Available online at https:
https://journal.uitm.edu.my/ojs/index.php/JIKM

e-ISSN: 2289-5337

Journal of Information and Knowledge Management (2026) Vol. 16, No. 1

Journal of Information and Knowledge Management

Declarable Integration of NIST AI Risk Management into AI-driven ISMS through Policy-as-Code

Dmitri Kharchevnikov[1*], Steven Robinett[1]

[1]*Great Falls College Montana State University, 2100 16th Avenue South Great Falls, MT, USA 59405, +1-406-771-4375*

Corresponding Authors' Email Address: d.kharchevnikov@gfcmsu.edu

ARTICLE INFO

Article history:
Received: 26 August 2026
Revised: 29 September 2026
Accepted: 8 January 2026
Online first
Published: 10 April 2026

Keywords:
AI risk management
Declarable security
Policy-as-Code
Cybersecurity

https://doi.org/10.24191/1pd5sq41

ABSTRACT

Despite growing AI integration into Information Security Management Systems (ISMS), organizations lack systematic methods to transform high-level AI governance frameworks into machine-executable security controls. Current standards like NIST AI Risk Management Framework (AI RMF) and ISO/IEC 27001:2022 provide principles but no actionable pathways for automated enforcement, creating compliance gaps and limiting trust in AI-driven systems. This study develops a unified framework for declarable cybersecurity risk assessment in AI-driven ISMS through Policy-as-Code integration. We introduce a novel four-criteria declarability schema to systematically evaluate which AI governance provisions can be automated, applying this to all 212 NIST AI RMF actions. Using mixed-methods analysis, we assessed extractability, classified actions by control logic (Preventive/Detective/Reactive), system layer (Model/Output/User), and Confidentiality-Integrity-Availability triad alignment, then conducted semantic crosswalk with ISO/IEC 27002:2022 operational domains. Results show 84.9% of AI governance actions are directly declarable for Policy-as-Code implementation, with Measure being the most automatable function (39.4%) and Detective controls dominating across functions (reaching 75% in Measure). Actions primarily target Model and Output layers (78% combined), with Integrity overwhelming other dimensions (75.6% overall). Crosswalk analysis reveals strong alignment with Governance (24.4%) and Threat Management (18.9%), but critical gaps in System Security (0%), Identity Management (1.1%), and Asset Management (1.7%). This research provides the first reproducible methodology for transforming AI governance frameworks into machine-actionable controls within existing ISMS architectures, enabling traceable, auditable, and standards-aligned security automation for AI systems.

INTRODUCTION

The contemporary cybersecurity landscape is undergoing profound transformation through the integration of Artificial Intelligence (AI) into Information Security Management Systems (ISMS). This shift represents more than incremental improvement—it fundamentally redefines how

organizations approach digital defense (Kaur et al., 2023; Mohamed, 2023). The AI cybersecurity sector is projected to grow at 21.9% annually between 2023-2028, reaching $60.6 billion by 2028, with nearly half of organizations already adopting AI technologies for cybersecurity enhancement (Polito and Pupillo, 2024; Mohamed, 2023). AI's capacity to process vast data volumes in real-time, identify subtle patterns, and automate complex responses significantly enhances threat detection, incident response, and predictive analytics (Akhtar and Rawol, 2024), transforming cybersecurity from reactive to proactive strategic function.

However, AI integration introduces unique security challenges. AI systems operate with opaque decision logic, exhibit model drift, face adversarial manipulation risks, and create complex deployment pipelines that complicate traditional access control (Kunle-Lawanson, 2022; Obisesan, 2024). These dynamic behaviors compromise data integrity assumptions underlying the CIA triad while demanding real-time monitoring and automated enforcement capabilities that traditional ISMS frameworks struggle to provide. As AI systems become increasingly autonomous in critical security functions, failure consequences escalate dramatically, potentially impacting operational continuity, financial stability, and organizational reputation (Flehmig et al., 2024; Badman, 2024).

The Core Problem: Despite the proliferation of AI governance frameworks—including NIST AI Risk Management Framework (AI RMF), ISO/IEC 27001:2022, and ISO/IEC 42001—organizations lack systematic methods to operationalize these high-level principles into machine-executable security controls. Current frameworks provide strategic guidance but no formal mechanisms to determine which provisions are automatable, how they align with existing cybersecurity controls, or where enforcement should occur within AI system architectures. This creates three critical gaps: (1) no systematic criteria exist to assess whether governance actions are sufficiently specific for automated enforcement, (2) the relationship between AI-specific risks and traditional CIA triad protections remains unclear, and (3) no validated crosswalk connects AI governance requirements to established ISMS operational capabilities. Without these connections, organizations face fragmented implementations, compliance gaps, and reduced trust in AI-enabled security systems.

Research Gap: While recent efforts have advanced AI governance conceptually—such as the Cloud Security Alliance's AI Controls Matrix (CSA, 2025) defining 243 control objectives, and industry tools like IBM AI FactSheets (Hind, 2020) and Google Model Cards (Mitchell et al., 2019) improving transparency—these remain largely descriptive. They lack integration into formal ISMS control environments and provide no pathway to Policy-as-Code (PaC) implementation. The gap between abstract governance documentation and concrete enforcement engines remains unbridged (Salako et al., 2024). Current research has not systematically evaluated which AI governance provisions can be declaratively specified, how they map to cybersecurity control logic, or where they align with ISO/IEC 27002 operational domains.

Purpose and Objectives: This study addresses these gaps by developing a unified framework for declarable cybersecurity risk assessment in AI-driven ISMS that enables Policy-as-Code integration. Specifically, we: (1) introduce and validate a novel declarability schema (R1-R4) to systematically assess which NIST AI RMF actions can be automated; (2) provide reproducible methodology for transforming these actions into PaC-ready statements aligned with the CIA triad and AI system layers; (3) conduct comprehensive crosswalk analysis between NIST AI RMF and

ISO/IEC 27002:2022 to identify alignment patterns and coverage gaps; and (4) stratify actions by control logic type (Preventive/Detective/Reactive) to guide enforcement strategy design.

Contribution and Approach: Through mixed-methods analysis of all 212 NIST AI RMF actions, combining qualitative framework analysis with quantitative extractability assessment, this research establishes the first systematic methodology for bridging AI governance principles and operational cybersecurity controls. By demonstrating that 84.9% of AI governance actions are declarable and by mapping these to established security frameworks, we provide organizations with a practical, standards-aligned pathway to implement traceable, auditable, and machine-executable AI security policies. This work lays the foundation for trustworthy AI deployment in regulated environments where explainability, accountability, and continuous compliance are mandatory.

BACKGROUND AND FOUNDATIONAL CONCEPTS

Evolution of AI-Integrated Information Security Management Systems

Information Security Management Systems (ISMS) provide structured frameworks for managing information security risks, policies, and controls through risk-based, cyclical assessment and continuous improvement (Al-Dhahri et al., 2017). Defined by ISO/IEC 27000:2018, an ISMS encompasses "the policies, procedures, guidelines, and associated resources and activities, collectively managed by an organization, in the pursuit of protecting its information assets" (ISO, 2018). Traditional ISMS frameworks apply risk management processes to protect the confidentiality, integrity, and availability (CIA triad) of information through systematic assessment, control implementation, monitoring, and improvement cycles.

The integration of AI technologies is fundamentally transforming ISMS architecture. AI-powered ISMS enhance traditional frameworks by employing machine learning, natural language processing, and intelligent automation to improve core functions including risk identification and prioritization through predictive analytics, continuous real-time monitoring of system behavior, automated policy enforcement via Policy-as-Code mechanisms, and intelligent audit and compliance management (Malik et al., 2025; Kunle-Lawanson, 2022; Obisesan, 2024). This evolution reflects what Kaur et al. (2023) characterize as intelligence-driven cybersecurity capable of providing dynamic defense through proactive and adaptive approaches to rapidly changing digital environments and regulatory expectations.

However, this transformation introduces distinctive security challenges. AI systems exhibit model drift, training data leakage, adversarial manipulation vulnerabilities, and opaque decision logic that must be explicitly addressed within ISMS lifecycles (Kunle-Lawanson, 2022; Obisesan, 2024). As Jada and Mayayise (2024) note, AI integration impacts the entire cybersecurity lifecycle, creating new attack surfaces and risk categories that traditional ISMS frameworks were not designed to address. AI-driven infrastructures challenge fundamental CIA triad assumptions: dynamic inference behavior can compromise data integrity, complex deployment pipelines complicate access control models, and autonomous decision-making systems create novel availability risks requiring real-time monitoring capabilities beyond traditional approaches.

Trust, Assurance, and the Emergence of Declarability

Trust and assurance form the foundation of information security management, particularly in AI-integrated environments where system decisions may lack transparency or predictability. Trust models in cybersecurity have evolved from subjective interpretations to structured assurance models supported by technical evidence through foundational anchors including authentication, authorization, access control, privacy protection, monitoring, auditing, encryption, and risk management (Pigola and de Souza Mierelles, 2024).

The principle of "declarability" has emerged as a critical requirement for AI-driven ISMS, shifting from implicit trust to explicit, verifiable assurance. Declarability requires that assessment processes, findings, and resulting security controls be formally stated, verifiably documented, and audit-ready (Hale and Gamble, 2019). This principle draws from three advanced AI and security paradigms: Explainable AI (XAI), which enables AI models to convey decision-making rationales, addressing black-box challenges inherent in many systems (Ali et al., 2023); Verifiable AI, which emphasizes transparent, auditable, and accountable systems allowing validation of data lineage, model governance, and output authenticity (Fok and Weld, 2024); and Declarative Programming, a paradigm specifying desired outcomes rather than procedural steps, translating to security policies defined as code enabling automated enforcement and clear audit trails (Kordjamshidi et al., 2022; Jothimani, 2022).

As Flehmig et al. (2024) emphasize, AI's growing autonomy in critical security functions escalates failure consequences dramatically. This heightened reliance demands robust mechanisms to establish trustworthiness beyond traditional security models. The operationalization of AI trustworthiness built on transparency, explainability, and accountability (Li et al., 2023) requires concrete, verifiable mechanisms. Declarability provides these mechanisms: transparency becomes declarable through explicit documentation of AI purpose, limitations, and decision-making; explainability is achieved when systems provide human-readable narratives supported by formal methods; and accountability is established through verifiable behavioral records and policy adherence (Raja and Zhou, 2023). Converting these abstract principles into machine-actionable governance is essential for regulatory compliance, transparency, stakeholder confidence, and automated governance in AI-driven systems (Jeffy and Bello, 2025).

Policy-as-Code: Bridging Governance and Technical Enforcement

Policy-as-Code (PaC) represents a methodological advancement enabling organizations to define and manage policies in machine-readable, executable formats. Unlike traditional static documentation, PaC enables automated policy enforcement, validation, testing, and integration into CI/CD pipelines and system workflows, ensuring consistency, scalability, and real-time compliance (Vakhula et al., 2024; Korrapati, 2024).

Tools such as Open Policy Agent (OPA) and HashiCorp Sentinel operationalize PaC by enabling organizations to encode access control, audit, and configuration policies in formal logic, enhancing auditability and enforcement within continuous deployment environments (Korrapati, 2024; Webster et al., 2023). NIST SP 800-204B demonstrates PaC application in service mesh

architectures for microservices, implementing fine-grained access control and zero-trust principles declaratively (NIST, 2021). Recent research demonstrates PaC's effectiveness in automating compliance checks across multi-cloud environments and ensuring alignment with standards like ISO/IEC 27001:2022 (Webster et al., 2023; Vakhula et al., 2024).

Despite these advances, significant implementation challenges remain. PaC requires governance actions to possess specific characteristics: machine-actionable behavior, lifecycle phase alignment, risk-targeted control specifications, and contextual specificity (Jothimani, 2022). Current AI governance frameworks largely lack these attributes, creating a fundamental gap between high-level principles and executable enforcement mechanisms.

Current AI Governance Frameworks: Capabilities and Limitations

Recent years have witnessed proliferation of AI governance frameworks attempting to address AI-specific risks. The NIST AI Risk Management Framework (AI RMF) structures risk management across four core functions—Govern, Map, Measure, and Manage—emphasizing lifecycle-specific controls such as transparency during model training and integrity during inference (NIST, 2023, 2024). ISO/IEC 42001:2023 provides structured AI management system requirements, while ISO/IEC 27001:2022 and 27002:2022 offer updated security controls incorporating cloud security and threat intelligence (ISO, 2022-a, 2022-b, 2023).

Habbal et al. (2024) propose the AI Trust, Risk and Security Management (AI TRiSM) framework as a comprehensive approach integrating trust, risk, and security concerns to ensure fairness, governance, efficacy, reliability, and privacy. These frameworks examine risk complexity in AI systems, proposing typologies for AI-specific hazards including hallucination, privacy leakage, adversarial misuse, and algorithmic bias. Research emphasizes automation's role in improving risk assessment effectiveness, enabling continuous monitoring and rapid responses (Habbal et al., 2024).

However, as Batool et al. (2025) critically observe, efforts to integrate cybersecurity, AI governance, and ISMS remain fragmented and at early stages. Most frameworks offer high-level principles and guidance but lack machine-executable actions or formal evaluative criteria. They describe what controls should exist but provide no specifications for how controls can be efficiently defined, verified, or enforced within automated governance systems. They typically lack mechanisms for assessing whether controls are effectively scoped for specific AI lifecycle phases or for addressing emerging risks like adversarial attacks. Furthermore, the absence of clear declarability criteria—specifications for policies interpretable by both humans and machines, traceable to threats or lifecycle stages, and adaptable to operational environments—creates fragmented implementations, compliance gaps, and reduced trust in AI-enabled security systems.

Industry initiatives like IBM AI FactSheets 360 (Hind, 2020) and Google Model Cards (Mitchell et al., 2019) have introduced transparency and accountability mechanisms for AI models, yet these tools remain primarily descriptive and are not integrated into formal ISMS control environments. Similarly, while OPA demonstrates machine-executable policy

enforcement (Webster et al., 2023; Korrapati, 2024), its application requires structured, formally defined governance actions currently absent from most AI frameworks.

Emerging Solutions and Positioning of Current Research

Recent developments show promising directions for bridging governance principles and operational enforcement. The Cloud Security Alliance's AI Controls Matrix (AICM), released in 2025, provides 243 control objectives specifically tailored for cloud-based AI systems, explicitly mapping to ISO 42001, ISO 27001, and NIST AI 600-1 (CSA, 2025). AICM's analysis dimensions, including control type, architectural relevance, and LLM lifecycle relevance demonstrate growing recognition of the need for granular, operationalizable AI governance.

However, critical gaps persist. No systematic methodology exists to evaluate which governance provisions are declarable and automatable. The relationships between AI-specific risks and traditional CIA triad protections remain underexplored. Most critically, no validated crosswalk connects AI governance requirements to established ISO/IEC 27002 operational capabilities, limiting practical ISMS integration.

This research addresses these gaps by providing the first systematic methodology for assessing AI governance action declarability, stratifying actions by control logic and system enforcement layers, and conducting comprehensive semantic alignment between NIST AI RMF and ISO/IEC 27002:2022. By establishing reproducible criteria for Policy-as-Code readiness and demonstrating practical pathways from abstract governance to executable controls, this work provides the foundation organizations need to implement trustworthy, standards-aligned AI security automation within existing ISMS architectures.

METHODOLOGY

Methodological Approach and Rationale

This study establishes a unified framework for declarable cybersecurity risk assessment in AI-driven ISMS by systematically mapping AI governance provisions from the NIST AI Risk Management Framework (NIST AI 600-1) to ISO/IEC 27001:2022 controls. This addresses the core challenge of making AI governance actionable within traditional cybersecurity frameworks through Policy-as-Code (PaC) implementation, thereby enhancing trust and assurance in AI-driven ISMS.

Justification for Mixed-Methods Approach

We employ a mixed-methods approach combining qualitative framework analysis with quantitative extractability and similarity assessments. This methodological choice is grounded in established practices for policy formalization research. Hale and Gamble (2019) demonstrated that combining semantic analysis with quantitative evaluation provides systematic extraction and formalization of compliance requirements from security control standards. Their work established

that mixed methods are essential when bridging the gap between high-level regulatory language and implementable technical controls, as different aspects of governance require both qualitative contextual interpretation and quantitative validation of patterns.

Similarly, Korrapati (2024) showed that automating compliance in CI/CD pipelines requires both qualitative policy interpretation and quantitative validation mechanisms. The mixed-methods approach enables us to: (1) qualitatively interpret the intent and context of NIST AI RMF actions through expert analysis, and (2) quantitatively assess their machine-enforceability and alignment with existing standards through statistical validation. This dual approach ensures both semantic fidelity and empirical rigor, which are critical when transforming abstract governance guidance into executable controls (Jothimani, 2022).

Justification for Binary Criteria (R1-R4)

We utilize binary criteria rather than scaled assessments for several theoretically grounded reasons rooted in the practical requirements of Policy-as-Code implementation.

First, Policy-as-Code implementations fundamentally require binary decisions at the enforcement stage: a policy check either passes or fails, and a provision is either sufficiently specified to be encoded as executable policy or it requires additional refinement. This binary enforcement logic is inherent to tools such as Open Policy Agent (OPA) and HashiCorp Sentinel, which operate on pass/fail validation (Korrapati, 2024; Vakhula et al., 2024). Our assessment criteria mirror this operational reality, determining whether governance provisions meet the threshold for machine-enforceability rather than rating them on a continuous scale.

Second, Vakhula et al. (2024) demonstrated that Security-as-Code methodologies for ISO/IEC 27001:2022 compliance rely on binary validation at the infrastructure level—configurations either meet defined security policies or they do not. This binary enforcement model provides a clear decision boundary that our assessment criteria emulate.

Third, binary classification reduces ambiguity in the assessment process by requiring clear yes/no determinations rather than subjective scoring along a continuum. This approach enhances reproducibility and inter-assessor consistency, as evaluators need only determine whether a criterion is met rather than calibrating numerical ratings. The middle category ("Requires Refinement" for scores of 2-3) acknowledges that some provisions are partially specified while still maintaining assessment clarity.

The specific four criteria (R1-R4) are grounded in established requirements for policy formalization and automated enforcement:

R1 (Specific Control Behavior): Derived from the practical requirement that machine-enforceable policies must specify concrete actions (e.g., "block", "require", "log") rather than normative suggestions (e.g., "consider", "promote"). This distinguishes operational controls from strategic guidance.

R2 (Lifecycle Scope): Derived from NIST's own AI RMF structure requiring lifecycle-specific controls for effective governance.

R3 (Measurable Parameters): Essential for automated policy enforcement as demonstrated by Korrapati (2024) in CI/CD compliance automation, where policies must define triggerable conditions.

R4 (Execution Context): Required for assigning responsibility and enabling enforcement, consistent with standard ISMS control implementation principles that require defined actors and processes.

Initial Extraction and Extractability Assessment for Policy-as-Code Implementation

To identify which AI governance provisions are amenable to operationalization as declarable cybersecurity controls, we first conducted a comprehensive extraction of all suggested actions from the NIST AI RMF (NIST AI 600-1). The framework contains extensive guidance distributed across its four core functions (Govern, Map, Measure, Manage), including recommendations, control statements, and suggested practices.

Through systematic review of the entire framework, we extracted all actionable governance statements, removing duplicates, consolidating overlapping provisions, and filtering out background information, definitions, and explanatory text. This extraction process yielded 212 unique suggested actions that represented discrete, potentially implementable governance controls.

Binary Assessment Approach

These 212 identified actions underwent a two-step evaluation process to assess their suitability for encoding within PaC systems. In the first, programmatic step, each action was evaluated against four binary criteria (R1-R4), designed to assess operational enforceability:

R1 — Specific Control Behavior: Does the action define a concrete, enforceable behavior (e.g., "block", "require") as opposed to vague or normative suggestions (e.g., "consider", "promote")?
R2 — Lifecycle Scope: Is the action clearly scoped to a specific AI lifecycle stage (e.g., data collection, model development, deployment)?
R3 — Measurable Parameters: Are measurable inputs, thresholds, or states included or implied in a way that allows for automated triggering or decision-making?
R4 — Execution Context: Is there a defined actor, process, or technical system responsible for the execution of the action?

Each action received a binary (1/0) score for each criterion. Based on the cumulative score across R1-R4 (range: 0 to 4), an initial extractability label was assigned using the following action-based logic:

Table 1: Declarability schema

Total Criteria Met	Assessment Result
4	Yes (Directly declarable)
2 or 3	Requires refinement
0 or 1	No (Not declarable)

This procedure was executed via a Python script to generate the first-pass classification for all 212 suggested actions.

Human-in-the-Loop Validation Approach

To improve semantic fidelity and correct misclassifications arising from ambiguous phrasing or complex semantics, a human-in-the-loop review was conducted. This validation approach follows established practices in compliance modeling research. Hale and Gamble (2019) conducted extensive validation studies with industry security experts, achieving high accuracy ratings (averaging 9.34 out of 10) for their pattern-based requirement extraction approach. Their research demonstrated that combining automated extraction with expert review yields reliable and repeatable compliance assessments.

A subject matter expert manually verified or corrected the programmatic classification by: reviewing each action in the context of its intended use and governance scope; adjusting the extractability classification where automation failed to capture implied operational detail; and reclassifying borderline actions based on judgment of contextual enforceability.

Multi-Dimensional Classification System

System Layer Stratification

To align governance interventions with AI system architecture, we implemented a two-step classification process combining automated keyword analysis with expert validation. This hybrid approach is justified by research demonstrating that while automated classification provides efficiency and consistency, human validation is essential for contextual accuracy.

Keyword analysis was selected over more complex NLP methods (such as deep learning classifiers) for several methodologically sound reasons. First, the NIST AI RMF uses consistent terminology when describing different system layers, making keyword-based classification highly effective for this specific corpus. Second, keyword approaches provide transparency and interpretability—each classification decision can be traced to specific terms, supporting auditability requirements for governance frameworks (Jothimani, 2022). Third, the pattern-based approach aligns with Hale and Gamble's (2019) finding that controlled vocabularies and consistent terminology in source documents enable effective requirement classification without requiring complex natural language processing.

NIST AI RMF-ISO/IEC 27002 Control Alignment

To evaluate the alignment between actionable generative AI (GAI) governance activities and existing cybersecurity control domains, we conducted a structured crosswalk between declarable actions from the NIST AI RMF and the operational capability domains defined by ISO/IEC 27002:2022. This crosswalk methodology is grounded in established practices for mapping between security standards. Hale and Gamble (2019) demonstrated that semantic hierarchy-based approaches effectively establish traceable and semantically justified mappings between different compliance frameworks, which is precisely what our crosswalk aims to achieve.

Validation and Reproducibility

Justification for Expert-Driven Validation Approach

To ensure the robustness and trustworthiness of our findings, all classification and mapping processes throughout the methodology employed human-in-the-loop validation to confirm accuracy and contextual appropriateness. This validation approach is justified by established practices in compliance modeling research.

Hale and Gamble (2019) conducted validation studies demonstrating that expert review enhances semantic accuracy in compliance requirement extraction, with their studies achieving high inter-rater reliability (67-71% internal consistency across independent expert assessments) and accuracy ratings (averaging 9.34 out of 10). Similarly, Jothimani (2022) emphasized that policy-as-code implementations require human oversight to validate that encoded policies accurately represent organizational security objectives.

Our validation process involved a single expert reviewer who applied the structured rubrics consistently across all assessments. While this approach ensured uniform application of criteria, the reliance on a single reviewer is acknowledged as a limitation.

Reliability Measures and Limitations

While a formal inter-rater reliability analysis (such as Cohen's kappa) was not conducted in this study, we acknowledge this as a limitation. Hale and Gamble (2019) did conduct inter-rater reliability analyses in their validation studies, achieving 67-71% consistency across independent expert assessments. Our study employed a consensus-based approach with structured rubrics as a practical alternative, though we recognize that formal inter-rater reliability metrics would strengthen the reproducibility claims of our methodology.

The structured rubrics and transparent review process were designed to minimize subjective bias. Each criterion (R1-R4) is defined with explicit decision rules, and all classification decisions are documented in publicly available datasets to enable external validation. This transparency approach aligns with best practices in policy-as-code research, where auditability and reproducibility are paramount (Korrapati, 2024; Vakhula et al., 2024).

Limitations of This Approach

Several methodological limitations should be acknowledged:

1. Expert availability and consistency: While expert validation enhances accuracy, it introduces dependency on expert availability and may introduce subtle variations in interpretation across different experts. We mitigated this through consensus review and structured rubrics.
2. Binary classification constraints: The binary (pass/fail) nature of R1-R4 criteria may oversimplify nuanced cases where provisions are partially declarable. Actions scored 2-3 on criteria were classified as "Requires Refinement," acknowledging this middle ground.
3. Keyword analysis limitations: Automated keyword-based classification, while transparent and efficient, may miss implicit contextual information. Human validation specifically addresses this limitation.
4. Temporal validity: The NIST AI RMF is subject to updates and evolution. Our framework is based on NIST AI 600-1 as published; future revisions may require reanalysis.
5. Generalizability: While our methodology is designed to be reproducible, application to other AI governance frameworks beyond NIST AI RMF would require validation of the R1-R4 criteria's applicability.
6. Lack of formal inter-rater reliability: The absence of formal inter-rater reliability metrics (e.g., Cohen's kappa) is a limitation. Future research should incorporate such measures to strengthen reproducibility claims.
7. All expert validation was conducted by a single reviewer rather than multiple independent reviewers. While structured rubrics were employed to ensure consistency, single-reviewer validation limits the ability to assess inter-rater reliability and may introduce individual bias. Future research should employ multiple independent reviewers to strengthen validation claims.

RESULTS

Declarability Assessment: Evaluating Extractable vs. Non-Extractable Actions

From Section 3.2 of NIST AI 600-1, 212 unique suggested actions were extracted and evaluated using the four binary declarability criteria (R1-R4). The analysis yielded three categories: 180 actions (84.9%) were directly declarable in Policy-as-Code (PaC), 24 (11.3%) were conditionally declarable, and 8 (3.8%) were non-declarable.

The high proportion of directly declarable actions demonstrates that PaC can natively capture the vast majority of NIST AI RMF provisions. Conditionally declarable actions require additional contextual metadata or threshold parameters to become machine-enforceable, while non-

declarable actions contain non-operational verbs (e.g., "consider," "encourage," "promote") or depend on unstructured human interpretation. Representative examples of non-declarable actions include:

- "Encourage diverse stakeholder engagement in AI system design discussions."
- "Consider the historical and cultural context in which training data was collected."
- "Promote a culture of continuous learning and ethical awareness among AI developers."

These actions are normative rather than operational, expressing strategic intent or ethical aspiration without measurable conditions or executable triggers. While they cannot be directly encoded in PaC frameworks, they remain critical for governance culture and may be transformed into compliance prompts, training criteria, or oversight indicators in hybrid governance models.

Distribution of Declarable Actions by AI RMF Function

The distribution of the 180 directly declarable actions across the four core AI RMF functions is presented in Table 2 and visualized in Figure 1.

Table 2: Distribution of Actions by Core RMF Function

Function	Number	Percent ± SE (%)
Govern	37	20.6 ± 3.0
Map	29	16.1 ± 2.7
Measure	71	39.4 ± 3.6
Manage	43	23.9 ± 3.2

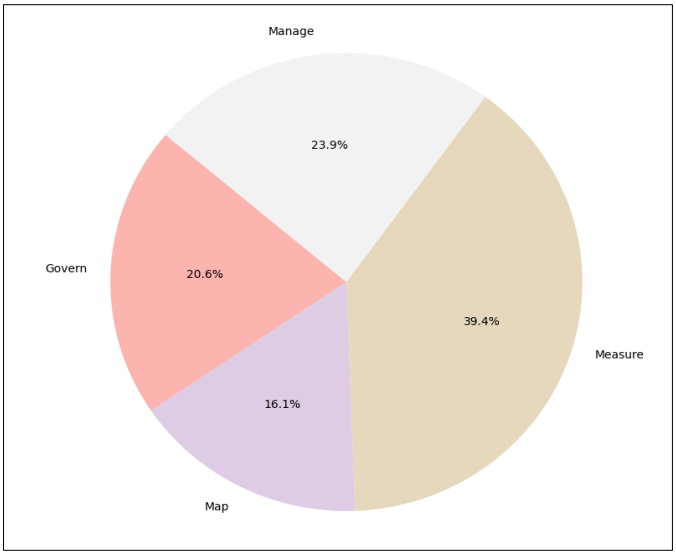

Figure 1: Proportion of Directly Declarable NIST AI RMF Actions by Core Function

The Measure function contains the largest share (39.4% ± 3.6%), indicating that monitoring and evaluation provisions translate most readily into automated enforcement. This pattern reflects the inherently quantitative and event-driven nature of measurement actions, which align naturally with PaC logic. In contrast, Map contains the smallest proportion (16.1% ± 2.7%), suggesting that contextual risk classification and model-output mapping require richer semantic models or manual parameterization. The Govern and Manage functions occupy intermediate positions, reflecting their blend of high-level policy directives and operational controls.

Stratification by AI System Layers (Model, Output, User)

To understand where declarable actions operationalize within AI systems, each of the 180 actions was classified by the architectural layer it primarily affects: Model, Output, or User. This stratification was performed using keyword-based Python logic derived from GAI Risk Categories, followed by human validation to ensure contextual accuracy. The results are presented in Table 3 and Figure 2.

Table 3: Stratification by AI System Layers

Layer	Count	Percent ± SE (%)
Model	68	37.8 ± 3.6
Output	72	40.0 ± 3.7
User	40	22.2 ± 3.1

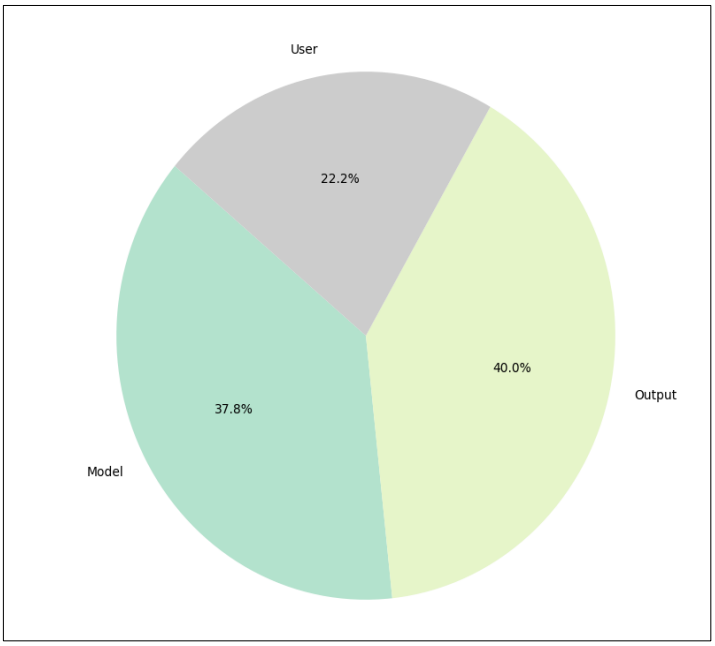

Figure 2: Proportional Distribution of Declarable Actions across AI System Layers

The Output layer contains the highest concentration of actions (40.0% ± 3.7), followed closely by the Model layer (37.8% ± 3.6). The User layer accounts for the smallest share (22.2% ± 3.1). This distribution reveals that declarable provisions predominantly target technical safeguards—model training controls and output validation—rather than user-facing governance mechanisms. The relatively low emphasis on user-layer controls suggests an opportunity to strengthen oversight protocols involving transparency disclosures, user consent workflows, and participatory risk management.

Control Logic Classification: Preventive, Detective, and Reactive

To characterize enforcement strategies embedded in the AI RMF, declarable actions were classified by control logic type: Preventive (constraints applied before execution), Detective (monitoring and evaluation during operation), and Reactive (responses triggered after detection). Table 4 and Figure 3 present this distribution.

Table 4: Control Logic Mapping

Function	Preventive % ± SE	Detective % ± SE	Reactive % ± SE
Govern	36.2 ± 6.3	44.8 ± 6.5	19.0 ± 5.1
Map	10.3 ± 5.7	62.1 ± 9.0	27.6 ± 8.3
Measure	20.8 ± 4.8	75.0 ± 5.1	4.2 ± 2.4
Manage	9.3 ± 4.4	62.8 ± 7.4	27.9 ± 6.8

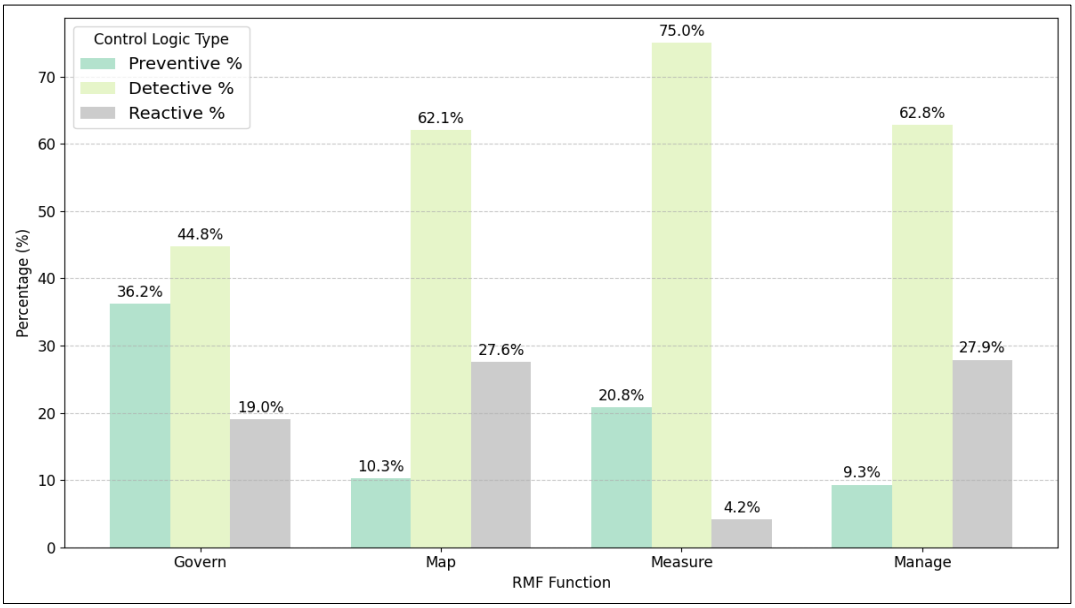

Figure 3: Distribution of Control Logic Across NIST AI RMF Functional Domains

Detective controls dominate across all functions, particularly in Measure (75.0%) and Manage (62.8%), emphasizing continuous evaluation and risk monitoring. Preventive controls are most prevalent in Govern (36.2%), where they enforce upstream constraints such as legal compliance frameworks and model documentation requirements. Reactive controls appear least frequently overall but are more common in Map (27.6%) and Manage (27.9%), supporting recovery and corrective workflows.

This pattern indicates that the AI RMF prioritizes runtime monitoring and post-hoc evaluation over pre-deployment constraint enforcement, reflecting a risk management philosophy centered on detection and response rather than prevention.

Alignment with CIA Triad

To assess how declarable actions address core cybersecurity objectives, each action was mapped to the CIA triad dimensions: Confidentiality, Integrity, and Availability. Table 5 and Figure 4 present this distribution.

Table 5: CIA Triad Component Mapping

Function	Confidentiality % ± SE	Integrity % ± SE	Availability % ± SE
Govern	7.4 ± 1.5	11.7 ± 1.9	5.4 ± 1.3
Map	9.0 ± 1.7	21.7 ± 2.4	5.7 ± 1.3
Measure	4.7 ± 1.2	14.0 ± 2.0	5.7 ± 1.3
Manage	4.7 ± 1.2	8.4 ± 1.6	1.7 ± 0.7

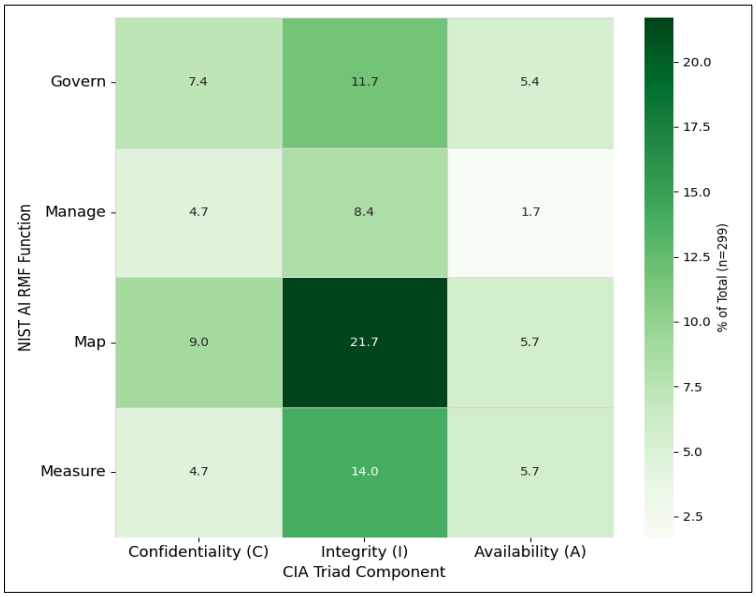

Figure 4: Heatmap of CIA Triad Coverage Across NIST AI RMF Functions

Integrity overwhelmingly dominates (75.6% ± 3.2), reflecting the AI RMF's emphasis on model accuracy, output correctness, and data validity. Confidentiality (11.7% ± 2.4) and Availability (12.8% ± 2.5) are comparatively underrepresented, suggesting potential blind spots in data protection and operational resilience.

Concentration on Integrity aligns with AI-specific risks such as adversarial attacks, training data poisoning, and output manipulation. However, the limited attention to Confidentiality raises concerns about privacy safeguards, particularly for systems handling sensitive personal data. Similarly, weak Availability coverage may leave AI systems vulnerable to denial-of-service conditions or insufficient failover mechanisms.

Crosswalk with ISO/IEC 27002:2022 Operational Capabilities

To evaluate how comprehensively the AI RMF addresses traditional cybersecurity domains, the 180 declarable actions were mapped to 12 applicable ISO/IEC 27002:2022 operational capabilities. The crosswalk results are presented in Table 6 and visualized in two figures: Figure 5 (pie chart showing overall distribution across ISO domains) and Figure 6 (heatmap showing function-level alignment).

Table 6: Frequency Distribution of Declarable AI RMF Actions by ISO/IEC 27002 Capability

ISO 27002 Operational Capability/ NIST AI RMF actions	**Govern N**	**Govern % ± SE**	**Map N**	**Map % ± SE**	**Measure N**	**Measure % ± SE**	**Manage N**	**Manage % ± SE**	**Total by capability N**	**Total by capability**
Asset Management	3	8.1 ± 4.5	0	0.0 ± 0.0	0	0.0 ± 0.0	0	0.0 ± 0.0	3	1.7 ± 1.0
Governance	8	21.6 ± 6.8	2	25.0 ± 15.3	17	20.2 ± 4.4	17	33.3 ± 6.6	44	24.4 ± 3.2
Human Resource Security	3	8.1 ± 4.5	0	0.0 ± 0.0	15	17.9 ± 4.2	2	3.9 ± 2.7	20	11.1 ± 2.3
Information Protection	2	5.4 ± 3.7	0	0.0 ± 0.0	9	10.7± 3.4	3	5.9 ± 3.3	14	7.8 ± 2.0
System and Network Security	0	0.0 ± 0.0	0	0.0 ± 0.0	0	0.0 ± 0.0	0	0.0 ± 0.0	0	0.0 ± 0.0
Application Security	1	2.7 ± 2.7	3	37.5 ± 17.1	9	10.7 ± 3.4	2	3.9 ± 2.7	15	8.3 ± 2.1
Identity and Access Management	0	0.0 ± 0.0	0	0.0 ± 0.0	1	1.2 ± 1.2	1	2.0 ± 1.9	2	1.1 ± 0.8
Continuity	3	8.1 ± 4.5	0	0.0 ± 0.0	8	9.5 ± 3.2	4	7.8 ± 3.8	15	8.3 ± 2.1
Legal and Compliance	3	8.1± 4.5	0	0.0 ± 0.0	2	2.4 ± 1.7	1	2.0 ± 1.9	6	3.3 ± 1.3
Supplier Relationships Security	3	8.1 ± 4.5	0	0.0 ± 0.0	1	1.2 ± 1.2	2	3.9 ± 2.7	6	3.3± 1.3

Threat and Vulnerability Management	5	13.5 ± 5.6	3	37.5 ± 17.1	18	21.4 ± 4.5	8	15.7 ± 5.1	34	18.9 ± 2.9
Information Security Event Management	6	16.2 ± 6.1	0	0.0 ± 0.0	4	4.8 ± 2.3	11	21.6 ± 5.8	21	11.7 ± 2.4

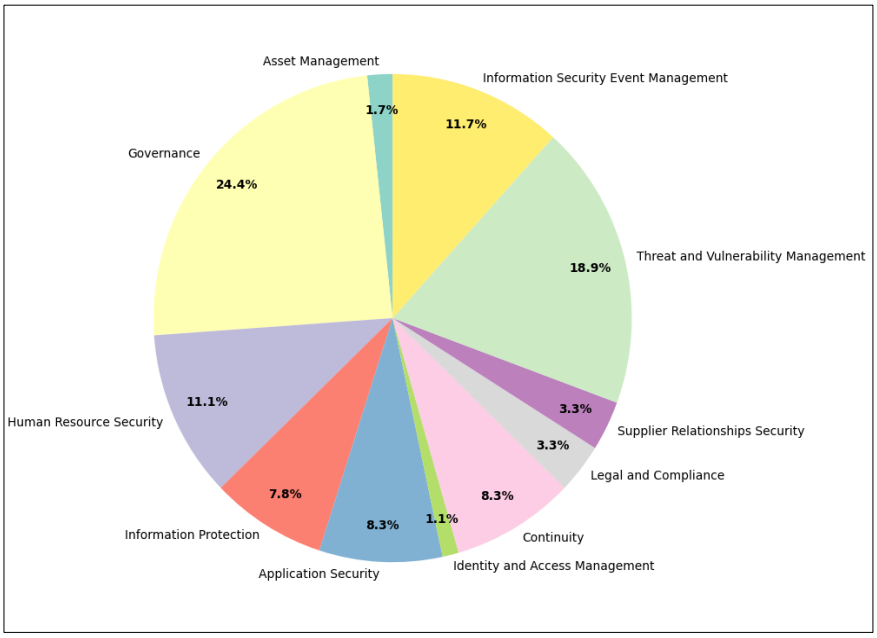

Figure 5: Pie Chart of AI RMF Declarable Actions Distributed Across ISO/IEC 27002 Domains

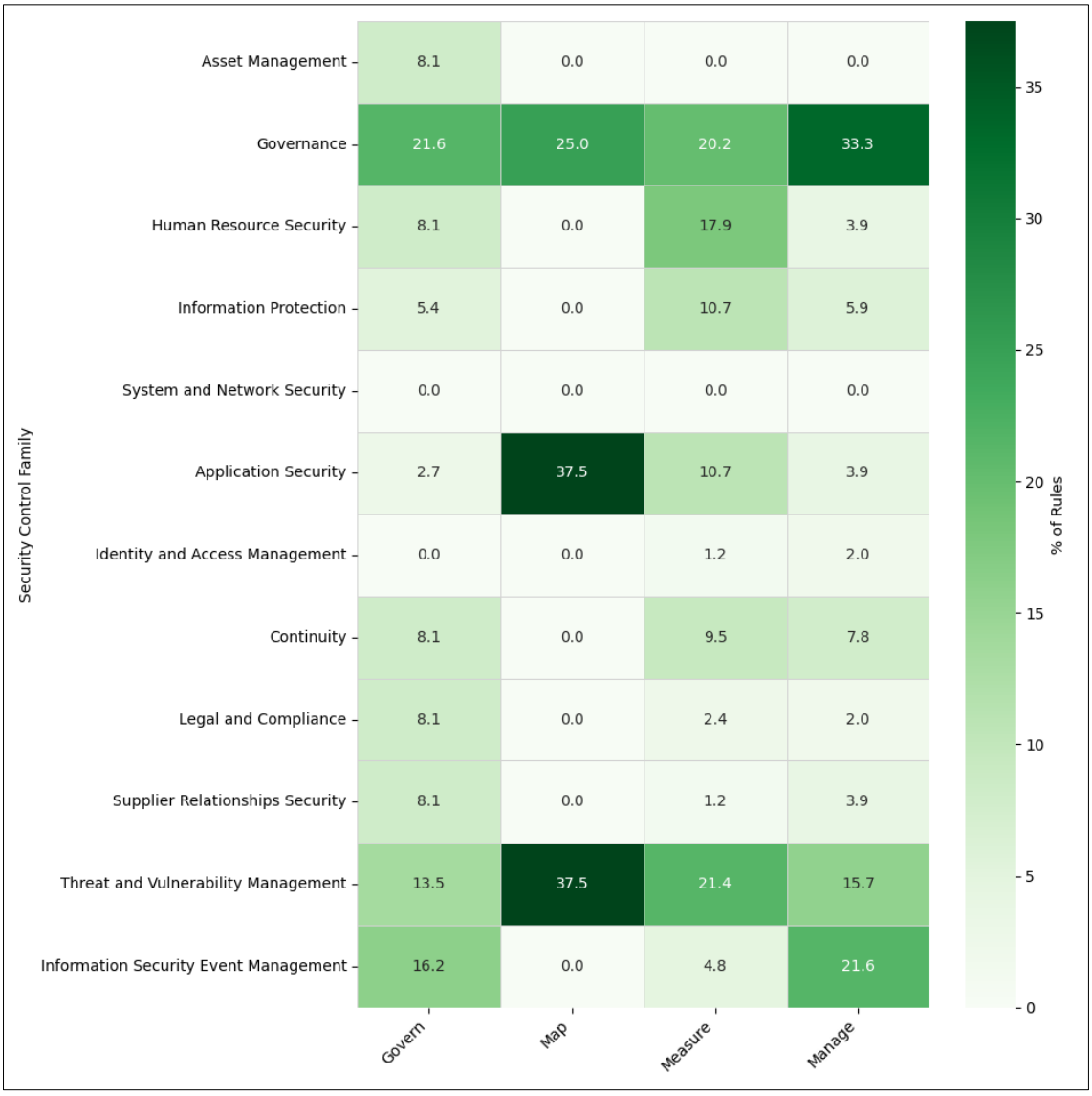

Figure 6: Heatmap of AI RMF Function-to-ISO Operational Capability Alignment

The crosswalk reveals highly uneven coverage. Governance (24.4%), Threat and Vulnerability Management (18.9%), and Information Security Event Management (11.7%) dominate, reflecting the AI RMF's emphasis on policy frameworks, risk monitoring, and incident response. In contrast, System and Network Security (0.0%), Identity and Access Management (1.1%), and Asset Management (1.7%) show weak or nonexistent representation.

This imbalance indicates that while the AI RMF provides strong policy-oriented and monitoring capabilities, it lacks direct operationalization of foundational technical controls

governing system configuration, access rights, and asset lifecycle management. These gaps may pose risks in environments where infrastructure-level defenses are critical for protecting AI systems against adversarial exploitation or unauthorized access.

The function-level breakdown in Table 6 reveals distinct coverage patterns across RMF functions. Governance shows consistent emphasis across all four functions (ranging from 20.2% to 33.3%), with particularly strong representation in Manage (33.3%). In contrast, Map function shows the highest concentration in Application Security (37.5%) and Threat and Vulnerability Management (37.5%). Information Security Event Management appears primarily in Govern (16.2%) and Manage (21.6%) but is absent or minimal in Map and Measure. System and Network Security receives zero coverage across all functions, and Identity and Access Management appears only minimally in Measure (1.2%) and Manage (2.0%).

These patterns underscore that the AI RMF's operational focus centers on organizational oversight and monitoring rather than technical infrastructure controls. While this aligns with the framework's risk management orientation, the absence of system-level security provisions suggests that AI governance efforts may inadvertently emphasize process-level oversight at the expense of technical and infrastructure-level safeguards. Such gaps could pose significant risks in mission-critical AI systems where model provenance, data access, and system-level hardening are essential for defending against adversarial exploitation or privacy compromise.

DISCUSSION

This study demonstrates that Policy-as-Code (PaC) can serve as a powerful enabler for embedding AI governance directly into existing Information Security Management Systems (ISMS), particularly when guided by structured crosswalks between AI-specific risk frameworks and general cybersecurity standards. By analyzing 212 suggested actions from the NIST AI Risk Management Framework (AI RMF 600-1) and aligning them with ISO/IEC 27002:2022 operational capabilities, we establish a replicable model for action formalization, standards interoperability, and technical implementation of AI governance at scale.

Declarability and Automation Feasibility

The high declarability rate (84.9%) of NIST AI RMF actions confirms that most AI governance provisions can be transformed into machine-readable and machine-enforceable policy artifacts. This finding directly addresses our first research objective: determining which AI governance requirements are amenable to automated enforcement through PaC. Our results validate the core hypothesis that AI governance is not inherently too complex for automation, contradicting common assumptions about the inscrutability of AI risk management.

This high automation potential aligns with recent empirical studies demonstrating PaC effectiveness in regulatory compliance contexts. Korrapati (2024) showed that embedding compliance checks directly into CI/CD pipelines using tools like Open Policy Agent (OPA) enables consistent regulatory adherence while minimizing human intervention, achieving near-real-time policy enforcement. Similarly, Webster et al. (2023) demonstrated that OPA-based automated compliance in multi-cloud environments significantly reduces compliance drift and

enables continuous compliance monitoring. Our findings extend these results by showing that the policy-as-code paradigm—proven effective for infrastructure and application security—is equally applicable to AI-specific governance requirements.

The 11.3% of conditionally declarable and 3.8% of non-declarable actions highlight the ongoing need for human-in-the-loop configurations and metadata-enhanced policies, reinforcing the importance of hybrid governance architectures. This finding resonates with Korrapati (2024), who demonstrated that Policy-as-Code frameworks using OPA and Sentinel achieve optimal compliance outcomes when automated policy enforcement is combined with continuous monitoring and validation mechanisms. Such hybrid approaches acknowledge that while automation provides scalability and consistency, human judgment remains essential for nuanced decision-making in ambiguous governance scenarios.

Functional Distribution and PaC Prioritization

The distribution of actions across RMF functions further informs PaC prioritization strategies. The Measure function's dominance (39.4%) offers immediate opportunities for integrating automated monitoring, auditing, and risk evaluation into ISMS workflows. This pattern reflects a broader characteristic of contemporary AI governance frameworks: their emphasis on continuous monitoring and performance assessment rather than solely preventive controls.

The concentration of declarable actions in the Measure function aligns with NIST AI RMF's core philosophy that trustworthy AI requires ongoing validation rather than one-time certification (NIST, 2023). The framework explicitly states that organizations must employ quantitative, qualitative, or mixed-method tools, techniques, and methodologies to analyze, assess, benchmark, and monitor AI risk and related impacts (NIST, 2023). Our findings demonstrate that this measurement-centric approach translates naturally into automated enforcement mechanisms, as evaluation criteria can be codified as binary pass/fail checks suitable for PaC implementation.

Conversely, the Map and Govern functions—while less directly declarable—require richer semantic modeling, taxonomic alignment, and metadata tagging, which are foundational elements in advanced PaC design. Research by Hale and Gamble (2019) on semantic compliance requirement extraction supports this finding, demonstrating that governance-level provisions require sophisticated semantic control hierarchies to capture contextual relationships between policies, actors, and enforcement boundaries. Their work on governance pattern extraction provides methodological precedent for the structured approaches needed to make high-level Map and Govern provisions machine-interpretable.

Control Logic Distribution and Architectural Implications

From a control logic perspective (Section 4.4), the dominance of Detective controls—especially within Measure (75%) and Manage (62.8%)—underscores that PaC is currently best suited to automating monitoring and evaluation logic. This finding has important theoretical implications for understanding the nature of AI governance frameworks and their relationship to traditional cybersecurity control taxonomies.

The emphasis on detective controls reflects a fundamental characteristic of AI risk management: the need for continuous observation and assessment in dynamic environments where model behavior may drift over time (Batool et al., 2025). Traditional cybersecurity frameworks like ISO 27001 emphasize preventive controls (firewalls, access controls, encryption), whereas AI governance necessarily prioritizes detective mechanisms because AI system behavior is probabilistic and context-dependent rather than deterministic.

This detective-heavy distribution aligns with Webster et al.'s (2023) findings that "continuous evaluation significantly reduced the window of exposure caused by compliance violations" in multi-cloud environments. Their research on OPA implementation demonstrated that real-time policy evaluation provides superior security outcomes compared to periodic manual audits. Our results extend this insight to AI-specific contexts, showing that the NIST AI RMF's emphasis on measurement and monitoring translates naturally into automated detective controls.

However, the relatively limited presence of Preventive and Reactive controls signals the need for PaC expansion into policy initiation and escalation workflows. Korrapati (2024) demonstrated that comprehensive compliance automation requires integrating preventive checks (policy-as-code validation before deployment) with detective monitoring (continuous compliance scanning) and reactive responses (automated remediation workflows). Our action-type stratification can guide ISMS architects in choosing appropriate enforcement logic for each RMF function, supporting layered and modular PaC development that combines all three control types.

CIA Triad Alignment and Security Coverage Analysis

Mapping RMF actions to the CIA triad (Section 4.5) shows that Integrity dominates AI-specific policy coverage, particularly in the Map function (21.7%). While this aligns well with PaC's strength in ensuring traceability and auditability (Jothimani, 2022), the underrepresentation of Confidentiality and Availability suggests that PaC implementations must incorporate broader data protection and continuity enforcement.

The Integrity emphasis reflects a core challenge in AI governance: ensuring that model outputs, training data, and decision processes remain trustworthy and unmanipulated. As the NIST AI RMF (NIST, 2023) emphasizes, AI trustworthiness characteristics include being valid and reliable and maintaining data integrity throughout the AI lifecycle. Vakhula et al. (2024) demonstrated that Security-as-Code approaches using Infrastructure as Code (IaC) frameworks naturally enforce integrity through version control, audit trails, and immutable infrastructure patterns—capabilities that translate directly to AI governance contexts.

However, the limited coverage of Confidentiality and Availability exposes potential blind spots in AI-centric PaC implementations. This gap is especially critical for mission-critical systems where adversarial attacks (threatening Confidentiality through model extraction or membership inference) or service disruptions (threatening Availability through denial-of-service or resource exhaustion) could undermine AI system trust and usability. As Obisesan (2024) note in their analysis of cybersecurity risk management for AI systems, both NIST and ISO frameworks emphasize the need to protect against loss of confidentiality, integrity and availability of information, yet our findings suggest that current AI governance action sets inadequately address two of these three pillars.

Organizations implementing AI-driven ISMS must therefore supplement NIST AI RMF actions with additional controls specifically targeting Confidentiality (e.g., model parameter protection, training data access controls) and Availability (e.g., redundancy mechanisms, performance monitoring, resource allocation policies). This supplementation strategy aligns with the broader observation by Batool et al. (2025) that existing cybersecurity frameworks like ISO 27001/27002 only partially address AI-specific security challenges, necessitating both new controls and modifications to existing ones.

Standards Crosswalk and Operational Integration

Our crosswalk between NIST AI RMF actions and ISO/IEC 27002 operational capabilities provides some of the most actionable insights for real-world PaC integration. Strong alignment with Governance (24.4%), Threat Management (18.9%), and Event Management (11.7%) domains supports PaC application in organizational control and risk oversight. These alignment patterns reveal where AI governance naturally complements existing ISMS capabilities and where critical gaps exist.

The strong Governance alignment confirms that NIST AI RMF's organizational and policy-level provisions map well to ISO 27002's governance domains, supporting integrated risk management approaches. This finding resonates with the NIST AI RMF's own guidance (NIST, 2023), which explicitly states that privacy and cybersecurity risk management considerations and approaches are applicable in the design, development, deployment, evaluation, and use of AI systems and that organizations should leverage available standards and guidance including ISO 27001 and related frameworks.

However, the absence or near absence of mappings to System and Network Security, Identity and Access Management, and Asset Management signals a serious blind spot: PaC logic derived from AI RMF actions is insufficient for enforcing foundational ISMS capabilities unless supplemented with additional cybersecurity controls. This gap has been identified by other researchers examining AI governance frameworks. Batool et al. (2025) note that while frameworks like NIST AI RMF and ISO/IEC 27001:2022 focus on a structured, process-oriented approach to managing information security, they differ significantly in coverage, with ISO 27001 providing more comprehensive technical controls while AI RMF emphasizes risk assessment and governance processes.

The function-to-capability heatmap (Figure 6) confirms that Governance and Security Event Management receive emphasis across multiple RMF functions, while Access Control, System Security, and Asset Lifecycle Management remain poorly covered. This distribution pattern reflects a fundamental architectural difference between AI governance frameworks and traditional ISMS standards: AI RMF prioritizes risk identification and measurement over technical enforcement mechanisms, whereas ISO 27002 provides detailed operational controls for infrastructure protection.

This imbalance highlights a key implementation risk: AI-centric PaC may reinforce policy and evaluation mechanisms without providing sufficient coverage of technical and infrastructure safeguards. For PaC to be a reliable security enforcement mechanism in AI-integrated ISMS, it

must bridge this operationalization gap by combining AI RMF governance provisions with ISO 27002 technical controls. Jothimani (2022) demonstrated that Policy-as-Code can automate compliance evaluation against industrial standards through declarative configuration, supporting our findings on the feasibility of standards-aligned automated AI governance.

Practical Implications and Implementation Guidance

Our findings directly address the practical challenge identified by multiple researchers: how to operationalize abstract AI governance principles within existing organizational security infrastructures. Webster et al. (2023) noted that the distributed and heterogeneous nature of multi-cloud environments exacerbates the difficulty in maintaining consistent compliance with internal policies and external regulatory mandates. Our framework provides a systematic methodology for addressing this challenge in AI-specific contexts by:

1. Establishing declarability criteria that enable organizations to systematically evaluate which AI governance provisions can be automated, rather than relying on ad-hoc judgments about automation feasibility.
2. Stratifying enforcement logic by RMF function, system layer, and control type to guide modular PaC design, allowing organizations to prioritize automation efforts based on technical feasibility and risk priorities.
3. Identifying operational domains (e.g., Governance, Threat Management) that are PaC-ready versus those (e.g., Network Security, Access Management) requiring complementary control logic, enabling targeted resource allocation for hybrid governance architectures.
4. Demonstrating that semantic modeling, not keyword matching, is critical for accurate, standards-compliant action generation—a finding supported by Hale and Gamble's (2019) work on semantic compliance requirement extraction.

These results advance the feasibility of machine-executable AI governance by providing a granular blueprint for PaC design and ISMS integration. As AI systems become more pervasive and regulatory scrutiny increases (with frameworks like the EU AI Act mandating specific controls for high-risk AI systems), this approach enables organizations to implement traceable, auditable, and standards-aligned governance policies using automation-first security architectures.

The framework's emphasis on crosswalk analysis between AI-specific and general cybersecurity standards addresses a critical gap identified by Kreutz and Jahankhani (2024), who found that most AI security challenges are either not addressed or only partially addressed by current ISO standards. By providing explicit mappings between NIST AI RMF actions and ISO 27002 capabilities, we enable organizations to systematically identify where existing ISMS controls can be extended to cover AI risks and where new controls must be developed.

Connection to Research Objectives

Returning to our original research objectives, this study successfully:

1. Determined which AI governance requirements are amenable to automated enforcement through systematic application of declarability criteria (R1-R4), identifying 84.9% of NIST AI RMF actions as directly declarable.
2. Stratified actions by multiple dimensions (RMF function, system layer, control logic, CIA alignment) to reveal enforcement patterns and guide PaC architecture decisions.
3. Established semantic crosswalks between AI governance frameworks and operational security standards, demonstrating how abstract governance principles can be mapped to concrete ISMS capabilities.
4. Provided a reproducible methodology for transforming AI governance frameworks into machine-actionable controls, with all analysis code, classification rubrics, and mapping matrices publicly available for validation and extension.

These contributions directly address the core problem articulated in our introduction: the lack of systematic methods to operationalize AI governance frameworks into machine-executable security controls within existing ISMS architectures.

CONCLUSION

This research addresses a critical gap in AI governance implementation by developing the first systematic methodology for transforming abstract AI risk management principles into machine-executable security controls within existing ISMS architectures. Through comprehensive analysis of all 212 NIST AI RMF actions, we demonstrate that 84.9% of AI governance provisions are amenable to automated enforcement through Policy-as-Code, challenging the assumption that AI governance is inherently too complex for automation.

Our multi-dimensional classification framework reveals important patterns in how AI governance maps to operational security capabilities. The dominance of Detective controls (75% in Measure function) and Integrity protections (21.7% in Map function) reflects AI governance's emphasis on continuous monitoring and data trustworthiness. However, critical gaps in Preventive and Reactive controls, as well as underrepresentation of Confidentiality and Availability protections, highlight areas where organizations must supplement AI RMF provisions with additional cybersecurity controls.

The semantic crosswalk between NIST AI RMF and ISO/IEC 27002:2022 provides practical guidance for ISMS integration, revealing strong alignment in Governance (24.4%) and Threat Management (18.9%) domains while exposing significant coverage gaps in System and Network Security, Identity and Access Management, and Asset Management. These findings enable organizations to strategically combine AI-specific governance with foundational cybersecurity controls, creating comprehensive security architectures for AI-driven systems.

This work makes three key contributions to the field. First, we establish reproducible declarability criteria (R1-R4) that enable systematic evaluation of automation feasibility for governance provisions. Second, we provide empirically-grounded stratification of AI governance actions by function, system layer, control logic, and security objectives, offering a blueprint for modular PaC design. Third, we demonstrate through rigorous crosswalk analysis that semantic modeling—not keyword matching—is essential for accurate standards alignment, with implications for both research methodology and practical implementation.

Limitations of this study include reliance on binary classification criteria that may oversimplify nuanced cases, single-reviewer validation that limits inter-rater reliability assessment, and focus on NIST AI RMF that may not generalize to other AI governance frameworks without validation. Additionally, our analysis reflects the current state of NIST AI 600-1; future framework updates may alter declarability patterns and require reanalysis.

Future research should extend this methodology to other AI governance frameworks, including the EU AI Act and ISO/IEC 42001, to validate generalizability and identify framework-specific automation patterns. Empirical validation through real-world PaC implementations would provide valuable insights into practical feasibility and performance characteristics. Development of automated tools for declarability assessment and standards crosswalk would enhance scalability and reduce manual effort in governance operationalization. Finally, investigation of hybrid approaches that combine automated policy enforcement with AI-powered adaptive learning could address the 15.1% of actions requiring human judgment or contextual interpretation.

As AI systems become increasingly embedded in critical infrastructure and regulatory frameworks mandate stronger governance controls, the ability to systematically operationalize abstract governance principles into enforceable security policies becomes essential. This research provides organizations with a practical, validated pathway to achieve that operationalization, enabling trustworthy AI deployment through automation-first security architectures that are both standards-aligned and auditable.

DATA AND CODE AVAILABILITY

The research results data and supporting code snippets for this study are publicly available on GitHub at https://github.com/shadyxur/AI_ISMS. A persistent, version-controlled archive of the materials corresponding to this publication is also available via Zenodo (DOI: https://doi.org/10.5281/zenodo.16620379).

REFERENCES

Akhtar, Z. B., & Rawol, A. T. (2024). Enhancing cybersecurity through artificial intelligence (AI)-powered security mechanisms. *IT Journal Research and Development (ITJRD)*, *9*(1). https://doi.org/10.25299/itjrd.2022.16852

Al-Dhahri, S., Al-Sarti, M., & Abdul, A. (2017). Information security management system. *International Journal of Computer Applications*, *158*(7), 29-33.

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable artificialiIntelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion*, *99*, 101805. https://doi.org/10.1016/j.inffus.2023.101805

Badman, A. (2024). *What is AI risk management?* IBM. Retrieved June 30, 2025, from https://www.ibm.com/think/insights/ai-risk-management

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. *AI and Ethics, 5*(3), 3265–3279. https://doi.org/10.1007/s43681-024-00653-w

CSA (2025). *AI Controls Matrix*. Cloud Security Alliance. https://cloudsecurityalliance.org/artifacts/ai-controls-matrix

Flehmig, N., Lundteigen, M. A., & Yin, S. (2024). Implementing artificial intelligence in safety-critical systems during operation: Challenges and extended framework for a quality assurance process. *IECON 2024 - 50th Annual Conference of the IEEE Industrial Electronics Society*. https://doi.org/10.1109/IECON55916.2024.10906021

Fok, R., & Weld, D. S. (2024). In search of verifiability: Explanations rarely enable complementary performance in AI-advised decision making. *AI Magazine*, *45*(3), 317-332. https://doi.org/10.1002/aaai.12182

Habbal, A., Ali, M. K., & Abuzaraida, M. A. (2024). Artificial intelligence trust, risk and security management (AI TRiSM): Frameworks, applications, challenges and future research directions. *Expert Systems with Applications, 240*, 122442. https://doi.org/10.1016/j.eswa.2023.122442

Hale, M. L., & Gamble, R. F. (2019). Semantic hierarchies for extracting, modeling, and connecting compliance requirements in information security control standards. *Requirements Engineering*, *24*, 365–402. https://doi.org/10.1007/s00766-017-0287-5

Hind, M. (2020). *IBM FactSheets Further Advances Trust in AI*. IBM. Retrieved July 9, 2025, from https://research.ibm.com/blog/aifactsheets

ISO (2018). *ISO/IEC 27000:2018. Information technology — Security techniques — Information security management systems — Overview and vocabulary (5th ed.)*. International Organization for Standardization. International Electrotechnical Commission.

ISO (2022-a). *ISO/IEC 27001:2022. Information security, cybersecurity and privacy protection — Information security management systems — Requirements* (3rd ed.). International Organization for Standardization. International Electrotechnical Commission.

ISO (2022-b). *ISO/IEC 27002:2022. Information security, cybersecurity and privacy protection — Information security controls* (3rd ed.). International Organization for Standardization. International Electrotechnical Commission.

ISO (2023). *ISO/IEC 42001:2023. Information technology — Artificial intelligence — Management system* (1st ed.). International Organization for Standardization. International Electrotechnical Commission.

Jada, I., & Mayayise, T. O. (2024). The impact of artificial intelligence on organizational cyber security: An outcome of a systematic literature review. *Data and Information Management, 8*(2), 100063. https://doi.org/10.1016/j.dim.2023.100063

Jeffy, M., & Bello, S. (2025). *AI governance in RPA: Ensuring compliance and transparency in automated decisions*. ResearchGate. Retrieved July 4, 2025, from https://www.researchgate.net/publication/391633567_AI_Governance_in_RPA_Ensuring_Compliance_and_Transparency_in_Automated_Decisions

Jothimani, A. P. (2022). *Enabling secure cloud governance using policy as code* [Master's thesis, Chalmers University of Technology. University of Gothenburg].

Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion, 97*, 101804. https://doi.org/10.1016/j.inffus.2023.101804

Kordjamshidi, P., Roth, D., & Kersting, K. (2022). Declarative learning-based programming as an interface to AI systems. *Frontiers in artificial intelligence*, *5*, 755361. https://doi.org/10.3389/frai.2022.755361

Korrapati, R. (2024). Automating compliance in CI/CD pipelines: A modern software development framework. *Available at SSRN 5139607*. https://dx.doi.org/10.2139/ssrn.5139607

Kreutz, H., & Jahankhani, H. (2024). Impact of artificial intelligence on enterprise information security management in the context of ISO 27001 and 27002: A tertiary systematic review and comparative analysis. In H. Jahankhani, G. Bowen, M. S. Sharif, & O. Hussien (Eds.), *Cybersecurity and artificial intelligence* (pp. 1–34). Springer. https://doi.org/10.1007/978-3-031-52272-7_1

Kunle-Lawanson, O. (2022). The role of Al in information security risk management. *World Journal of Advanced Engineering Technology and Sciences*, *7*(2), 308-319. https://doi.org/10.30574/wjaots.2022.7.2.0128

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys*, *55*(9). https://doi.org/10.1145/3555803

Malik, A., Arshid, K., Noonari, N., & Munir, R. (2025). Artificial intelligence-driven cybersecurity framework using machine learning for advanced threat detection and prevention. *Scholars Journal of Engineering and Technology*. https://doi.org/10.36347/sjet.2025.v13i06.005

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *FAT* '19: Conference on Fairness, Accountability, and Transparency, January 29–31, 2019, Atlanta, GA, USA*. https://doi.org/10.1145/3287560.3287596

Mohamed, N. (2023). Current trends in AI and ML for cybersecurity: A state-of-the-art survey. *Cogent Engineering*, *10*(2). https://doi.org/10.1080/23311916.2023.2272358

NIST (2021). *NIST AI 800-204B. Attribute-based access control for microservices-based applications using a service Mesh* U.S. Department of Commerce. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-204B

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

NIST (2024). *NIST AI 600-1. Artificial intelligence risk management framework: Generative artificial intelligence profile*. U.S. Department of Commerce. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

Obisesan, S. M. (2024). Integrating artificial intelligence and cybersecurity frameworks: Challenges and opportunities in e-commerce cybersecurity management. *Available at SSRN 5070108*. https://dx.doi.org/10.2139/ssrn.5070108

Pigola, A., & De Souza Mierelles, F. (2024). Unraveling trust management in cybersecurity: Insights from a systematic literature review. *Information Technology and Management*. https://doi.org/10.1007/s10799-024-00438-x

Polemi, N., Praça, I., Kioskli, K., & Bécue, A. (2024). Challenges and efforts in managing AI trustworthiness risks: a state of knowledge. *Frontiers in Big Data*, *7*. https://doi.org/10.3389/fdata.2024.1381163

Polito, C., & Pupillo, L. (2024). Artificial intelligence and cybersecurity. *Intereconomics*, *59*(1), 10-13. https://doi.org/10.2478/ie-2024-0004

Raja, A. K., & Zhou, J. (2023). AI accountability: approaches, affecting factors, and challenges. *Computer*, *56*(4), 61–70. https://doi.org/10.1109/MC.2023.3238390

Salako, A. O., Fabuyi, J. A., Aideyan, N. T., Selesi-Aina, O., Dapo-Oyewole, D. L., & Olaniyi, O. O. (2024). Advancing information governance in AI-driven cloud ecosystem: Strategies for enhancing data security and meeting regulatory compliance. *Asian Journal of Research in Computer Science, 17*(12), 66–88. https://doi.org/10.9734/ajrcos/2024/v17i12530

Vakhula, O., Kurii, Y., Opirskyy, I., & Susukailo, V. (2024). Security-as-code concept for fulfilling ISO/IEC 27001:2022 requirements. *Proceedings of the Workshop Cybersecurity Providing in Information and Telecommunication Systems (CPITS 2024)*, *3654*. https://ceur-ws.org/Vol-3654/paper6.pdf

Webster, N., Burton, A., Hawkins E., Pum, M., & Watson, J. (2023). *automated compliance checks with open policy agent (OPA) in multi-cloud*. ResreachGate. Retrieved July 9, 2025, from https://www.researchgate.net/publication/392163378_Automated_Compliance_Checks_with_Open_Policy_Agent_OPA_in_Multi-Cloud