Human–AI complementarity needs augmentation, not emulation

Zhicheng Lin
Department of Psychology, Yonsei University

Correspondence
Zhicheng Lin, Department of Psychology, Yonsei University, Seoul, 03722, Republic of Korea (zhichenglin@gmail.com; X/Twitter: @ZLinPsy)

Acknowledgments
I used Claude Sonnet 4.5 and Gemini 2.5 Pro to proofread the manuscript, following the prompts described at https://www.nature.com/articles/s41551-024-01185-8.

Competing interests
The author declares no competing interests.

Abstract
Gonzalez and Heidari propose cognitive AI—systems that emulate human cognitive processes—as essential for human–AI complementarity in dynamic decision-making. I argue this framework rests on two questionable premises. First, the distinction between cognitive AI and data-driven approaches lacks practical significance: modern AI trained on behavioral data already exhibits emergent human-like properties through implicit modeling of statistical regularities in human decision-making. Second, the framework assumes complementarity requires AI to mirror human cognition, including human limitations and constraints. Yet if noise and systematic biases fundamentally characterize human cognition, complementary AI should compensate for these limitations rather than reproduce them. I propose that effective human–AI complementarity requires design principles emphasizing appropriate role allocation, transparent uncertainty communication, adaptive personalization that improves decision quality, and mutual modeling of functionally relevant features without necessarily replicating cognitive mechanisms. These principles can be instantiated through various technical approaches and should be evaluated by team outcomes rather than adherence to cognitive theories. Complementarity requires AI that augments human capabilities, not cognitive architectures that reproduce human limitations.

In their Perspective, Gonzalez and Heidari propose cognitive AI—systems that "emulate and simulate the human mind as an information-processing system"—as essential for achieving human–AI complementarity in dynamic decision-making[1]. However, this framework rests on questionable premises: that cognitive AI represents a meaningful practical distinction from data-driven approaches, and that effective human–AI teams require AI systems to replicate human cognitive processes.

The authors distinguish cognitive AI from data-driven AI by its grounding in formal cognitive models that "explain and replicate how and why decisions are made." Yet modern AI trained on behavioral data already exhibits emergent human-like properties, including biases, through implicit modeling of statistical regularities in human decision-making[2-4]. The authors acknowledge cognitive AI must ultimately integrate with data-driven approaches, raising the question: what essential work do cognitive architectures accomplish beyond what adaptive learning already achieves?[5,6] The framework may label valuable design principles—transparency, adaptability, personalization—but these can be realized through various technical approaches.

More fundamentally, the framework assumes human–AI complementarity requires AI to mirror human cognition—that "cognitive AI must mimic human limitations and constraints, because doing so will allow us to interpret human behaviour and to predict (and prevent) human error." Yet as Kahneman observes, "humans do so poorly and are so noisy that, just by removing the noise, you can do better than people."[7] If noise and other limitations fundamentally characterize human cognition, why should complementary AI systems faithfully reproduce them? The value proposition appears reversed: we need AI that compensates for cognitive limitations, not systems replicating them.

Consider the authors' disaster management scenario. Should AI model human biases in probability assessment to predict evacuee behavior, or provide debiased estimates to improve emergency managers' decisions? An AI that "mimics human limitations" for interpretability may sacrifice the very complementarity—correction of human weaknesses—motivating collaboration[8,9].

The proposal conflates two distinct AI roles with different requirements[10]. For decision support, predictive accuracy about human behavior matters, achievable through various modeling approaches. For autonomous teammates, functional complementarity matters—can AI accomplish what humans cannot, or accomplish it better? Neither necessarily requires algorithmic-level mirroring of human cognition.

Rather than privileging cognitive architectures, design principles for human–AI complementarity should emphasize 1) appropriate role allocation (e.g., AI handles pattern recognition where humans are noisy; humans oversee ethical judgments and novel situations); 2) transparent uncertainty communication enabling calibrated reliance; 3) adaptive personalization improving decision quality rather than achieving human-like errors; and 4) mutual modeling capturing functionally relevant features—capabilities, reliability, typical strategies—without necessarily replicating cognitive mechanisms.

These principles can be instantiated through cognitive architectures, hybrid systems, or learned models. Frameworks should be evaluated by outcomes in human–AI teams, not adherence to cognitive theories. The authors acknowledge "substantially more research is needed to demonstrate [cognitive AI's] broader utility." I suggest this research should comparatively evaluate whether cognitive approaches offer practical advantages over alternatives in achieving complementarity. Complementarity requires AI that augments human capabilities, not cognitive architectures that reproduce human limitations.

References

1 Gonzalez, C. & Heidari, H. A cognitive approach to human–AI complementarity in dynamic decision-making. *Nature Reviews Psychology* (2025). https://doi.org/10.1038/s44159-025-00499-x

2 Binz, M. *et al.* A foundation model to predict and capture human cognition. *Nature* **644**, 1002-1009 (2025). https://doi.org/10.1038/s41586-025-09215-4

3 Chen, Y., Kirshner, S. N., Ovchinnikov, A., Andiappan, M. & Jenkin, T. A manager and an AI walk into a bar: does ChatGPT make biased decisions like we do? *Manufacturing & Service Operations Management* **27**, 354-368 (2025). https://doi.org/10.1287/msom.2023.0279

4 Hagendorff, T., Fabi, S. & Kosinski, M. Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. *Nat. Comput. Sci.* **3**, 833-838 (2023). https://doi.org/10.1038/s43588-023-00527-x

5 Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. Building machines that learn and think like people. *Behav. Brain Sci.* **40**, e253 (2017). https://doi.org/10.1017/S0140525X16001837

6 Laird, J. E., Lebiere, C. & Rosenbloom, P. S. A standard model of the mind: toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine* **38**, 13-26 (2017). https://doi.org/10.1609/aimag.v38i4.2744

7 Kahneman, D. in *The Economics of Artificial Intelligence: An agenda* (eds Ajay Agrawal, Joshua Gans, & Avi Goldfarb) 608-610 (University of Chicago Press, 2019).

8 Bansal, G. *et al.* in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* Article 81 (Association for Computing Machinery, Yokohama, Japan, 2021).

9 Vaccaro, M., Almaatouq, A. & Malone, T. When combinations of humans and AI are useful: a systematic review and meta-analysis. *Nat. Hum. Behav.* **8**, 2293-2303 (2024). https://doi.org/10.1038/s41562-024-02024-1

10 O'Neill, T., McNeese, N., Barron, A. & Schelble, B. Human–autonomy teaming: a review and analysis of the empirical literature. *Hum. Factors* **64**, 904-938 (2020). https://doi.org/10.1177/0018720820960865