Zizhao Hu

CS PhD Student, University of Southern California

zizhaoh@usc.edu | 213-675-4878 | zizhaohu.com | zizhao-hu.github.io | linkedin.com/in/zizhaohu

Research Interests

Post-training for LLM agentic memory (continual learning, in-context adaptation, unlearning, and world models); **Low-latency AI** (efficient attention, KV-cache compression, recurrent transformers); and **AI safety** (synthetic data, multi-agent risks, post-training guardrails).

Education

University of Southern California — Los Angeles, CA
Ph.D. in Computer Science (Artificial Intelligence) — 2022 – Present (expected 2027)
Advised by Dr. Jesse Thomason and Dr. Mohammad Rostami
M.S. in Computer Science — 2020 – 2022
Coursework: ML, CV, NLP, RL, continual learning, synthetic data, reasoning models, model fine-tuning.

Georgia Institute of Technology — Atlanta, GA
B.S. in Physics, Highest Honor — 2016 – 2020
Coursework: remote sensing, optical signal processing, quantum mechanics, robotics.

Research Experience

USC GLAMOR Lab — Graduate Research Assistant
Advised by Dr. Jesse Thomason — 2022 – Present

- Research self-improving agentic LLMs and the training pipelines that shape their memory and adaptation. Developed PRISM and Experience Tuning, post-training pipelines for adaptive LLM agents.
- Investigate safety alignment and guardrails in synthetic fine-tuning of VLMs and text-to-image models.
- Develop efficient neural architectures and cross-modal alignment methods for vision and vision-language tasks (text-to-image diffusion, VLMs).

USC iLab — Graduate Research Assistant
Advised by Dr. Laurent Itti — 2021 – 2022

- Developed GalilAI, an RL algorithm for AI agents to understand the physics of environments.

Georgia Tech Nanophotonics Lab — Undergraduate Research Assistant
Advised by Dr. Ali Adibi — 2019 – 2020

- Autoencoders for reverse-design of phononic metasurfaces.

Georgia Tech Agile Systems Lab — Undergraduate Research Assistant
Advised by Dr. Simon Sponberg — 2017 – 2018

- Image-segmentation pipeline for hawk-moth flight analysis.

Professional Experience

Researcher — IARPA Bengal Program (U.S. Government) — Jan 2026 – Present

- Developed SHRED, an LLM unlearning method to unlearn private IP documents and knowledge from pretrained models using on-policy self-distillation with logit demotion. PI: Dr. Robin Jia; Co-PI: Dr. Jesse Thomason.

Domain Lead — Handshake AI Fellowship, Project Canary — 2025

- Domain lead (AI & Machine Learning, PhD level) for LLM and vision-language post-training data curation; managed 78 PhD domain experts, automation-tool development, and logistics & communications with

teams from Meta, Anthropic, and OpenAI for curating reasoning datasets targeting Chatbot Arena and Humanity's Last Exam.

Domain Expert — Scale AI 2024

- STEM SFT data curation team; STEM domain expert labeling graduate-level STEM data for OpenAI.

Selected Publications

- **SHRED: Document Unlearning via Self-Distillation and Entropy Demotion.** Under review, NeurIPS 2026.
- **Expert Personas Improve LLM Alignment but Damage Accuracy: Bootstrapping Intent-Based Persona Routing with PRISM.** Under review, EMNLP 2026.
- **Multi-modal Synthetic Data Training and Model Collapse: Insights from VLMs and Diffusion Models.** Presented at ACM ICMI 2025.
- **Lateralization MLP: A Simple Brain-inspired Architecture for Diffusion.**
- **Intermediate Adapter: Efficient Alignment of Text in Diffusion Models.**
- **Rethinking Attention in Vision Tasks: Is Dynamic Parameterization Always Necessary?** IEEE TPAMI.
- **GalilAI: Out-of-Task Distribution Detection using Causal Active Experimentation for Safe Transfer RL.** AISTATS 2022.

Teaching

Teaching Assistant at USC; supervised, lectured, and mentored 1000+ graduate students on deep-learning projects in CV and NLP.

- **DSCI 552** (Machine Learning for Data Science): Summer 2025, Summer 2023.
- **CSCI 576** (Multimedia Systems Design): Spring 2025, Fall 2024, Fall 2022.
- **CSCI 567** (Machine Learning): Summer 2024, Summer 2022.
- **CSCI 566** (Deep Learning and Its Applications): Spring 2024.
- **CSCI 544** (Applied Natural Language Processing): Fall 2023, Spring 2023.

Awards, Leadership & Service

- **Joyce M. and Glenn A. Burdick Prize**, Georgia Tech (2018).
- **2 Gold Medals**, 27th & 28th Chinese Mathematics Olympiads.
- **Reviewer:** NeurIPS (2024, 2025, 2026), ICML (2024, 2025), ICLR (2024, 2025).

Skills

Fast adapter to new technologies and tools, with a track record of quickly mastering emerging models, frameworks, and workflows to prototype and ship.

- **Expertise:** Multimodal AI, synthetic data, VLM, VLA, LLM, diffusion, post-training, continual learning.
- **Languages:** English, Mandarin Chinese; Python, Java, JavaScript, SQL, C/C++.
- **Tools:** PyTorch, TensorFlow, CUDA, Git, SLURM, GNU/Linux, Cursor, Claude Code, Codex, Unity.
- **Agentic coding:** context management, agentic harness design, agentic skills, agentic coding workflows.