empirically examine the effectiveness of control-based AI governance models across organizational contexts, sectors, and regulatory environments. Quantitative and mixed-method studies may assess how integrating administrative, technical, and operational controls influences governance outcomes, such as auditability, risk mitigation, and compliance consistency throughout the AI lifecycle. Additional research is required to evaluate the practical application of control mappings to frameworks, such as the NIST AI Risk Management Framework and NIST SP 800-53, particularly with respect to control testing, assurance processes, and governance maturity. Longitudinal analyses can further examine how governance controls respond to system modifications, retraining, and retirement over time. Comparative studies across jurisdictions and industries may also inform the development of harmonized governance approaches and identify contextual factors that shape the effectiveness of control-based AI governance.

Conclusion

This study proposes a control-based AI governance model that addresses the ongoing gap between principle-oriented guidance and operational governance practice. The model provides a systematic approach to risk ownership, compliance mapping, and continuous assurance across the AI lifecycle by structuring AI governance around administrative, technical, and operational controls and integrating these controls with enterprise GRC functions. Control mapping to the NIST AI Risk Management Framework and NIST SP 800-53 demonstrates the model's compatibility with established governance and assurance frameworks while preserving implementation flexibility. Collectively, these findings contribute to a practical, auditable, and technology-agnostic governance approach that advances the operationalization of AI oversight in complex organizational environments.

References

Avianti, I., Yunita, I., Sarumpaet, S., & Yuniasih, N. (2025). A bibliometric analysis of governance, risk, and compliance (GRC): Trends, themes, and future directions. *Humanities and Social Sciences Communications*, *12*, Article 1945. https://doi.org/10.1057/s41599-025-06194-9

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, *104*(3), 671–732. https://doi.org/10.2139/ssrn.2477899

Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: Themes, knowledge gaps and future agendas. *Internet Research*, *33*(7), 133–167. https://doi.org/10.1108/INTR-01-2022-0042

Bromley, P., & Powell, W. W. (2012). From smoke and mirrors to walking the talk: Decoupling in the contemporary world. *Academy of Management Annals*, *6*(1), 483–530. https://doi.org/10.1080/19416520.2012.684462

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, *81*(5), 825–836. https://doi.org/10.1111/puar.13293

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, *10*, Article 567. https://doi.org/10.1057/s41599-023-02079-x

Committee of Sponsoring Organizations of the Treadway Commission. (2013). *Internal control — integrated framework*. COSO. https://www.coso.org/guidance-on-ic

Committee of Sponsoring Organizations of the Treadway Commission. (2017). *Enterprise risk management — integrating with strategy and performance*. COSO. https://www.coso.org/guidance-on-erm

Cremer, F., Sheehan, B., Fortmann, M., Kia, A. N., Mullins, M., Murphy, F., & Hasan, S. (2022). Cyber risk and cybersecurity: A systematic review of data availability. *The Geneva Papers on Risk and Insurance — Issues and Practice*, *47*, 698–736. https://doi.org/10.1057/s41288-022-00266-6

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160. https://doi.org/10.2307/2095101

Emery, F. E., & Trist, E. L. (1960). Socio-technical systems. In C. W. Churchman & M. Verhulst (Eds.), *Management sciences: Models and techniques* (Vol. 2, pp. 83–97). Pergamon.

Ghyati, S., & Kasbaoui, T. (2023). Internal control in organization theories: A review of theoretical literature. *African Scientific Journal*, *03*(21), 326–354. https://doi.org/10.5281/zenodo.10375588

Heier, J. R., Dugan, M. T., & Sayers, D. L. (2005). A century of debate for internal controls and their assessment: A study of reactive evolution. *Journal of Management History*, *11*(3), 239–262. https://doi.org/10.1108/17511340510600145

Jankensgård, H. (2019). A theory of enterprise risk management. *Corporate Governance: The International Journal of Business in Society*, *19*(3), 565–579. https://doi.org/10.1108/CG-02-2018-0092

Kamiya, S., Kang, J.-K., Kim, J., Milidonis, A., & Stulz, R. M. (2021). Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics*, *139*(3), 719–749. https://doi.org/10.1016/j.jfineco.2019.05.019

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, *44*(6), Article 101976. https://doi.org/10.1016/j.telpol.2020.101976

Liu, C., & Babar, M. A. (2024). Corporate cybersecurity risk and data breaches: A systematic review of empirical research. *Australian Journal of Management*. Advance online publication. https://doi.org/10.1177/03128962241293658

Lu, Q., Zhu, L., Xu, X., Whittle, J., Douglas, D., & Sanderson, C. (2022). Responsible AI pattern catalogue: A collection of best practices for AI governance. *ACM Computing Surveys*, *56*(7), Article 180. https://doi.org/10.1145/3626234

Mahoney, J., & Thelen, K. (Eds.). (2010). *Explaining institutional change: Ambiguity, agency, and power*. Cambridge University Press.

Manias, G., Apostolopoulos, D., Athanassopoulos, S., Borotis, S., Chatzimallis, C., Chatzipantelis, T., Corrales Compagnucci, M., Zdolsek Draksler, T., Fournier, F., Goralczyk, M., Gucek, A., Karabetian, A., Kefala, S., Kotios, D., Kovacic, M., Kyrkou, D., Limonad, L., Magopoulou, S., Mavrogiorgos, K., . . . Kyriazis, D. (2023). AI4Gov: Trusted AI for transparent public governance fostering democratic values. In *2023 19th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)* (pp. 548–555). IEEE. https://doi.org/10.1109/DCOSS-IoT58021.2023.00090

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics*, *2*(4), 603–609. https://doi.org/10.1007/s43681-021-00084-x

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, *83*(2), 340–363. https://doi.org/10.1086/226550

Mikes, A., & Kaplan, R. S. (2012). Managing risks: A new framework. *Harvard Business Review*, *90*(6), 48–60.

Mikes, A., & Kaplan, R. S. (2013). *Towards a contingency theory of enterprise risk management* (Working Paper No. 13-063). Harvard Business School. https://www.hbs.edu/ris/Publication%20Files/13-063_5e67dffe-aa5e-4fac-a746-7b3c07902520.pdf

Minkkinen, M., Laine, J., & Mäntymäki, M. (2022). Continuous auditing of artificial intelligence: A conceptualization and assessment of tools and frameworks. *Digital Society*, *1*, Article 7. https://doi.org/10.1007/s44206-022-00022-2

Minkkinen, M., Niukkanen, A., & Mäntymäki, M. (2022). What about investors? ESG analyses as tools for ethics-based AI auditing. *AI and Society*, *37*, 1307–1317. https://doi.org/10.1007/s00146-021-01261-6

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Mökander, J., Morley, J., Taddeo, M., & Floridi, L. (2021). Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. *Science and Engineering Ethics*, *27*(4), Article 44. https://doi.org/10.1007/s11948-021-00319-4

Mökander, J., Sheth, M., Gersbro-Sundler, M., Blomgren, P., & Floridi, L. (2022). Challenges and best practices in corporate AI governance: Lessons from the biopharmaceutical industry. *Frontiers in Computer Science*, *4*, Article 1068361. https://doi.org/10.3389/fcomp.2022.1068361

Mulej, M., Potocan, V., Zenko, Z., Kajzer, S., Ursic, D., Knez-Riedl, J., Lynn, M., & Ovsenik, J. (2004). How to restore Bertalanffian systems thinking. *Kybernetes*, *33*(1), 48–61. https://doi.org/10.1108/03684920410514346

National Institute of Standards and Technology. (2020). *Security and privacy controls for information systems and organizations* (NIST Special Publication 800-53 Rev. 5). U.S. Department of Commerce. https://doi.org/10.6028/NIST.SP.800-53r5

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

Okhmatovskiy, I., & David, R. J. (2012). Setting your own standards: Internal corporate governance codes as a response to institutional pressure. *Organization Science*, *23*(1), 155–176. https://doi.org/10.1287/orsc.1100.0642

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *Journal of Strategic Information Systems*, *34*(2), Article 101885. https://doi.org/10.1016/j.jsis.2024.101885

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Thiefalaine, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 33–44. https://doi.org/10.1145/3351095.3372873

Ramage, M., & Shipp, K. (2009). *Systems thinkers*. Springer/Open University Press.

Reis, J. F., & Pinheiro Junior, L. P. (2025). Institutional theory (IT) and diffusion of innovation (DOI): A theoretical approach on artificial intelligence (AI). *Brazilian Administration Review*, *22*(4), e250060. https://doi.org/10.1590/1807-7692bar2025250060

Romanosky, S. (2016). Examining the costs and causes of cyber incidents. *Journal of Cybersecurity*, *2*(2), 121–135. https://doi.org/10.1093/cybsec/tyw001

Schneider, J., Abraham, R., Meske, C., & vom Brocke, J. (2023). Artificial intelligence governance for businesses. *Information Systems Management*, *40*(3), 229–249. https://doi.org/10.1080/10580530.2022.2085825

Scott, W. R. (2014). *Institutions and organizations: Ideas, interests, and identities* (4th ed.). SAGE.

Stettinger, M., Veledar, O., Schleiß, P., & Steger, C. (2024). Trustworthiness assurance assessment for high-risk AI-based systems. *IEEE Access*, *12*, 28947–28965. https://doi.org/10.1109/ACCESS.2024.3364387

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, *40*(2), 137–157. https://doi.org/10.1080/14494035.2021.1928377

Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations*, *4*(1), 3–38. https://doi.org/10.1177/001872675100400101

von Bertalanffy, L. (1968). *General system theory: Foundations, development, applications*. George Braziller.

Waltersdorfer, L., & Sabou, M. (2025). Leveraging knowledge graphs for AI system auditing and transparency. *Journal of Web Semantics*, *84*, Article 100849. https://doi.org/10.1016/j.websem.2024.100849

Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. *Government Information Quarterly*, *39*(4), Article 101685. https://doi.org/10.1016/j.giq.2022.101685

Zhang, X. (2014). Evolution of Western internal control theory. In X. Zhang, *Enterprise management control systems in China* (pp. 67–85). Springer. https://doi.org/10.1007/978-3-642-54715-7_5

Appendix A

Expanded Mapping of Governance Control Categories to NIST AI RMF Functions and NIST SP 800-53 Control Families Across the AI Lifecycle

This appendix presents the complete control mapping underlying Table 1. The expanded table provides additional coverage across governance control categories and AI lifecycle phases while preserving the same column structure and control logic used in the main text.

Control Category	Control Objective	NIST AI RMF Function	NIST SP 800-53 Control Family	AI Lifecycle Phase	Example Governance Artifacts
Administrative	Establish formal AI governance policy and scope	GOVERN	PL (Planning)	Strategy & Design	AI Governance Charter; Enterprise AI Policy
Administrative	Define organizational accountability for AI risk	GOVERN	PM (Program Management)	Strategy & Design	AI RACI Matrix; Governance Role Definitions
Administrative	Assign AI risk ownership and escalation authority	GOVERN	PM (Program Management)	Strategy & Design	Risk Ownership Register; Escalation Protocol
Administrative	Require AI risk assessments prior to deployment	MAP	RA (Risk Assessment)	Development	AI Risk Assessment Report; Risk Register Entry
Administrative	Establish AI ethics and oversight committee	GOVERN	PM (Program Management)	Organization-Wide	Committee Charter; Terms of Reference
Administrative	Define acceptable AI use and prohibited practices	GOVERN	PL (Planning)	Strategy & Design	Acceptable Use Policy; Prohibited Use Register
Administrative	Require documentation of model purpose and limitations	MAP	SA (System & Services Acquisition)	Development	Model Card; System Purpose Statement
Administrative	Mandate third-party AI risk due diligence	MAP	SR (Supply Chain Risk Management)	Procurement	Vendor AI Risk Review Checklist; Third-Party Risk Policy
Administrative	Establish AI incident response policy	MANAGE	IR (Incident Response)	Deployment & Operations	AI Incident Response Policy; Escalation Thresholds
Administrative	Define audit and review cadence for AI systems	GOVERN	CA (Assessment, Authorization & Monitoring)	Operations	AI Governance Review Plan; Audit Schedule
Administrative	Require decommissioning criteria for AI systems	MANAGE	CP (Contingency Planning)	Retirement	AI Retirement Criteria; Decommissioning Checklist
Technical	Implement data governance and quality controls	MAP	IA (Identification & Authentication), SI (System & Information Integrity)	Data Preparation	Data Quality Report; Lineage Documentation
Technical	Enforce access controls on AI training data	MANAGE	AC (Access Control), IA (Identification & Authentication)	Data Preparation	Access Control Matrix; Privileged Access Review
Technical	Monitor model performance drift	MEASURE	SI (System & Information Integrity)	Operations	Drift Monitoring Dashboard; Performance Metrics Report
Technical	Validate models prior to production release	MEASURE	CA (Assessment, Authorization & Monitoring), SA (System & Services Acquisition)	Development	Model Validation Report; Testing Summary
Technical	Log AI system decisions for traceability	MEASURE	AU (Audit & Accountability)	Operations	Audit Logs; Decision Trace Records
Technical	Implement bias testing and evaluation mechanisms	MEASURE	RA (Risk Assessment), SI (System & Information Integrity)	Development	Bias Risk Assessment; Fairness Evaluation Report
Technical	Secure AI pipelines and infrastructure	MANAGE	SC (System & Communications Protection), SI	Development	Pipeline Security Assessment; Infrastructure Hardening Log

Control Category	Control Objective	NIST AI RMF Function	NIST SP 800-53 Control Family	AI Lifecycle Phase	Example Governance Artifacts
			(System & Information Integrity)		
Technical	Enable explainability or interpretability mechanisms	**MAP**	SI (System & Information Integrity)	Deployment	Explainability Documentation; Interpretability Test Results
Technical	Detect anomalous AI behavior	**MEASURE**	SI (System & Information Integrity)	Operations	Anomaly Detection Alerts; Behavioral Monitoring Report
Technical	Protect AI models from unauthorized modification	**MANAGE**	CM (Configuration Management), SI (System & Information Integrity)	Operations	Configuration Baseline; Change Control Log
Operational	Require human-in-the-loop for high-risk decisions	**MANAGE**	PL (Planning), SA (System & Services Acquisition)	Deployment	Human Review Protocol; Risk Threshold Register
Operational	Define human override and escalation procedures	**MANAGE**	IR (Incident Response)	Deployment	Override Procedure Document; Escalation Flowchart
Operational	Train personnel on AI risk and governance controls	**GOVERN**	AT (Awareness & Training)	Organization-Wide	Training Completion Records; Curriculum Documentation
Operational	Conduct periodic AI risk reviews	**GOVERN**	CA (Assessment, Authorization & Monitoring)	Operations	Governance Review Report; Risk Reassessment Summary
Operational	Integrate AI risks into enterprise risk register	**GOVERN**	RA (Risk Assessment), PM (Program Management)	Organization-Wide	Enterprise Risk Register; AI Risk Entries
Operational	Execute AI incident response playbooks	**MANAGE**	IR (Incident Response)	Operations	AI Incident Playbook; After-Action Report
Operational	Manage AI system changes through formal change control	**MANAGE**	CM (Configuration Management)	Operations	Change Request Log; Change Advisory Board Records
Operational	Validate continued compliance post-deployment	**MEASURE**	CA (Assessment, Authorization & Monitoring)	Operations	Compliance Validation Report; Control Test Evidence
Operational	Document lessons learned from AI incidents	**MANAGE**	IR (Incident Response)	Operations	Lessons Learned Register; Incident After-Action Report
Operational	Retire AI systems that exceed risk tolerance	**MANAGE**	SA (System & Services Acquisition), CM (Configuration Management), MP (Media Protection)	Retirement	AI Retirement Plan; Risk Closure Documentation

The listed control objectives are illustrative and designed to support governance development and assurance alignment. Organizations can tailor their control selection and implementation based on risk appetite, regulatory environment, and system importance.

Appendix A1
Guidance for Interpreting and Applying the Control Mapping

Control mapping in Appendix A supports the design and evaluation of AI governance programs by connecting governance control objectives to established risk management and assurance frameworks. Each row represents a control objective that organizations can adopt selectively, based on risk tolerance, regulatory environment, and AI system importance. The table serves as a reference framework rather than a strict implementation model, allowing alignment with existing enterprise governance, risk, and compliance (GRC) practices. Controls are grouped by governance control category to reflect the complementary roles of organizational oversight, technical safeguards, and operational processes. The mapped NIST AI Risk Management Framework functions show how each control supports AI-specific risk governance, while the

associated NIST SP 800-53 control families demonstrate compatibility with established internal control and assurance mechanisms. The AI lifecycle phase indicates the context where the control is most relevant, recognizing that many governance controls apply across multiple stages.

This mapping can be used by governance leaders, risk managers, and auditors to evaluate coverage, identify gaps, and support traceability between AI governance objectives and enterprise assurance structures. Controls may be tailored, combined, or mapped to multiple lifecycle phases or control families as needed, allowing the model to remain flexible while ensuring consistent, auditable governance throughout the AI lifecycle.