Qeios

Peer Review

Review of: "Rethink Your Mental Model in the Age of Generative AI: A Triadic Framework for Human-AI Collaboration"

Nineta Polemi[1]

1. Informatics, University of Piraeus, Greece

Review of the Paper

"Rethink Your Mental Model in the Age of Generative AI: A Triadic Framework for Human-AI Collaboration"

1. Overview and Contribution

This paper is well-written; I found it very useful with a plethora of related references to support its analysis. It offers a comprehensive, theory-driven synthesis of research on human–AI collaboration in the era of generative AI. Its central contribution is the Triadic Framework and it proposes seven actionable routines designed to improve calibration, preserve human agency, and manage the nature of AI capabilities. The paper addresses a timely and important problem: the mismatch between inherited mental models (from deterministic technologies and human teamwork) and the probabilistic, opaque, and rapidly evolving nature of generative AI. By shifting the focus from better prompting to better mental models, the authors make a conceptual contribution to research in human–AI interaction.

2. Strengths The Triadic Framework is clearly articulated and well-motivated. The paper draws on an impressive range of recent empirical and theoretical work (largely 2023–2025), integrating findings from HCI, IS, psychology, organizational studies, and AI research. The seven propositions (P1–P7) translate abstract theory into practical routines, such as verification rituals, mode declarations, provenance tracking, and reset practices. This design-oriented turn makes the paper valuable not only for researchers but also for practitioners, educators, and organizations. By framing generative AI as a "pre-cognitive System 0," the authors convincingly argue that AI shapes human cognition before conscious evaluation begins. This is a powerful lens for understanding automation bias, over-trust, and persuasion risk.

doi.org/10.32388/I4RTBX

3. Weaknesses The work is primarily conceptual and synthetic. While this is clearly acknowledged, the seven propositions remain untested design hypotheses. The paper would benefit from at least illustrative case studies, pilot deployments, or experimental sketches showing how the routines work in practice. At over 40 pages, the paper is extremely dense. While the depth is admirable, some sections (especially the extensive system-level review) could be streamlined without losing conceptual force. While the paper explicitly states it is not an ethics treatise, issues such as labor displacement, power asymmetries, and institutional accountability are only briefly touched upon. These factors strongly shape mental models in real organizations and could be more tightly integrated.

4. Overall evaluation - Recommendations to the author

This paper can be **accepted with comments** since it reframes a central problem of the generative AI era: not how to make AI smarter, but how to make humans' mental models more adequate. It stands as a strong theoretical foundation for future work on hybrid intelligence and human-AI collaboration.

Comments to the author:

1. Reduce size
2. Add examples or empirical testing results
3. Consider developing a shorter, practitioner-oriented version focusing on the framework and the seven propositions. Provide examples.
4. Explore how the framework interacts with ethical, legal, and governance structures in organizations.

Declarations

Potential competing interests: No potential competing interests to declare.