

Cheng Chang

chc012@alumni.stanford.edu

linkedin.com/in/chc012

github.com/chc012

Link to Google Scholar

Research Interests

LLM/VLM-based agents for GUI and web automation; benchmark design and evaluation; scalable data pipelines for foundation model training; model-based planning for agents.

Education

Stanford University

MS in Computational and Mathematical Engineering

Stanford, CA, USA

Sep 2022 – Jun 2024

University of California San Diego

BS in Cognitive Science; BA in Philosophy

La Jolla, CA, USA

Sep 2018 – Jun 2022

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

Experience

Uniphore Technologies North America (acquired Orby AI)

Senior AI Scientist

Palo Alto, CA, USA

Feb 2026 – Present

- Lead development of production-grade software system that evaluate AI agents with an LLM-as-a-judge approach.
- Research on benchmark for evaluating fully automated applied scientist and software engineering AI agents.

Senior Software Engineer / AI

Aug 2025 – Jan 2026

- Develop coding AI agents for computer automation tasks; create benchmarks and train SOTA web automation models.
- Build services for selecting SLMs for enterprise use cases.

Orby AI

Software Engineer

Mountain View, CA, USA

Jul 2024 – Aug 2025

- Trained generative models and built web automation agents achieving SOTA on public benchmarks.
- Integrated LLMs into backend business automation tools.

Software Engineer Intern

Apr 2024 – Jun 2024

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

SALT Lab, Stanford University

Graduate Student Researcher

Stanford, CA, USA

Oct 2023 – Mar 2024

- Developed CARE 2.0, an LLM-based counselor training platform; implemented UI, backend APIs, and database.

GrubMarket, Inc.

Data Science Intern

San Francisco, CA, USA

Jun 2023 – Aug 2023

- Designed and implemented recommender systems for produce.

Publications

WARC-Bench: Web Archive Based Benchmark for GUI Subtask Executions

ICLR 2026

<https://arxiv.org/abs/2510.09872>

Sanjari Srivastava, Gang Li, **Cheng Chang**, Rishu Garg, Manpreet Kaur, Charlene Y. Lee, Yuezhong Li, Yining Mao, Ignacio Cases, Yanan Xie, Peng Qi

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

TMLR 2025

<https://arxiv.org/abs/2411.06559>

Yu Gu, Kai Zhang, Yuting Ning, Boyuan Zheng, Boyu Gou, Tianci Xue, **Cheng Chang**, Sanjari Srivastava, Yanan Xie, Peng Qi, Huan Sun, Yu Su

Navigating the Digital World as Humans Do: Universal Visual Grounding for GUI Agents

ICLR 2025

<https://osu-nlp-group.github.io/UGround>

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

EMNLP 2024

<https://roleplay-doh.github.io/>

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 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

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, Machine Learning (2023)

Stanford University: Teaching Assistant, NLP with Deep Learning (2024)

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

Tools & Skills

Python; HuggingFace; Ray; Databricks; AWS; Kubernetes; PyTorch; evaluation/benchmarks; data pipelines; LLM/VLM agents.