Journal of Ethics and Emerging Technologies

IEET

Book Review

Review of Human-AI Interaction and Collaboration

Dan Wu and Shaobo Liang, *Human-AI Interaction and Collaboration* (Cambridge University Press)

Michael Umeokoli[1]

[1] Independent AI Engineer/Researcher
MSc, Manchester Metropolitan University, United Kingdom (2025)
* Correspondence: michaelumeokoli@gmail.com

Abstract Dan Wu, Shaobo Liang, and their co-authors make a strong case in Human-AI Interaction and Collaboration that the days of considering artificial intelligence (AI) as a simple computing tool are over. The authors instead propose a paradigm change in favor of Human-AI Collaboration (HAIC), a collaborative model in which AI and humans function as a Joint Cognitive System (JCS) in symbiotic harmony. The authors propose that we may address complicated societal and scientific issues without compromising human autonomy by combining AI's computational power with human biological and cognitive intelligence. This book effectively provides a thorough, human-centered paradigm for the future of AI integration, making it an essential resource for engineers, academics, developers, and ethicists.

Keywords: Artificial Intelligence; AI Systems; Human-AI Interaction; Large Language Models; AI Ethics, Data Privacy

Citation: Umeokoli, Michael, 2026. Review of Human-AI Interaction and Collaboration. *Journal of Ethics and Emerging Technologies* 36: 1. https://doi.org/10.55613/jeet.v36i1.218

Received: 04/03/3036
Accepted: 06/03/2026
Published: 31/03/2026

Publisher's Note: IEET stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

The main argument of the book is that Human-Centered AI (HCAI) design, which puts user trust, privacy, and explainability ahead of unbridled algorithmic autonomy, provides the foundation for successful Human-AI interaction. By breaking the work up into theoretical underpinnings, practical procedures, and high-risk applications, the authors methodically develop this case. Utilizing a variety of sources, including user-interaction studies and real-world datasets (such as Reddit and Twitter/X), the text offers strong, empirical support for its assertions by combining information science, computer science, and psychology.

The book's unwavering analysis of the moral dilemmas raised by AI advancement is one of its strongest points. The writers grapple with the "anthropomorphism paradox" throughout the first few chapters. While giving AI human characteristics like warmth or cognitive empathy might boost perceived credibility and user acceptability, it also runs the danger of encouraging "dishonest personification," which can result in emotional dependence or the appearance of external validation. The authors cleverly contend that HAIC aims for "calibrated trust," in which human confidence precisely corresponds to the system's actual reliability, rather than blind trust.

The book places a strong emphasis on Explainable AI (XAI) and privacy measures to establish this trust. By emphasizing that openness in protective measures is the absolute baseline for user trust, Chapter 3 distinctively moves the privacy discussion away from generic anonymization and toward a dynamic, context-aware Privacy Type Model. The authors also successfully show that an AI's credibility is multifaceted, encompassing fairness, intelligibility, and ethical alignment in addition to accuracy.

The book's middle portion skillfully demonstrates how AI functions as a "cognitive extension" in collaborative settings by moving from theory to practice. In Chapter 5, a particularly noteworthy case study looks at how the Zhongxun platform combats child trafficking by using both human crowdsourcing and face aging algorithms. Here, the authors assess how AI effectively filters data, mitigating the "Tragedy of the Commons" and enabling human participants to concentrate their localized, selfless efforts where they have the greatest impact. Similar to this, Chapter 7 examines how emotional cues such as fear versus surprise drive the dissemination of false material on social media using the Stimuli-Organism-Response (SOR) framework. The authors come to an interesting conclusion: AI should be used as a "emotion-aware" system to control human behavioral reactions and lessen harmful social bonding around incorrect information, rather than just as a fact-checker.

The final chapters thoroughly evaluate the HAIC architecture in high-stakes fields as healthcare, data annotation, and scientific discovery. Chapter 9, which describes a "Expert-Machine Co-development" strategy to determine the health information needs of caregivers for Alzheimer's Disease and Related Dementias (ADRD), has a particularly significant influence. The authors offer a scalable remedy for the severe lack of human medical specialists by proving that AI-generated simulated data can train classification models more successfully than noisy, real-world data. The writers, however, maintain their ethical foundation by emphasizing time and again that AI is meant to supplement human intelligence rather than replace it, particularly in safety-critical fields where human judgment is essential.

The chapter on scientific discovery is equally compelling. It demonstrates the scope of AI's developing role in research without reducing everything to LLMs by covering mathematics, physics, chemistry, and biological sciences. The examples, which range from protein structure prediction to theorem proof to LHC (The Large Hadron Collider) data filtering, illustrate the idea that AI can foster creativity and discovery rather than merely automate repetitive analysis.

The human role is also well preserved in this chapter. It highlights the fundamental scientific duties of data curation, model selection, and training design. Current narratives that accept AI outputs as self-validating are welcomely challenged by that. This chapter delivers one of the clearest assertions in the volume concerning synergy over substitution, but I believe it might have gone farther on reproducibility and research inequality, particularly the hazards of concentrating discovery power in a small number of well-funded labs.

Although the book mostly succeeds in living up to its expectations, a critical assessment identifies a few places where the analysis may be expanded. To tackle the "black box" problem and provide ethical oversight, the authors regularly emphasize the importance of "Human-in-the-loop" (HITL) systems and Interactive Machine Learning (IML). The practical conflict between the enormous engineering resources needed for these transparent systems and the tech industry's demand for quick scaling, however, is barely touched upon in passing. Furthermore, even though the book recognizes the "global data divide" and the presence of national representative biases in training data, the suggested remedies like ongoing user feedback seem a little inadequate to overcome the structural, financial obstacles to fair access to AI.

To sum up, Human-AI Interaction and Collaboration is a well-researched and well-argued work that transcends the straightforward "man vs. machine" debate. The book's strongest arguments are supported by chapters that demonstrate real-world examples of meaningful human-AI collaboration, including trade-offs, in fields like healthcare, search,

crowdsourcing, and science. Wu, Liang, and their collaborators have produced a crucial roadmap for AI's future by skillfully striking a balance between technical approaches and profound psychological and ethical insights. Researchers, tech developers, and legislators who are dedicated to creating intelligent systems that enhance human agency rather than diminish it are strongly encouraged to read it.

Use of AI Tools: Use of Grammarly for spell checks and Pangram/Gemini for paraphrasing and formatting.

Funding: No external funding was received for this book review.

Conflicts of Interest: The author declares no conflict of interest.

References

Wu, D., & Liang, S. (Eds.). (2025). Human-AI Interaction and Collaboration. Cambridge: Cambridge University Press, https://doi.org/10.1017/9781009587877.

Fragiadakis, G., Diou, C., Kousiouris, G., & Nikolaidou, M. (2025). EVALUATING HUMAN-AI COLLABORATION: A REVIEW AND METHODOLOGICAL FRAMEWORK. https://arxiv.org/pdf/2407.19098