

ANDREW BAI

andrewbai@cs.ucla.edu | Github | Google Scholar

I am a fifth year PhD student at UCLA Computer Science Department advised by Prof. Cho-Jui Hsieh. I study high quality **data selection (synthesis)** and how data impacts skill **acquisition** and **forgetting**. Recently I focus on applications in large language model (LLM) **post-training**.

Research projects: SFT, reward modeling, LLM agents, prompt optimization, model interpretability

ML coding skills: RLHF in `trl`, implement transformers in Pytorch, optimize sparse matrix ops in C++

WORK EXPERIENCE

Nvidia

Sep 2025 – Dec 2025 (expected)

LLM Technologist Intern

- Characterized the severity of catastrophic forgetting when fine-tuning LLMs with different reinforcement learning algorithms, expert demonstrations, and reward models for reasoning tasks.

Google

Research Intern @ DeepMind

Jun 2025 – Sep 2025

- Architected an end-to-end prototype with cross-team collaboration for tool calling in the Gemini App using SLMs loaded on Android devices. Achieved an average of 98% accuracy and sub-1s E2E latency.
- Designed a benchmark for LLM response formatting inspired by pedagogical principles. Leveraged LLM persona role-play for online multi-turn auto-eval. Identified 30%+ losses with LLM-as-a-judge.

Student Researcher @ Bard

Jun 2024 – Oct 2024

- Developed novel early-stopping metrics for supervised fine-tuning on instruction data to maximize downstream DPO performance. Validated hypothesis across 7 instruction and 3 alignment datasets.

Student Researcher @ Cloud

Apr 2023 – Aug 2023

- Designed a computationally-free rehearsal scheme to mitigate catastrophic forgetting by increasing the likelihood of sampling “useful” samples. Achieved equal performance with up to 50% less computation.

Amazon

Jun 2022 – Sep 2022

Applied Scientist Intern

- Implemented and optimized factorization machine training and inferencing in C++, increasing the training speed by 43x compared to `libffm` (see open-source code PECOS for details).

EDUCATION

University of California, Los Angeles

Sep 2021 – Dec 2025 (expected)

Ph.D. in Computer Science

National Taiwan University

Sep 2016 – Jan 2021

B.S. in Computer Science and Information Engineering

SELECTED PUBLICATIONS

- On the Loss of Context-awareness in General Instruction Fine-tuning. Under review.
- An Efficient Rehearsal Scheme for Catastrophic Forgetting Mitigation during Multi-stage Fine-tuning. In *Findings of the Association for Computational Linguistics, Conference of the Nations of the Americas Chapter (NAACL)*, Apr 2025.
- Concept Gradient: Concept-based Interpretation Without Linear Assumption. In *Proceedings of the 11th International Conference on Learning Representations (ICLR)*, May 2023.