# Cheng Chang

chc012@alumni.stanford.edu chc012.github.io github.com/chc012 Google Scholar

## Research Interests

Large language model (LLM) agents for computer task automation; trustworthy LLM evaluation; scalable data collection for foundation model training; LLM agent design.

## Education

**Stanford University** Stanford, CA, USA
*MS in Computational and Mathematical Engineering* Sep 2022 – Jun 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); Triton Research & Experiential Learning Scholars (2022); Guided Engineering Apprenticeship in Research (2019)

## Experience

**Uniphore (acquired Orby AI)** Palo Alto, CA, USA
*Senior AI Scientist* Feb 2026 – Present

- Drive research on software-engineering agent evaluation and LLM-as-a-judge meta-evaluation.
- Lead development of a production-grade software system that evaluates AI agents end-to-end using an LLM-as-a-judge approach.

*Senior Software Engineer* Aug 2025 – Jan 2026

- Developed coding agents for search automation; created an open-source graphical user interface (GUI) automation benchmark and trained web navigation models.
- Built services to help users select small language models (SLMs) for custom enterprise use cases.

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

- Trained generative models and built web automation agents achieving state-of-the-art results on public benchmarks.
- Integrated large language models into backend automation tools for code generation and web interaction.

*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, a counselor training platform powered by large language models; implemented the user interface, backend APIs, and database.

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

- Designed and implemented recommender systems for fresh produce.

## Publications

**AppSci-Bench: Can LLMs Build AI Agents and Make Them Better?**
*Under review*
Sanjari Srivastava, Manpreet Kaur, **Cheng Chang**, Sofía Ramos Antón, Oumaima Essamadi, Rishu Garg, Alejandro Lavale Jiménez, Marcos Caballero Yus, Mercedes García Martínez, Peng Qi

**Agent Behavior as Code: Efficient and Robust LLM Agents with Programmatic Specifications**
*Under review*
Peng Qi, Chunliang Lyu, Gang Li, Fabian Chan, **Cheng Chang**, Ignacio Cases, Will Lu

**PADMÉ: Preference Alignment Data Synthesis for Meta-Evaluation of LM Agent Evaluators**
*(To appear) Trust-AI-Eval (TAE) Workshop, Conference on Neural Information Processing Systems (NeurIPS), 2026*
https://openreview.net/forum?id=Xn765SbEmW
**Cheng Chang**, Yining Mao, Peng Qi

**WARC-Bench: Web Archive Based Benchmark for GUI Subtask Executions**
*International Conference on Learning Representations (ICLR), 2026*
https://openreview.net/forum?id=Hgw56DUFzD
Sanjari Srivastava, Gang Li, **Cheng Chang**, Rishu Garg, Manpreet Kaur, Charlene Y. Lee, Yuezhang Li, Yining Mao, Ignacio Cases, Yanan Xie, Peng Qi

**[WebDreamer] Is Your LLM Secretly a World Model of the Internet? Model-Based Planning for Web Agents**
*Transactions on Machine Learning Research (TMLR), 2025*
https://openreview.net/forum?id=c6l7yA0HSq
Yu Gu, Kai Zhang, Yuting Ning, Boyuan Zheng, Boyu Gou, Tianci Xue, **Cheng Chang**, Sanjari Srivastava, Yanan Xie, Peng Qi, Huan Sun, Yu Su

**[UGround] Navigating the Digital World as Humans Do: Universal Visual Grounding for GUI Agents**
*International Conference on Learning Representations (ICLR), 2025*
https://openreview.net/forum?id=kxnoqaisCT
Boyu Gou, Ruohan Wang, Boyuan Zheng, Yanan Xie, **Cheng Chang**, Yiheng Shu, Huan Sun, Yu Su

**Roleplay-doh: Enabling Domain Experts to Create LLM-Simulated Patients**
*Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024*
https://aclanthology.org/2024.emnlp-main.591/
Ryan Louie, Ananjan Nandi, William Fang, **Cheng Chang**, Emma Brunskill, Diyi Yang

**Integrative Machine Learning Approaches for Predicting Disease Risk Using Multi-Omics Data from the UK Biobank**
*bioRxiv preprint, 2024*
https://www.biorxiv.org/content/10.1101/2024.04.16.589819v1
Oscar Aguilar, **Cheng Chang**, Elsa Bismuth, Manuel A. Rivas

**Affective Response to Volitional Input Perturbations in Obstacle Avoidance and Target Tracking Games**
*Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2021*
https://ieeexplore.ieee.org/abstract/document/9630523
Aashish N. Patel, Geeling Chau, **Cheng Chang**, Allan Sun, Jingya Huang, Tzyy-Ping Jung, Vikash Gilja

## Teaching

**Stanford University:** Teaching Assistant, Natural Language Processing with Deep Learning (2024)

**Stanford University:** Teaching Assistant, Machine Learning (2023)

**University of California San Diego:** Teaching Assistant, Neurobiology of Cognition (2021)

## Services

**Reviewer:** ICLR 2027; ACL ARR 2026; NeurIPS TAE Workshop 2026; NeurIPS SLM Agent Workshop 2026; NeurIPS UserSim Workshop 2026; PLOS ONE (2026–present)