

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Toward Conscious Agency in Artificial Intelligence: Evaluating and Bridging the Gap Between a Philosophical Principle and Technological Capabilities

Permalink

<https://escholarship.org/uc/item/78h3h30v>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Hotz, Lothar

Petzold, Thomas

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

[Toward Conscious Agency in Artificial Intelligence: Evaluating and Bridging the Gap Between a Philosophical Principle and Technological Capabilities

Lothar Hotz (lothar.hotz@hitec-hamburg.de)

Hamburger Informatik Technologie-Center e.V.
c/o University of Hamburg, Computer Science Departement
Bundesstraße 56 b, 20146 Hamburg, Germany

Thomas Petzold (t.petzold@media-university.de)

Trust Machines Lab
Media University of Applied Sciences
Ackerstraße 76, 13355 Berlin, Germany

]

Abstract

Large language models (LLMs) display linguistic fluency but lack human-like understanding. Philosophical theories of consciousness, rooted in subjective experience, and AI research, which evaluates observable performance, often diverge methodologically. We reconcile these views by treating conscious agency as an emergent property of adaptive dyadic communication (Solomon, Brinberg, Bodie, Jones, & Ram, 2021), using Tyler's Ten Testable Properties of Consciousness as functional constraints (Tyler, 2020) rather than evidence of inner experience. An analysis of current LLM-based dialogue systems shows they can mimic isolated properties yet cannot satisfy all constraints simultaneously, due to weak persistent interaction states, limited interlocutor modeling, and absent temporally structured episodic experience. To move beyond technology-deterministic approaches, we propose a framework in which agency-like properties arise from three adaptive network layers: (i) a User-Model Network encoding goals and intentions, (ii) an Agent-State Network encoding semantic-pragmatic-experiential state, and (iii) a Communicative-Reasoning Layer monitoring coherence and alignment. We instantiate this design as the Large Communication Model (LCM).

Keywords: Artificial consciousness, cognitive architecture, generative AI, machine awareness, Tyler's properties, AI ethics, cognitive modeling

References

- Solomon, D. H., Brinberg, M., Bodie, G. D., Jones, S. M., & Ram, N. (2021). A Dynamic Dyadic Systems Approach to Interpersonal Communication. *Journal of Communication*, 71, 1001–1026.
- Tyler, C. W. (2020). Ten Testable Properties of Consciousness. *Frontiers in Psychology*, 11, 1144. doi: 10.3389/fpsyg.2020.01144