JOURNAL OF INTERDISCIPLINARY RESEARCH IN ARTIFICIAL INTELLIGENCE AND SOCIETY

Journal of Interdisciplinary Research in Artificial Intelligence and Society,
2025, 1(1), 1
eISSN: xxxx-xxxx

lectito

A Review of Human-Centered AI: A Multidisciplinary Perspective for Policy-Makers, Auditors, and Users

Ratan Sarkar

Tezpur University, INDIA

Corresponding Author: ratan91@tezu.ernet.in

Citation: Sarkar, R. (2025). A review of human-centered AI: A multidisciplinary perspective for policy-makers, auditors, and users. *Journal of Interdisciplinary Research in Artificial Intelligence and Society, 10*(1), xx. https://doi.org/10.20897/jirais/xxxxx

Published: December 13, 2025

ABSTRACT

This review essay critically evaluates *Human-Centered AI: A Multidisciplinary Perspective for Policy-Makers, Auditors, and Users* (Régis et al., 2024), a landmark contribution that reframes artificial intelligence as a social, ethical, and political project rather than a purely technical pursuit. The volume's strength lies in its methodological pluralism, integrating perspectives from law, healthcare, education, culture, and governance to foreground human flourishing as the central design imperative. By combining conceptual insights with pragmatic frameworks, it advances the human-centered AI (HCAI) paradigm from principle to practice. Limitations include its Euro-American focus and limited Global South perspectives, but the book nonetheless makes a timely and rigorous contribution to AI governance debates, offering policymakers, educators, and researchers critical tools for shaping democratic and inclusive AI futures.

Keywords: human-centered AI, governance, ethics, accountability, human flourishing.

The edited volume *Human-Centered AI (HCAI)* is a landmark contribution to contemporary debates on artificial intelligence, offering a rich, multidisciplinary exploration of what it means to embed human values, accountability, and inclusivity at the core of AI design and governance. At a time when public discourse oscillates between techno-utopian promises and dystopian anxieties, this book reframes AI not as a purely technical project but as a profoundly ethical and social undertaking. By foregrounding human flourishing as the ultimate measure of success, editors Catherine Régis, Jean-Louis Denis, Maria Luciana Axente, and Atsuo Kishimoto present a compelling intellectual architecture for AI that prioritizes people, communities, and democracy over efficiency and profit.

The book opens with a foreword by Yoshua Bengio, one of AI's leading figures, who warns of the risks of concentrated technological power undermining democratic values. His call for proactive regulation and cross-disciplinary collaboration sets the stage for the volume's central concern: making AI systems genuinely human-centered rather than narrowly utilitarian. Shannon Vallor's interview in the introduction is particularly striking. She distinguishes between *responsible AI*, which emphasizes trust and governance, and *human-centered AI*, which prioritizes human flourishing as the ultimate goal of technology. Vallor critiques systems like ChatGPT, noting that their development was guided by scalability and market efficiency rather

than the needs of learners or educators (p. 17). This sharp distinction illustrates the gap between innovation and human-centered outcomes—a recurring theme throughout the book.

The first section lays the theoretical foundations for HCAI. Malwina Anna Wójcik critiques Western rationalist traditions that normalize algorithmic inequality, proposing Ubuntu, the African philosophy of dignity and relationality, as an alternative ethical framework. Oshri Bar-Gil examines recommender systems, highlighting how platforms such as Google subtly reshape autonomy, memory, and rationality by privileging machine-curated choices. Christopher Quintana explores user experience design, arguing that HCAI must embed values directly into technological interfaces rather than treating them as afterthoughts. Pierre Larouche connects ethics with law, illustrating tensions between permissive innovation-driven frameworks and protective rights-based regimes. Giada Pistilli examines large language models, highlighting risks of opacity, bias, and hallucination, and urging philosopher–engineer collaboration to align these systems with human flourishing. Collectively, these chapters expose the challenge of operationalizing values into technical design, reflecting the broader "implementation gap" between aspirational AI ethics principles and real-world practices.

The second section applies these frameworks to institutional domains. In healthcare, Da Silva, Denis, and Régis propose dual governance strategies that balance normative safeguards with contextual flexibility. Jennifer Garard and colleagues extend HCAI principles to agriculture and sustainability, highlighting inclusive data stewardship. Takehiro Ohya explores judicial applications, warning of the dangers when algorithmic reasoning begins to displace democratic values. Bruno Poellhuber and colleagues examine adaptive learning and generative AI in higher education, identifying opportunities for innovation but cautioning against de-skilling. Similarly, Christian Lévesque and his team explore AI in the workplace, warning against disempowerment and emphasizing autonomy, solidarity, and participatory policy. These chapters illustrate that HCAI is not a single template but a flexible, context-dependent framework that adapts to diverse domains.

The third section expands into cultural and civic life. Alistair Knott, Tapabrata Chakraborti, and Dino Pedreschi critique opaque recommender systems in social media, where transparency deficits undermine trust and human-centered design. Sandra Rodriguez (pp. 189–202) highlights the project *CHOM5KY vs. CHOMSKY*, which uses art to demystify AI, promote public literacy, and encourage critical reflection. These chapters emphasize that AI governance cannot be confined to technical or legal discourses alone—cultural practices, creativity, and civic engagement are integral to building human-centered systems.

The final section provides practical frameworks for operationalizing HCAI. Karine Gentelet and Sarit Mizrahi advocate for citizen participation as an antidote to algorithmic bias, reframing governance as a civic process rather than a technical fix. Vicky Charisi and Virginia Dignum propose regulatory sandboxes for children's rights, echoing UNICEF's calls for child-centered AI. Marina Teller critiques top-down European regulation, warning against "algocracy" and advocating participatory approaches. Mario Passalacqua and colleagues situate HCAI in Industry 5.0, emphasizing worker well-being and system trust. Jakob Kappenberger and Heiner Stuckenschmidt provide a pragmatic public policy framework organized around prerequisites, processes, and outputs. Finally, Clementine Collett, Gina Neff, and Maria Axente identify three major "implementation gaps"- engagement, translation, and dialogue - that undermine organizational adoption of inclusive AI. Their emphasis on diverse teams and cross-stakeholder collaboration underscores the participatory ethos of the entire volume.

The volume's greatest strength lies in its methodological pluralism and participatory ethos. By weaving together scholarly essays, empirical case studies, and practitioner reflections, the editors successfully bridge the often-cited gap between normative AI ethics and the practical realities of implementation. This combination provides readers with a holistic understanding of human-centered AI (HCAI) as both an analytical framework and an actionable governance paradigm. The insistence on co-design, stakeholder engagement, and inclusivity reflects broader trends within contemporary AI governance scholarship, where the procedural legitimacy of AI systems—how decisions are made, whose voices are included, and how accountability is distributed—matters as much as the technical performance of the systems themselves.

The volume is also well aligned with existing theoretical foundations in the field. Floridi and Cowls's (2021) unified framework—beneficence, non-maleficence, autonomy, justice, and explicability—serves as a conceptual backdrop to many contributions. The chapters extend these principles into domain-specific

analyses, showing how abstract ethics can translate into governance structures. Similarly, Crawford's Atlas of AI (2021) provides a powerful reminder of the material and ecological consequences of AI, underscoring the importance of contextualizing technological governance within broader political economies. Shneiderman's (2022) call to advance "human-centered AI" as a design science is echoed throughout the volume, which repeatedly emphasizes value-sensitive and participatory methodologies as essential components of responsible AI.

Yet, despite these strengths, notable limitations remain. While the editors position the book as a globally relevant intervention, its empirical grounding is heavily weighted toward Europe, North America, and Japan. This geographical concentration risks reinscribing a familiar imbalance in AI discourse, where high-income regions serve as normative reference points while the Global South is engaged only marginally or symbolically. The inclusion of Ubuntu as an alternative ethical lens is a welcome gesture, yet its treatment is comparatively brief and not thoroughly integrated into the book's broader conceptual scaffolding. More sustained engagement with indigenous epistemologies, decolonial scholarship, and postcolonial critiques would have strengthened the theoretical diversity and global legitimacy of the volume's arguments.

Furthermore, although several chapters examine the risks and limitations of large language models (LLMs), the analysis occasionally struggles to keep pace with the rapid acceleration of generative AI technologies in the post-2023 period. For instance, concerns about hallucination, bias, and opacity are well articulated but do not fully grapple with emerging issues such as synthetic media proliferation, data provenance challenges, computational inequality, or frontier-model governance debates. The book therefore captures a snapshot of the field at a critical juncture but inevitably falls short of addressing developments that have since reshaped the global conversation on AI regulation.

Another limitation arises from the structure of the volume itself. While domain-specific case studies are informative, they sometimes take a descriptive rather than deeply analytical form. This can produce uneven theoretical depth across chapters. The practical orientation is undoubtedly valuable—particularly for policy-makers and practitioners—but it occasionally comes at the expense of integrating insights into a more unified theoretical synthesis. As a result, the book excels in demonstrating *how* HCAI might operate in particular sectors but is somewhat less clear about the conditions required for systemic transformation across institutional and geopolitical contexts.

Nonetheless, despite these shortcomings, *Human-Centered AI* remains a timely, rigorous, and necessary contribution to debates about the future of AI. It reframes human-centeredness not as a rhetorical or branding exercise but as an essential ethical condition for the legitimacy, trustworthiness, and long-term sustainability of AI systems. For policymakers, the volume offers tools to resist the concentration of technological power and to cultivate participatory governance structures that foreground democratic values. For educators, it underscores the importance of AI literacy, critical pedagogy, and civic engagement in preparing future generations for AI-mediated societies. For researchers, it maps fertile territory for interdisciplinary inquiry, highlighting the need for collaborations between technologists, social scientists, philosophers, and affected communities.

Most importantly, the book insists that AI must be understood as a means rather than an end—one that must ultimately enrich human life, strengthen democratic institutions, and foster solidarity. In an era where AI often advances faster than societies can respond, this volume offers a compelling reminder that human-centeredness is not merely an aspiration but a prerequisite for designing equitable and just AI futures.

Acknowledgements

The reviewer thanks the journal's editorial team for the opportunity to evaluate *Human-Centered AI: A Multidisciplinary Perspective for Policy-Makers, Auditors, and Users*, edited by Catherine Régis, Jean-Louis Denis, Maria Luciana Axente, and Atsuo Kishimoto, and published by CRC Press (2024). The reviewer also appreciates colleagues whose brief comments helped refine this piece. All views and interpretations expressed are the reviewer's own.

Author Contributions

Dr. Ratan Sarkar conceptualized, analyzed, and authored the review. Dr. Sarkar led the critical evaluation and synthesis, contributed to the interpretive analysis and language refinement, version and share equal responsibility for the content.

Declarations:

- Ethical compliance: Follows academic and ethical standards.
- Consent to participate/publication: Not applicable.
- Conflict of interest: None declared.
- Funding: No external funding.
- Data availability: Not applicable.
- Use of AI Tools: *Grammarly* and *Paperpal* (Premium) were used only for language refinement and readability improvement.

REFERENCES

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Floridi, L., & Cowls, J. (2021). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1), 11-14. https://doi.org/10.1162/99608f92.8cd550d1

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press. https://doi.org/10.1093/oso/9780192845290.001.0001