Shuying (Sue) Cao

Tel: +1 (510)-333-0784, Email: shuyingcao7@gmail.com

EDUCATION

Stanford University **Stanford, CA, US**
Visiting Student Researcher *June. 2026 – Present*
Research Area*:* Medical AI agents, neuroimaging, MRI/ fMRI data analysis

University of Southern California (USC) **Los Angeles, CA, US**
M.S. in Computer Science *Sept. 2024 – Dec. 2026*
Research Interests*:* Trustworthy Large Language Models, LLM Diversity, Benchmark & Evaluation

University of California, Berkeley (UCB) **Berkeley, CA, US**
Visiting Student, SKY Lab *Jan. 2023 – Jul. 2023*

Wuhan University (WHU) **Wuhan, Hubei, China**
B.E. in Geodesy and Geomatics Engineering, GPA **3.81/4.0** (rank **top 2%**) *Sept. 2020 – June 2024*
- **Awards**: Qinghai Li Scholarship (top 2%, 2022); First Class Scholarship of Wuhan University (2021&2022); Meritorious Winner in Mathematical Contest in Modelling (2021).

RESEARCH EXPERIENCE

Investigating Diversity Collapse in Large Language Models *Jan. 2026 – Present*
*RA, advised by **Prof. Sai Praneeth Karimireddy** (USC)*
- Investigate diversity collapse as a failure mode in LLM generation, where models concentrate on generic high-probability responses while suppressing alternative valid outputs.
- Build an evaluation framework for quantifying diversity–validity trade-offs across prompting, ICL, RAG, temperature, and sampling settings.
- Perform token- and sequence-level distributional analysis to study why generic responses dominate over diverse but valid alternatives.

Towards Generalizable Intelligence: Leveraging Differentially Private Large Language Models for Synthetic Mobility Generation *Jan. 2026 – Present*
*RA, advised by **Prof. Cyrus Shahabi** (USC), **Prof. Sai Praneeth Karimireddy** (USC)*
- Developing a novel framework integrating LLMs and differential privacy techniques for generating high quality synthetic human trajectory data;
- Designing a differential privacy reward model to fine-tune LLMs to generate controlled trajectories under realistic spatio-temporal constraints;
- Proposed a differential privacy context learning strategy that greatly reduces privacy risk and computational cost while maintaining data quality;
- Conducted rigorous membership inference attacks, empirically audited and quantified privacy leakage, and demonstrated strong privacy protection.

Auditing Privacy Preserved In-context Learning Methods *Dec. 2024 – May 2025*
*RA, advised by: **Prof. Sai Praneeth Karimireddy** (USC)*
- Developed ContextLeak, a black-box auditing framework for private in-context learning that uses canary insertion and adversarial query construction to probe worst-case privacy leakage and estimate lower bounds on the privacy budget.
- Designed a suite of canary types and query strategies, including hex, rare-unigram, and false-fact canaries, together with GEPA-based prompt optimization for stronger attack generation.
- Conducted unified evaluations across classification and open-ended generation settings, benchmarking RNM, ESA, and heuristic defenses under a common auditing framework.

- Showed that empirical leakage increases monotonically with theoretical privacy budget across models and datasets, and identified clear privacy-utility limitations in both heuristic and differential privacy defenses.

Publication: Jacob Choi#, Xingjian Dong#, **Shuying Cao#** (equal contribution), *Sai Praneeth Karimireddy**, **Auditing Privacy Preserved In-context Learning Methods**, accepted at L2M2 workshop in ACL 2025.

Development of a Chatbot for Analyzing and Interpreting Data from Academic Paper Images *May 2023 – Sept. 2023*
*RA, advised by **Prof. Joseph E. Gonzalez** (SKY Lab, UCB), PhD Student Tianjun Zhang (SKY Lab, UCB)*

- Used GPT-4 to assist with generating prompts for training and modified the prompts;
- Designed a model based on the Stanford Alpaca model and LLaMA 2 to assist researchers in reading papers by analyzing graphs in the relevant papers;
- Optimized the new model based on the Large Language and Visual Assistant (LLaVA) model for visual and language comprehension purposes.

PROJECTS & INTERSHIPS

Soul Soul - Persona Layer for OpenClaw Agents *March 2026 – Present*

- Led the design of the "soul" specification for agents, determining how persona attributes should be translated into structured prompts and interaction behavior.
- Designed a user questionnaire to capture persona signals and enable agents to understand user preferences, communication style, and behavioral tendencies.
- Launched the SoulSoul website and attracted 4K+ page views within the first 3 days of release, providing early validation for user interest in persona-driven agent experiences.

Website: https://soulsoul.live/

AskTheMap - WhatsApp Agent (Gemini 3 + Google APIs) *Feb. 2026 – April 2026*

- Integrated Gemini 3 into WhatsApp group chats and connected multiple Google APIs (e.g., Maps/ Places/ Directions/ Calendar) to enable tool-using agent workflows: restaurant recommendations, route planning, event creation, calendar scheduling, and reminders.
- Built message routing and tool orchestration (intent/command parsing, structured parameterization, tool invocation, and result synthesis) to deliver end-to-end task completion through chat.
- Designed a safer tool-access pattern: map capabilities run through a server-side MCP-style registry with no public endpoints and no direct network access granted to the LLM.

"GIS Banker" Personal Credits Evaluation Platform Based on Space Intelligence *Dec. 2021 - Jun. 2022*
Co-Founder, Rematrix

- Constructed a platform to reduce the problem of information asymmetry faced by the traditional credit evaluation model; accomplished this using remote sensing technology and machine learning;
- Added elements such as spatial information, remote sensing images, and carbon emissions as the basis for judging credit;
- Optimized the model with discriminant analysis and neural networks to assess personal economic status and credit status and composed a more comprehensive business plan;

SKILLS

- Computer Programming: C/C++, Python, Java;
- ML / LLM: PyTorch, Transformers, Prompt Engineering, In-Context Learning, RAG, LLM Evaluation;
- Data & Research: Experiment Design, Statistical Analysis, Benchmarking, Privacy Auditing;
- Language: Native Chinese; Fluent English (TOEFL 101);