Who Should Trust What? Role-based Adaptive Governance for Trust Calibration in Industry 5.0

Svetlana Norkin, Jussi Ville Mikkonen and Kathrin Kirchner
Department of Engineering Technology and Didactics, Technical University of Denmark, Denmark

slwno@dtu.dk
kakir@dtu.dk
jusmi@dtu.dk

Abstract: Existing AI governance approaches define strategic commitments, operating responsibilities, and technical controls, but provide less guidance on how organizations can capture and translate what happens in human–AI collaboration in situated work contexts. Industry 5.0 makes this gap critical because human-centricity depends not only on AI system control, but also on how reliance, oversight, responsibility, accountability, and control are managed when humans and AI systems collaborate. We propose a role-based adaptive governance model in which shifting roles within human–AI collaboration are linked with governance responsibilities for review, escalation, and adjustment. The model conceptualizes the relational layer as a governance-informing layer where observations from human–AI collaboration are reviewed, recurring patterns are identified as trust calibration signals, and these signals are translated into governance-relevant input through a relational translation loop. This governance model specifies how trust calibration signals from situated work contexts can be captured and translated into governance-relevant input for adjustments to strategic commitments, operating responsibilities, and technical controls.

Keywords: Industry 5.0, AI governance, Governance adaptation, Trust calibration, Socio-technical governance

1. Introduction

Industry 5.0 emerges as a human-centred, sustainable, and resilient industrial paradigm in which advanced technologies are integrated into industrial work and production (van Erp et al. 2024; Ghobakhloo et al. 2023). Human-centricity means that humans remain at the core of work processes, while technologies augment, enhance, and support human judgment, competence, responsibility, accountability, and well-being (van Erp et al. 2024; Ghobakhloo et al. 2023). Within this paradigm, human–AI collaboration refers to work arrangements in which humans and AI systems work interdependently to conduct tasks, make decisions, or learn in situated work contexts (Misra, Barik & Kvalvik 2025; Passalacqua et al. 2025). Here, AI systems are understood as socio-technical arrangements that embed AI models in organizational workflows (cf. Urbach et al. 2026). However, as AI systems become more autonomous, human–AI collaboration may redistribute roles, responsibilities, expertise, and accountability across tasks. This redistribution creates uncertainty about contribution, responsibility, human judgment, and whether outcomes of human–AI collaboration can be trusted. Human–AI collaboration may therefore support or undermine human-centricity in Industry 5.0 contexts (Passalacqua et al. 2025). Thus, human-centricity is not guaranteed by the adoption of AI systems; it must be actively governed. The question therefore is how companies within Industry 5.0 contexts can keep humans at the core of human-AI collaboration and translate human-centricity into situated work contexts where roles, responsibilities, expertise, and accountability change.

Human–AI collaboration increasingly involves not only task execution and decision-making, but also human–AI co-learning. In this process, humans develop judgment and competence through interaction with AI system outputs, while AI systems are refined through human feedback and adaptation (cf. Passalacqua et al. 2025). However, such co-learning may emerge without being visible, reviewed, or controlled by the organization (Zampirolo et al. 2026). Organizations may not know what humans learn from AI system outputs, how they question or correct these outputs, or how AI systems are refined in response (Passalacqua et al. 2025; Zampirolo et al. 2026). Thus, human–AI co-learning may drive continuous adaptation, but this adaptation can remain uncontrolled and may support or undermine human-centricity. If human–AI co-learning is to support human-centricity, it must be organized and reviewed as part of governing human–AI collaboration.

The limited organizational visibility and control over human–AI co-learning shifts the focus from AI system control to organizational AI governance. Organizational AI governance is a system of rules, practices, processes, and technological tools through which an organization ensures that the use of AI systems aligns with strategy, values, legal requirements, ethical principles, and stakeholder expectations (Birkstedt et al. 2023; Mäntymäki et al. 2022). In Industry 5.0 contexts, organizational AI governance must therefore be understood as more than compliance; it must also support continuous adaptation generated through human–AI co-learning (van Erp et al. 2024). Regulations, standards, AI governance literature, and practice-facing industry frameworks provide

important guidance on security, transparency, accountability, human oversight, risk management, lifecycle controls, and incident response (Anthropic 2026; Birkstedt et al. 2023; Gartner 2025; Misra, Barik & Kvalvik 2025; OpenAI 2026; Urbach et al. 2026). Yet a recurring challenge is that organizations may be aware of existing regulations, standards, and ethical risks while lacking the operational clarity needed to allocate responsibility, support oversight, and act on this awareness in situated work contexts where human-AI collaboration takes place (Birkstedt et al. 2023; Zampirolo et al. 2026). Thus, existing AI governance provides important guidance for strategic commitments, operating responsibilities, and technical controls, but lacks sufficient operationalization of how human–AI collaboration in situated work contexts is reviewed and translated into governance-relevant input and reported back to those involved in the collaboration.

Within Industry 5.0 contexts, this governance challenge becomes especially important when human–AI collaboration is assessed through trust calibration. The issue is not only whether humans trust AI systems, but whether outcomes from human–AI collaboration can be trusted when human and AI contributions, learning, and roles change over time (cf. Passalacqua et al. 2025). These changing roles may include, for example, who acts as decision-maker, learner, or teacher in human–AI collaboration. Within human–AI collaboration in situated work contexts, those involved in the collaboration continuously make trust decisions: when to trust AI system outputs, when to question, correct, or override them, when to escalate uncertainty, and how to allocate responsibility for outcomes (Zampirolo et al. 2026). Trust calibration is therefore not about increasing trust in AI systems, but about making reliance, oversight, responsibility, accountability, and control governable in human–AI collaboration. Thus, the governance challenge is how these trust decisions are captured, reviewed, and translated into governance-relevant input (Zampirolo et al. 2026). Because these trust decisions concern how reliance, review, correction, override, learning, and decision-making are distributed between humans and AI systems, trust calibration also requires clarity about how changing roles in human–AI collaboration are captured within organizational AI governance and who is responsible for acting on them.

To address this challenge, we propose a role-based adaptive governance model for human–AI collaboration in the Industry 5.0 paradigm. The paper contributes to Industry 5.0 and organizational AI governance literature by conceptualizing human–AI collaboration in situated work contexts as a source of governance-relevant input, rather than merely an implementation context.

2. Theoretical Point of Departure

This section develops the theoretical point of departure for the role-based adaptive governance model by explaining how human-centric human–AI collaboration depends on governing changing roles, reliance, responsibility, and accountability in Industry 5.0 contexts.

2.1 Human-centric Human–AI Collaboration and Human-AI co-learning

In Industry 5.0 contexts, humans and AI systems contribute different capabilities to human–AI collaboration (van Erp et al. 2024). Humans contribute contextual judgment, ethical sensitivity, experience, creativity, responsibility, and understanding of situated work contexts, while AI systems contribute speed in information processing, algorithmic consistency, prediction, and pattern recognition (Zárate-Torres et al. 2025). Value and risk do not emerge from these capabilities separately, but through their interaction within human-AI collaboration (Passalacqua et al. 2025; Zárate-Torres et al. 2025). Therefore, human–AI collaboration changes not only how tasks are conducted, but also how decisions are made and how learning develops within work arrangements. These changes also reshape role relations, understood here as how advising, reviewing, deciding, executing, monitoring, learning, and feedback provision are distributed between humans and AI systems (cf. Ghobakhloo et al. 2023; Passalacqua et al. 2025).

In situated work contexts, this interaction turns AI system outputs into objects of human judgment: humans must interpret them, decide whether to trust, question, correct, or override them, and remain responsible for outcomes. Sergeeva, Leonardi, and Faraj (2026) argue that AI outputs are grounded in computational modelling and large-scale information processing, while human judgment is grounded in situated judgment, embodied experience, and professional knowledge. This difference matters for governance because AI system outputs may change whose judgment is treated as authoritative, how responsibility is allocated, and how decisions become legitimate in human–AI collaboration.

Existing work points to human–AI co-learning as involving human correction of AI outputs, competence development, adjusted reliance, adaptive design, and feedback-based AI system refinement (van Erp et al. 2024; Ghobakhloo et al. 2023; Passalacqua et al. 2025). Human–AI co-learning explains how role relations change over time and represents ongoing adaptation in human judgment, human competence, and AI system refinement (cf.

Passalacqua et al. 2025). As these role relations change, governance must clarify responsibility, accountability, whose judgment is treated as authoritative, and what should happen when AI outputs and human judgment conflict (Sergeeva, Leonardi & Faraj 2026). Thus, Industry 5.0 governance must address not only AI systems, but also human–AI collaboration in situated work contexts, where changing role relations shape how judgment, learning, responsibility, and accountability are managed.

2.2 Organizational AI Governance

Organizational AI governance pertains to how organizations structure, coordinate, and control the use of AI systems, alongside IT governance and data governance (Mäntymäki et al. 2022; Norton 2025). It is therefore not only a matter of compliance or technical control, but also a management and organizational design challenge (Norton 2025). Existing AI governance research and practice-facing frameworks have moved beyond abstract principles by translating governance principles and rules into organizational mechanisms, including responsibility allocation, decision rights, risk-based controls, AI model monitoring, auditability, incident response, and workforce-related measures (Birkstedt et al. 2023; Gartner 2025; Urbach et al. 2026). Urbach et al. (2026) conceptualize AI governance through structural, procedural, and relational mechanisms that align AI use with organizational, ethical, and regulatory requirements: structural mechanisms define roles and responsibilities; procedural mechanisms define rules, assessments, documentation, and lifecycle controls; and relational mechanisms support communication, coordination, training, and shared understanding. These mechanisms show that AI governance is already organizational, but they do not fully explain how feedback from human–AI collaboration in situated work contexts becomes governance-relevant input for reviewing and adjusting AI governance.

In Industry 5.0 contexts, this limitation becomes critical because governance must connect strategic commitments, operating responsibilities, and technical controls with what happens when humans and AI systems collaborate in situated work contexts. This connection requires continuous feedback between design, operation, monitoring, and adaptation (van Erp et al. 2024). Weak feedback loops between AI system users and governance structures may create operational risk (Zampirolo et al. 2026). The gap is therefore not the absence of governance principles, but the lack of operational mechanisms that connect human–AI collaboration in situated work contexts with broader organizational governance arrangements. This gap matters for Industry 5.0 because human-centricity cannot be secured only through board-level oversight, formal policies, technical controls, or pre-deployment assessments. Thus, organizational AI governance in Industry 5.0 contexts must capture feedback from human–AI collaboration in situated work contexts and translate it into governance-relevant input so that strategic commitments, operating responsibilities, and technical controls can be reviewed and adjusted.

2.3 Trust Calibration as the Governance Objective

In human–AI collaboration, organizations cannot fully control or predict every AI system output or every human interaction with AI systems (cf. Passalacqua et al. 2025). Therefore, trust matters as a way of organizing work under uncertainty (cf. McEvily, Perrone & Zaheer 2003). For governance, the central issue concerns how reliance, oversight, responsibility, and accountability are adjusted when human judgment, AI system outputs, and AI system refinement are distributed across human–AI collaboration (Passalacqua et al. 2025; Sergeeva, Leonardi & Faraj 2026). Because human–AI co-learning involves ongoing adaptation in human judgment, reliance, correction practices, role relations, and AI system refinement, trust must be continuously calibrated.

We understand trust calibration as the ongoing adjustment of reliance, oversight, responsibility, accountability, and control in human–AI collaboration. The aim is not to increase trust in AI systems, but to support appropriate reliance and oversight so that human–AI collaboration supports human-centricity (Passalacqua et al. 2025). Trust calibration therefore becomes an Industry 5.0 governance objective because inappropriate reliance can undermine human judgment, responsibility, accountability, competence, and well-being as roles and responsibilities change.

Trust calibration depends on knowing who relies, who reviews, who corrects or overrides, who escalates uncertainty, and who remains accountable. To sustain human-centricity, governance must capture how roles, reliance, responsibility, and oversight are interpreted and adjusted in human–AI collaboration in situated work contexts, rather than allowing these adjustments to remain informal or invisible (Zampirolo et al. 2026). We therefore view trust calibration as the governance objective for a role-based adaptive governance model, because such calibration requires linking clarity about role relations within human–AI collaboration with governance responsibility for receiving, interpreting, escalating, and acting on uncertainties about roles,

reliance, responsibility, oversight, and accountability. This requires a governance model that connects changing role relations in human–AI collaboration with strategic commitments, operating responsibilities, and technical controls, and specifies how uncertainties about trust calibration can be reviewed, escalated, and translated into governance adjustment.

3. A Role-based Adaptive Governance Model for Industry 5.0

Building on the synthesis of the literature above, we develop a role-based adaptive governance model that specifies how observations from human–AI collaboration in situated work contexts can be reviewed, how recurring patterns can be identified as trust calibration signals, and how these signals can be translated into governance-relevant input.

3.1 Governance Model Overview

The governance model distinguishes four governance layers that together connect strategic commitments, operating responsibilities, and technical controls with human–AI collaboration in situated work contexts. *The strategic layer* sets the direction for AI governance by defining the organization's principles, human-centric commitments, and legitimate boundaries for AI use. *The operating layer* translates this direction into operating responsibilities, including decision rights, training, and escalation routes. *The technical layer* concerns technical controls for AI systems, including monitoring, documentation, and technical assurance needed to ensure that AI systems are fit for use. *The relational layer* is where humans and AI systems collaborate in situated work contexts, where trust decisions are made and where role relations, reliance, oversight, responsibility, and accountability are interpreted and adjusted. The governance model is role-based in the sense that it links shifting roles within human–AI collaboration with governance responsibilities across the strategic, operating, technical, and relational layers.

The central claim of the governance model is that the relational layer should be treated as governance-informing. The layers are analytically distinct but interdependent. Strategic commitments require operating responsibilities to become actionable; operating responsibilities require technical controls to make AI systems governable; and technical controls require relational feedback to assess how AI systems are actually used in situated work contexts. The relational layer is therefore the adaptive core of the governance model: it is where observations from human–AI collaboration can be reviewed and where patterns indicating possible misalignment can first be detected. Such patterns become trust calibration signals when they indicate that reliance, oversight, responsibility, accountability, or control may be misaligned in human–AI collaboration (cf. Passalacqua et al. 2025; Zampirolo et al. 2026). When trust calibration signals are interpreted as requiring governance attention, they become governance-relevant input. Through the relational translation loop, trust calibration signals are identified from patterns of observations, translated into governance-relevant input, and used to inform adjustments to strategic commitments, operating responsibilities, and technical controls.

The governance model does not require time-consuming governance procedures in situated work contexts. Rather, it introduces a proportionate review and feedback process for identifying when reliance, oversight, responsibility, accountability, or control may be misaligned in human–AI collaboration. In this way, the governance model helps detect when human–AI collaboration may no longer support human-centricity. It is conceptual and intended to guide the design, review, and adaptation of human–AI collaboration in Industry 5.0 contexts. Figure 1 illustrates how the relational translation loop reviews observations from the relational layer, identifies recurring patterns as trust calibration signals, translates these signals into governance-relevant input, and informs adjustments to strategic commitments, operating responsibilities, and technical controls.

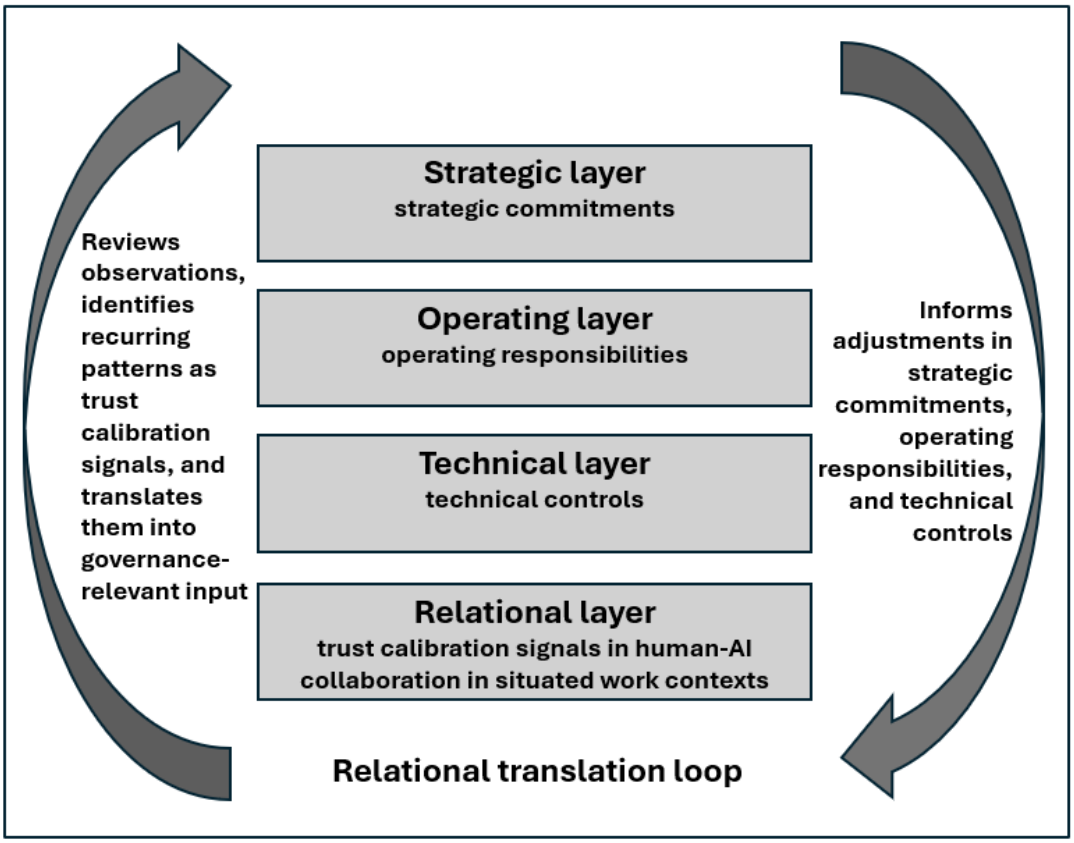

Figure 1: The role-based adaptive governance model in Industry 5.0 contexts

3.2 The Relational Translation Loop

The relational translation loop consists of five steps through which observations from human–AI collaboration in situated work contexts are reviewed, recurring patterns are identified as trust calibration signals, translated into governance-relevant input, escalated through responsibility holders, used for governance adjustment, and reported back to those involved in the collaboration.

First, situated review takes place when relevant managers, team leads, process owners, professional users, AI champions, and AI product owners review how AI systems are used in situated work contexts. This review asks whether human–AI collaboration remains human-centred and whether reliance, oversight, responsibility, accountability, and control remain appropriate.

Second, pattern formation aggregates observations and distinguishes recurring or consequential issues from isolated events. At this stage, recurring or consequential patterns become trust calibration signals. Examples include role ambiguity, accountability gaps, high override rates, unclear explanations, inappropriate reliance, underuse of useful AI support, declining human competence, increased stress, biased outputs, unsafe recommendations, or AI behaviour outside the intended use context.

Third, escalation through named responsibility holders occurs when trust calibration signals are interpreted as governance-relevant input and escalated to the relevant governance layer. The purpose is not to escalate everything to the board, but to ensure that the relevant governance layer receives input it can act on.

Fourth, governance adjustment occurs when the relevant layer adjusts strategic commitments, operating responsibilities, or technical controls, including policies, training, escalation procedures, monitoring, risk thresholds, or system design.

Fifth, downward reporting communicates these adjustments back to the employees, managers, and teams involved in human–AI collaboration. Downward reporting closes the loop by making governance adjustments visible to those working in the same situated work contexts where the human–AI collaboration takes place.

For example, a design-team lead in an Industry 5.0 toy factory notices that designers repeatedly override AI warnings because structural calculations block otherwise acceptable creative choices. The lead and designated AI owner interpret this pattern as a trust calibration signal, define conditions for safe deviation, adjust the relevant thresholds, and report new guidance to the team.

In this way, the loop specifies the governance value of trust calibration signals by clarifying what changed, why it changed, who is responsible, and how human–AI collaboration should be conducted going forward.

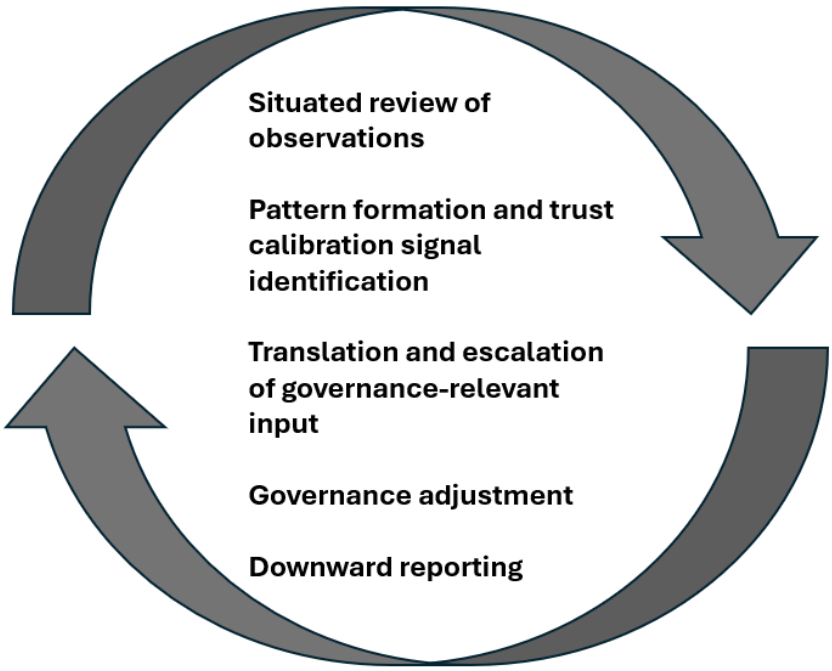

Figure 2: Relational translation loop

Thus, the relational translation loop makes the relational layer governance-informing by reviewing observations, identifying trust calibration signals, translating them into governance-relevant input, and linking this input to governance adjustment. It helps strategic, operating, and technical governance layers assess whether role relations, reliance, oversight, responsibility, accountability, and control remain appropriately calibrated to support human-centricity.

3.3 Governance Responsibilities Across Governance Layers

Building on the need to connect role clarity in human–AI collaboration with governance responsibility, the model specifies who should receive, interpret, escalate, and act on governance-relevant input. Role clarity is especially important in Industry 5.0 because human–AI collaboration may redistribute responsibility, expertise, oversight, learning, and accountability across tasks. Table 1 summarizes how the four governance layers differ in their key governance questions, governance responsibility holders, governance-relevant input, and governance adjustment.

Table 1: Governance responsibilities across four governance layers

Layer	Key governance question	Governance responsibility holders	Governance-relevant input	Governance adjustment
Strategic	What forms of human–AI collaboration are legitimate?	Board of directors CEO/Executive committee	Evidence of whether human–AI collaboration supports or undermines human-centricity, sustainability, resilience, workforce responsibility, and long-term value creation	Adjust risk appetite, strategic priorities, human-centric commitments, investment decisions, and legitimate-use boundaries

Layer	Key governance question	Governance responsibility holders	Governance-relevant input	Governance adjustment
Operating	Who is responsible for governing human-AI collaboration, and how are decisions, responsibilities, and accountabilities distributed?	CIO/CDO/CAIO or AI council, or equivalent governance owner	Gaps between policy and practice: evidence of role ambiguity, accountability gaps, insufficient training, worker concerns, unclear escalation routes, misuse, or shadow AI	Adjust governance responsibility descriptions, policies, training, approval processes, escalation routes, governance forums, and decision rights
Technical	Is the AI system technically fit for this specific human–AI collaboration context?	CISO/AI product owner, data/ML owner, or technical system owner	Evidence of system failures, drift, high override rates, poor explanations, hallucinations, bias, privacy concerns, unsafe outputs, inappropriate reliance, vendor issues, or incidents	Adjust AI models, AI system controls, monitoring, explanations, data practices, technical documentation, thresholds, and system design
Relational	Are human and AI roles, reliance, oversight, responsibility, and accountability calibrated to support human-centricity?	Middle managers/team leads, process owners, professional users, AI champions/super users	Trust calibration signals, reliance problems, role ambiguity, oversight gaps, accountability concerns, competence effects, stress, and well-being concerns	Conduct situated review, distinguish patterns from isolated events, identify trust calibration signals, escalate governance-relevant input, and receive and use downward reporting on governance adjustments

This distribution of governance responsibilities prevents governance from remaining either too centralized or too informal. It preserves clear ownership for governance-relevant input across layers. By assigning distinct responsibilities to each layer, the governance model clarifies how trust calibration can be supported without placing responsibility for human–AI collaboration on a single governance layer.

4. Discussion

This paper contributes to AI governance research by proposing a role-based adaptive governance model with a relational translation loop. Existing approaches emphasize organizational structures, decision rights, risk controls, lifecycle management, and technical assurance (Birkstedt et al. 2023; Mäntymäki et al. 2022; Norton 2025; Urbach et al. 2026; Zampirolo et al. 2026). We extend these approaches by specifying how observations from human-AI collaboration are reviewed, recurring patterns and consequential issues are detected, trust calibration signals are identified and translated into governance-relevant input and used to adjust existing governance arrangements.

This governance model also contributes to Industry 5.0 literature by operationalizing human-centricity (cf. Passalacqua et al. 2025). Human-centricity can remain a broad value statement unless organizations have ways to assess whether it is sustained in human–AI collaboration in situated work contexts. Our governance model treats human-centricity as something that must be reviewed and sustained through governance adjustment in situated work contexts. The relational layer makes visible whether human–AI collaboration supports or undermines role clarity, human judgment, competence development, autonomy, workload, well-being, responsibility, accountability, and appropriate human oversight. This specifies the existing Industry 5.0 literature (Ghobakhloo et al. 2023; Misra, Barik & Kvalvik 2025; Passalacqua et al. 2025) by showing how human-centricity can be translated into governance questions, governance responsibility holders, and relational translation loops at the level where human–AI collaboration takes place.

The governance model further contributes to trust research by shifting the trust question from whether workers trust AI systems to whether organizations can trust human–AI collaborative outcomes when role relations change. We argue that trust in Industry 5.0 human–AI collaboration is not only an individual attitude toward

technology. It is also a governance problem concerning reliance, oversight, responsibility, accountability, and control. The model extends research on trust as an organizing principle by conceptualizing trust calibration as a governance objective for human–AI collaborative outcomes (McEvily, Perrone & Zaheer 2003).

Although our governance model is developed for Industry 5.0 contexts, it may also be relevant for other organizations in which AI systems are embedded in work processes and human judgment, reliance, oversight, responsibility, and accountability are distributed across humans and AI systems. Outside Industry 5.0, the relational translation loop may primarily support risk management and compliance. In Industry 5.0, however, the loop helps to assess whether human–AI collaboration supports human-centricity. This differentiates the model in Industry 5.0 contexts: it does not only make AI-related risks visible but also makes human-centricity governable.

For practice, the governance model suggests that managers should not wait for incidents or rely only on formal policies or pre-deployment assessments. They need regular and structured reviews that capture observations from human–AI collaboration, identify recurring patterns as trust calibration signals, and translate them into governance-relevant input for governance adjustment in roles, training, accountability, oversight, and technical controls. Relational translation loops can be organized by embedding them into existing work arenas, such as team meetings, risk reviews, quality reviews, sprint reviews, supervision meetings, or process owner meetings. Managers can use short and repeated check-ins focused on a limited set of trust calibration questions: Where did employees rely on AI outputs? Where were outputs questioned, corrected, or overridden? Were responsibilities and accountabilities clear? Did anything need escalation?

The loop also requires a lightweight procedure capturing observations from human–AI collaboration, identifying recurring patterns as trust calibration signals, and routing governance-relevant input upward. A named responsibility holder, such as a team lead, process owner, AI champion, or AI product owner, should distinguish patterns from isolated events, identify trust calibration signals, and route governance-relevant input to the relevant governance responsibility holders. This upward routing can be done through short meeting notes, lightweight templates, risk logs, issue trackers, or existing quality-management systems. The aim is not to document every interaction with AI, but to ensure that recurring patterns concerning reliance, oversight, responsibility, accountability, competence, workload, or well-being are identified as trust calibration signals and made visible as governance-relevant input. For such loops to work, organizations also need conditions under which employees can raise concerns about reliance, uncertainty, or competence without being blamed.

Consistent with Norton's (2025) argument that AI governance should be robust but commensurate with risk, low-risk applications may require lighter local review, while high-risk human–AI collaboration requires stronger oversight, escalation, and cross-functional accountability. The practical implication is that organizational AI governance should be embedded where human–AI collaboration takes place, while ensuring that trust calibration signals can be translated into governance-relevant input for strategic, operating, and technical governance layers.

Because our governance model is conceptual, empirical research should explore the loop across AI use cases, risk levels, and organizational settings. Studies could assess how organizations distinguish isolated observations from recurring patterns, identify trust calibration signals, and translate them into governance-relevant input, as well as whether downward reporting changes local practice. They could also examine when the loop creates reporting fatigue or managerial resistance. Comparative work could test whether the model applies beyond Industry 5.0 when the primary objective is compliance or risk management rather than human-centricity.

5. Conclusion

Human–AI collaboration in Industry 5.0 reshapes judgment, expertise, responsibility, and accountability. The proposed role-based adaptive governance model connects these changes to organizational action. Its relational translation loop turns observations from collaborative work into trust calibration signals and governance-relevant input, routes this input to responsible holders, supports adjustments to strategic commitments, operating responsibilities, or technical controls, and reports those adjustments back to the people affected. The model therefore provides a way to adapt AI governance while keeping human-centricity central.

Future research should explore how this governance model works in empirical Industry 5.0 settings. The governance model can be tested through student simulations, design-based pilots, or organizational case studies where AI autonomy, role relations, task risk, and accountability vary across human–AI collaboration tasks. Such testing can assess whether the relational translation loop helps organizations detect trust calibration signals, preserve human-centricity, and make governance adjustments before problems become incidents.

Ethics and AI Declaration: Ethical clearance was not required for this study. The authors used ChatGPT and Microsoft Copilot to assist with language editing of the manuscript. After using ChatGPT and Microsoft Copilot, the authors reviewed the manuscript and take full responsibility for the content of the publication.

References

Anthropic 2026, Policy on the AI Exponential.

Birkstedt, T, Minkkinen, M, Tandon, A & Mäntymäki, M 2023, 'AI governance: themes, knowledge gaps and future agendas', Internet Research, vol. 33, no. 7, pp. 133–167.

van Erp, T, Carvalho, NGP, Gerolamo, MC, Gonçalves, R, Rytter, NGM & Gladysz, B 2024, 'Industry 5.0: A new strategy framework for sustainability management and beyond', Journal of Cleaner Production, vol. 461, Elsevier, p. 142271.

Gartner 2025, Reference Guide for AI Governance.

Ghobakhloo, M, Iranmanesh, M, Tseng, M-L, Grybauskas, A, Stefanini, A & Amran, A 2023, 'Behind the definition of Industry 5.0: a systematic review of technologies, principles, components, and values', Journal of Industrial and Production Engineering, vol. 40, no. 6, pp. 432–447.

Mäntymäki, M, Minkkinen, M, Birkstedt, T & Viljanen, M 2022, 'Defining organizational AI governance', AI and Ethics, vol. 2, no. 4, pp. 603–609.

McEvily, B, Perrone, V & Zaheer, A 2003, 'Trust as an Organizing Principle', Organization Science, vol. 14, no. 1, pp. 91–103.

Misra, S, Barik, K & Kvalvik, P 2025, 'A Comprehensive Review of Human-Centric AI, Regulatory Frameworks, and Their Role in Shaping Industry 5.0', Procedia Computer Science, vol. 259, Elsevier, pp. 1672–1681.

Norton, LW 2025, 'Artificial intelligence and organizational strategy: Ethical and governance implications', Consulting Psychology Journal, vol. 77, Educational Publishing Foundation, US, no. 2, pp. 131–141.

OpenAI 2026, OpenAI's Frontier Governance Framework.

Passalacqua, M, Pellerin, R, Magnani, F, Doyon-Poulin, P, Del-Aguila, L, Boasen, J & Léger, P-M 2025, 'Human-centred AI in industry 5.0: a systematic review', International Journal of Production Research, vol. 63, no. 7, pp. 2638–2669.

Sergeeva, AV, Leonardi, PM & Faraj, S 2026, 'Beyond the tool view of AI: Intelligent technologies and the emergence of new epistemic regimes', Strategic Organization, vol. 24, no. 2, pp. 463–478.

Urbach, N, Feulner, D, Feulner, S, Schüll, M & Mayer, V 2026, 'Governance and Management of AI', in N Urbach & D Feulner (eds), Managing Artificial Intelligence, Future of Business and Finance, Springer Nature Switzerland, Cham, pp. 181–210, viewed 26 June 2026, <https://link.springer.com/10.1007/978-3-032-13308-3_9>.

Zampirolo, G, Burtscher, F, Mariotti, Y, Rudnik, D & deremi-Makinde, M 2026, 'The AI Governance Paradox: Ethical Awareness Without Operational Clarity', in Artificial intelligence ethics, law, and policy, IntechOpen.

Zárate-Torres, R, Rey-Sarmiento, CF, Acosta-Prado, JC, Gómez-Cruz, NA, Rodríguez Castro, DY & Camargo, J 2025, 'Influence of Leadership on Human–Artificial Intelligence Collaboration', Behavioral Sciences, vol. 15, Multidisciplinary Digital Publishing Institute, no. 7, p. 873.