Cognitive Computing in Sociotechnical Systems, Vol. 205, 2026, 95–103
https://doi.org/10.54941/ahfe1007364

Towards Fair Representation in AI-Mediated Decision-Making: A Conceptual Model for Socio-Technical Contexts

Noushin Qeybi[1], Khadija Sabiri[2], and Gustavo Vieira[3]

[1]Human Technology Centre, RWTH Aachen University, 52062 Aachen, Germany
[2]MORE – Laboratório Colaborativo Montanhas de Investigação – Associação, 5300-358 Bragança, Portugal
[3]CeDRI, SusTEC, Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal

ABSTRACT

AI-mediated decision-making enhances human performance and efficiency in small and medium-sized enterprises (SMEs). However, compared with human decision-making, it raises concerns about human-centered principles such as transparency, fairness, and representation. In particular, formal worker representatives, such as works councils, who have relied on conventional oversight, often lack the technical entry points required to influence, oversee, or negotiate black-box AI-mediated decisions. This mismatch creates a representation gap that challenges the legitimacy of algorithmic outcomes and raises doubts about whether workers' voices are adequately considered. Consequently, despite ongoing efforts to balance biased AI, it remains unclear how fair representation can be conceptualized and practically realized within digitally mediated decision-making. Building on this foundational challenge, the study asks how workers' voices can shape AI-mediated decisions concerning the shop floor; how fair representation can be conceptually defined and embedded as a human-centered principle. To achieve this, the study first evaluates existing socio-technical and institutional attempts to embed human-centered principles into algorithmic decision-making, from social forms of representation, such as collective bargaining, to technical forms, such as Multi-Agent Systems (MAS). Secondly, it addresses the limitations of these existing approaches and comprehensively bridges the representation gap by developing a novel five-layered model, grounded in a case study and shop-floor insights. The model spans Level 0 (problem recognition) to Level 4 (AI-mediated decisions via MAS), integrating worker interests into AI-driven processes to reduce centralized, unilateral decision-making and ensure a substantive role for workers and their representatives. The study shows that bridging the representation gap requires more than algorithmic systems; the proposed framework highlights a deeper, trust-building form of worker involvement at the shop-floor level. Since AI is limited in fulfilling the social and legitimate representational functions of works councils, their participation at the proposed layers becomes essential. This interdependence forms a "nested representation," integrating workers' needs into AI tools to strengthen the human-centered foundations of AI-supported decisions in SMEs.

Keywords: AI-mediated decision-making, Representation gap, SME, Conceptual model, nested representation

Received March 30, 2026; Revised April 28, 2026; Accepted May 17, 2026; Available online July 20, 2026

INTRODUCTION

The increasing use of AI in industry has driven significant transformations within SMEs, reshaping work routines, job roles, and operational decision-making (Amanollahnejad et al., 2026; OECD, 2025). While these developments promise enhanced efficiency and performance, they also introduce critical challenges regarding the legitimacy, transparency, and fairness of AI-mediated processes. Recent research has extensively examined socio-technical bias and transparency in AI (Bahangulu & Owusu-Berko, 2025; Sarker et al., 2026); however, issues of worker representation and institutional legitimacy have received comparatively little attention, leaving a significant gap in our understanding of how AI can facilitate socio-technical systems that meaningfully reflect workers' interests and needs.

Since the Industrial Revolution, the inclusion of workers' voices in organizational decision-making has constituted a pivotal concern. Historically, this need for worker voice led to the emergence of formal structures, such as works councils, designed to institutionalize employee representation within organizations (Fischer-Daly & Anner, 2023; Nelson, 1993). This role is especially critical in SMEs, where shop-floor decisions directly determine operational performance. However, in the digital transformation of these structures, current AI-mediated systems often lack mechanisms to incorporate these representatives (Qeybi & Böschen, 2025a; Romo et al., 2024; Shao et al., 2025). When AI-generated decisions are assumed to reflect workers' perspectives without their active input, questions of fair representation naturally arise.

This creates a representation gap, particularly at the shop-floor level where workers first encounter operational issues. This gap raises fundamental questions: Do AI-mediated decisions sufficiently reflect workers' interests? Do their consequences align with workers' expectations? And can workers trust and act upon these decisions? To address these questions, this paper first evaluates existing socio-technical and institutional attempts to embed human-centered principles into algorithmic systems, ranging from social forms of representation (e.g., collective bargaining) to technical forms (e.g., MAS). Building on these insights, the study introduces a novel conceptual five-layered model designed to support democratic, AI-supported decision-making. Grounded in a case study and shop-floor observations, the model illustrates how workers and their representatives can be integrated into the technical architecture of these algorithmic systems.

The remainder of the paper is structured as follows: The Background section reviews the literature on socio-technical frameworks and human-centered AI. The subsequent section outlines the methodological approach, followed by the presentation of the five-layered model in detail. Finally, the paper concludes with a discussion of the proposed model and directions for future research.

BACKGROUND

A growing body of academic literature and industrial research projects examines the socio-technical effects of integrating AI into the workplace. To date, these efforts have primarily focused on enhancing transparency, mitigating bias, or in some cases, leveraging traditional collective bargaining to regulate algorithmic decision-making. However, while the theoretical frameworks for "how AI should be applied" are well-established, there remains a significant gap regarding the practical, meaningful, and direct socio-technical involvement of workers in AI-mediated decision-making. Specifically, the transition from purely automated decisions toward participatory human-AI arrangements, and the operationalization of fair representation within these systems remain underexplored.

Recent years have seen institutions move beyond theory to leverage collective bargaining rights, ensuring that workers and their representatives play a role in overseeing AI-mediated decision-making. Key examples include the Danish Hilfr2 agreement (2024) and the Spanish Riders' Law (2021), both of which empower workers to influence and challenge algorithmic decisions. These frameworks establish rights such as objecting to algorithmic matching, pursuing legal challenges in labor courts, and institutionalizing representation through protected virtual spaces. Furthermore, they mandate that representatives be informed of the automated decision logic affecting working conditions, thereby opening the "black box" of algorithmic decisions to human oversight (Brunnerová et al., 2024; Eurofound, 2025; Ilsøe, 2025; Todoli-Signes, 2021).

Beyond collective bargaining, the necessity of human validation in digital decision-making is reflected in a broad range of frameworks, from Responsible Innovation to addressing the Responsibility and Governance Gaps. Specifically, parallel efforts focus on an evolution from technical Human-in-the-Loop (HITL) architectures toward the normative goal of Meaningful Human Control (MHC). These frameworks ensure that human intervention remains the primary driver of the system's logic. While these concepts initially emerged from safety and security debates, they have increasingly permeated industrial organizations, where they are now vital for governing complex machinery and shop-floor decision systems (Balamurugan et al., 2025; Kakko, 2022; Khomkhunsorn, 2025; Matthias, 2004; Owen, 2025; Qeybi & Böschen, 2025b; Santoni de Sio & van den Hoven, 2018).

Integrating humans and machines for decision-making, particularly in SMEs, which do not function as "linear phenomena," requires a return to Socio-Technical Systems (STS) theory (Amanollahnejad et al., 2026). Moving beyond joint optimization to focus on embedding ethical values, modern STS approaches consider technical efficiency alongside a critical question: Do system outcomes align with workers' needs and safety? Crucially, how can workers and their representatives keep pace with rapid technological advancements without losing their agency? To address this, Participatory Design (PD) involves workers and unions in the iterative design and implementation of technology, ensuring that systems enhance human labor. In modern STS, PD has evolved further, incorporating a broader range of stakeholders to ensure that systems remain flexible and adaptable in the face of constant digital transformation (Ang et al., 2025; Klein, 2014; Sony & Naik, 2020; Trist & Bamforth, 1951; Wacnik et al., 2025).

Drawing upon these socio-technical foundations, decentralized MAS provide a robust technical framework for operationalizing the negotiation power of unions within AI-mediated decision-making. Their inherent ability to negotiate and process data in real-time provides the infrastructure for transparent, non-hierarchical decision-making on the shop floor. By encoding the negotiated interests and expectations of both workers and union representatives into autonomous agents, MAS can facilitate a specific, firm-related variant of democratic AI governance. This transformation presents a pathway to translate the principle of collective bargaining into the technical architecture itself, representing a significant step toward making AI-mediated decision-making more legitimate, participative, and worker-centered (Ahmed et al., 2024; Lin, 2025; Qeybi & Böschen, 2025b; Schmitt et al., 2023; Shin, 2026; Vinitsky et al., 2023).

Building on this background, this study presents a conceptual model that integrates socio-technical requirements for worker representation into AI-mediated decision-making. To detail the development of this framework, the following section outlines the methodology employed, followed by a comprehensive presentation of the five-layered model.

METHODS

To address the socio-technical representation gap in AI-mediated decision-making within SMEs, this study adopts an informed Grounded Theory approach (Thornberg, 2012). This exploratory method allows for the development of a novel conceptual model that operationalizes worker voice where previous frameworks have struggled. The empirical foundation is based on a two-step case study conducted in 2023 and 2024 as part of the EU-FAIRWork project.

The first phase consisted of an on-site visit, document analysis and preliminary worker interviews at a selected SME to establish the empirical baseline and to identify the contextual factors influencing the implementation of a Decision Support System (DSS). The second phase sought deeper insights into worker perspectives through a comprehensive workshop and follow-up interviews (Gheibi & Böschen, 2024; Qeybi & Böschen, 2025a). Analysis of this data revealed divergent insights regarding participation. While shop-floor routines showed no evidence of participatory decision-making, management viewed worker involvement as essential for transparency and technology acceptance. Furthermore, the Workers' Council expressed clear concerns regarding replacement by digital tools, emphasizing that human discretion and the "four-eyes principle" must be preserved to prevent the pure delegation of decisions to AI systems.

Recognizing these insights, the five-layered model was designed as a socio-technical solution to bridge the identified representation gap and address the project's requirement for a democratic DSS. By synthesizing these empirical findings with the technical capabilities of MAS, the model ensures that shop-floor workers can be involved, both directly and through their representatives, in algorithmic processes. This architecture leads to more legitimate and fair representation within AI-mediated decision-making (Vieira et al., 2025).

The following section details the nuanced levels of the developed model, explaining the rationale for each layer and its contribution to participatory socio-technical decision-making.

NOVEL CONCEPTUAL MODEL FOR NESTED REPRESENTATION

To bridge the identified representation gap and empower workers to negotiate within AI-mediated decision-making, this study proposes a five-layered model based on case study insights, which function as technical and procedural entry points. The model spans from the worker level (Level 0), ensuring involvement from the outset, to participatory AI-mediated decisions via MAS (Level 4). In this framework, MAS are interpreted as a technical form of representation embedded in a specific socio-technical setting. Although MAS enhance opportunities for representation through a broad range of data parameters, legitimacy challenges arise as the quality of the technical representation depends on the social reality of the SME. Shop-floor workers are rooted in specific socio-material practices, and they require tailored representation to address their unique needs. Achieving fair representation depends on balancing the tension between the social aims of representation and the technical forms of representation. Thus, representation must be established as a socio-technical practice with and through MAS. To resolve these tensions, we employ the concept of legitimation to focus on the socio-technical logistics and structures essential for introduction and monitoring within the SME. Hence, it is imperative to address the tensions indicated by the arrows (See Figure 1). Whether an expressed worker need or a systemic MAS logic, each requires undergoing a "legitimation run." This process is essential to fairly transform these exchanges into valid, accepted realities. Representation within this context occurs via two primary channels: workers autonomously articulating their needs, or experts identifying and expressing these requirements on their behalf.

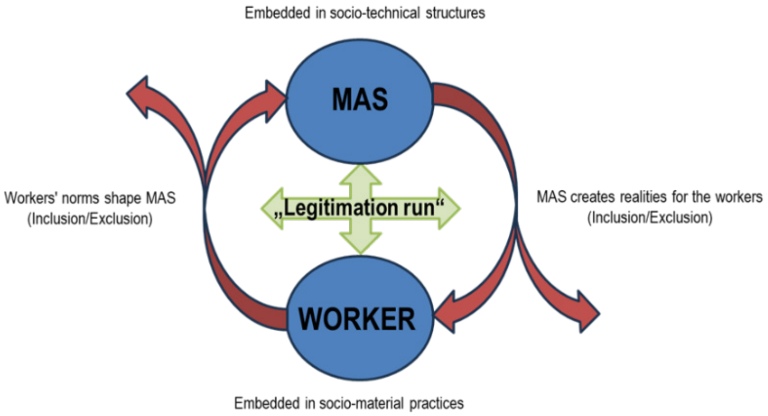

Figure 1: Analytical exploration space for far representation and legitimation.

To operationalize this, our model utilizes a five-layered funnel diagram, facilitating reflexive cycles of opening-up and closing-down representations. Our conceptual structure is based on three crucial elements. Firstly, the MAS functions as a form of indirect representation, acting as a mediation instance between workers and the specific production situations. Secondly, this indirect representation is developed as a layered model, recognizing that appropriately constituted mediation processes are necessary to maintain organizational decision-making capacity. Thirdly, these layers elaborate the relationship between the technical MAS parameters and worker requirements.

This "nested representation" structure enables fair representation by balancing the desire for representation with the technical form of being represented. The process begins at Level 0 (Problem Recognition), where workers identify issues and seek collective solutions. Moving to Level 1 (Consultation), representatives convey these issues to experts to deliberate on potential approaches. At Level 2 (Participation), workers select preferred solutions from those suggested. Before technical integration, these solutions undergo Level 3 (Human Negotiation), where representatives and experts deliberate on which solutions align with system goals. This interaction assesses the precision of the solution while enhancing legitimacy and perceived justice. Once this three-stage legitimation process is complete, the information is integrated into Level 4 (MAS). Here, human input, is modeled into the MAS and aligned with specific process goals. A negotiation dynamic between agents is then designed to meet system goals within the boundaries of each task. These agent interactions play a decisive role in balancing stakeholder parameters, achieving fair representation (See Figure 2). Such a nested structure demonstrates the socio-technical complexity of operationalizing fair representation in SMEs. While these levels can be technically mapped, they require a socio-technical arrangement where coordination is anchored institutionally. Consequently, within this framework, fair representation is conceptually defined as the procedural alignment of worker interests and systemic logics through a legitimate, multi-layered mediation process.

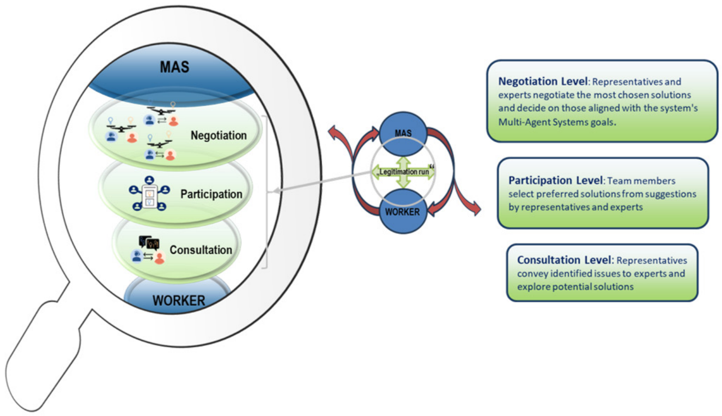

Figure 2: Five-layered model for a fair nested representation: Operationalizing the legitimation process in AI-mediated decision-making.

CONCLUSION

The conceptual goal of this study was to bridge the representation gap in AI-mediated decision-making in socio-technical environments. For this purpose, this study initially reviewed the existing institutional attempts to integrate human-centered principles into algorithmic decision-making. These approaches ranged from collective bargaining frameworks to responsible innovation, STS, and decentralized MAS. However, these approaches lack operationalizing workers' needs and their fair representation in everyday algorithmic operations. Therefore, it was decisive to model how worker voice can be embedded into these systems. The solution was the presented five-layered framework grounded in empirical insights from the FAIRWork project. This architecture defines fair representation conceptually as a socio-technical process of legitimation where the direct and indirect representation of workers, particularly through the presence of works councils, ensures that AI outcomes are governed by shop-floor needs rather than purely algorithmic processes. By doing so, the model demonstrates how workers' voices can meaningfully shape shop-floor AI-decisions, thereby enhancing decision precision and representational fairness.

However, a critical tension remains regarding the source of this requirement. While this structure provides the infrastructure for human-centered AI, our empirical evidence indicates that a proactive interest in decision-making participation was mainly expressed by management rather than shop-floor workers. This raises a fundamental question: who truly benefits from encouraging fairness and democratic participation in AI decisions? In the absence of explicit shop-floor worker demands in our case study, it is necessary to consider whether Fairness and Democracy are being implemented as top-down impositions or forced fairness rather than bottom-up necessities. Therefore, future research and cross-project analysis are required to investigate the true source of willingness for participatory principles in AI tools. If there is no immediate worker demand, should we continue to consider a fair and democratic AI approach as a necessary ethical principle?

ACKNOWLEDGMENT

This work has been supported by the FAIRWork project (www.fairwork-project.eu) and has been funded within the European Commission's Horizon Europe Programme under contract number 101069499. This paper expresses the opinions of the authors and not necessarily those of the European Commission. The European Commission is not liable for any use that may be made of the information contained in this paper.

REFERENCES

Ahmed, I., Syed, M. A., Maaruf, M., & Khalid, M. (2024). Distributed computing in multi-agent systems: a survey of decentralized machine learning approaches. Computing, 107(1), 2. https://doi.org/10.1007/s00607-024-01356-0

Amanollahnejad, A., Fosso-Wamba, S., Shabbir, M. S., & Pakseresht, A. (2026). Aligning socio-technical systems: Rethinking AI adoption and digital transformation in SMEs. Information Systems Management, 1–15.

Ang, K. C. S., Sankaran, S., & Liu, D. (2025). Advancing sociotechnical systems theory: New principles for human-robot team design and development. Applied Ergonomics, 129, 104604. https://doi.org/https://doi.org/10.1016/j.apergo.2025.104604

Bahangulu, J. K., & Owusu-Berko, L. (2025). Algorithmic bias, data ethics, and governance: Ensuring fairness, transparency and compliance in AI-powered business analytics applications. World Journal of Advanced Research and Reviews, 25(2), 1746–1763.

Balamurugan, M., Shanmugasamy, K., & Balaguru, S. (2025). Humans in the Loop, Lives on the Line: AI in High-Risk Decision Making. European Journal of Computer Science and Information Technology, 13(51), 27–31. https://doi.org/10.37745/ejcsit.2013/vol13n512731

Brunnerová, S., Ceccon, D., Holubová, B., Kahancová, M., Lukáčová, K., & Medas, G. (2024). Collective bargaining practices on ai and algo rithmic management in european services sectors. In: Central European Labour Studies Institute and Wage-Indicator Foundation Report.

Eurofound. (2025). Collective bargaining on artificial intelligence at work. https://www.eurofound.europa.eu/en/publications/all/collective-bargaining-on-artificial-intelligence-at-work

Fischer-Daly, M., & Anner, M. (2023). Worker Voice: A Literature Review. Center for Global Workers' Rights, The Pennsylvania State University.

Gheibi, N., & Böschen, S. (2024). Democratization in Industry via Multi-Agent Systems, The case of a production company. In AHFE International, Cognitive Computing and Internet of Things. AHFE International Conference, France.

Ilsøe, A. (2025). Digital rights as collective rights: The new Hilfr2 agreement on artificial intelligence ensures both the managerial prerogative and workers' rights to unionize. Tijdschrift Recht En Arbeid, 1.

Kakko, Y. (2022). Meaningful Human Control as an Exceptional Concept: Added Value and Common Ground for CCW Discussions on Lethal Autonomous Weapons Systems 1-24. https://www.nonproliferation.eu/wp-content/uploads/2022/01/KakkoMeaningfulHumanControl2022-1.pdf (EU Non-Proliferation and Disarmament Consortium)

Khomkhunsorn, T. (2025). Meaningful Human Control and Responsibility Gaps in AI: No Culpability Gap, but Accountability and Active Responsibility Gap. Journal of Integrative and Innovative Humanities, 5(1), 35–57.

Klein, L. (2014). What do we actually mean by 'sociotechnical'? On values, boundaries and the problems of language. Applied Ergonomics, 45(2), 137–142.

Lin, C.-T. (2025). Blockchain-based shop floor control system for small and medium-sized enterprise evolution to industry 4.0. Scientific Reports, 15(1), 28601. https://doi.org/10.1038/s41598-025-13809-3

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics and Information Technology, 6, 175–183. https://doi.org/10.1007/s10676-004-3422-1

Nelson, D. (1993). Employee representation in historical perspective. Employee representation: alternatives and future directions, 371–390.

OECD. (2025). AI adoption by small and medium-sized enterprises. In.

Owen, B. (2025). Human in the Loop Decision Making in AI Driven Business Intelligence Systems.

Qeybi, N., & Böschen, S. (2025a). Exploring Democratization in Industry via Multi-Agent Systems: A firm-based Case Study. In: Lucas Paletta (eds) Cognitive Computing and Internet of Things. AHFE (2025) International Conference, USA.

Qeybi, N., & Böschen, S. (2025b). Integrating perspectives on democratization in industry via multi-agent systems: insights from a firm-based case study. Human-Intelligent Systems Integration. https://doi.org/10.1007/s42454-025-00077-9

Romo, Ó. M., Caprile, M., Arasanz, J., de Miguel, P. S., & Moral-Martín, D. (2024). Worker participation and representation:: the impact on risk prevention of AI worker management systems.

Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. Front Robot AI, 5, 15. https://doi.org/10.3389/frobt.2018.00015

Sarker, I. H., Janicke, H., Mohsin, A., & Maglaras, L. (2026). SME-TEAM: leveraging trust and ethics for secure and responsible use of AI and LLMs in SMEs. npj Artificial Intelligence, 2(1), 12.

Schmitt, R. H., Kiesel, R., Buschmann, D., Cramer, S., Enslin, C., Fischer, M., Gries, T., Hopmann, C., Huebser, L., & Janke, T. (2023). Improving Shop Floor-Near Production Management Through Data-Driven Insights. In Internet of Production: Fundamentals, Applications and Proceedings (pp. 1–23). Springer.

Shao, Y., Zope, H., Jiang, Y., Pei, J., Nguyen, D., Brynjolfsson, E., & Yang, D. (2025). Future of work with ai agents: Auditing automation and augmentation potential across the us workforce. arXiv preprint arXiv:2506.06576.

Shin, M. (2026). Agent-Based Decentralized Manufacturing Execution System via Employment Network Collaboration. Applied Sciences, 16(1), 386. https://www.mdpi.com/2076-3417/16/1/386

Sony, M., & Naik, S. (2020). Industry 4.0 integration with socio-technical systems theory: A systematic review and proposed theoretical model. Technology in society, 61, 101248.

Thornberg, R. (2012). Informed grounded theory. Scandinavian journal of educational research, 56(3), 243–259.

Todoli-Signes, A. (2021). Spanish riders law and the right to be informed about the algorithm. European Labour Law Journal, 12(3), 399–402.

Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. Human relations, 4(1), 3–38.

Vieira, G., Zeiner, H., Paletta, L., Fernandes, R., Rosse, H. V., Sabiri, K., F De Paz, J., & Barbosa, J. (2025). Fairness in Designing Decision-Making Processes with Multi-Agent Systems and Human Factors. AHFE International.

Vinitsky, E., Köster, R., Agapiou, J. P., Duéñez-Guzmán, E. A., Vezhnevets, A. S., & Leibo, J. Z. (2023). A learning agent that acquires social norms from public sanctions in decentralized multi-agent settings. Collective Intelligence, 2(2), 26339137231162025. https://doi.org/10.1177/26339137231162025

Wacnik, P., Daly, S. R., & Verma, A. (2025). Participatory design: a systematic review and insights for future practice. Design Science, 11, e21, Article e21. https://doi.org/10.1017/dsj.2025.10009