HAO LIU

https://haoliu4391.github.io/personal-website/ | liuh36@uci.edu

EDUCATION

University of California, Irvine (UCI) Sep 2026- Present
Cognitive Sciences Ph.D. with concentration (M.S.) in Cognitive Neuroscience
- Doctoral advisor: Mark Steyvers, Aaron Bornstein (co-supervise)

New York University (NYU) Sep 2024– May 2026
Master of Science (Cognitive Psychology)
- Major GPA: 3.76 /4.0

Tsinghua University (THU) June 2021– June 2024
Bachelor of Engineering (Computer Science and Technology)
- Computer Science Major GPA: 3.23 /4.0
- Core Courses: Fundamentals of Programming: 4.0
  Discrete Mathematics:3.6
  Linear Algebra: 3.6
  Foundation of Object-Oriented Programming: 4.0
  Artificial Neural Networks: 4.0
  Advances in Autonomous Driving and Intelligent Vehicles: 4.0 (A+)
  English Reading and Writing for academic purpose: 4.0 (A)
  English listening and speaking for academic purpose: 4.0 (A+)

Tsinghua University (THU) Sep 2020– June 2021
Electronic Engineering

Beijing Institute for General Artificial Intelligence (BIGAI) Sep 2022– June 2024
Intern (Multiagent Group)

Massachusetts Institute of Technology (MIT) June 2023– September 2023
Summer Internship (Department of Brain and Cognitive Sciences)

PUBLICATIONS

Huang, Y., Wang, X., **Liu, H.**, Kong, F., Qin, A., Tang, M., Zhu, S., Bi, M., Qi, S., & Feng, X. (2024). AdaSociety: An adaptive environment with social structures for multi-agent decision-making. *Advances in Neural Information Processing Systems, 37* (Datasets and Benchmarks Track).

Liu, H., Mattar, M., & Michelmann, S. (2026). Learning distributed representations with complementary components: A generative approximation framework for episodic memory encoding. *bioRxiv preprint doi:10.64898/2026.10.18.759853*.

Yu, J., Wu, X., **Liu, H.**, Guo, W., & Xing, X. (2025). BlockScan: Detecting Anomalies in Blockchain Transactions. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.

Liu, H., & Liu, J. (2026). The Gordian Knot for VLMs: Diagrammatic Knot Reasoning as a Hard Benchmark. *arXiv preprint arXiv:2605.09900*.

You, S., Wang, Z., **Liu, H.**, Liu, Y., He, Y., Hao, X., & Wang, L. (2026). Beyond Belief Nesting: Rethinking Higher-Order Theory of Mind in Large Language Models. *arXiv preprint arXiv:2610.17777*.

CONFERENCE CONTRIBUTIONS

NeurIPS 2025 Poster: BlockScan: Detecting Anomalies in Blockchain Transactions.
(https://arxiv.org/pdf/2410.04039)
Advisor: Xinyu Xing, Wenbo Guo / Northwestern (Remote) March 2024– June 2024
- Help build and test LLM based anomaly detection algorithm on blockchain transaction.
- Carefully clean and analyze the transaction data, including developing multi-thread algorithms to efficiently preprocess the data.
- Train and test different LLM models on fitting the dataset, LoRA for example.

NeurIPS 2024 Benchmark Track, Poster: AdaSociety: An Adaptive Environment with Social Structures for Multi-Agent Decision-Making
(https://neurips.cc/virtual/2024/poster/97511)
Advisor: Siyuan Qi, Xue Feng, Songchun Zhu / Peking University June 2023– June 2024
- Built a Grid world Based environment benchmark for studying generalized social behavior. The main dynamic of the environment consists of 3 parts: resource, event and agent. Agents could gather resources to make higher products through engaging events.
- Using RLLib to deploy classic algorithms like MAPPO, MADDPG to see the result.

AWARDS

YINGCAIJIHUA (Research training for science and technology innovation for high school students) October 2019

- Outstanding paper award
- Outstanding Student award

RESEARCH EXPERIENCE

PhD project: Stuckness and AI intervention in learning

Advisor: Mark Steyvers / UCI Sep 2026– Present

- Studying when and how an AI should step in to help people learn, starting from how productive struggle differs from being stuck.
- Built an online Sokoban experiment that logs every move, with puzzle difficulty set by the number of box changes in the shortest solution (computed for about 80,000 levels).
- Analyzing attempts as paths through each level's state graph to separate progress from stuckness; later experiments will test AI interventions at different moments.

PhD project: Group foraging and the edge of chaos

Advisor: Aaron Bornstein / UCI Sep 2026– Present

- Asking what individual foragers need to predict, or hear from others, for a group to approach the ideal free distribution when resources change.
- Building agent-based simulations with yield and crowding forecasts, depleting and regrowing patches, and word-of-mouth information sharing.
- Mapping the parameter space between groups that freeze and groups that never settle, to test whether groups do best near the edge of chaos.

Main MA project: Learning distributed representations with complementary components

Advisor: Sebastian Michelmann, Marcelo Mattar/ NYU October 2024– May 2026

- We simulate distributed memory representations with a generative neural network model (Stable Diffusion). These representations are learned with various levels of encoding strength that result in more accurate memory representations.
- Neo-cortical statistical knowledge is modeled by the weights of the network and hippocampus is implemented as a Modern Hopfield Network that stores the starting points to re-generate distributed representations from a low-dimensional latent.
- The model can perform the Mnemonic Similarity Task (MST) with various depth of encoding. Its output patterns mirror findings from memory impairment in human behavior.
- Domain-finetuned "expert" models need fewer encoding steps for images from their own domain than for images from other domains.
- I related the representational geometry of the model's latent codes to the Natural Scenes Dataset (NSD): the latent starting points resemble the medial temporal lobe (MTL), while reinstatement traverses the visual hierarchy in reverse order.
- I tested model predictions by designing a human behavioral experiment in which participants reconstruct masked images in a modified mnemonic similarity task.

Summer research project: Watch-Talk-Help – Multiagent communication based on Large Language Model and Theory of Mind

Advisor: Tianmin Shu, Joshua Tenenbaum / MIT June 2023– September 2023

- My project aimed to determine when to engage in natural language communication to bring the goals between humans and robots into alignment and to effectively plan together.
- I improved the model underlying the previous articles "watch-and-help" (Puig, X., et al., 2020) and "online-watch-and-help" (Puig, X., et al., 2023) by adding natural language communication parsed using LLM based agents' believes.
- Based on this natural language interaction, I defined rules for updating the agent's belief and goals.

OTHER RESEARCH PROJECTS

Pairwise learning / Functional embedding

Advisor: Zhiting Hu / UCSD(Remote) June 2022– September 2022

- Developed a unified framework of data & parameter space by learning pairs of data (leveraging CNNs & Transformers)

Stock prediction based on large scale pretrained models February 2022–September 2022

Advisor: Maosong Sun / Tsinghua University

- Built a transformer-based Chinese financial model by crawling 15GB of market text and designing a fine-tuning pipeline across Macro, Meso, and Micro levels.

A Self-supervised pretraining model based on Encoder-Decoder-Masker architecture December 2021–February 2022

Advisor: None

- Built a self-supervised encoder–decoder–unmasker on ImageNet-1K (hand-implemented ViT backbone), achieving competitive transfer to classification and low-level vision tasks.

A Q-A model based on query mechanism September 2021–November 2022

Advisor: Minlie Huang / Tsinghua University

- Curated 1M+ Zhihu QA pairs, built two different query mechanisms (BM25 and BERT-based relevance), integrated retrieval into LongLM pretraining, and surpassed baselines on BLEU/ROUGE.

Pollution Image classification system December 2017–October 2019(High school)

Advisor: Mingming Cheng / Nankai University

- Crawled 10k+ pollution images and trained an SVM classifier using hand-crafted features (dark channel, information entropy, and HSV color histograms).

Additional open coursework

- Open course: Stanford CS231n, Computer Vision — September 2020–December 2020
- Open course: Stanford CS224n/Ling284, Natural language processing with deep learning — January 2022–February 2022
- Open course: Introduction to reinforcement learning — May 2021– August 2021
- Open course: Stanford CS224w, Graph neural networks — May 2021– August 2021
- Open course: UCB CS285, Deep reinforcement learning — May 2021– August 2021
- Open course: Next Step of Machine Learning — January 2021– February 2021
- Open course: Geometric Deep Learning — June 2022 – September2022
- Open course: Stanford CS330 Deep Multi-Task and Meta Learning — September2022 – December 2022
- Open course: Categories for AI — March 2023 – December 2023

SKILLS

Programming: C, C++, Python (Pytorch, Gym, Ray(RLlib), Triton, PsychoPy), CUDA, HTML, FPGA, JavaScript
Applications: Matlab, WordPress
Languages: English (TOEFL BEST SCORE 2025: 113/120). Mandarin Chinese (First language)