Benjamin Yu

714-507-6312 | yubenjamin2004@gmail.com | linkedin.com/in/benjamin-yu | yubenjamin2022.github.io/

Research Interests

Natural Language Processing, Explainable AI, Large Language Models and Multimodal Models, multilingual and cross-lingual NLP, reinforcement learning for language models

Education

University of California, Los Angeles — Los Angeles, CA
Bachelor of Science in Data Theory — *Sept. 2023 - June 2026*

Publications

Smart-GRPO: Smartly Sampling Noise for Efficient RL of Flow-Matching Models — Preprint
***Benjamin Yu**, Ziyang Liu*, Justin Cui*

Compressing Datasets for Machine Learning Interatomic Potentials Using Information Theory — Preprint
***Benjamin Yu**, Daniel Schwalbe-Koda**

Video Text Preservation with Synthetic Text-Rich Videos — ICCV 2025 Workshop
*Ziyang Liu, **Benjamin Yu**, Justin Cui*

** Equal Contribution*

Research

Undergraduate Researcher — April 2024 – Present
Digital Synthesis Lab — *Los Angeles, CA*

- Developed information-theoretic compression algorithms of datasets for machine learning interatomic potentials (MLIPs) in Python
- Developed Python sub-package of baseline and information-theoretic algorithms
- Evaluated the information-theoretic compression algorithm against baseline models to assess dataset compression efficiency and downstream training performance
- Implemented parallelization of sampling algorithms, increasing speed of sampling by 200%
- Presented research and results at the 2025 APS Global Physics Summit

Undergraduate Researcher — April 2025 – Sept. 2025
Computational Machine Learning Group — *Los Angeles, CA*

- Designed and tested gradient-free methods to efficiently sample noise for reward tuning in reinforcement learning algorithms such as GRPO, applied to flow-matching generative image models
- Explored optimization techniques for aligning flow-matching diffusion transformers with human preference signals through efficient GRPO-based reward tuning
- Studied supervised fine-tuning strategies to improve text rendering accuracy in text-to-image and video generation pipelines

Experience

Machine Learning Intern — June 2025 – August 2025
Johns Hopkins Applied Physics Laboratory — *Laurel, MD*

- Developed retrieval-augmented generation (RAG) capabilities on large language models using LangChain and transformers to answer questions related to large (300+ page) documents
- Fine-tuned encoder models (CrossEncoders, SBERT) with LoRA for failure log classification
- Optimized prompts for LLMs using DSPy for classification of failure logs, increasing classification accuracy by 50%
- Developed MLP, CNN and diffusion models to predict gravitational fields using PyTorch and diffusers

Software Engineering Intern — June 2024 – August 2024
U.S. Naval Research Laboratory — *Washington, D.C.*

- Constructed a matrix representation of points in polar space for machine learning applications
- Designed a CNN model with PyTorch to find targets in the matrix representation
- Implemented complex neural networks for classification of raw satellite data, reaching 97% accuracy on validation data

Projects

Dad Joke Transformer | *Python, PyTorch, einops, transformers* September 2024 – December 2024

- Implemented and trained a GPT-2 model from scratch to generate dad jokes in PyTorch
- Studied and implemented techniques such a top-k sampling to diversity responses of the GPT2 model

Reward-tuning LLM for Mathematics | *Python, PyTorch, Huggingface, transformers* March 2025 – June 2025

- Fine-tuned the Qwen language model using GRPO to enhance chain-of-thought reasoning and mathematical problem-solving, leveraging the Hugging Face ecosystem
- Designed, implemented, and evaluated multiple reward functions to assess their impact on model performance and alignment

Leadership & Extracurriculars

ACM AI at UCLA April 2025 – Present

Workshop Officer *Los Angeles, CA*

- Developed and taught workshops on various AI concepts such as neural networks, natural language processing and reinforcement learning to UCLA undergraduate students
- Created and led a Jupyter Notebook workshop on training neural networks and reinforcement learning models using PyTorch to teach technical ML skills

Reading Group

- Read and discussed papers with other undergraduate and graduate students regarding current topics in ML such as RLHF, generative images and videos, singular learning theory, and natural language processing

Statistics Club at UCLA April 2024 – May 2025

Workshop Chair *Los Angeles, CA*

- Prepared several workshops and presented to over 100 students in various data science concepts such as machine learning, Python, and data analysis
- Planned and coordinated a data science hackathon for UCLA freshmen for over 20 teams
- Led a team of 3 interns to create a data science course for UCLA students

Technical Skills

Languages: Java, Python, C/C++, SQL, MATLAB, HTML/CSS, R
Developer Tools: Git, Docker, VS Code, Visual Studio, PyCharm
Libraries: pandas, NumPy, Matplotlib, PyTorch, TensorFlow, diffusers, transformers, LangChain