# Toward Identifying Role-Aligned AI Governance Needs in Mixed-Trust Environments

**Bijesh Shrestha, Nicholas B. Harrell, Matthew L. Corbett and Michael D. Quigg**
Army Cyber Institute at West Point, New York, USA

bijesh.shrestha@westpoint.edu
nicholas.harrell@westpoint.edu
matthew.corbett@westpoint.edu
michael.quigg@westpoint.edu

**Abstract:** The rapid emergence of AI is driving governance challenges in security-sensitive organizational settings. For example, deploying AI in these contexts typically requires embedding automated or semi-automated decision-making into mixed-trust environments characterized by legacy systems, uneven access to and visibility into governance-relevant artifacts, and strict operational, security, and compliance constraints. In such environments, governance decisions, system design, and day-to-day operations are often handled by distinct organizational roles, leading to different expectations and dependencies among stakeholders. Despite growing work on AI governance, there is limited, role-specific guidance that clarifies what governance leaders and oversight bodies should require, what engineers and implementers must ensure, and what operators and system administrators can rely on when deploying and operating AI in mixed-trust environments. This leaves organizations without a shared basis for translating general governance principles into concrete deployment requirements, compliance controls, risk assessments, and ongoing operational accountability. Current and emerging AI governance and risk management frameworks largely provide lifecycle-based approaches to AI risk, establishing a common vocabulary and baseline for AI use within organizations. However, they provide limited guidance on how responsibilities should be interpreted and enacted across organizational roles, particularly in constrained, security-sensitive deployments. In this paper, we analyze widely adopted AI governance and risk-management frameworks and prior synthesis literature through the lens of three stakeholder layers: governance leaders and oversight bodies; engineers and implementers; operators and system administrators. For each layer, we examine what current frameworks explicitly and implicitly address and where role expectations remain underspecified. Building on this analysis, we propose an initial set of role-aligned governance needs and highlight cross-cutting socio-technical factors that shape the enactment of AI governance in practice. The goal is to inform future research, policy work, standards development, and operational planning by moving AI adoption in mixed-trust contexts from aspirational frameworks toward practical, accountable, and auditable implementations.

**Keywords:** AI governance, Audit frameworks, Trust, Risk-management, Mixed-trust environments

## 1. Introduction

Artificial intelligence (AI) capabilities are increasingly integrated into organizational workflows, supporting tasks such as analytics, decision support, automation, and reporting. As these capabilities mature, organizations are deploying AI in more complex and higher-stakes settings, where systems must operate reliably within existing infrastructure, policies, and operational constraints. Existing governance frameworks, standards, and policies establish a foundation for addressing AI risks and building trust through transparency, accountability, and responsible use. However, they offer little guidance on how lifecycle-based governance maps to organizational roles and responsibilities.

Building on definitions of governance, IT governance, and prior research on AI in organizations, we define *AI governance* as an organization's allocation of decision rights and accountability for AI systems to align AI-related activities with organizational objectives. AI governance differs from traditional IT governance because AI systems can exhibit autonomy, learning, and inscrutability, complicating assurance, oversight, and accountability (Berente et al. 2021). These challenges are amplified in *Mixed-trust environments,* organizational and operational contexts in which AI capabilities are deployed across multiple access and control regimes, resulting in uneven access to governance-relevant artifacts across roles due to security, regulatory, or organizational constraints. In such settings, no single role has sufficient visibility into the evidence required to justify decisions, assess risk, or demonstrate compliance. Governance leaders may be accountable for deployment or risk acceptance without direct access to technical or operational evidence, while engineers and operators may generate artifacts they cannot fully contextualize or escalate across access boundaries. Common examples include enclave segmentation, cross-domain transfer constraints, differential audit visibility, and contractor or partner access boundaries (Rose et al., 2020; Goswami 2021).

Despite a growing body of AI governance frameworks and guidance, prior research indicates that organizations receive limited practical direction on how responsibilities should be interpreted and enacted across decision-making, implementation, and operations. Reviews of responsible AI governance tools show limited attention to

governance leaders, deployers, and operators, complicating role alignment across AI lifecycle stages (Kuehnert et al. 2025), while systematic reviews of stakeholder motivations suggest that many guidelines remain disconnected from concrete organizational decision-making (Heymans and Heyman 2024).

In this paper, we analyze AI governance through three stakeholder lenses: (1) governance leaders and oversight bodies, who set policy and risk expectations; (2) engineering and implementation personnel, who translate those expectations into design, documentation, and technical controls; and (3) operators and system administrators, who operate and monitor systems to maintain alignment with policy and infrastructure constraints. Prior work shows that governance gaps often arise when guidance is not clearly connected to such role-specific concerns (Kuehnert et al. 2025; Heymans and Heyman 2024). While existing reviews identify these gaps, they do not decompose them by organizational role or examine how mixed-trust conditions amplify them. This paper breaks the problem down by role: what frameworks tell each role to do, what they assume each role can do, and what they leave unclear, particularly in mixed-trust settings where no single role sees the full picture. Accordingly, we synthesize widely used AI governance guidance to examine how current frameworks address these roles and to derive an initial set of role-aligned governance needs for mixed-trust environments.

## 2. Problem Context and Related Work

For our review, we assembled AI governance materials spanning research, standards, and practitioner-facing guidance. We found that AI governance is defined broadly and inconsistently across sources, complicating systematic comparison and obscuring how high-level governance expectations translate into role-specific responsibilities. This lack of role-level specificity is consequential in practice: when a developer is instructed to "be accountable" or "manage risk," it is often unclear what artifacts they should produce and how such evidence should persist beyond deployment—reflecting the recurring "who does what, when, and how" problem identified in prior reviews (Kuehnert et al. 2025; Batool et al. 2024). These observations motivate our role-based approach, which examines what existing guidance explicitly assigns to each role, what it implies, and what remains underspecified, particularly in mixed-trust environments where evidence, access, and accountability are fragmented across organizational boundaries.

### 2.1 Related Work

We organize related work into four clusters that inform the paper's role-based analysis: (1) AI governance frameworks and synthesis reviews, (2) role and stakeholder-centered governance needs, (3) evidence, audit, and accountability practices, and (4) mixed-trust environments where segmented access complicates oversight and coordination. Together, these clusters situate our analysis and foreground where role-specific responsibilities remain underspecified.

#### *2.1.1 Cluster 1: Frameworks and baselines*

AI governance frameworks are increasingly moving from voluntary to mandatory instruments. Across both forms, a recurring characteristic is abstraction at the level of organizational roles and responsibilities. Voluntary frameworks such as the NIST AI Risk Management Framework (AI RMF) adopt a lifecycle orientation—Govern, Map, Measure, and Manage—and emphasize accountability and auditability across governance, data, performance, and monitoring (NIST 2023b; NIST 2024), as reinforced by GAO's accountability framing, without prescribing how responsibilities should be distributed across governance, engineering, and operational roles (U.S. GAO 2021). Mandatory instruments are characterized by clearer definitions of roles and responsibilities. The EU's AI Act, for example, defines "providers" and "deployers" and specifies documentation and post-deployment monitoring obligations (EU 2024). ISO/IEC 42001 similarly defines requirements for AI management systems while leaving internal role assignments to implementers (International Organization for Standardization 2023). International instruments such as the OECD AI Recommendation remain more abstract and are intended to be operationalized through downstream instruments (OECD 2024). U.S. federal AI governance illustrates structural instability, with three major policy shifts in 15 months as Executive Orders 14110 and 14179 mandated and rescinded AI governance requirements (Executive Office of the President, 2023; 2025), accompanied by corresponding shifts in agency-level implementation guidance (Office of Management and Budget, 2024; 2025). Schuett's Three Lines of Defense framework represents progress toward role specialization but still lacks a clear demarcation between governance bodies, engineers, and operators (Schuett, 2023).

#### *2.1.2 Cluster 2: Roles and responsibility*

Governance frameworks provide structural guidance, yet a central question remains poorly specified: who is responsible for implementing governance requirements across different stages of the AI lifecycle defined in

prevailing frameworks? A systematic review of responsible AI governance found that only five of 61 studies explicitly addressed core governance questions—who does what, when, and how (Batool et al. 2024). Similarly, an analysis of more than 220 responsible AI tools showed a strong focus on designers and developers, with limited support for governing authorities, deployers, and users (Kuehnert et al. 2025). Although frameworks such as the NIST AI RMF and ISO/IEC 42001 are intentionally broad, existing work offers little guidance on how governance responsibilities should be assigned in practice. Prior research shows that unclear role definitions can enable more powerful actors to shift responsibility onto contractors, users, or subordinate roles, particularly at organizational interfaces (Heymans and Heyman 2024). While the role-definition problem is increasingly recognized, how accountability is coordinated across governance, engineering, and operational layers remains underexamined, especially in mixed-trust environments.

#### *2.1.3 Cluster 3: Trust, reliance, and oversight in human-AI teams*

Governance must account for how trust, reliance, and supervision shape the gap between formal accountability and practice. Prior work argues that trust in AI systems should be deliberately designed rather than assumed to emerge from system performance, shifting attention from controlling AI behavior to understanding how humans perceive, interpret, and rely on AI outputs (Ezer et al. 2019). Empirical studies further show that affective trust in human–AI teams is consistently lower than in human-only teams, even when cognitive trust is comparable, undermining assumptions that performance alone will generate appropriate reliance (Ulfert et al. 2024; Georganta and Ulfert 2024). The CHAI-T model highlights the need to continuously adjust trust throughout a system's development and operational contexts, a challenge that becomes more pronounced when stakeholders operate with uneven access across different stages of deployment (McGrath et al. 2025). The most applicable study in this case highlights the need for four factors to ensure effective human oversight of an AI system: sufficient levels of causal efficacy, epistemic access, self-regulation, and intention aligned with the goals set for the system or algorithm in question (Sterz et al., 2024). These four different factors combine to create the concept of moral responsibility—a feature that cannot be achieved in most cases with humans in the loop systems due to automation bias and access levels (Parasuraman and Manzey, 2010).

#### *2.1.4 Cluster 4: Assurance, auditability, and verifiability*

The literature identifies evidence requirements for AI assurance and auditability but provides limited guidance on how responsibilities are divided among governance leaders, engineers, and operators. Audit-enabling attributes—traceability, accessibility, reproducibility, transparency, understandability, and explainability—are identified across lifecycle stages but remain role-agnostic and do not specify duties by role category (Li and Goel, 2024). Empirical evidence further indicates that new competencies are required for AI-driven process auditing, particularly in data governance and explainability, as most IT auditors are not equipped to address these demands (Li and Goel, 2025). Other approaches link safety, security, and performance evidence through artifact hierarchies but rely on organizational structures that do not translate to mixed-trust environments (Kapusta et al., 2025). Related work also argues for governance structures that distinguish between pre-deployment and post-deployment assurance, a distinction not addressed by other frameworks (Kwiatkowska and Zhang, 2023). Across this literature, three gaps persist: unclear role-level division of assurance responsibilities, unclear access requirements for auditability, and limited guidance on sustaining these requirements across governance, engineering, and operational categories.

## 3. Methodology

### 3.1 Selection of Governance Sources and Synthesis Reviews

We use exploratory role-based qualitative content analysis (Krippendorff, 2018) to examine how role expectations are specified, implied, or left ambiguous in AI governance guidance. Governance texts are treated as analytical artifacts through which expectations about organizational roles and associated outputs are articulated. Coding focuses on: (1) the targeted role (governance leaders and oversight bodies; engineers and implementers; operators and system administrators), (2) expected outputs (e.g., documentation, logs, tests, monitoring), and (3) whether boundaries between decision-making, implementation, and operations are explicit or ambiguous.

Rather than conducting an exhaustive systematic review, we focus on established governance sources and recent synthesis reviews for exploratory analysis (Coners and Matthies, 2014; Aromataris et al., 2024). We use two complementary source bins: (1) primary governance and risk guidance (frameworks, standards, and policies

referenced in organizational practice), and (2) prior synthesis reviews summarizing role-related governance challenges and stakeholder needs.

To identify candidate sources and keep the scope grounded, we conducted a literature search between November 2025 and January 2026 using Semantic Scholar, Scopus, and OpenAlex with the keywords 'AI,' 'artificial intelligence,' 'governance,' 'regulation,' 'policy,' 'ethics,' and 'framework,' yielding 2,020 studies. After removing 122 duplicates and restricting publications from this period, 126 studies remained. Three were excluded due to inaccessibility, leaving 123 studies for screening. Inclusion criteria were published 2018–2025, English-language, addressing AI governance or risk management at an organizational or policy level, and accessible in full text. Studies were excluded if they were purely technical without a governance framing, or editorials and commentaries under three pages. Titles and abstracts were manually reviewed for governance-related terms (e.g., governance, policy, regulation, accountability, oversight, framework, compliance). We further filtered to studies with at least 5 citations, yielding 14 studies. A supplementary Google search for 'AI governance' identified 4 additional publications, resulting in 18 studies for inclusion.

To support analytical organization, sources were grouped into clusters aligned with governance roles and analytical purpose, each anchored by a conceptually representative "seed" source. We applied BERT-based topic modeling (BERTopic) to the abstracts of the 123 screened sources, generating candidate topic clusters that served as a navigational heuristic for grouping sources by thematic affinity during early organization. These clusters did not determine final cluster assignments, source inclusion decisions, or code assignments.

We apply a three-role analytical lens to code how governance guidance assigns—or implies—expectations across AI deployment work (Deshpande and Sharp, 2022; Kuehnert et al., 2025; Batool et al., 2024). The unit of analysis is a discrete governance statement or recommendation, interpreted in its surrounding textual context. Coding interpretations were iteratively refined through comparison across sources, with attention to boundary cases. For consistency across heterogeneous sources, each recommendation was coded against the following role layers:

- *Governance leaders and oversight bodies:* set policy and governance expectations, define acceptable risk and use constraints, approve deployment conditions, and determine evidence requirements for compliance and oversight.
- *Engineers and implementers:* translate governance expectations into system design and implementation decisions, produce required documentation, and build technical and procedural controls.
- *Operators and system administrators:* deploy, operate, and maintain systems in real environments, monitor performance and incidents, manage access and configuration, and ensure availability of audit-relevant records.

## 4. Role-Based Analysis of Governance Frameworks

This section analyzes and synthesizes AI governance sources through a role-based lens to surface role-aligned governance needs and recurring gaps across governance, engineering, and operational roles. Figure 1 illustrates how governance functions flow from governance leaders to engineering and operational roles within a mixed-trust environment and where underspecified needs emerge at the interfaces between these layers. Figure 2 consolidates the findings of this role-based analysis into a structured mapping of governance needs across the three role layers.

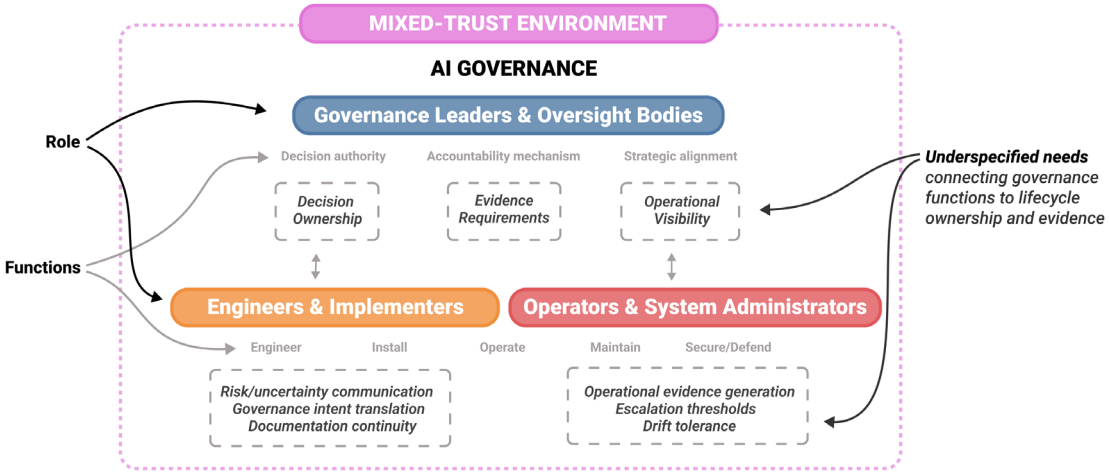


**Figure 1: The figure illustrates how AI governance is exercised within mixed-trust environments by governance leaders and oversight bodies through decision authority, accountability mechanisms, and strategic and policy alignment, and how these governance functions interact with engineering, implementation, operation, and system administration activities. For each organizational role, the figure highlights areas that are commonly underspecified in existing AI governance frameworks**

### 4.1 Governance Leaders and Oversight Bodies

*What is most explicit* across AI governance sources is that governance leaders and oversight bodies are responsible for establishing and sustaining AI governance, including defining risk management programs, organizational policies, and accountability structures. This role is most clearly articulated in the NIST AI RMF's 'Govern' function, which emphasizes organizational oversight, defined processes, and accountability mechanisms (NIST, 2023b). The AI RMF Playbook similarly stresses the definition of human oversight roles as part of operationalizing governance practices (NIST, 2023a), and U.S. GAO's accountability framework reinforces role clarity and accountability mechanisms as central to responsible AI oversight (U.S. GAO, 2021).

*What is implied*, rather than specified, is how decision authority is exercised in practice. Governance sources describe required actions but rarely assign ownership for key decisions such as accepting residual risk, determining deployment readiness, or authorizing system suspension. While the NIST AI RMF Playbook encourages organizations to distinguish among AI actors (e.g., developer, owner, operator), it leaves the allocation of decision ownership to organizational discretion. This pattern aligns with synthesis findings that highlight persistent ambiguity about responsibility even when governance expectations are clear (Kuehnert et al., 2025).

*What remains underspecified* is the evidence basis for accountable, auditable decision-making (Schuett, 2023). The NIST AI RMF offers flexibility as an outcomes-based framework that organizations can adapt to their context, rather than a prescriptive checklist. However, it does not specify what concrete evidence decision-makers should consistently require from engineering and operations at key decision points (e.g., risk acceptance, readiness to deploy, or escalation thresholds) (NIST 2023b).

*In mixed-trust environments*, these gaps are amplified because governance-relevant evidence is fragmented across teams, tools, and access boundaries (Schuett, 2023; Kuehnert et al., 2025). Governance leaders may remain accountable for risk and readiness decisions without direct access to the technical or operational artifacts needed to justify those decisions. Binding regulations for high-risk systems, such as the EU AI Act, partially address this gap by specifying baseline obligations for logging and post-market monitoring and tying them to legally defined actor categories (EU, 2024). However, how governance leaders access and evaluate such evidence across trust boundaries remains largely unspecified.

### 4.2 Engineers and Implementers

*What is most explicit* across AI governance sources is that engineers and implementers are responsible for translating governance intent into system design, technical controls, and documentation. The NIST AI RMF frames risk management as a lifecycle activity that must be operationalized through traceable documentation linking system purpose, data use, development or integration decisions, and accepted performance and risk

trade-offs (NIST, 2023b). For high-risk systems, binding instruments such as the EU AI Act make this translation mandatory by requiring technical documentation prior to deployment and throughout system operation (EU, 2024). U.S. GAO's accountability framework similarly implies that engineers must generate and sustain evidence across governance, data, performance, and monitoring dimensions, not solely model-level metrics (U.S. GAO, 2021).

*What is implied* is that engineers, implementers, and deployers possess sufficient organizational authority and epistemic access to fully document governance intent. In practice, engineers may have deep access to technical artifacts (e.g., code, data, models) but limited visibility into governance considerations such as risk tolerance, policy constraints, or deployment context. Prior synthesis shows that responsible AI tools disproportionately target designers and developers during development stages, reinforcing this imbalance (Kuehnert et al., 2025). Epistemic access is a prerequisite for effective oversight, yet engineers may lack insight into how components are integrated or operated across segmented or classified systems (Sterz et al., 2024).

*What remains underspecified* is how information about uncertainty, limitations, and residual risks is conveyed beyond development. Although governance sources require documentation, they provide little guidance on documentation granularity, continuity, or handoff across roles and lifecycle stages. While auditability attributes such as reproducibility and traceability are well established conceptually (Li and Goel, 2024), empirical work shows that auditors often lack the specialized competencies needed to assess AI-specific artifacts (Li and Goel, 2025).

***In** mixed-trust environments*, these gaps are amplified because documentation and provenance may be created at one trust level but remain inaccessible to operators or auditors at another. This disrupts documentation continuity and weakens sustained review and accountability across roles over time. Governance sources rarely specify what contextual information must persist across trust boundaries, leaving organizations to manage documentation handoffs through informal or ad hoc practices.

### 4.3 Operators and System Administrators

*What is most explicit* across AI governance sources is that operators and system administrators are responsible for post-deployment monitoring and oversight. The EU AI Act requires deployers of high-risk systems to monitor operation, retain system logs, and assign personnel with appropriate competence, authority, and awareness of automation bias (EU, 2024). Similarly, the NIST AI RMF calls for documented thresholds to bypass or deactivate system components (NIST, 2023b), and U.S. GAO's accountability framework treats monitoring and corrective action documentation as core accountability practices (U.S. GAO, 2021). Across both voluntary and binding guidance, continuous monitoring is positioned as a central operational responsibility.

*What is implied* is that operators possess sufficient access, understanding, and authority to detect anomalies and intervene appropriately. Effective oversight presumes epistemic access to system behavior, yet this access may be constrained when system internals are opaque or partitioned across environments. Prior work identifies epistemic access as a condition for meaningful human oversight, while empirical evidence shows that automation complacency can persist under operational load and cannot be mitigated through training alone (Sterz et al., 2024; Parasuraman and Manzey, 2010). GAO acknowledges that "AI systems pose unique challenges to such oversight because their inputs and operations are not always visible" (U.S. GAO 2021).

*What remains underspecified—and central to this paper—*is how operational evidence (e.g., logs, incidents, and performance metrics) should flow back to governance leaders and engineers to support accountable decision-making. While auditability attributes such as traceability, accessibility, reproducibility, transparency, understandability, and explainability are well defined conceptually (Li and Goel, 2024), governance sources provide little guidance on who is responsible for generating, interpreting, and consuming operational evidence across roles. Frameworks frequently reference "specialized competencies," yet do not clarify what those competencies entail in practice (Li and Goel, 2025). As a result, operators may generate logs they cannot fully interpret, while governance leaders may request evidence they cannot directly access. Synthesis reviews confirm that responsible AI tools disproportionately target development stages, leaving deployment and operation comparatively under-supported (Kuehnert et al., 2025). This gap contributes to unclear escalation thresholds, drift tolerance, and documentation standards, making it difficult for organizations to reduce AI and ethics risks in a sustained and meaningful way (Batool et al., 2024).

*In mixed-trust environments*, when governance requirements flow downward, limited shared visibility and fragile handoffs hinder operators and administrators from reporting evidence upward accurately and consistently across enclaves. Operational evidence may be generated within a trust boundary but fail to transfer

in a form accessible to governance leaders or engineers for reporting, risk review, or audit. As a result, escalation, drift management, and compliance follow-through depend on ad hoc coordination rather than role-defined evidence pathways. Figure 2 maps the governance needs surfaced across all three roles, distinguishing between what current frameworks explicitly assign, what they presuppose but do not assign, and what remains unaddressed.

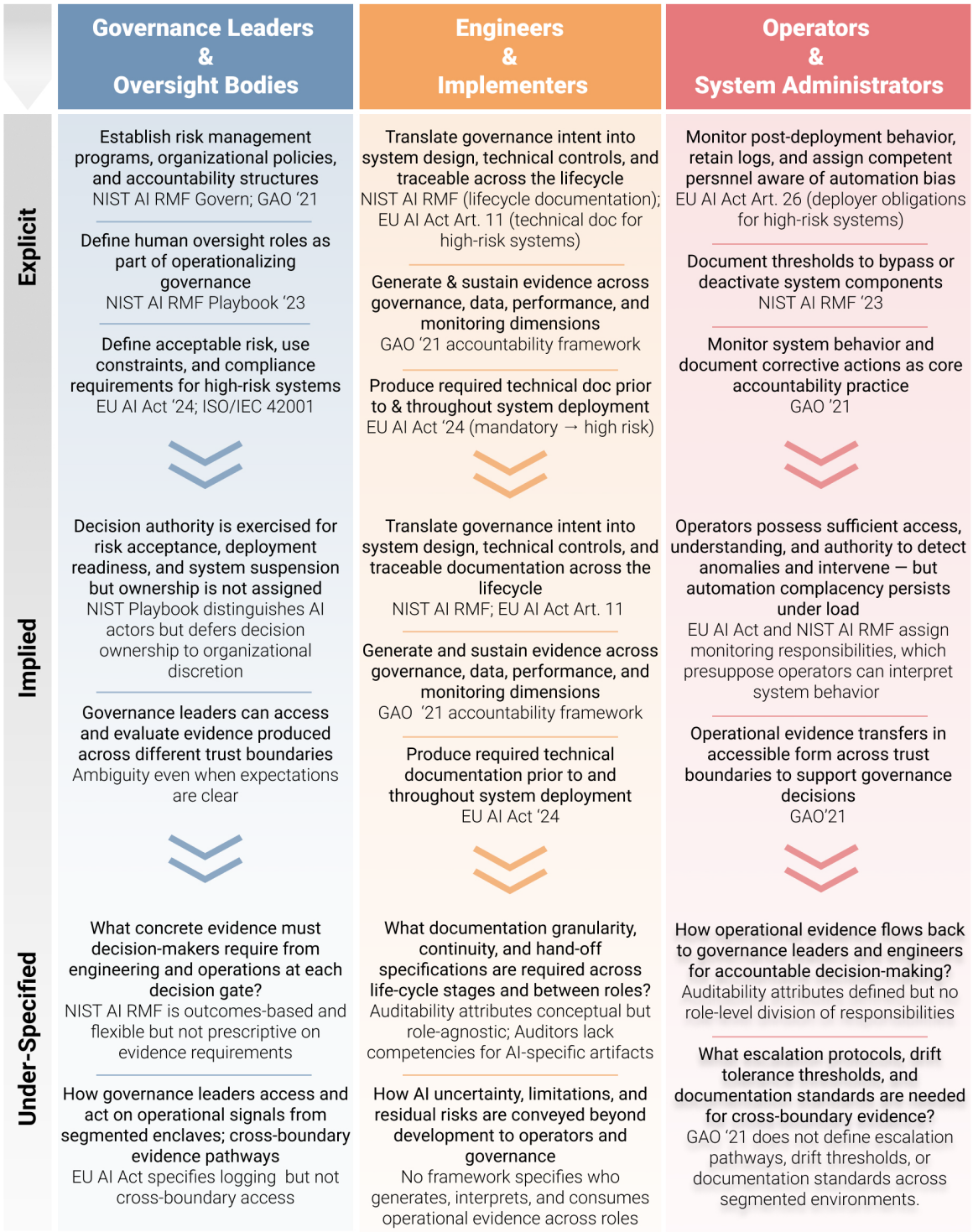


**Figure 2: Role-aligned AI governance needs across three organizational roles, categorized by the degree of specification found in current frameworks from explicit requirements to implied and under-specified gaps**

## 5. Discussion and Future Work

Governance frameworks, policies, and standards establish a common language and overarching principles for AI governance. Still, a persistent challenge remains in effectively translating these high-level mandates into actionable responsibilities across diverse roles. The lack of clear delineation of who is responsible for implementing, monitoring, and enforcing governance across the AI lifecycle can diffuse accountability, with responsibilities shifting among governance bodies, engineers, and operators. Kuehnert et al. (2025) defined this phenomenon as an "ownership problem".

For example, consider a scenario where an AI-enabled tool is developed and validated within one enclave but deployed operationally in another access-controlled enclave. The transfer, provenance metadata, including annotations and boundary-level audit results, may be stripped or degraded to comply with cross-domain constraints. Operators in the receiving environment inherit a system whose known limitations are documented elsewhere and potentially inaccessible, weakening their capacity for meaningful oversight (Sterz et al., 2024). Governance leaders reviewing the deployment may approve continued operation based on summary documentation that omits risk mitigations recorded at the originating level. In this scenario, each role 's evidence basis appears adequate in isolation, yet end-to-end accountability cannot be reconstructed, precisely the kind of fragmentation that current frameworks leave unaddressed.

The ambiguity can diminish the efficacy of governance, underscoring the importance of role-aware requirements that specify who does what, when, and how, as highlighted by recent literature reviews (Kuehnert et al., 2025; Schuett, 2023; Floridi et al., 2018). The ambiguity around guidance and frameworks becomes more evident in evidence for audit and accountability in mixed-trust environments. As noted by Schuett (2023), the same implementers end up generating artifacts for audits, raising questions about audit and trust.

Governance also has a critical socio-technical dimension that many frameworks only acknowledge in passing. Effective documentation and audit mechanisms are essential, but their success heavily relies on organizational factors such as training, management, protocols, and responsibility handoffs (Salako et al., 2024). This paper positions the role-based analysis as a starting point for understanding where these responsibilities and checks break down in practice. By making ownership gaps visible, the goal is to begin a discussion on what must be anchored across roles before auditing and oversight can be reliable in mixed-trust settings.

A limitation of our current approach is that our analysis is scoped to a small set of governance sources. It remains primarily conceptual rather than validated through in-situ organizational practice. Our immediate aim is to surface and frame this role-based accountability problem clearly for the broader community. Future work should aim to explicitly refine role-based mapping across a broader range of frameworks, standards, and operational guidance. This will enable a clearer identification of where requirements are assigned and where gaps or ambiguities in ownership persist. Additionally, a socio-technical study using empirical methods can help understand how evidence moves across roles in mixed-trust settings. It should also capture governance leaders' perspective on what operational signals support oversight and how they would act on them.

## 6. Conclusion

AI governance guidance provides high-level direction but often leaves ambiguity about how responsibilities are distributed among governance leaders, engineers, and operators. Mixed-trust settings further complicate this because evidence and authority are split across enclaves, limiting consistent visibility, segmented access, and legal and organizational restrictions, further complicating accountability. We use a role-based lens to begin understanding where responsibility and evidence pathways remain unclear and to motivate more actionable, role-aware governance requirements that can effectively address these contextual complexities.

**Ethics Declaration**: This research did not involve human participants or the collection of personal data and therefore did not require institutional ethical approval. The study is based on publicly available documents, doctrinal publications, and secondary sources.

**AI Declaration**: We used generative AI tools to support literature clustering, interrogate and summarize source files, and assist with grammar and rephrasing.

## References

Aromataris, E., Lockwood, C., Porritt, K., Pilla, B. and Jordan, Z. (eds.) (2024) *JBI Manual for Evidence Synthesis*, *JBI*. Available at: https://synthesismanual.jbi.global.

Batool, A., Hussain, M. and Zowghi, D. (2024) "*Responsible AI Governance: A Systematic Literature Review*". In: *arXiv.* DOI:10.48550/arXiv.2401.10896.

Berente, N., Gu, B., Recker, J. and Santhanam, R., (2021) "Managing artificial intelligence". *MIS Quarterly*, *45*(3), pp. 1433–1450.

Coners, A. and Matthies, B. (2014) "A Content Analysis of Content Analyses in is Research: Purposes, Data Sources, and Methodological Characteristics". *Pacific Asia Conference on Information Systems*. Available at: https://api.semanticscholar.org/CorpusID:8319295.

Deshpande, A. and Sharp, H. (2022) "Responsible AI Systems: Who are the Stakeholders?" In: *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society.* doi:10.1145/3514094.3534187.

EU (2024). *Regulation (European Union) 2024/1689 - Artificial Intelligence Act.* Available at: https://eurlex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689.

Executive Office of the President (2025) Executive Order 14179: Removing Barriers to American Leadership in AI. Available at: https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to- american-leadership-in-artificial-intelligence/.

Executive Office of the President (2023) *Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. Available at: https://www.federalregister.gov/documents/2023/ 11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial- intelligence.

Ezer, N., Bruni, S.,Cai, Y., Hepenstal, S.J., Miller, C.A., and Schmorrow, D.D. (2019) "Trust engineering for human AI teams". *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 63(1), pp. 322–326. doi:10.1177/1071181319631264.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F. et al. (2018) "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations". *Minds and Machines,* 28(4), pp. 689–707.

Georganta, E. and Ulfert, A.-S. (2024) "Would you trust an AI team member? Team trust in human–AI teams", *Journal of Occupational and Organizational Psychology,* 97(3), pp. 1212–1241. doi:10.1111/joop.12504.

Goswami, M. (2021)."Challenges and solutions in integrating AI with multi-cloud architectures". *International Journal of Enhanced Research in Management & Computer Applications, ISSN*, pp. 2319–7471. doi:10.48550/ arXiv.2504.12170.

Heymans, F. and Heyman, R. (2024) "Identifying stakeholder motivations in normative AI governance: a systematic literature review for research guidance", *Data & Policy* 6, e58. doi:10.1017/dap.2024.66.

International Organization for Standardization (2023) *ISO/IEC 42001:2023 - AI Management System*. Available at: https://www.iso.org/standard/42001

Kapusta, A.S., Jin, D., Teague, P.M., Houston, R.A., Elliott, J.B., Park, G.Y. and Holdren, S.S. (2025) "A Framework for the Assurance of AI-Enabled Systems". *Proceedings of SPIE*, 13476, 134760C. doi:10.1117/12.3056719

Krippendorff, K. (2018) *Content Analysis: An introduction to its methodology*. *Sage Publications, INC*. Available at: https://doi.org/10.4135/9781071878781

Kuehnert, B., Kim, R.M., Forlizzi, J., and Heidari, H. (2025) "The "Who", "What", and "How" of Responsible AI Governance: A Systematic Review and Meta-Analysis of (Actor, Stage)-Specific Tools". *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. doi:10.1145/3715275.3732191.

Kwiatkowska, M. and Zhang, X. (2023) "When to Trust AI: Advances and Challenges for Certification of Neural Networks". *Proceedings of the 18th Conference on Computer Science and Intelligence Systems (FedCSIS)*. 35. pp. 25–37. doi:10.15439/2023F2324.

Li, Y. and Goel, S. (2024) "Making It Possible for the Auditing of AI: A Systematic Review of AI Audits and AI Auditability". *Information Systems Frontiers,* 27, pp. 1121–1151. d o i :10.1007/s10796-024-10508-8.

Li, Y. and Goel, S. (2025) "Bridging IT Auditors and AI Auditing: Understanding Pathways to Effective IT Audits of AI-Driven Processes". *Advances in Accounting,* 69, p. 100842. d o i :10.1016/j.adiac.2025.100842.

McGrath, M.J., Duenser, A., Lacey, J., and Paris, C. (2025) "Collaborative human-AI trust (CHAI-T): A process framework for active management of trust in human-AI collaboration". *Computers in Human Behavior: Artificial Humans* 3, p. 100-200. doi:10.1016/j.chbah.2025.100200.

NIST (2023a) *AI RMF Playbook*. Available at: https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook.

NIST (2023b) *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. doi:10.6028/NIST.AI.100-1.

NIST (2024) *Artificial Intelligence Risk Management Framework: Generative AI Profile*. NIST AI 600-1. Available at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf.

Office of Management and Budget (2024) *OMB Memorandum M-24-10: Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence*. Available at: https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-GovernanceInnovation-and-Risk-Management-for- Agency-Use-of-Artificial-Intelligence.pdf

Office of Management and Budget (2025) *M-25-21: Accelerating Federal Use of AI*. Available at: https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf

OECD (2024) *OECD Recommendation on Artificial Intelligence*. Available at: https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.

Parasuraman, R. and Manzey, D.H. (2010) "Complacency and bias in human use of automation: An attentional integration". *Human Factors,* 52(3), pp. 381–410. doi:10.1177/0018720810376055.

Rose, S., Borchert, O., Mitchell, S., and Connelly, S. (2020) "Zero trust architecture". *NIST Special Publication* 800-207, Available at: https://doi.org/10.6028/NIST.SP.800-207

Salako, A.O., Fabuyi, J.A., Aideyan, N.T., Selesi-Aina, O., Dapo-Oyewole, D.L., and 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), pp. 66–88.

Schuett, J. (2023) "Three Lines of Defense Against Risks from Artificial Intelligence". *AI & Society*. doi:10.1007/s00146-023-01811-0.

Sterz, S., Baum, K., Biewer, S., Hermanns, H., Lauber-Rönsberg, A., Meinel, P., and Langer, M. (2024) "On the quest for effectiveness in human oversight: Interdisciplinary perspectives". *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. pp. 2495–2507. doi:10.1145/3630106.3659051.

U.S. Government Accountability Office (2021) *Artificial Intelligence: An Accountability Framework for Federal Agencies*. Available at: https://www.gao.gov/assets/gao-21-519sp.pdf.

Ulfert, A.-S., Georganta,. E., Centeio Jorge. C., Mehrotra, S., and Tielman, M. (2024) "Shaping a multidisciplinary understanding of team trust in human-AI teams: A theoretical framework". *European Journal of Work and Organizational Psychology* 33(2), pp. 158–171. doi:10.1080/1359432X.2023.2200172.