

Cheng Chang

chc012@alumni.stanford.edu | linkedin.com/in/chc012 | github.com/chc012 | Google Scholar

Education

Stanford University Stanford, CA, USA
MS in Computational and Mathematical Engineering Sep 2022 – Jun 2024

– **Teaching Assistantships:** Machine Learning (2023); Natural Language Processing with Deep Learning (2024)

University of California San Diego La Jolla, CA, USA
BS in Cognitive Science; BA in Philosophy Sep 2018 – Jun 2022

– **Scholarships:** Eric Paul Allison Memorial Prize Fellowship (2022); Guided Engineering Apprenticeship in Research (2019)

Experience

Uniphore Palo Alto, CA, USA
Senior AI Scientist Feb 2026 – Present

- **Agent Evaluation:** Lead development of a production-grade system that evaluates AI agents end-to-end using an LLM-as-a-judge approach.
- **Benchmarking:** Drive research on benchmarks for evaluating fully autonomous applied-scientist and software-engineering AI agents.

Senior Software Engineer / AI Aug 2025 – Jan 2026

- **Agentic AI Research:** Develop coding AI agents for computer automation; create open-source benchmarks and train SOTA web automation models.
- **Software Engineering:** Develop LLM-based service for agent evaluation. Build services to help users select small language models (SLMs) for custom enterprise use cases.
- **Tech Stack:** Python, HuggingFace, LiteLLM, W&B, AWS, OpenAI, Anthropic, Git, Fireworks.ai

Orby AI Mountain View, CA, USA
Software Engineer Jul 2024 – Aug 2025

- **Multimodal Model Research:** Trained generative models and built web automation agents achieving SOTA on public benchmarks.
- **Software Engineering:** Integrated LLMs into backend automation tools for code generation and web interaction.
- **Tech Stack:** Python, Ray, Databricks, HuggingFace, W&B, AWS, Kubernetes

Software Engineer Intern Apr 2024 – Jun 2024

- Built a distributed web UI data crawler for foundation model training.

SALT Lab, Stanford University Stanford, CA, USA
Graduate Student Researcher Oct 2023 – Mar 2024

- Developed CARE 2.0 (LLM-based counselor training platform); implemented UI, backend APIs, and database.
- **Tech Stack:** Python, Flask, SQL, PyTorch, HuggingFace, Linux, Figma

GrubMarket, Inc. San Francisco, CA, USA
Data Science Intern Jun 2023 – Aug 2023

- Designed and implemented recommender systems for fresh produce.
- **Tech Stack:** Python, pandas, scikit-learn, PyTorch, FastAPI, SQL, AWS

Selected Publications

WARC-Bench: Web Archive Based Benchmark for GUI Subtask Executions. *ICLR 2026.*

Navigating the Digital World as Humans Do: Universal Visual Grounding for GUI Agents. *ICLR 2025.*

Is Your LLM Secretly a World Model of the Internet? Model-Based Planning for Web Agents. *TMLR 2025.*

Roleplay-doh: Enabling Domain Experts to Create LLM-Simulated Patients. *EMNLP 2024.*