

Sheer Karny

sheer.karny@gmail.com | sheerkarny.com | [LinkedIn](#) | [Google Scholar](#)

EDUCATION

Massachusetts Institute of Technology (MIT)

Graduate Student, Media Lab

Cambridge, MA

Aug. 2025 – Present

- Advisor: Pat Pataranutaporn, Secondary Advisor: Pattie Maes.
- Research Group: Cyborg Psychology

University of California, Berkeley

Bachelor of Arts in Cognitive Science (Honors), Minor in Data Science

Berkeley, CA

2020 – 2022

Irvine Valley College

Honors Program

Irvine, CA

2017 – 2019

RESEARCH EXPERIENCE

Graduate Researcher

MIT Media Lab

Aug. 2025 – Present

Cambridge, MA

- Research Focus: Human-AI Collaboration, Human Resilience, AI Behavior, Mechanistic Interpretability, Explainability, and Design.
- Contributor to the Advancing Humans with AI (AHA) Program.

Junior Specialist

University of California, Irvine

Sep. 2023 – Aug. 2025

Irvine, CA

- Advisor: Mark Steyvers PhD.
- Research Focus: Human-AI teaming, Large Language Models (LLMs), Social and Intuitive AI, Theory of Mind.
- Worked in collaboration with Honda Research Institute USA.

Undergraduate Researcher

University of California, Berkeley

May 2021 – Aug. 2022

Berkeley, CA

- Advisors: Richard Ivry PhD., Jonathan Tsay PhD.
- Conducted research on implicit and explicit motor learning processes and differential aging effects.

INDUSTRY EXPERIENCE

Software Engineering Internship

Apple

June – Aug. 2026

Cupertino, CA

- Working on the future of work and agents in Apple's Product Incubation team.

PUBLICATIONS

(* denotes equal contribution)

Pre-prints, Under Review

- **(under review)** Karny, S.*, Baez, A.*, Pataranutaporn, P. (2026). *Multi-Turn Neural Transparency: Surfacing Neural Activations Improves User Calibration to LLM Behavioral Drift*. arXiv:2605.15455 (2026).
- **(pre-print)** Pataranutaporn, P., Karny, S., Archiwaranguprok, C., Albrecht, C., Liu, A.R. (2025). *"My Boyfriend is AI": A Computational Analysis of Human-AI Companionship in Reddit's AI Community*. arXiv:2509.11391 (2025).
- **(pre-print)** Archiwaranguprok, C., Chen, A., Karny, S., Ishii, H., Maes, P., Pataranutaporn, P. (2025). *Atlas of Human-AI Interaction (v1): An Interactive Meta-Science Platform for Large-Scale Research Literature Sensemaking*. arXiv:2509.25499 (2025).

Journal Articles

- Cisneros, E., Karny, S., Ivry, R. B., Tsay, J. S. (2026). *A systematic investigation reveals dissociable effects of ageing on implicit and explicit components of sensorimotor learning*. Nature Human Behaviour.
- Mayer, L. W.*, Karny, S.*, Ayoub, J., Song, M., Tian, D., Moradi-Pari, E., Steyvers, M. (2026). *Human-AI collaboration: Trade-offs between performance and preferences*. Cognitive Research: Principles and Implications, 11(1), 18.

- Steyvers, M., Tejeda, H., Kumar, A., Belem, C., **Karny, S.**, Hu, X., Mayer, L. W., Smyth, P. (2025). *What large language models know and what people think they know*. Nature Machine Intelligence, 7(2), 221–231.

Conference Proceedings

- Pataranutaporn, P., Albrecht, C., Archiwaranguprok, C., **Karny, S.**, Baez, A. (2026). *What If AI Lived Inside Your Mind? Simulating “Neural Integration” of Human and AI through Mechanistic Interpretability as Provocation*. In Proceedings of the Augmented Humans International Conference 2026 (AHs '26), 223–238.
- **Karny, S.***, Baez, A.*, Pataranutaporn, P. (2025). *Neural Transparency: Mechanistic Interpretability Interfaces for Anticipating Model Behaviors for Personalized AI*. In Proceedings of the 31st International Conference on Intelligent User Interfaces (IUI '26). Association for Computing Machinery, New York, NY, USA, 868–884.
<https://doi.org/10.1145/3742413.3789120>.
- **Karny, S.**, Mayer, L. W., Ayoub, J., Song, M., Su, H., Tian, D., Moradi-Pari, E., Steyvers, M. (2024). *Learning with AI Assistance: A Path to Better Task Performance or Dependence?* In Proceedings of the ACM Collective Intelligence Conference (CI '24), Boston, MA, 10–17.

Workshop Papers and Posters

- Baez, A., **Karny, S.**, Pataranutaporn, P. (2026). *Dissociating the Internal Representations of Sycophancy in LLMs*. ICML '26 Mechanistic Interpretability Workshop, Seoul, Korea. arXiv:2607.07003.
- Fang, C.M., **Karny, S.**, Archiwaranguprok, C., Samaradivakara, Y., Pataranutaporn, P., Maes, P. (2026). *AI-Wrapped: Participatory, Privacy-Preserving Measurement of Longitudinal LLM Use In-the-Wild*. CHI '26 Bi-Align Workshop. arXiv:2602.18415.
- Cisneros, E., Tsay, J., **Karny, S.**, Ivry, R. *Differential effect of age on implicit and explicit motor learning processes*. Neural Control of Movement Conference (NCM '23), Victoria, BC.

Presentations

- “Neural Transparency.” Lightning Talk at Google Deep Mind. Cambridge, MA. (5 min).
- “Towards Nutritional Labels for Human-AI Interaction.” Talk at Advancing Humans and AI (AHA) Seminar Series. March '26. Cambridge, MA. (25 min).
- “Learning with AI Assistance.” Talk at the ACM Collective Intelligence Conference (CI '24); Human-AI Teaming plenary session. (15 min).

SELECTED MEDIA COVERAGE

The New York Times <i>I Fell in Love With an A.I. Chatbot</i>	Nov. 2025
The Guardian <i>'I'm suddenly so angry': my strange, unnerving week with an AI friend</i>	Oct. 2025
Observer <i>MIT Study Finds Chatbot Love Is Real—and It's Often Unintentional</i>	Sep. 2025
MIT Technology Review <i>People are finding relationship AI without even seeking it</i>	Sep. 2025

TEACHING & LEADERSHIP

Workshop Facilitator	March 2023
<i>Tibetan Library of Works and Archives</i>	<i>Dharamshala, India</i>
• Taught the foundations of neuroscience to 25 Tibetan Buddhist nuns over 10 days with live translation.	
Co-Lead Facilitator	Nov. 2020 – June 2022
<i>Meditation and Mindfulness Course (UC Berkeley)</i>	<i>Berkeley, CA</i>
• Led a course introducing 70 students per semester to meditation practice over 4 semesters.	
• Conducted research showing that the course produces persisting benefits to students' well-being.	

TECHNICAL SKILLS

Languages: Python, R, JavaScript/TypeScript, C/C++
Applications: Jupyter Hub, Firebase, Excel, Blender
Technical: Machine Learning, Statistical Analysis, Cognitive Modeling, Experimental Design, LLM Prompting, Web Development, Database Design, Scientific Writing
Reviewing: CogSci 2024, 2025; CHI 2025