

Matthew Sebastian Ho

mah048@ucsd.edu, (650) 460-0082

github.com/matt-seb-ho,

[linkedin.com/in/matthew-s-ho/](https://www.linkedin.com/in/matthew-s-ho/)

Education

PhD Computer Science | University of California, San Diego

Beginning September 2024

Isolating, understanding, and enhancing LLM reasoning capability; advised by Professor Lianhui Qin

B.S. Computer Science | University of California, Santa Barbara

October 2020 - December 2023

GPA: 4.00, Dean's Honors All Quarters Offered

- Scholarships: National Merit Scholarship, Raytheon MathMovesU Scholarship
- Coursework: Natural Language Processing, Computer Vision, Deep Learning, Machine Learning, Statistics, Data Structures, Algorithms, Object-Oriented Programming, Linear Algebra, Vector Calculus, Discrete Math
- Honors: **Chancellor's Award for Excellence in Undergraduate Research 2023**

Experience

Benchling Software Engineer Intern (Big Data Infrastructure Team)

June 2023 - September 2023

- Implemented batching to improve Data Warehouse sync pipeline reliability (**92%** reduction in retried jobs)
- Reduced sync latency via PostgreSQL query optimization and Celery job scheduling interventions
- Prototyped an AI assistant/natural language query interface for the Notebook product (**Hackathon Winner!**)
- Deployed Llama 2 model to AWS EC2 and harnessed retrieval augmented generation with chromaDB

Amazon Software Dev Engineer Intern

June 2022 - September 2022

- Created data revision viewing service to reduce operational burden and save dev hours across **6+** teams
- Tested through beta, gamma stages; launched into production with plans to release to entire organization
- Leveraged AWS Lambda, API Gateway, and Cloud Development Kit for backend; React for frontend

UCSB NLP Group Undergraduate Researcher

September 2021 - Present

- Led team on LLM reasoning benchmark research project, advised by Professor William Wang
- Crowdsourced 10,000+ entry reasoning dataset with MTurk using custom frontend + JS validation + Rails
- Leveraged PyTorch Lightning Modules, HuggingFace Metrics, and Pyserini Toolkit for model evaluation
- **First-authored Top 1.2% (Rank 51/4019, Oral Presentation) paper at top AI conference ICLR 2023**
- Project Repository: github.com/matt-seb-ho/WikiWhy | Paper: openreview.net/pdf?id=vaxnu-Utr4I

ServiceNow Software Engineer Intern

June - September 2021

- Developed debugging log utility; Added new security functionality; Increased test coverage

UCSB Robotics Software Lead

November 2020 - June 2022

- Developed oncoming vehicle detection via openCV preprocessing, YOLO classification & centroid tracking
- Trained reinforcement learning agent for VEX Change Up Simulation computer opponent

Stanford (Pulmonary, Allergy, and Critical Care) Medicine Research Intern

June - August 2019

- Researched signaling pathways in Pulmonary Hypertension under Professor Maya Kumar
- Created R program to identify, tabulate and visualize receptor-ligand pair scores
- Used R for Principal Component Analysis and t-SNE Dimension Reduction for single cell sequence data

Selected Projects

Free Food Forecast

May 2023

- Integrated with G Suite to summarize email digests and parse food event invitations into calendar entries
- Utilized LLM Parse prompting with OpenAI API for information extraction and summarization
- Mitigated interpretability problem by visualizing source spans and model confidence with streamlit frontend

Browser-based Virtual Assistant: Dingus (Overall 2nd place winner at SBHacks VIII)

February 2022

- Employed AssemblyAI's neural speech recognition for Google search and YouTube player voice commands
- Fine-tuned DialoGPT model with *Friends* dialog dataset as an aligned, personable chatbot

Publications

Ho et al. *WikiWhy: Answering and Explaining Cause-and-Effect Questions*. ICLR 2023, Oral (Top 1.2%)

Ho et al. *Proof Flow: Preliminary Study on Generative Flow Network Language Model Tuning for Formal Reasoning*. NeurIPS 2024 System 2 Reasoning at Scale Workshop.

Skills

Languages: Python, C++, Java, JavaScript, OCaml, Ruby, SQL, HTML, CSS

Tech: Pandas, Matplotlib, Pytorch, Pytorch Lightning, HuggingFace, Flask, ChromaDB, AWS, React, Git, Maven