

• SAURABH RANJAN •

Gainesville, FL, U.S.A. • ranjan.saurabh@outlook.com • saurabhr.github.io

Ph.D. Cognitive Neuroscientist | AI / ML Researcher | Cognitive and Machine Psychology | NeuroAI

EDUCATION

Ph.D. Psychology — Behavioral & Cognitive Neuroscience University of Florida, USA | 2025

Dissertation: [Reality Monitoring in Humans and Artificial Intelligence](#)

M.Sc. Cognitive Science University of Allahabad, India | 2018

Thesis: [Intentional Binding in Future-Directed Intentions](#)

B.Sc. Physics Birla Institute of Technology, Mesra, India | 2016

CORE TECHNICAL STACK

AI / ML & LLMs: evaluation pipelines for LLM safety and hallucination detection, agentic workflows, retrieval-augmented generation (RAG), prompt engineering; PyTorch, HuggingFace, LangChain, LangGraph, Pydantic, scikit-learn

Interpretability & Representational Analysis: linearizing encoding models and temporal generalization analysis (linear probing of internal representations across layers and time), psychological network analysis for comparing representational structure across human and model populations, large-scale multivariate signal decoding (MEG/EEG with MNE, fMRI with Nilearn/SPM), predictive modeling on biobank-scale data (UK Biobank RS-fMRI → phenotype prediction)

Statistical Modeling & Evaluation Design: Bayesian hierarchical models (PyMC3, brms), signal detection theory, generalized mixed-effects models (lme4), multinomial processing trees, network analysis (bootnet, qgraph), causal & metacognitive modeling

Data Engineering & MLOps: Python (expert), R (advanced), MATLAB, Bash; pandas, NumPy, Git/GitHub, SLURM/HPC, containerized environments, large-scale behavioral data collection (Prolific, Qualtrics, PsychoPy/JS)

Visualization & Communication: seaborn, matplotlib, plotly, ggplot2, tidyverse — publication-ready figures, dashboards, and stakeholder reporting

RESEARCH EXPERIENCE

Researcher — AI for Biomedical Health Outcomes 2026

University of Florida, Gainesville, FL

- Building AI agents with planning, reflection, tool use, and multi-agent coordination, using clinical and medical decision-making as the test domain for agentic-system design and evaluation.
- Engineering reproducible research pipelines (Python, PyTorch, LangChain/LangGraph, SLURM/HPC) for running and evaluating agentic workflows at scale.
- Co-developing UF's Master's-level curriculum on Data Science and Agentic LLMs for [UF AI for Biomedical Health Sciences \(AIBHS\)](#).

Graduate Researcher 2020 – 2025

Odegaard Lab & UF Department of Psychology, Gainesville, FL

- **Hallucination & Self-Knowledge in LLMs:** Ran a two-experiment, six-model study on whether LLMs can tell their own generated content from user-provided content (“reality monitoring,” the capability that fails during hallucination in both humans and LLMs).

Source-attribution accuracy hit ceiling when conversational-memory demands were minimal, but reversed into a fragile external-item advantage once an episodic delay removed that shortcut, evidence that near-perfect single-turn benchmarks can mask an unreliable underlying mechanism.

- **Mechanistic & Metacognitive Failure Modes:** Found that corrective feedback split models into two distinct failure patterns: some fully swapped their internal and external judgments, others gained raw accuracy while their confidence decoupled from correctness, a miscalibration that is arguably more dangerous than being wrong outright. The cross-model pattern tracked active, not aggregate, parameter count, suggesting engaged compute rather than scale drives introspective reliability.
- Designed and ran the underlying behavioral experiments with 100+ in-person human participants matched against LLM simulations across conversation-history and feedback conditions, producing a standardized human-LLM benchmarking pipeline built to be reused across models.
- **Neuroimaging as an Interpretability Method:** Engineered ML pipelines for [MEG temporal generalization](#) and [fMRI linearizing encoding models](#), predicting phenotype from RS-fMRI on UK Biobank-scale data using the same linear-probe logic now used to interpret internal representations in LLMs.
- **Imagination Networks Expose a Representational Gap:** Built psychological network models of imagination vividness from 2,743 human participants across Florida, Poland, and London, then compared them against six LLMs (12B-272B parameters). Human networks stayed highly consistent across populations ($r = 0.31-0.93$), while most LLM configurations collapsed into a single undifferentiated cluster regardless of scale, evidence that parameter count alone doesn't give a model the layered representational structure humans build from embodied experience.
- **LLM Evaluation Infrastructure:** Authored [Psych Scanner](#), a Python framework (LangChain + LangGraph + Pydantic) for running cognitive experiments on LLMs at scale with persona management, conversation-history manipulation, and population-level synthesis, and [metasignal](#), a signal-detection-theory and metacognition library for scoring the resulting data; together, functionally similar to the internal evaluation pipelines research labs build to benchmark model behavior.
- Mentored 7 undergraduate researchers, co-authored conference posters (1 award-winning), and taught a full segment of Physiological Psychology.

Research Assistant 2019 – 2020

Center of Behavioral and Cognitive Sciences, University of Allahabad, India

- Studied the sense of agency for future-directed intentions using the intentional-binding paradigm: participants formed an intention for an outcome reached via an intermediate action, and predictive intermediate outcomes produced shorter time estimates and faster responses, both markers of stronger binding, consistent across three delays (300/500/700ms) and two contingency levels. Also contributed to human-agency and global-local perception studies using arrays of patches. Published as an [OSF preprint](#); the same construct, monitoring and correctly attributing one's own actions, now underlies how agentic AI systems track and self-correct across multi-step plans.

Research Assistant 2018

Homi Bhabha Centre for Science Education, TIFR, Mumbai, India

- Studied how gesture-based interaction with digital tools shapes conceptual understanding in STEM, using vectors as a test case: students usually meet them as static textbook arrows, never something to touch or rotate. The working idea behind the two studies below was that letting students manipulate a vector by hand changes what they understand about it, with implications for how embodied interaction shapes mathematical cognition.

- Contributed statistical analysis for Touchy-Feely Vectors (TFV), a gesture-based geometric environment built for direct interaction rather than the video/animation-style “physical access” typical of edtech. TFV targeted students’ well-documented difficulty with the geometric aspects of vectors in grade-11 physics and math, letting them manipulate vector quantities directly on the premise that bodily interaction shapes reasoning. Across six classrooms (3 experimental, 3 control; 266 students total), the experimental group reasoned differently about vectors and showed higher engagement, what the paper terms “epistemic access.” Published as [Touchy Feely Affordances of Digital Technology for Embodied Interactions Can Enhance ‘Epistemic Access’](#) (2019 IEEE Tenth International Conference on Technology for Education (T4E)).
- Contributed statistical analysis for this 3-year design study testing two TFV iterations. Pre-post testing of the first, a computer-based prototype, showed it helped students form a more integrated vector concept and reason better. The second, a textbook-linked touch-screen version, was field-tested across six classrooms (3 control, N=135; 3 experimental, N=131), boosting engagement in average students and model-based reasoning (geometry-algebra integration) in stronger ones. Published as [Touchy feely vectors: A compensatory design approach to support model-based reasoning](#) (Journal of Computer Assisted Learning, 2021).

PAPERS

Working Papers

- [2] **Ranjan, S.**, & Odegaard, B. Generalization of generation effect in reality monitoring.

Software

- [2] **Ranjan, S.**, Makwana, M., Sokratous, K., & Odegaard, B. metasignal: A Python package for comprehensive metacognitive analysis and decision-making. Can be viewed on Preprint on [Google Drive](#).

- [1] **Ranjan, S.**, Sokratous, K., & Makwana, M. [Psych Scanner: A Framework for Systematic Cognitive Evaluation of Large Language Models](#).

Theory and Empirical Peer-Reviewed & Preprints

- [10] [Preprint] **Ranjan, S.**, Sokratous, K., & Odegaard, B. (2026). Reality Monitoring in Large Language Models: Self-Knowledge That Transforms with Conversation Memory. arXiv:2607.23927. <https://arxiv.org/abs/2607.23927>
- [9] [Preprint] **Ranjan, S.**, & Odegaard, B. (2025). Psychological Imagination Networks Show Cross-Population Centrality and Clustering Alignment in Humans That Large Language Models Fail to Replicate. arXiv:2510.04391. <https://arxiv.org/abs/2510.04391>
- [8] **Ranjan, S.**, & Odegaard, B. (2024). Heterarchy or hierarchy? Insights from a new model of visual imagination. *Physics of Life Reviews*, 49, 74–76.
- [7] **Ranjan, S.**, & Odegaard, B. (2024). Reality monitoring and metacognitive judgments in a false-memory paradigm. *Neuroscience Research*, 201, 3–17.
- [6] Maynes, R., Faulkner, R., Callahan, G., Mims, C. E., **Ranjan, S.**, Stalzer, J., & Odegaard, B. (2023). Metacognitive awareness in the sound-induced flash illusion. *Philosophical Transactions of the Royal Society B*, 378(1886), 20220347.
- [5] Chiasson, P., Boylan, M. R., Elhamiasl, M., Pruitt, J. M., **Ranjan, S.**, Riels, K., . . . & Keil, A. (2023). Effects of neurofeedback training on performance in laboratory tasks: A systematic review. *International Journal of Psychophysiology*.
- [4] Aggarwal, A., & **Ranjan, S.** (2022). How do undergraduate students reason about ethical and algorithmic decision-making? *Proc. 53rd ACM Technical Symposium on Computer Science Education*, 488–494.
- [3] Karnam, D., Agrawal, H., Parte, P., **Ranjan, S.**, Borar, P., Kurup, P. P., . . . & Chandrasekharan, S. (2021). Touchy feely vectors: A compensatory design approach to support model-based reasoning. *Journal of Computer Assisted Learning*, 37(2), 446–474.

[2] Karnam, D., Agrawal, H., Parte, P., **Ranjan, S.**, Sule, A., & Chandrasekharan, S. (2019). Touchy feely affordances of digital technology for embodied interactions can enhance ‘epistemic access’. *2019 IEEE Tenth International Conference on Technology for Education (T4E)*, 114–121.

[1] [Preprint] **Ranjan, S.**, & Srinivasan, N. (2019). Sense of agency for future-directed intentions. doi.org/10.31234/osf.io/qa93k.

INVITED TALKS

[2] Ranjan, S. (May, 2026). *The Spark of Artificial Neuroscience: Psychologically-Grounded Evaluation of Large Language Models*. Virtual invited talk, [Autonomous Empirical Research Group + Laboratory for Automated Scientific Discovery of Mind and Brain](#) (PI: Dr. Sebastian Musslick), Osnabrück University.

[1] Ranjan, S. (March, 2026). *Controlling Imagery Generation and its Awareness*. Virtual invited talk, [Robert Reinhart Lab](#), Boston University.

SELECTED CONFERENCE POSTERS & TALKS

[8] **Ranjan, S.**, & Odegaard, B. (2025). Psychological Imagination Networks in Humans and LLMs. *Frontiers in NeuroAI*, Kempner Institute Symposium, Harvard University.

[7] **Ranjan, S.**, & Odegaard, B. (2025). Visual Imagination Networks in Humans and LLMs. *Vision Sciences Society Annual Meeting*, St. Pete, FL.

[6] **Ranjan, S.**, & Odegaard, B. (2024). The Fragility of Reality Monitoring under Extraneous Factors. *Psychonomic Society 65th Annual Meeting*, NYC.

[5] **Ranjan, S.**, & Odegaard, B. (2024). Reality Monitoring, Fast and Slow [Poster + Flash Talk]. *Association for Psychological Science*, San Francisco. *Awarded Scott O. Lilienfeld APS Travel Award*.

[4] Maw, M., Zhuang, L., Baltes, J., **Ranjan, S.**, & Odegaard, B. (2024). Task demands and sensory externalization in reality monitoring. *PGSO Undergraduate Research Forum*, UF. *Mentee won 3rd Prize*.

[3] Dundigalla, S., Johnson, D., Roh, A., **Ranjan, S.**, & Odegaard, B. (2024). Cognitive strategies in reality monitoring. *PGSO Undergraduate Research Forum*, UF.

[2] **Ranjan, S.**, Baltes, J., Roh, A., & Odegaard, B. (2023). Confidence in reality monitoring judgments. *Journal of Vision*, 23(9), 4844.

[1] **Ranjan, S.**, & Srinivasan, N. (2018). Intentional binding and future-directed intentions [Talk]. *Annual Conference of Cognitive Science*, IIT-Guwahati, India.

AWARDS & FELLOWSHIPS

- 2025 — College of Liberal Arts and Sciences Travel Award, University of Florida
- 2025 — Threadgill Dissertation Fellowship, University of Florida
- 2024 — Scott O. Lilienfeld APS Travel Award, Association for Psychological Science
- 2024–2025 — UF Department of Psychology Travel Awards (×4)
- 2020–2025 — Graduate Student Fellowship, UF Department of Psychology
- 2016–2018 — Graduate Merit Scholarship, Centre of Behavioral & Cognitive Sciences, University of Allahabad