Towards Fourth-Order Cybernetics in AI Governance: SocioTechnical Perspectives on AI Risk Management in Knowledge-Based Organizations

Ludmila Jiříčková[1]

[1] *Prague University of Economics and Business, W. Churchill Sq. 1938/4, Prague 3, Czech Republic*

Abstract

As artificial intelligence (AI) becomes increasingly embedded in knowledge-based small and medium-sized enterprises (SMEs), especially in consulting, education, and finance, managing AI-related risks requires more than static compliance. This paper introduces a novel governance framework by integrating socio-technical systems (STS) theory with fourth-order cybernetics. Unlike conventional approaches that treat governance as a one-time setup, this paper conceptualizes AI governance as a reflexive, multi-layered, and adaptive feedback process embedded in organizational context. It is critical review four major AI governance frameworks (EU AI Act, OECD AI Principles, IEEE Ethically Aligned Design, and UNESCO Recommendation), highlighting their limitations in addressing dynamic socio-technical risks. In response, it is proposed a cybernetic model designed for SMEs, combining participatory co-design, continuous monitoring, contextual awareness, and adaptive compliance. Practical use cases from consulting, EdTech, and fintech illustrate how this model supports responsible AI innovation while enhancing resilience and trust. This contribution offers a forward-looking pathway for aligning AI systems with both human values and evolving operational realities.

Keywords

Artificial Intelligence, AI Governance, Fourth-Order Cybernetics, Socio-Technical Systems, Knowledge-Based Organizations, SMEs, Adaptive Risk Management, Reflexivity

CEUR-WS.org/Vol-4134/paper30.pdf

1. Introduction

As AI systems permeate organizations, especially knowledge-intensive SMEs in consulting, education, and finance, managing the socio-technical risks of AI becomes critical. These knowledge-based organizations rely heavily on expert human capital and information flows, making AI integration both an opportunity and a challenge. Traditional governance frameworks emphasize trustworthy, rightsrespecting AI, but they often treat AI as a bounded technical system. For example, scholars note that "AI systems are not just technical artifacts – they are embedded in social structures and organizations" (Baxter & Sommerville, 2011; Cheong et al., 2024). This insight motivates a sociotechnical governance approach. In this paper, is argued that integrating Fourth-Order Cybernetics with socio-technical systems (STS) theory offers a deeper risk-management framework. In this paper is reviewed the history of STS in IS development, analyze four leading AI governance frameworks, critique their limitations, and propose a novel adaptive governance model tailored to Czech/European knowledge-based SMEs. It is concluded with use-case implications illustrating how the new model can improve responsible AI deployment.

STPIS'25: The 11th International Workshop on Socio-Technical Perspectives in IS (STPIS'25) September 17-18 2025 Skopje, North Macedonia.

[1] Corresponding author.

xmall31@vse.cz (L. Jiříčková)

0009-0004-3761-0993 (L. Jiříčková)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2025-12-20

Copyright © 2025 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Background: Socio-Technical Systems and Cybernetics

The socio-technical systems (STS) perspective originated in the mid-20th century (Tavistock Institute, Scandinavia) to reconcile technology and work. Ropohl (1999) explains that STS theory was designed "to stress the reciprocal interrelationship between humans and machines" and to shape both the technical and social conditions of work. In other words, STS aims to jointly optimize tools and human organization, improving both efficiency and human well-being. Enid Mumford and colleagues extended these ideas into information systems; for instance, Mumford's ETHICS method explicitly integrated employee needs into system design. The core notion is that technology must be understood in its human/organizational context, not in isolation.

Parallel to STS in organizational studies, the field of cybernetics has evolved through multiple "orders" of system dynamics. First-order cybernetics (Wiener, 1948) viewed control systems from an external perspective, while second-order cybernetics (von Foerster, 1974) introduced self-observation and reflexivity (systems as both objects and observers). More recently, scholars have conceptualized third-order and fourth-order cybernetics to address increasingly complex systems. In particular, fourth-order cybernetics asks what happens when a system can "redefine itself" within its environment. Chiolerio (2020) notes that fourth-order cybernetics "focuses on the integration of a system within its larger, co-defining context" and implies the system will "immerge" into its environment. In practical terms, this suggests viewing organizations (and their governance) as dynamic, self-modifying entities rather than static frameworks.

STS and cybernetics intersect naturally in the AI era. Knowledge-based firms now rely on data-intensive automation and digital collaboration, increasing the complexity of socio-technical issues (e.g. algorithmic impacts on work, learning, and privacy). Several authors argue that conventional STS approaches—often focusing on one-time design optimizations—may not suffice for such dynamic contexts. Instead, ongoing adaptation and reflexive learning are needed. Bednar and Welch (2020) argue that smart working environments require a socio-technical redesign that sustains both technological performance and human meaning, thus emphasizing the need for systems thinking at the organizational level. This has prompted interest in higher-order cybernetic concepts (sociocybernetics) to model organizations and technologies co-evolving. The combination of STS (emphasizing human–tech relationships) with fourth-order cybernetics (emphasizing system reconfiguration and context-embedding) provides a rich theoretical basis for next-generation AI governance.

3. Critical Review of Socio-Technical Development

Over the past decades, STS research has broadened from local workplace design to enterprise and societal levels. Early STS interventions (e.g. Tavistock, Scandinavian participative design) targeted joint optimization of tools and tasks. Mumford and Legge (1978) illustrated STS in practice by redesigning offices and workflows to enhance both job satisfaction and productivity. However, critiques have emerged. Some researchers note that STS can be too idealistic, potentially neglecting external pressures like market forces and power relations. In practice, purely socio-technical interventions may falter if they ignore wider constraints. Nevertheless, the fundamental insight persists: technology and organization are inseparable; socio-technical integration can make systems more sustainable and humane.

Digital transformation and AI add new dimensions to this phenomenon. In knowledge-based SMEs, data-driven tools and AI agents interact with experts and clients in complex ways, leading to emergent risks (bias, mission creep, deskilling). Traditional STS methods typically treat governance

design as a discrete phase. By contrast, sociocybernetic thinking suggests continuous adaptation. Fourth-order cybernetics has been invoked to fill this gap: it envisages systems (and meta-systems) that continuously self-observe and reconfigure. Chiolerio (2020) explicitly describes fourth-order cybernetics as a realm where a system "redefines itself" within its context. From this perspective, one can critique existing STS-based governance as often too static; a truly socio-technical governance must enable organizations to learn and transform their own rules. In summary, the STS tradition provides rich insights into human–tech integration, but applying it to modern AI challenges calls for higher-order reflexivity – an adaptive, context-sensitive governance loop.

4. Theoretical Contribution: Fourth-Order Socio-Technical Perspective

Building on the above, this theoretical contribution is to synthesize STS theory with Fourth-Order Cybernetics to reconceptualize AI governance. It is proposed treating AI governance frameworks themselves as socio-technical cybernetic systems. In this view, the governance apparatus (policies, committees, monitoring tools) is not a one-way controller but a reflexive agent embedded in the organization's context. It has a dual nature: on one level, it imposes rules and standards on AI systems; on another level, it continually observes AI outcomes, stakeholder feedback, and environmental changes, then adapts its own rules and processes accordingly.

P2P Foundation's description of a fourth-order system captures this: "The 4th Order system is contextualized, embedded and integrated into the context… It operates both as a system in its context, and as a system that is part of the context". Applied to governance, this means an organization's AI risk processes must not only influence the technical system, but also evolve from lessons learned. For example, rather than treating risk categories as fixed, a fourth-order approach would allow the organization to revise what it considers "high risk" based on emerging information. In short, the model envisions governance as a meta-system with self-awareness: it steers itself via feedback loops. This self-regulatory, embedding quality (cf. Chiolerio's "immergence" into the environment) sets the stage for more robust, sustainable AI oversight.

5. Comparative Analysis of Leading Frameworks

Four prominent AI governance frameworks illustrate the current landscape. The OECD AI Principles (2019; updated 2024) offer voluntary guidelines emphasizing trust and democracy. They call for AI that is "innovative and trustworthy" while respecting human rights and democratic values. The OECD enumerates values such as inclusive growth, transparency, robustness, fairness and accountability, and urges international cooperation in governance. These principles serve as a non-binding baseline: many countries (including EU members) cite them in national policies. While broad, the OECD framework lacks specific enforcement mechanisms, relying on governments and industry to interpret it.

In contrast, the EU Artificial Intelligence Act (Regulation (EU) 2024/1689) is a binding law. The European Commission bills it as the "first-ever comprehensive legal framework on AI," aimed at fostering "trustworthy AI in Europe". The Act adopts a risk-based approach: certain "unacceptable" AI uses are banned outright (e.g. social scoring systems or biometric surveillance in public spaces). "High-risk" systems (such as AI for hiring, credit scoring, or medical devices) must undergo strict premarket checks: providers must conduct risk assessments, ensure high-quality training data, maintain detailed documentation, and implement human oversight and robustness measures. Lesser-risk tools (e.g. AI chatbots) are subject to transparency requirements (users must be notified they interact with AI). To ease SME compliance, the Act includes supportive measures: for instance, SMEs get priority

access to regulatory AI "sandboxes" and benefit from reduced fees and lighter reporting obligations. Critics have raised concerns: some warn that the Act's broad definition of AI might inadvertently capture simple software, potentially "stifling further innovation" with unclear boundaries.

The IEEE Ethically Aligned Design (EAD) documents are voluntary industry guidelines produced by the IEEE Global Initiative on the Ethics of Autonomous and Intelligent Systems. They are not legally enforceable but are influential in engineering practice. EAD is explicitly human-centric. It is subtitled "A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems"(Floridi & Cowls, 2019). Version I of EAD articulates high-level ethical principles and concrete recommendations, aiming to ensure AI "provably aligns with and improves holistic societal wellbeing". EAD covers a wide range of issues—from privacy and accountability to ecological sustainability—and it has spurred related standards (e.g. the IEEE P7000 series) and certification efforts. However, as a self-regulatory vision, EAD leaves actual implementation up to organizations; it relies on technologists and policymakers to translate principles into practice.

The UNESCO Recommendation on the Ethics of AI (2021) is the first global intergovernmental AI ethics framework. It explicitly centers human rights and dignity: "the protection of human rights and dignity is the cornerstone of the Recommendation," with specific emphasis on transparency, fairness, and the importance of human oversight of AI. UNESCO's document extends beyond general values by defining ten Policy Action Areas (e.g. data governance, environment, education, health, culture) that member states should address in AI policies. In practice, UNESCO's approach provides a normative standard for states and organizations, though it lacks legal binding force. The UNESCO framework underscores broad societal considerations (e.g. equity, literacy, gender inclusion) alongside technical ethics.

In comparing these frameworks, common themes emerge: all stress trustworthy, human-centric AI (rights, transparency, fairness, accountability), and call for multi-stakeholder cooperation. However, they differ in scope and mechanism. The EU Act is detailed law with penalties, while OECD and UNESCO are non-binding principles. IEEE EAD is a private-sector initiative focusing on design. Notably, none of these frameworks explicitly incorporates fourth-order systemic dynamics. In practice, each provides static goals or rules but offers limited guidance on how organizations should adapt their governance processes over time. From a fourth-order perspective, this is a key limitation: governing AI only through fixed principles or checklists may miss how risks evolve. A fourth-order view suggests that we need governance models that themselves learn and reconfigure based on experience. For example, whereas the EU Act mandates documentation, a 4th-order approach would also ask how those documents feed back into policy revision and organizational learning. In summary, existing frameworks supply important content but largely assume that organizations will implement them as given. They do not prescribe how organizations can co-evolve with their environments, which is precisely the gap our approach seeks to fill.

6. Proposed Socio-Technical Cybernetic Governance Model

To address these gaps, this paper proposes a novel governance model grounded in socio-technical systems theory and fourth-order cybernetics. The model treats AI governance as an adaptive socio-technical feedback process, not a one-time compliance checklist. Its key components are:

- Layered Reflexive Governance: Risk management is structured in layers mirroring organizational scales. At the technical layer, firms implement continuous monitoring (e.g. automated logging, explainability and bias-detection tools) on their AI systems. At the human/organizational layer, cross-functional teams (engineers, domain experts, ethicists,

and end-users) hold regular review meetings. These teams interpret the technical data, assess social implications, and decide adjustments. At a higher "meta" layer, industry consortia or regulator-led coalitions aggregate lessons from multiple firms. In effect, each layer provides feedback to the others, creating double-loop learning.

- Participatory Co-Design and Oversight: Consistent with STS principles, the model embeds stakeholder engagement throughout the AI lifecycle. For example, when developing an AI tool, an SME would hold co-design workshops with employees and even clients to gather requirements and detect potential harms early. During deployment, affected parties (users, customers) can report issues (e.g. through surveys or community forums). Governance bodies then incorporate this feedback. Such practices echo sociotechnical guidance: companies should "co-design technical solutions with the communities who stand to benefit" and engage domain experts and users to inform design. In essence, oversight is shared: the organization and its stakeholders jointly steer the AI's evolution.

- Context-Integrated Monitoring: The governance system continuously integrates external context signals. For instance, it watches for regulatory changes (like updates to the EU AI Act), market shifts, or public concerns. If a new law is proposed or a public controversy erupts (say, about student data privacy), the SME's governance team revises its risk matrix accordingly. This reflects the 4th-order idea that a system is "embedded in its context". Practically, this might involve automated news monitoring and scheduled policy reviews. The key is that governance does not react only to internal incidents but remains sensitive to the broader environment, "immerging" into it.

- Adaptive Compliance Processes: Rather than static checklists, policies and controls are treated as living documents. An internal AI ethics code or risk policy is periodically reviewed and updated based on new experiences. Technical controls (e.g. model retraining frequency, privacy thresholds) are adjusted via ongoing metrics and feedback. For example, if an audit finds a persistent bias, the model is retrained with more diverse data. The organization documents not just compliance records, but also lessons learned and process changes, feeding them back into a continuous improvement loop. This cybernetic self-regulation means the governance "rules of the game" evolve as the game is played.

Together, these features create a socio-technical cybernetic loop: technical performance and social inputs inform each other iteratively. The model leverages existing structures where available. For instance, an SME can use the EU AI Act's classification as a starting taxonomy, but then enrich it with internal risk categories based on its context. Czech and European knowledge-based SMEs can plug into national AI programs: they might test their AI in an EU-regulator-backed sandbox (as SMEs receive priority access) and share findings through Digital Innovation Hubs or industry consortia.

The goal is to situate each organization within a learning ecosystem. In effect, the firm becomes both a subject and object of governance: it helps shape norms even as it follows them, exemplifying a 4th-order "meta-system" stance.

7. Discussion: Implications and Use Cases

This fourth-order socio-technical governance model has several implications. It operationalizes responsible innovation principles by embedding reflexivity, inclusion, and anticipation into AI risk management. In practice, it means organizations are proactive learners, not just rule-followers. For example, by institutionalizing stakeholder feedback, firms can anticipate unintended harms (e.g.

algorithmic bias) before they escalate. By continuously scanning the environment, they can adapt faster than regulatory cycles. Policy-wise, this suggests that regulators and standards bodies should encourage dynamic compliance (for example, by recognizing continuous improvement reports or iterative certification), not just one-time audits. The model also builds trust: when employees and clients see that an organization takes feedback seriously and adjusts its technology, confidence in AI grows.

The following use cases illustrate how knowledge-based SMEs might apply the model:

- Consulting SME: A small consulting firm develops an AI analytics tool for clients. Using our model, the firm first co-designs the tool with senior consultants and pilot clients, uncovering initial assumptions (e.g. which market data matters). Upon deployment, the firm collects feedback from consultants using the tool – perhaps clients report that certain recommendations seem biased or irrelevant. The governance team then investigates, retrains the model on better data, and updates the tool. The team monitors technical metrics (accuracy, fairness) alongside business metrics (client satisfaction, revenue impact), and holds monthly review sessions. If a new data privacy regulation is announced, the firm revises its data handling policy immediately. In this way, the tool evolves based on both quantitative logs and qualitative user insights, embodying a continuous STS feedback loop.

- Educational SME: An EdTech startup offers an AI-driven personalized learning platform. Following our approach, it pilots the system in classrooms and gathers teacher and student input on the learning content. Teachers note if the AI's suggestions are pedagogically sound and culturally appropriate. The startup's oversight board (including educators) uses this feedback to adjust the recommendation algorithms and to define new usage guidelines. The platform also transparently informs students when content is AI-generated, addressing UNESCO's call for transparency and human oversight. Importantly, the governance team tracks changes in curriculum standards or public concerns about AI in education. For example, if an education authority updates its digital literacy curriculum, the firm updates the AI content to match. In sum, the educational SME governs its AI system by tightly coupling technological adjustments with social context and stakeholder oversight.

- Financial SME: A small fintech company deploys an AI system for SME loan approvals. Recognizing the "high-risk" nature of credit decisions, the firm applies the EU Act's rules (strict testing, traceability, human review) and layers on this model. The AI system includes extensive logging and fairness metrics (fulfilling the Act's documentation and accuracy requirements). An internal ethics committee (with finance experts and customer advocates) reviews any contentious decisions. The company participates in an EU regulatory sandbox (which offers free priority access for SMEs) to test its model under supervision. It also scans financial news and policy forecasts: if the economy shifts or if new laws on algorithmic lending are proposed, the model parameters or risk thresholds are adjusted accordingly. Thus, governance is not a single compliance project but an ongoing adaptive process, with the firm constantly co-evolving its AI model, aligning with the fourth-order notion of a self-modifying system.

In each case, the model synthesizes legal rules and ethical ideals into practice. It leverages formal mechanisms (the EU Act's obligations, OECD best-practice, UNESCO's human-rights lens, IEEE's wellbeing focus) but situates them in a reflexive organizational process. Practical supports like EU sandboxes and Digital Innovation Hubs become nodes in the feedback network. Crucially, the

emphasis is on how the organization uses these tools: co-design workshops, iterative testing, and open communication channels operationalize the abstract principles. This is aligned with sociotechnical best-practices: for example, a known recommendation is to involve future technology users early and iteratively, which this model institutionalizes.

8. Conclusion

This paper has proposed that bringing Fourth-Order Cybernetics into socio-technical AI governance can address key limitations of existing approaches. Our review showed that while frameworks like the EU AI Act, OECD Principles, IEEE EAD, and UNESCO Recommendations provide valuable guidance on what values to uphold, they often assume governance is static. By contrast, a fourth-order STS perspective treats governance itself as an evolving, embedded system. Our novel model makes this explicit, framing AI risk management as a continuous, multi-layered feedback loop. This deepens the connection between technology and context, ensuring that Czech and European knowledge-based SMEs can adapt their AI usage responsibly over time.

In doing so, the model also resonates with responsible innovation discourse: it embeds reflexivity, anticipatory thinking, and stakeholder inclusion at the core of governance. Policymakers and standards bodies should note this orientation, perhaps by incentivizing adaptive compliance (e.g. by recognizing "learning" reports or dynamic risk assessments). Future research should empirically evaluate this framework: case studies could document how SMEs implement layered governance and whether it leads to fewer AI-related incidents. Overall, embracing a fourth-order socio-technical outlook offers a promising pathway to more resilient and responsible AI deployment in practice.

Declaration on Generative AI

During the preparation of this work, the author used GPT-4 in order to: Grammar and spelling check. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] Ropohl, G. (1999). Philosophy of socio-technical systems. Science, Technology & Society, 4(3), 59–76.

[2] Leitch, S., & Warren, M. J. (2010). ETHICS: The past, present and future of socio-technical systems design. In P. Trischler & H. Schtzer (Eds.), History of Computing (IFIP AICT, Vol. 325, pp. 189–197).

[3] OECD. (2024). AI Principles Overview. Retrieved from OECD.AI.

[4] European Commission. (2024). Shaping Europe's Digital Future: The AI Act. Niehaus & Wiesche, 2021.

[5] ArtificialIntelligenceAct.eu (Statworx). (2025). Small Businesses' Guide to the AI Act. Kumar et al., 2023.

[6] IEEE Global Initiative on Ethics of A/IS. (2019). Ethically Aligned Design (v1).

[7] UNESCO. (2021). Recommendation on the Ethics of AI. Hagendorff, 2020.

[8] Bogen, M., & Winecoff, A. (2024). Applying Sociotechnical Approaches to AI Governance in Practice. Center for Democracy & Technology.

[9] Chiolerio, A. (2020). Liquid Cybernetic Systems: The Fourth-Order Cybernetics. Advanced

[10] “Fourth Order Cybernetics” (2023). P2P Foundation.

[11] Clifford Chance. (2023, April). The EU AI Act: Concerns and Criticism. Retrieved from Clif-

[12] European Commission. (2023). Czech Republic AI Strategy Report. Retrieved from AI Watch

[13] Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. Interacting with Computers, 23(1), 4–17. https://doi.org/10.1016/j.intcom.2010.07.003

[14] Cheong, B. C., et al. (2024). The sociotechnical entanglement of AI and values. AI & Society. https://doi.org/10.1007/s00146-023-01852-5

[15] Bednar, P.#M., & Welch, C. (2020). Socio$technical perspectives on smart working: Creating meaningful and sustainable systems. Information Systems Frontiers, 22(2), 281–298. https://doi.org/10.1007/s10796-019-09921-1

[16] Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing arti"icial intelligence into the organization. Journal of Business Research, 120, 262–273. https://doi.org/10.1016/j.jbusres.2020.07.045

[17] Chiolerio, A. (2020). Liquid cybernetic systems: The fourth-order cybernetics. Advanced Intelligent Systems, 2(12), 2000120. https://doi.org/10.1002/aisy.202000120

[18] Dalpiaz, F., Giorgini, P., & Mylopoulos, J. (2013). Adaptive socio-technical systems: A requirements-based approach. Requirements Engineering, 18(1), 1–24. https://doi.org/10.1007/s00766-011-0132-1

[19] Niehaus, F., & Wiesche, M. (2021). A socio-technical perspective on organizational interaction with AI: A literature review. Proceedings of the European Conference on Information Systems (ECIS 2021). https://aisel.aisnet.org/ecis2021_rp/156

[20] Kumar, A., Krishnamoorthy, B., & Bhattacharyya, S. S. (2023). Machine learning and arti"icial intelligence-induced technostress in organizations. International Journal of Organizational Analysis, 32(4). https://doi.org/10.1108/IJOA-01-2023-3581

[21] Xu, W., & Gao, Z. (2024). An intelligent sociotechnical systems (iSTS) framework. https://doi.org/10.48550/arXiv.2401.03223

[22] Dean, S., Gilbert, T. K., Lambert, N., & Zick, T. (2021). Axes for sociotechnical inquiry in AI research. https://doi.org/10.48550/arXiv.2105.06551

[23] Ehsan, U., & Riedl, M. O. (2020). Human-centered explainable AI. https://doi.org/10.48550/arXiv.2002.01092

[24] Binder, T., & Sommerville, I. (2011). Socio-technical systems. Interacting with Computers, 23(1), 4–17. https://doi.org/10.1016/j.intcom.2010.07.003

[25] Hagendorff, T. (2020). The ethics of AI ethics. Minds and Machines, 30(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

[26] Floridi, L., & Cowls, J. (2019). A unified framework for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

[27] Hazenberg, J., & Zwitter, A. (2024). Cybernetic governance. Ethics and Information Technology. https://doi.org/10.1007/s10676-024-09763-9

[28] Burton-Jones, A., & Grange, C. (2013). Powers of action in socio-technical systems. Journal of Management Information Systems, 30(4), 13–48. https://doi.org/10.2753/MIS0742-

[29] Burton-Jones, A., & Straub, D. (2006). Reconceptualizing system usage. Information Systems Research, 17(3), 228–246. https://doi.org/10.1287/isre.1060.0096