Allan Zhang

Last updated in May 2025

Los Angeles, CA allanzhang440@gmail.com
609-943-8429 https://giyushino.github.io giyushino

Education

University of California, Los Angeles *Sept 2024 – June 2028*
BS in Applied Mathematics

- **Relevant Coursework:** Multivariable Calculus, Linear Algebra, Discrete Structures, Probability Theory

Experience

Undergraduate Research Assistant | BigML *Los Angeles, CA*
UCLA *Nov 2024 – Present*

- Aided PhD student Siddharth Joshi in training and evaulating custom, lightweight VLMs on spatial reasoning tasks using mechanistic interpretability techniques (i.e. linear probing, ablation studies, and feature visualization)
- Wrote functions to create highly modular datasets, testing the effects of modality mismatch, spurious correlations, alignment between visual and textual context, and number of unique images per class on model's performance
- Helped implement variation of GRPO to improve LLMs reasoning capatabilities, wrote framework to detect sparse errors in model trajectories, validating hypothesis that sparse errors accounted for noisy gradient updates

Projects

Neural Network Evolution *NeuroEvolution*

- Designed and implemented a project to explore the effectiveness of evolutionary principles(sexual reproduction, mutation, and natural selection) in producing accurate CNNs and ViTs
- Developed a flexible Python framework to apply custom transformations to any PyTorch or Jax neural networks, including layer-wise merging and controlled injection of randomness into model weights
- Compared neural networks trained via gradient descent and evolution-based optimization, finding that both approaches converged to models with highly similar weight configurations

TinyMathLLM *TinyMathLLM*

- Researched how datasets, model architecture, compute power, prompting, low-rank adapation, etc can be optimized to finetune LLMs. Demonstrated effectiveness of using LLMs to extract data from questions and using tools
- Finetuned TinyLlama (1.1B) on custom synthetic datasets entirely on RTX 4070, demonstrating effective math agents can be trained with limited resources. Despite training 0.0024% of params using LoRA and using 4-bit quantization, model experienced significant improvment in performance

Doodle Guesser *MyOwnDoogleGuesser*

- Finetuned OpenaAI CLIP-Vit-Large-Patch-14 and custom vision transformer on 6 different animal drawings. Collected and processed 1,000,000+ images of the different classes from web to create training set
- Developed a custom training and inference pipeline that overcame memory-related crashes in the previous method, boosting accuracy from 54% to 87%.
- Created GUI using Pygame to allow users to draw on a canvas. After drawing, screenshot was taken, processed, and fed into model, softmax taken of the outputted logits to predict the most likely class

Skills

Programming Languages and Frameworks: Python, PyTorch, NumPy, Matplotlib, TensorFlow, scikit-learn, OpenCV, Hugging Face, LaTeX, C++, Linux (Arch btw), Vim, Neovim

Languages: English, Korean