

AI Governance and Ethical Implications for Organizational Oversight in Cybersecurity

Irlenys Josefina Tersek Rodríguez

ijtersek@gmail.com

Independent Researcher

Perspective

Keywords: AI governance, cybersecurity governance, organizational oversight, accountability, explainability, human oversight

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10662896/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Artificial intelligence (AI) is increasingly embedded in cybersecurity processes such as anomaly detection, threat assessment, alert prioritization, and decision support. This development expands analytical capacity, but it also changes how organizational oversight is exercised and how responsibility for cybersecurity decisions is allocated. This conceptual paper develops an integrative framework for examining AI governance and its ethical implications for organizational oversight in cybersecurity. Drawing on AI governance, cybersecurity governance, organizational oversight, and Dynamic Capabilities Theory, the paper distinguishes the operational use of AI from the governance arrangements through which AI-supported outputs are interpreted, prioritized, escalated, supervised, and legitimized. It conceptualizes AI governance in cybersecurity through four interrelated dimensions: monitoring and analytical support, prioritization and decision support, coordination and oversight integration, and ethical and governance alignment. Four propositions explain how AI-supported mechanisms may reshape organizational oversight, increase the salience of accountability, transparency, explainability, and meaningful human oversight, and produce different outcomes depending on dynamic capabilities and cybersecurity governance maturity. The framework contributes a management-oriented perspective that moves beyond technology-centered accounts of AI adoption and provides a foundation for future construct development, empirical testing, and organizational governance practice.

1 Introduction

Organizations increasingly operate in environments characterized by digital interdependence, expanding cyber exposure, and growing reliance on data-driven technologies for monitoring, analysis, and decision support. Cybersecurity therefore extends beyond a technical or operational function and increasingly constitutes a governance concern involving strategic alignment, accountability, risk ownership, coordination, and board-level oversight (AlGhamdi et al. 2020; Gale et al. 2022). At the same time, AI is becoming embedded in cybersecurity processes including anomaly detection, threat assessment, alert prioritization, and decision support. The central management question is no longer only whether AI can improve cybersecurity performance, but how AI-supported mechanisms reshape organizational oversight and what ethical consequences follow from that transformation.

Responsible AI research emphasizes accountability, transparency, explainability, fairness, and human oversight as central governance concerns (Wirtz et al. 2020; Papagiannidis et al. 2025). These concerns are also reflected in operational governance frameworks. The NIST AI Risk Management Framework treats AI risk management as a socio-technical and organizational activity and identifies accountability, transparency, explainability, security, resilience, and fairness among the characteristics of trustworthy AI (Tabassi 2023). ISO/IEC 42001:2023 similarly frames responsible AI as an organizational management-system challenge requiring policies, responsibilities, risk treatment, monitoring, and continual improvement (ISO/IEC 2023). Yet these governance perspectives are rarely integrated with cybersecurity governance and organizational oversight in a way that explains how AI-supported outputs become consequential within managerial decision processes.

This paper addresses a threefold research gap. First, the literature lacks an integrated, management-oriented explanation of how AI governance shapes organizational oversight specifically in cybersecurity. Second, the ethical implications of embedding AI-supported mechanisms into cybersecurity governance remain insufficiently developed, particularly with respect to accountability, transparency, explainability, responsibility allocation, and meaningful human oversight. Third, limited attention has been paid to why these effects may vary across organizations according to their adaptive capabilities and governance maturity.

Accordingly, this paper develops an integrative conceptual framework grounded in four complementary streams: AI governance, cybersecurity governance, organizational oversight, and Dynamic Capabilities Theory. Dynamic capabilities are treated as an enabling organizational condition, while cybersecurity governance maturity is introduced as a contextual conditioning factor. The paper contributes by clarifying the construct of AI governance in cybersecurity, distinguishing governance from mere AI tool use, systematizing the principal ethical implications of AI-supported oversight, and advancing four propositions for future empirical research.

2 Theoretical foundations

2.1 AI governance

AI governance concerns the principles, structures, roles, processes, and controls through which organizations direct and supervise the development and use of AI. Existing work has highlighted the need to translate high-level ethical principles into organizational mechanisms that can support accountability, transparency, explainability, fairness, and human oversight (Wirtz et al. 2020; Papagiannidis et al. 2025). This shift from principles to organizational capability is increasingly visible in formal frameworks: NIST AI RMF 1.0 organizes AI risk management around Govern, Map, Measure, and Manage functions (Tabassi 2023), while ISO/IEC 42001:2023 specifies requirements for establishing, implementing, maintaining, and continually improving an AI management system (ISO/IEC 2023).

These developments support a view of AI governance as more than compliance or ethics documentation. Governance becomes consequential when AI outputs influence organizational decisions, risk prioritization, escalation, or resource allocation. In cybersecurity, this is particularly important because AI-supported judgments may affect high-consequence decisions under uncertainty, where speed and automation must coexist with traceability, challenge, and human responsibility.

2.2 Cybersecurity governance and organizational oversight

Cybersecurity governance focuses on the structures, responsibilities, policies, and oversight mechanisms through which cybersecurity is directed, controlled, and aligned with organizational objectives (AlGhamdi et al. 2020; Gale et al. 2022). Its scope includes risk ownership, strategic prioritization, reporting, escalation, control assurance, and board or executive oversight. Organizational

oversight adds a broader managerial perspective by emphasizing monitoring, supervision, coordination, accountability, and the legitimacy of decision processes.

AI-supported cybersecurity creates a new interface between these domains. A model may rank alerts, classify threats, or recommend response options, but the governance question concerns how those outputs enter organizational routines: who can rely on them, who can challenge them, what evidence must be retained, when escalation is mandatory, and who remains accountable when automated or semi-automated recommendations influence action. The distinction is therefore between AI as an operational tool and AI as an input embedded in governance.

2.3 Dynamic capabilities as an enabling condition

Dynamic Capabilities Theory explains how organizations sense changes, seize opportunities, and reconfigure resources and processes in uncertain environments (Teece 2007). Capability-based research also links adaptive capacity to organizational resilience (Duchek 2020; Su and Junge 2023; Biswakarma and Bohora 2025). This perspective is relevant to AI governance because the responsible integration of AI-supported cybersecurity mechanisms requires organizations to revise roles, controls, skills, escalation routines, and assurance practices as technologies and risks evolve.

Dynamic capabilities are therefore not treated here as an outcome of AI adoption. Rather, they represent an enabling organizational condition that may explain why organizations exposed to similar technologies achieve different governance outcomes. Stronger sensing, learning, coordination, and reconfiguration capabilities should make it easier to identify emerging AI-related risks, redesign oversight routines, and preserve accountability as AI use changes.

3 Conceptualizing AI governance for organizational oversight in cybersecurity

AI governance in cybersecurity is defined here as a multidimensional organizational governance framework through which AI-supported cybersecurity mechanisms are structured, supervised, and aligned with organizational objectives, oversight arrangements, risk responsibilities, and ethical expectations. This definition positions the construct at the organizational and managerial level and distinguishes it from both traditional cybersecurity governance and the mere operational use of AI.

Traditional cybersecurity governance remains concerned with the direction and control of cybersecurity through roles, policies, risk structures, and oversight mechanisms. AI governance in cybersecurity extends these arrangements when AI-supported mechanisms begin to shape how information is interpreted, prioritized, escalated, coordinated, supervised, and justified. It does not replace cybersecurity governance; it represents an evolution of governance in response to increased automation, data intensity, and socio-technical complexity.

Likewise, the use of AI for anomaly detection, threat classification, alert prioritization, or predictive analysis does not by itself constitute governance. Governance begins when these mechanisms are incorporated into organizational arrangements that define authority, accountability, evidence, challenge, escalation, and human decision rights. This distinction is consistent with contemporary responsible-AI approaches that emphasize organizational governance and lifecycle risk management rather than model performance alone (Tabassi 2023; ISO/IEC 2023).

3.1 Dimensions of AI governance in cybersecurity

Table 1
Proposed dimensions of AI governance for organizational oversight in cybersecurity.

Dimension	Organizational focus	Illustrative governance questions
Monitoring and analytical support	Visibility, signal interpretation, pattern detection	What AI-supported evidence enters oversight? How are quality, drift, and uncertainty monitored?
Prioritization and decision support	Ranking incidents, risks, and response options	How are recommendations validated, challenged, and documented before action?
Coordination and oversight integration	Escalation, communication, supervision, cross-functional alignment	Which roles receive AI-supported outputs? When is escalation or independent review required?
Ethical and governance alignment	Accountability, transparency, explainability, human oversight	Who remains accountable? What level of explanation and human control is proportionate to the decision?

The four dimensions are analytically distinct but mutually reinforcing. Monitoring without governance integration may increase information without improving accountability. Prioritization without explainability may accelerate decisions while weakening their legitimacy. Coordination without clearly allocated responsibility may diffuse accountability. Ethical alignment without operational monitoring may remain aspirational. The construct therefore captures the organizational integration of AI-supported mechanisms rather than the presence of isolated controls.

4 Conceptual model and propositions

The conceptual model proposes that AI-supported mechanisms reshape organizational oversight by changing how cybersecurity information is detected, interpreted, prioritized, escalated, and supervised. As AI becomes more deeply embedded in these processes, oversight shifts from a predominantly human-centered arrangement toward a socio-technical governance structure in which human judgment and AI-supported inputs interact. This transformation can improve speed, scale, and analytical consistency, but it can also obscure decision rationales, create automation bias, diffuse responsibility, and make effective challenge more difficult.

The model also recognizes that ethical consequences intensify as organizational reliance on AI-supported mechanisms increases. Accountability becomes more difficult when decision influence is distributed across model developers, vendors, security analysts, managers, and governance bodies. Transparency and explainability become relevant not as abstract technical properties, but as organizational requirements for challenge, auditability, escalation, and justification. Meaningful human oversight requires more than nominal approval: decision-makers must have sufficient authority, information, competence, and time to question or override AI-supported outputs when appropriate.

Proposition 1

The integration of AI-supported mechanisms into cybersecurity governance reshapes organizational oversight by influencing how cybersecurity-related information is interpreted, prioritized, escalated, coordinated, and supervised.

Proposition 2

The greater the organizational reliance on AI-supported mechanisms in cybersecurity oversight, the greater the governance relevance of accountability, transparency, explainability, responsibility allocation, and meaningful human oversight.

Dynamic capabilities are incorporated as an enabling condition because responsible AI integration requires continuous organizational adaptation. Organizations with stronger sensing and learning mechanisms should be better able to identify new failure modes, recognize regulatory or technological change, and update governance arrangements. Seizing and reconfiguration capabilities should support the redesign of roles, workflows, controls, and skills required to integrate AI without losing organizational accountability.

Proposition 3

Stronger dynamic capabilities positively condition an organization's ability to adapt, supervise, and responsibly govern AI-supported cybersecurity oversight processes.

Cybersecurity governance maturity is introduced as a contextual conditioning factor. Mature governance environments are more likely to provide clearly defined responsibilities, escalation paths, risk ownership, assurance routines, and strategic oversight into which AI-supported mechanisms can be embedded. Where governance maturity is low, AI may amplify fragmentation by adding technically sophisticated outputs to already weak decision structures.

Proposition 4

Cybersecurity governance maturity positively conditions the quality and coherence of AI-supported organizational oversight, such that accountability and effective supervision are more likely under higher levels of governance maturity.

5 Discussion

5.1 Theoretical contributions

The first contribution is conceptual. The paper positions AI governance in cybersecurity as an organizational governance phenomenon rather than a synonym for AI-enabled security operations. This distinction clarifies the unit of analysis: the central concern is not whether an organization uses AI, but how AI-supported outputs become embedded in structures of authority, oversight, accountability, and strategic decision-making.

The second contribution is theoretical integration. AI governance, cybersecurity governance, organizational oversight, and Dynamic Capabilities Theory have each addressed important parts of the problem, but they offer a more complete explanation when considered together. AI governance supplies the ethical and control concerns; cybersecurity governance locates those concerns within risk ownership and security oversight; organizational oversight explains supervision and responsibility; and Dynamic Capabilities Theory explains variation in organizations' ability to adapt governance arrangements over time.

The third contribution is a contingent rather than deterministic account of AI-enabled governance. The framework does not assume that AI automatically improves oversight. Instead, it proposes that the governance value and ethical robustness of AI-supported mechanisms depend on organizational capabilities and governance maturity. This provides a theoretical basis for explaining heterogeneous outcomes across organizations using similar technologies.

5.2 Practical implications

For practitioners, the framework suggests that organizations should govern AI-supported cybersecurity decisions as part of their broader governance architecture rather than treating AI controls as an isolated technical layer. This includes explicit decision rights, traceable escalation paths, documented accountability, proportionate explainability requirements, mechanisms for human challenge and override, and ongoing monitoring of performance and emerging risks. NIST AI RMF and ISO/IEC 42001 provide complementary operational reference points for translating these principles into risk-management and management-system practices (Tabassi 2023; ISO/IEC 2023).

The framework also implies that governance design should be risk-sensitive. Not every AI-supported output requires the same degree of explanation or human intervention. Higher-consequence decisions, uncertain model outputs, novel threat contexts, or actions with significant organizational impact may justify stronger review, evidence retention, and escalation. The objective is not to maximize human intervention, but to preserve meaningful and accountable human judgment where governance consequences are material.

5.3 Research agenda

Future research can develop and validate a measurement instrument for the four proposed dimensions and examine whether they form a higher-order construct. Empirical studies could test the propositions using survey designs, case studies, mixed methods, or longitudinal research across sectors with different regulatory and cyber-risk profiles. Researchers could also investigate whether regulatory pressure, sectoral complexity, organizational risk culture, AI literacy, third-party dependence, or the criticality of AI-supported decisions further condition the relationship between AI governance and oversight outcomes.

A second research stream should examine outcomes. Potential dependent variables include decision quality, incident-response effectiveness, auditability, accountability clarity, trust in AI-supported recommendations, governance resilience, and the frequency or severity of oversight failures. Longitudinal research would be particularly valuable because governance maturity and dynamic capabilities are likely to evolve as organizations gain experience with AI-supported cybersecurity.

6 Limitations

This paper is conceptual and does not empirically test the proposed relationships. The four dimensions and propositions therefore require operationalization and validation across organizational contexts. The framework also focuses on organizational oversight rather than the technical performance or security of specific AI models. Finally, although contemporary standards and governance frameworks are used to strengthen the organizational interpretation, the framework is not intended as a compliance mapping to any single regulatory regime or standard.

7 Conclusion

AI-supported mechanisms are becoming part of the infrastructure through which cybersecurity information is interpreted, prioritized, escalated, and acted upon. Their significance therefore extends beyond technical performance. When AI influences consequential cybersecurity decisions, it also changes the conditions under which oversight, responsibility, and legitimacy are produced within organizations.

This paper develops an integrative framework that conceptualizes AI governance in cybersecurity through monitoring and analytical support, prioritization and decision support, coordination and oversight integration, and ethical and governance alignment. It further proposes that dynamic capabilities enable responsible adaptation and that cybersecurity governance maturity conditions the coherence of AI-supported oversight. By shifting the focus from tools to organizational governance, the framework provides a foundation for empirical research and for governance practices that seek to combine AI-enabled analytical capacity with accountability, explainability, and meaningful human judgment.

Declarations

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Ethics approval and consent to participate

Not applicable. This conceptual article does not involve human participants, human data, or animals.

Consent for publication

Not applicable.

Competing interests

The author declares no competing interests.

Funding

The author received no specific funding for this work.

Author contributions

I.T. conceived the study, developed the conceptual framework and propositions, conducted the literature-based conceptual synthesis, and wrote and revised the manuscript.

Acknowledgements

Not applicable.

References

1. AlGhamdi S, Win KT, Vlahu-Gjorgievska E. Information security governance challenges and critical success factors: Systematic review. *Computers Secur.* 2020;99:102030. <https://doi.org/10.1016/j.cose.2020.102030>.
2. Biswakarma G, Bohora B. Dynamic capabilities and organizational performance: The mediating role of organizational resilience in IT sector. *Future Bus J.* 2025;11(1):1–19. <https://doi.org/10.1186/s43093-025-00592-w>.

3. Duchek S. Organizational resilience: A capability-based conceptualization. *Bus Res.* 2020;13(1):215–46. <https://doi.org/10.1007/s40685-019-0085-7>.
4. Gale M, Bongiovanni I, Slapnicar S. Governing cybersecurity from the boardroom: Challenges, drivers, and ways ahead. *Computers Secur.* 2022;121:102840. <https://doi.org/10.1016/j.cose.2022.102840>.
5. ISO/IEC. ISO/IEC 42001:2023 Information technology—Artificial intelligence—Management system. International Organization for Standardization; 2023.
6. Papagiannidis E, Mikalef P, Conboy K. Responsible artificial intelligence governance: A review and research framework. *J Strateg Inf Syst.* 2025;34(2):101885. <https://doi.org/10.1016/j.jsis.2024.101885>.
7. Su W, Junge S. Unlocking the recipe for organizational resilience: A review and future research directions. *Eur Manag J.* 2023;41(6):1086–105. <https://doi.org/10.1016/j.emj.2023.03.002>.
8. Tabassi E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
9. Teece DJ. Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strateg Manag J.* 2007;28(13):1319–50. <https://doi.org/10.1002/smj.640>.
10. Wirtz BW, Weyerer JC, Sturm BJ. The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *Int J Public Adm.* 2020;43(9):818–29. <https://doi.org/10.1080/01900692.2020.1749851>.

Figures

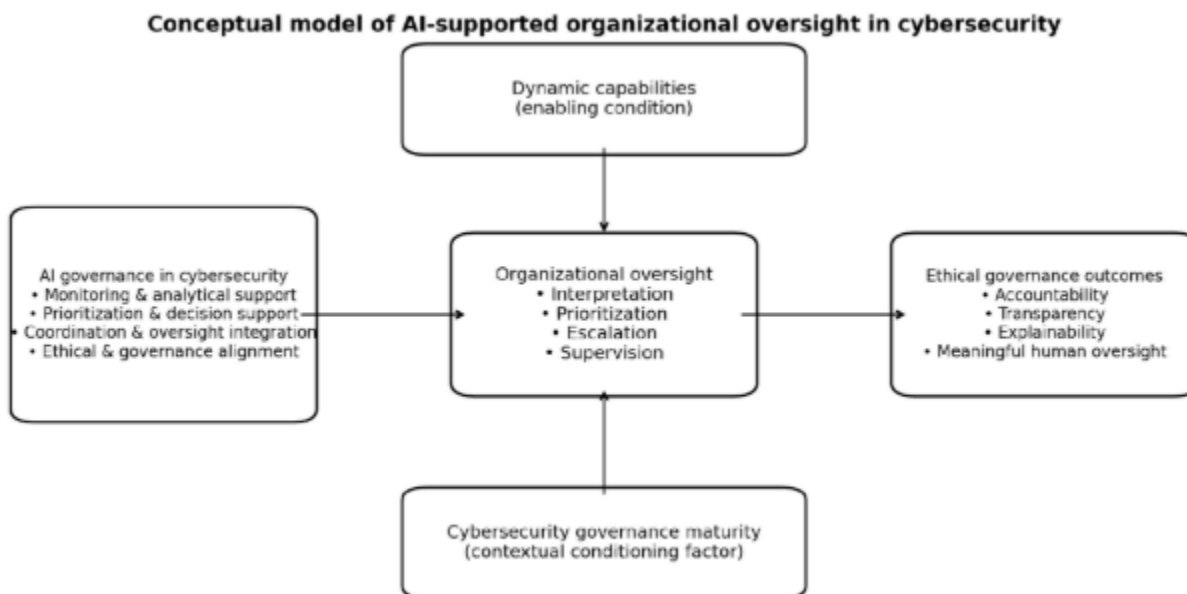


Figure 1

