*Corresponding Author's Email: w.sanjeewa@yorksj.ac.uk,
https://orcid.org/0000-0001-7880-1500
Proceedings of the International Conference on Advanced Research in Management, Economics and Accounting, Vol. 4, Issue. 1, 2026, pp. 1-13
DOI: https://doi.org/10.33422/armea.v4i1.2204
Copyright © 2026 Author(s)
ISSN: 3030-0290 online

Inclusive AI in the Intelligent Economy: An Intersectional Framework for Equality, Diversity, and Ethical Innovation

Winojith Sanjeewa
York St John University, UK

Abstract

Artificial intelligence (AI) increasingly informs consequential workplace decisions, including recruitment, task allocation, performance monitoring, promotion, and dismissal. Although AI systems may improve efficiency and support organisational innovation, they can also reproduce structural inequalities when data, design assumptions, and governance arrangements fail to account for intersecting social positions. This article develops an intersectional framework for governing workplace AI in the intelligent economy. Using an integrative literature-review and conceptual framework-development approach, the study synthesises scholarship on algorithmic fairness, workplace AI, algorithmic management, explainability, accountability, auditing, and participatory governance. The analysis identifies four interdependent requirements for inclusive AI governance: equitable distributive outcomes, meaningful procedural safeguards, continuous technical accountability, and clearly assigned institutional responsibility. The article argues that bias mitigation and nominal human review are insufficient when workers cannot understand, contest, or influence consequential automated decisions. Inclusive governance instead requires lifecycle-based risk management, accessible explanation and redress, substantive human oversight, and meaningful participation by affected workers and communities. The proposed framework offers practical guidance for organisations and policymakers seeking to embed equality, dignity, accountability, and ethical innovation in AI-enabled workplace decision-making.

Keywords: artificial intelligence; algorithmic fairness; intersectionality; workplace AI; AI governance

1. Introduction

Artificial intelligence (AI) is becoming embedded in recruitment, task allocation, performance monitoring, and other workplace decisions. AI adoption may improve efficiency and productivity, but the effects on job design, worker well-being, and organisational power remain contested (Bankins, Ocampo, et al., 2023) (Enholm et al., 2021). Recent reviews link workplace AI to questions of autonomy, stress, mental health, and employee well-being,

while noting that evidence remains uneven across sectors and social groups (Soulami et al., 2024) (Kinowska & Sienkiewicz, 2022). Algorithmic hiring is particularly consequential because it operates throughout the recruitment pipeline and can shape access to employment for groups already exposed to structural disadvantage (Fabris et al., 2024) (Rigotti & Fosch-Villaronga, 2024).

An intersectional perspective is necessary because unfair outcomes cannot be understood through gender, race, disability, or class in isolation. People positioned across overlapping social categories may experience distinct forms of disadvantage that are concealed by broad group-level assessments. Critical interdisciplinary scholarship cautions that dominant technical approaches to machine-learning fairness can abstract away structural injustice, racialised classification, non-consensual data practices, and the absence of participatory design (Weinberg, 2022). Fairness in machine learning is also not a single technical property: it involves different pre-processing, in-processing, and post-processing approaches, alongside unresolved dilemmas about how fairness should be defined and evaluated in context (Caton & Haas, 2023) (Pagano et al., 2023).

This article examines workplace AI as a sociotechnical governance challenge within the intelligent economy. Algorithms increasingly structure major social institutions and may reinforce actuarial forms of classification that shape opportunity, mobility, and social reproduction (Burrell & Fourcade, 2021). A sociotechnical perspective recognises that AI systems are shaped by institutional context, organisational narratives, human control, and existing inequalities rather than by technical design alone (Sartori & Theodorou, 2022). The article therefore argues that inclusive AI requires more than bias mitigation or nominal human review. Organisations must consider how automated systems distribute opportunities and burdens, how decisions can be questioned and corrected, how technical systems are documented and monitored, and who is responsible when harm occurs.

1.1 Research Questions

This study examines how intersectional fairness can be operationalised in the governance of AI systems used across the intelligent economy. It addresses three questions:

1. How do AI-enabled decision systems create or reproduce unequal outcomes for people positioned at the intersection of disability, race, gender, class, and other social characteristics?

2. Which organisational, technical, and regulatory governance mechanisms support more inclusive and accountable AI deployment?

3. How can intersectional fairness, explainability, human oversight, and stakeholder participation be integrated into a practical framework for ethical AI innovation?

Together, these questions shift the analysis beyond whether an AI system is technically accurate or fair for a single demographic group. They instead focus on how AI distributes opportunities, risks, and decision-making power across overlapping social positions and institutional settings.

2. Literature Review

This literature review brings together four related areas: intersectional algorithmic fairness, AI-enabled workplace decision-making, accountability and oversight, and participatory governance. Taken together, these strands show why inclusive AI cannot be treated solely as a technical design issue.

2.1 Intersectionality and Algorithmic Fairness

Intersectionality explains how disadvantage can arise through the interaction of social positions rather than through a single characteristic considered in isolation. In algorithmic decision-making, conventional fairness approaches may focus on differences between broad groups while overlooking within-group inequality, structural disadvantage, and unequal distributions of power (Kasy & Abebe, 2021). Formal fairness criteria may also conflict with one another, making it necessary to assess whether an AI system advances substantive equality in its real-world context rather than merely satisfying a mathematical measure (Green, 2022a). Systematic evidence further shows that bias can arise through datasets, algorithms, user interaction, chosen metrics, and mitigation methods, with no single metric producing uniform results across use cases (Pagano et al., 2023). The review therefore examines how data selection, modelling choices, evaluation practices, and deployment conditions can obscure or reproduce compounded disadvantage across disability, race, gender, class, and other intersecting characteristics.

2.2 AI Decision-Making in the Intelligent Economy

AI increasingly shapes employment through recruitment, task allocation, performance monitoring, and algorithmic management. Research on algorithmic hiring finds that these systems operate throughout the recruitment pipeline and should be assessed in relation to structural inequalities, legal obligations, datasets, mitigation strategies, and stakeholder interests (Fabris et al., 2024). AI-supported hiring may process candidate information efficiently and potentially broaden candidate pools, but recruiters have also reported concerns about data accuracy and limited control over algorithmic matching (Li et al., 2021). Experimental research further suggests that explainable candidate-management systems can affect hiring discrimination differently across demographic contexts, underscoring that explanation is not a universal solution (Hofeditz et al., 2022). More broadly, workplace AI affects workers at individual, group, and organisational levels through human–AI collaboration, employee attitudes, management control, and labour-market change (Bankins et al., 2023).

2.3 Explainability, Accountability, and Human Oversight

Explainability, accountability, and oversight are frequently proposed as safeguards for high-stakes AI. Trustworthy-AI frameworks identify human agency and oversight, technical robustness, privacy and data governance, transparency, diversity and non-discrimination, societal well-being, and accountability as interconnected requirements across the AI life cycle (Díaz-Rodríguez et al., 2023). Explainability research distinguishes data, model, post-hoc, and explanation-evaluation dimensions, while emphasising that explanation should be tailored to the user and application context (Ali et al., 2023). Reviews of black-box AI likewise identify interpretability as a practical constraint on deployment in mission-critical settings (Hassija et al., 2023). However, human oversight alone does not guarantee meaningful control: reviewers need sufficient information, authority, time, and institutional support to identify and correct harms (Green, 2022b) (Alon-Barkat & Busuioc, 2022).

2.4 Inclusive Innovation and Multilevel Governance

Inclusive innovation requires attention to the sociotechnical conditions in which AI is designed and deployed. Participatory AI can help represent the needs, values, and perspectives of historically marginalised communities, but participation can also be co-opted or reduced to

symbolic consultation (Birhane et al., 2022). Research on the participatory turn in AI similarly shows substantial variation in whether affected stakeholders receive genuine agency in design and governance (Delgado et al., 2023). Community-led intersectional practice demonstrates both the value and the practical difficulty of building structures that empower marginalised participants (QueerInAI et al., 2023). Participatory approaches to AI for social good further argue that communities affected by systems should help define what social benefit means and should be treated as partners throughout the project (Bondi et al., 2021).

The literature identifies a central gap. Technical fairness research offers methods for detecting and mitigating bias, while organisational and governance research focuses on accountability, institutional design, participation, and remedy. Recent reviews of AI governance distinguish team-, organisation-, industry-, national-, and international-level governance arrangements, illustrating that responsibility cannot be located solely in the technical system (Batool et al., 2025). Responsible-AI scholarship also cautions that ethical issues arise across the wider development and deployment lifecycle, not solely at the level of a model or fairness metric (Lu et al., 2023) (Camilleri, 2023). This study addresses that gap by developing an intersectional framework for workplace AI governance that links distributive outcomes, procedural safeguards, technical accountability, and institutional responsibility.

3. Method

This article uses a qualitative integrative literature-review and conceptual framework-development approach. The approach is appropriate because the research questions require synthesis across computer science, organisational studies, ethics, law, and social-inequality research rather than testing a single causal relationship. The purpose is to identify how intersectional harms may emerge in workplace AI and to derive governance principles capable of addressing them.

3.1 Research Design

The review followed a staged conceptual synthesis. First, it examined scholarship on intersectional fairness, algorithmic hiring, workplace AI, explainability, auditing, accountability, and participatory governance. Second, it compared how these literatures define harms, affected groups, governance responsibilities, and safeguards. Third, it synthesised the findings into a framework for inclusive workplace-AI governance. This design treats AI as a sociotechnical system whose consequences depend on technical features as well as organisational practices and power relations.

3.2 Search Strategy and Selection Criteria

This integrative review drew on 51 scholarly sources published between 2021 and 2025. The corpus spans management, human-resource management, information systems, computer science, ethics, law, and public governance, reflecting the interdisciplinary character of ethical and inclusive AI. Literature identification was structured around AI-related terms including “artificial intelligence”, “algorithmic decision-making”, “machine learning”, and “algorithmic management” and terms relating to equality and governance, such as “fairness”, “discrimination”, “inclusion”, “accountability”, “explainability”, “impact assessment”, and “participatory AI governance”. A representative search string was: (“artificial intelligence” OR “algorithmic decision-making” OR “machine learning”) AND (“fairness” OR “discrimination” OR “inclusion” OR “accountability” OR “explainability”) AND (“workplace” OR “employment” OR “recruitment” OR “management”). Backward and forward citation

searching was used to identify relevant conceptual, legal, policy, and empirical contributions that might not be captured through keyword searches alone.

Sources were included where they: (1) addressed AI-enabled decision-making in employment or another comparable high-stakes setting; (2) examined fairness, discrimination, inclusion, accountability, transparency, participation, or human oversight; and (3) offered conceptual, empirical, legal, or policy-relevant insight for workplace-AI governance. Sources focused exclusively on technical performance were excluded unless they directly addressed bias measurement, fairness evaluation, or explainability. The final corpus was synthesised across technical, organisational, ethical, labour, legal, and public-governance perspectives. The review also considered the emerging regulatory context for high-impact AI, including the European Union's Artificial Intelligence Act (EU Artificial Intelligence Act, n.d.) and New York City's Automated Employment Decision Tools (AEDT) Law (New York City Council, 2021).These instruments highlight the practical importance of governance arrangements that address bias, transparency, accountability, and meaningful human oversight in AI-supported employment decisions.

3.3 Data Extraction and Analysis

A structured data-extraction table will be used for each included source. The table will record the publication details, research design, AI application, decision context, population or stakeholder group, intersecting inequalities considered, source of harm or bias, proposed governance mechanism, and stated limitations. This approach will make it possible to compare technical and organisational evidence systematically across the selected literature.

The extracted material was analysed through thematic synthesis. Initial coding identified: (a) forms of intersectional disadvantage; (b) data, model, and deployment risks; (c) effects on workers' opportunities, autonomy, dignity, and participation; and (d) governance responses, including auditing, impact assessment, explanation, contestability, and stakeholder engagement. AI impact assessments can support early examination of positive and negative social and ethical effects, though existing approaches vary in scope, process, transparency, and implementation (Stahl et al., 2023). Risk-management guidance similarly emphasises that organisations that design, develop, deploy, or use AI should manage risks throughout the system lifecycle (Tabassi, 2023). Bias-management guidance recognises that harmful impacts may arise across technology processes even without deliberate discriminatory intent (Schwartz et al., 2022).

3.4 Framework Development

The final stage will develop an intersectional AI-governance framework for workplace settings. The framework will comprise four dimensions: **distributive outcomes** (who receives benefits, burdens, and opportunities); **procedural safeguards** (how decisions are explained, reviewed, and challenged); **technical accountability** (how data, models, and outcomes are documented, tested, audited, and monitored); and **institutional responsibility** (which actors are answerable for design, deployment, oversight, and remedy). The framework will be used to organise the results and discussion, while recognising that the study is conceptual and does not itself evaluate a specific workplace AI system.

3.4.1 An Intersectional AI-Governance Framework

The synthesis indicates that workplace AI should be governed across four connected dimensions rather than through a single fairness metric or compliance mechanism.

3.4.1.1 Distributive Outcomes

Organisations should assess who receives the benefits, burdens, and opportunities created by an AI system. This includes examining whether recruitment, allocation, evaluation, or monitoring tools unequally affect access to work, advancement, autonomy, privacy, and meaningful participation. Algorithmic systems can arbitrarily exclude people from opportunities when similar data-driven decisions operate across multiple contexts (Creel & Hellman, 2022). Fairness also becomes more complex in multistakeholder, ranked, and personalised information environments (Ekstrand et al., 2022). Assessment should therefore consider intersecting characteristics, material workplace conditions, and cumulative organisational effects rather than relying only on averages for broad demographic groups.

3.4.1.2 Procedural safeguards.

Workers and applicants need understandable information about when and how AI affects consequential decisions. Meaningful safeguards include timely explanation, accessible communication, correction of inaccurate information, human review with authority to alter outcomes, and channels to challenge or appeal a decision. Empirical research on algorithmic fairness perceptions shows that public views vary according to algorithmic characteristics, human characteristics, comparative human-versus-algorithm effects, and decision consequences (Starke et al., 2022). Transparency and accountability research also identifies technical, legal, ethical, and multistakeholder dimensions of responsible AI governance (Cheong, 2024). These mechanisms matter because automated decisions can otherwise become difficult to question, particularly where systems are technically or organisationally opaque.

3.4.1.3 Technical accountability.

Technical governance should cover the full AI lifecycle: defining the problem, selecting and documenting data, testing models, monitoring outcomes after deployment, and recording system changes. Bias testing should be repeated as organisational contexts, datasets, and system uses change. Responsible-AI engineering research recommends combining multilevel governance patterns, trustworthy processes, and responsible-AI-by-design practices (Lu et al., 2023). Third-party audit research cautions that audits alone are unlikely to establish accountability without careful institutional design and sustained external oversight (Raji et al., 2022). AI governance should therefore treat technical evaluation as necessary but insufficient, integrating it with clear processes for response, remediation, and escalation (Camilleri, 2023).

3.4.1.4 Institutional responsibility

Responsibility should be allocated across employers, human-resource teams, system developers, vendors, managers, and relevant regulators. Accountability requires answerability and depends on recognised authority, the ability to interrogate decisions, and limits on the exercise of power (Novelli et al., 2023). Digital-governance research further distinguishes governance modes that rely on different combinations of control, coordination, incentives, and trust (Hanisch et al., 2023). This is especially important where workplace systems are procured through multi-actor supply chains, which can diffuse responsibility and reduce visibility over decision processes. Organisations should identify who may approve deployment, investigate harms, suspend a system, and provide remedy.

These dimensions are mutually dependent. For example, a technically accurate system can still be exclusionary if affected workers cannot understand, contest, or influence its use. Similarly, participatory processes are insufficient if organisations do not act on the concerns they raise. The framework consequently treats fairness as an ongoing organisational responsibility rather than a one-time model-testing exercise.

4. Discussion

4.1 From Technical Fairness to Substantive Inclusion

The review shows that technical fairness is necessary but insufficient for inclusive workplace AI. Existing technical solutions have often concentrated on scoring and classification problems, particularly fairness and unwanted bias, while broader socially responsible AI requires attention to additional sources of harm (Cheng et al., 2021). Critiques of dominant fairness approaches further identify risks of technological solutionism, ethics washing, non-transparent data collection, and predatory inclusion of marginalised groups (Weinberg, 2022). Fairness methods can identify particular disparities, but their effectiveness depends on how organisations define the decision problem, collect data, choose performance objectives, and act on identified risks. A system can satisfy a selected fairness measure while remaining insensitive to workplace power relations, accessibility needs, or the lived experiences of intersectionally marginalised workers.

The proposed framework shifts attention from isolated model outputs to the broader distribution of power and opportunity. In recruitment, organisations should ask not only whether shortlisting rates differ between groups, but also whether job criteria, data sources, assessments, and appeal procedures disadvantage applicants at particular intersections. Research on algorithmic inclusion in hiring identifies data, design, and decisions as connected sites where organisational fit, categorisation, stakeholder power, auditing, and regulation require attention (Kelan, 2023). In algorithmic management, organisations should assess whether automated monitoring or allocation changes workers' autonomy, dignity, privacy, and access to meaningful work (Bankins & Formosa, 2023). Studies of app-based work show that algorithmic HRM can coordinate tasks, performance, and rewards while creating pervasive forms of control experienced directly by workers (Duggan et al., 2023) (Huang, 2022).

4.2 Participation and Meaningful Oversight

Stakeholder participation is central to identifying harms that may not be visible in technical documentation or aggregate performance data. Participatory AI approaches can better represent the needs and perspectives of historically marginalised communities, but participation can also become symbolic, extractive, or disconnected from decision-making authority (Birhane et al., 2022). Participatory governance has been proposed as a mechanism for ensuring that AI for social good accounts for uneven benefits and risks across communities (Moon, 2023). Human-centric governance likewise requires moving beyond a narrow user-centred perspective toward community- and society-centred approaches that support social sustainability (Sigfrids et al., 2023). Organisations should therefore involve workers and affected communities early, provide accessible routes for input, explain how that input shapes decisions, and compensate participation where appropriate.

Human oversight must likewise be substantive. Assigning a person to review an AI recommendation does not itself establish accountability. Effective oversight requires sufficient information, practical capacity, authority to intervene, and organisational support to challenge an output. Research on AI decision-making shows that human interaction with algorithmic advice can be affected by selective adherence, including greater willingness to accept advice aligned with social stereotypes (Alon-Barkat & Busuioc, 2022). High-level governance principles should therefore be translated into specific organisational requirements for transparency and explainability (Balasubramaniam et al., 2023). Institutions should complement individual review with independent auditing, clear escalation routes, and governance bodies able to pause or withdraw harmful systems.

4.3 Implications for Practice and Policy

For organisations, the framework offers a practical sequence: conduct an intersectional impact assessment before deployment; document data sources, intended uses, limitations, and decision owners; test outcomes across relevant intersections where lawful and feasible; create accessible explanation and redress mechanisms; monitor outcomes throughout deployment; and involve workers in periodic review. These measures should be proportionate to the system's impact, with stronger controls for systems influencing hiring, discipline, pay, promotion, or dismissal. This lifecycle orientation is consistent with responsible-AI and risk-management approaches addressing the responsibilities of actors involved in designing, developing, deploying, and using AI (Díaz-Rodríguez et al., 2023) (Tabassi, 2023). Organisational adoption also depends on how managers understand AI's value-generating mechanisms, enablers, inhibitors, and second-order effects (Enholm et al., 2021).

For policymakers, the analysis supports governance that extends beyond transparency duties. Regulation and guidance should require evidence that high-impact workplace systems are appropriate for their intended purpose, subject to meaningful review, and accountable to affected people. External oversight requires careful institutional design; audits alone do not guarantee algorithmic accountability (Raji et al., 2022). Collective bargaining may also provide an important mechanism for addressing risks to non-discrimination, privacy, and freedom of association arising from algorithmic management (Stefano & Taes, 2022). Governance should remain attentive to inequalities in access, representation, and decision-making authority across gender, race, geography, income, and other social positions (Ayana et al., 2024).

5. Conclusion

This article has argued that inclusive AI in the intelligent economy requires an intersectional approach to workplace governance. AI systems do not operate independently of social structures: their effects are shaped by data practices, design choices, organisational incentives, management decisions, and the opportunities available to workers to question and influence automated processes.

The proposed framework brings together four requirements: equitable distributive outcomes, meaningful procedural safeguards, continuous technical accountability, and clearly assigned institutional responsibility. It offers a way to move beyond narrow assessments of whether an algorithm is accurate or compliant toward a fuller evaluation of whether workplace AI supports equality, diversity, dignity, and ethical innovation.

This article is conceptual and based on an integrative review; it does not test the framework in a particular organisation or sector. Future research should evaluate the framework through comparative workplace case studies, worker-centred qualitative research, and longitudinal assessments of AI systems after deployment. Research should also examine how algorithmic management shapes worker well-being, job autonomy, psychological contracts, and resistance in varied organisational contexts (Kinowska & Sienkiewicz, 2022) (Tomprou & Lee, 2021) (Cini, 2022). Further research is needed to determine how intersectional harms can be identified without creating additional privacy, legal, or data-governance risks. These questions are increasingly important as generative AI creates workplace opportunities alongside concerns about bias, context insensitivity, privacy, misinformation, security, and worker well-being (Budhwar et al., 2023) (Dwivedi et al., 2023).

Ultimately, ethical innovation is not achieved by adding fairness checks after a system has been designed. It depends on embedding inclusion, contestability, and responsibility throughout the

AI lifecycle and on ensuring that workers most affected by automated decisions have a meaningful role in shaping the systems that govern their working lives.

References

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso, J. M., Confalonieri, R., Guidotti, R., Ser, J. D., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion, 99*, Article 101805. https://doi.org/10.1016/j.inffus.2023.101805

Alon-Barkat, S., & Busuioc, M. (2022). Human-AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory, 33*(1), 153–169. https://doi.org/10.1093/jopart/muac007

Ayana, G., Dese, K., Nemomssa, H. D., Habtamu, B., Mellado, B., Badu, K., Yamba, E. I., Faye, S. L., Ondua, M., Nsagha, D. S., Nkweteyim, D., & Kong, J. D. (2024). Decolonizing global AI governance: Assessment of the state of decolonized AI governance in Sub-Saharan Africa. *Royal Society Open Science, 11*(8), Article 231994. https://doi.org/10.1098/rsos.231994

Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkanen, K., & Kujala, S. (2023). Transparency and explainability of AI systems: From ethical guidelines to requirements. *Information and Software Technology, 159*, Article 107197. https://doi.org/10.1016/j.infsof.2023.107197

Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2023). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of Organizational Behavior, 45*(2), 159–182. https://doi.org/10.1002/job.2735

Bankins, S., & Formosa, P. (2023). The ethical implications of artificial intelligence (AI) for meaningful work. *Journal of Business Ethics, 185*(4), 725–740. https://doi.org/10.1007/s10551-023-05339-7

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. *AI and Ethics, 5*(3), 3265–3279. https://doi.org/10.1007/s43681-024-00653-w

Birhane, A., Isaac, W. M., Prabhakaran, V., Díaz, M., Elish, M. C., Gabriel, I., & Mohamed, S. (2022). Power to the people? Opportunities and challenges for participatory AI (pp. 1–8). https://doi.org/10.1145/3551624.3555290

Bondi, E., Xu, L., Acosta-Navas, D., & Killian, J. A. (2021). Envisioning communities: A participatory approach towards AI for social good (pp. 425–436). https://doi.org/10.1145/3461702.3462612

Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R., Boselie, P., Cooke, F. L., Decker, S., DeNisi, A. S., Dey, P. K., Guest, D., Knoblich, A. J., Malik, A., Paauwe, J., Papagiannidis, S., Patel, C., Pereira, V., Ren, S., … Varma, A. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal, 33*(3). https://doi.org/10.1111/1748-8583.12524

Burrell, J., & Fourcade, M. (2021). The society of algorithms. *Annual Review of Sociology, 47*(1), 213–237. https://doi.org/10.1146/annurev-soc-090820-020800

Camilleri, M. A. (2023). Artificial intelligence governance: Ethical considerations and implications for social responsibility. *Expert Systems, 41*(7). https://doi.org/10.1111/exsy.13406

Caton, S., & Haas, C. (2023). Fairness in machine learning: A survey. *ACM Computing Surveys, 56*(7), 1–38. https://doi.org/10.1145/3616865

Cheng, L., Varshney, K. R., & Liu, H. (2021). Socially responsible AI algorithms: Issues, purposes, and challenges. *Journal of Artificial Intelligence Research, 71*, 1137–1181. https://doi.org/10.1613/jair.1.12814

Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics, 6*. https://doi.org/10.3389/fhumd.2024.1421273

Cini, L. (2022). Resisting algorithmic control: Understanding the rise and variety of platform worker mobilisations. *New Technology, Work and Employment, 38*(1), 125–144. https://doi.org/10.1111/ntwe.12257

Creel, K., & Hellman, D. (2022). The algorithmic leviathan: Arbitrariness, fairness, and opportunity in algorithmic decision-making systems. *Canadian Journal of Philosophy, 52*(1), 26–43. https://doi.org/10.1017/can.2022.3

Delgado, F., Yang, S., Madaio, M., & Yang, Q. (2023). The participatory turn in AI design: Theoretical foundations and the current state of practice (pp. 1–23). https://doi.org/10.1145/3617694.3623261

Díaz-Rodríguez, N., Ser, J. D., Coeckelbergh, M., Prado, M. L. de, Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion, 99*, Article 101896. https://doi.org/10.1016/j.inffus.2023.101896

Duggan, J., Carbery, R., McDonnell, A., & Sherman, U. (2023). Algorithmic HRM control in the gig economy: The app-worker perspective. *Human Resource Management, 62*(6), 883–899. https://doi.org/10.1002/hrm.22168

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Ekstrand, M. D., Das, A., Burke, R., & Díaz, F. (2022). Fairness in information access systems. *Foundations and Trends® in Information Retrieval, 16*(1–2), 1–177. https://doi.org/10.1561/1500000079

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). Artificial intelligence and business value: A literature review. *Information Systems Frontiers, 24*(5), 1709–1734. https://doi.org/10.1007/s10796-021-10186-w

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union.* https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Fabris, A., Baranowska, N., Dennis, M., Graus, D., Hacker, P., Saldivar, J., Borgesius, F. Z., & Biega, A. J. (2024). Fairness and bias in algorithmic hiring: A multidisciplinary survey. *ACM Transactions on Intelligent Systems and Technology, 16*(1), 1–54. https://doi.org/10.1145/3696457

Green, B. (2022a). Escaping the impossibility of fairness: From formal to substantive algorithmic fairness. *Philosophy & Technology, 35*(4). https://doi.org/10.1007/s13347-022-00584-6

Green, B. (2022b). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review, 45*, Article 105681. https://doi.org/10.1016/j.clsr.2022.105681

Hanisch, M., Goldsby, C., Fabian, N. E., & Oehmichen, J. (2023). Digital governance: A conceptual framework and research agenda. *Journal of Business Research, 162*, Article 113777. https://doi.org/10.1016/j.jbusres.2023.113777

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2023). Interpreting black-box models: A review on explainable artificial intelligence. *Cognitive Computation, 16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

Hofeditz, L., Clausen, S., Rieß, A., Mirbabaie, M., & Stieglitz, S. (2022). Applying XAI to an AI-based system for candidate management to mitigate bias and discrimination in hiring. *Electronic Markets, 32*(4), 2207–2233. https://doi.org/10.1007/s12525-022-00600-9

Huang, H. (2022). Algorithmic management in food-delivery platform economy in China. *New Technology, Work and Employment, 38*(2), 185–205. https://doi.org/10.1111/ntwe.12228

Kasy, M., & Abebe, R. (2021). Fairness, equality, and power in algorithmic decision-making (pp. 576–586). https://doi.org/10.1145/3442188.3445919

Kelan, E. (2023). Algorithmic inclusion: Shaping the predictive algorithms of artificial intelligence in hiring. *Human Resource Management Journal, 34*(3), 694–707. https://doi.org/10.1111/1748-8583.12511

Kinowska, H., & Sienkiewicz, Ł. (2022). Influence of algorithmic management practices on workplace well-being: Evidence from European organisations. *Information Technology and People, 36*(8), 21–42. https://doi.org/10.1108/ITP-02-2022-0079

Li, L., Lassiter, T., Oh, J., & Lee, M. K. (2021). Algorithmic hiring in practice (pp. 166–176). https://doi.org/10.1145/3461702.3462531

Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2023). Responsible AI pattern catalogue: A collection of best practices for AI governance and engineering. *ACM Computing Surveys, 56*(7), 1–35. https://doi.org/10.1145/3626234

Moon, M. J. (2023). Searching for inclusive artificial intelligence for social good: Participatory governance and policy recommendations for making AI more inclusive and benign for society. *Public Administration Review, 83*(6), 1496–1505. https://doi.org/10.1111/puar.13648

New York City Council. (2021). Local Law 144 of 2021: Automated Employment Decision Tools. *The New York City Council*. https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page.

Novelli, C., Taddeo, M., & Floridi, L. (2023). Accountability in artificial intelligence: What it is and how it works. *AI & Society, 39*(4), 1871–1882. https://doi.org/10.1007/s00146-023-01635-y

Pagano, T. P., Loureiro, R. B., Lisboa, F. V. N., Peixoto, R. M., Guimarães, G. A. de S., Cruz, G. O. R., Araujo, M. M., Santos, L. L. dos, Cruz, M. A. S., Oliveira, E. L. S., Winkler, I., & Nascimento, E. G. S. (2023). Bias and unfairness in machine learning models: A systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. *Big Data and Cognitive Computing, 7*(1), Article 15. https://doi.org/10.3390/bdcc7010015

QueerInAI, O. of, Ovalle, A., Subramonian, A., Singh, A., Voelcker, C., Sutherland, D. J., Locatelli, D., Breznik, E., Klubička, F., Yuan, H., Hetvi, J., Zhang, H., Shriram, J., Lehman, K., Soldaini, L., Sap, M., Deisenroth, M. P., Pacheco, M. L., Ryskina, M., … Stark, L. (2023). Queer in AI: A case study in community-led participatory AI (pp. 1882–1895). https://doi.org/10.1145/3593013.3594134

Raji, I. D., Xu, P., Honigsberg, C., & Ho, D. E. (2022). Outsider oversight: Designing a third-party audit ecosystem for AI governance (pp. 557–571). https://doi.org/10.1145/3514094.3534181

Rigotti, C., & Fosch-Villaronga, E. (2024). Fairness, AI & recruitment. *Computer Law & Security Review, 53*, Article 105966. https://doi.org/10.1016/j.clsr.2024.105966

Sartori, L., & Theodorou, A. (2022). A sociotechnical perspective for the future of AI: Narratives, inequalities, and human control. *Ethics and Information Technology, 24*(1). https://doi.org/10.1007/s10676-022-09624-3

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence* [Working paper]. https://doi.org/10.6028/NIST.SP.1270

Sigfrids, A., Leikas, J., Salo-Pöntinen, H., & Koskimies, E. (2023). Human-centricity in AI governance: A systemic approach. *Frontiers in Artificial Intelligence, 6*, Article 976887. https://doi.org/10.3389/frai.2023.976887

Soulami, M., Benchekroun, S., & Galiulina, A. (2024). Exploring how AI adoption in the workplace affects employees: A bibliometric and systematic review. *Frontiers in Artificial Intelligence, 7*, Article 1473872. https://doi.org/10.3389/frai.2024.1473872

Stahl, B. C., Antoniou, J., Bhalla, N., Brooks, L., Jansen, P., Lindqvist, B., Kirichenko, A. S., Marchal, S., Rodrigues, R., Santiago, N., Warso, Z., & Wright, D. (2023). A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review, 56*(11), 12799–12831. https://doi.org/10.1007/s10462-023-10420-8

Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society, 9*(2). https://doi.org/10.1177/20539517221115189

Stefano, V. D., & Taes, S. (2022). Algorithmic management and collective bargaining. *Transfer: European Review of Labour and Research, 29*(1), 21–36. https://doi.org/10.1177/10242589221141055

Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* [Working paper]. https://doi.org/10.6028/NIST.AI.100-1

Tomprou, M., & Lee, M. K. (2021). Employment relationships in algorithmic management: A psychological contract perspective. *Computers in Human Behavior, 126*, Article 106997. https://doi.org/10.1016/j.chb.2021.106997

Weinberg, L. (2022). Rethinking fairness: An interdisciplinary survey of critiques of hegemonic ML fairness approaches. *Journal of Artificial Intelligence Research, 74*, 75–109. https://doi.org/10.1613/jair.1.13196