# Benjamin Yu

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

## 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** — AAAI 2026 AIR-FM Workshop
***Benjamin Yu**, Ziyang Liu*, Justin Cui* — *arXiv:2510.02654*

**Maximizing Efficiency of Dataset Compression for Machine Learning Potentials With Information Theory** — Preprint
***Benjamin Yu**, Vincenzo Lordi, Daniel Schwalbe-Koda** — *arXiv:2511.10561*

## Research Experience

**UCLA NLP Group** — September 2024 – Present
- Surveyed unified and multimodal model literature to examine how generation objectives may support multimodal understanding beyond generation quality.
- Analyzed existing architectures to identify gaps in leveraging generation learning for understanding-focused objectives.
- Reproduced and experimented with open-source unified model codebases to study representation learning.

**UCLA Digital Synthesis Lab** — April 2024 – October 2025
- Implemented a greedy Shannon information-theoretic subsampling algorithm to maximize entropy and coverage of atomistic datasets for more efficient machine learning
- Designed baseline subsampling methods and evaluation pipelines using information-theoretic coverage metrics.
- Trained ML interatomic potentials on subsampled datasets and showed up to 80% reduced variance in low-data regimes.
- Co-authored a manuscript, presented at the APS Global Physics Summit, and submitted to the *Journal of Chemical Physics*.

**UCLA Computational Machine Learning Group** — April 2025 – October 2025
- Led a project on optimizing noise inputs for RLVR of flow-matching generative models to stabilize learning.
- Ran proof-of-concept experiments showing differentiable reward-optimized noise prevents reward degradation.
- Developed a beam-search noise optimization method for non-differentiable rewards that improves stability and performance over rewards between 8% to 28%.
- Co-authored a workshop paper for the 2026 AAAI AIR-FM Workshop.

## 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
- Developed MLP, CNN and diffusion models to predict gravitational fields using PyTorch and diffusers

**Software Engineer Intern** — June 2024 – August 2024
*U.S. Naval Research Laboratory* — *Washington, DC*
- Devised a keypoint detection model to find a matrix of targets, attaining a difference of less than 5% for 95% of data
- Implemented neural networks for imaginary-valued data for classification of raw satellite data, reaching 97% accuracy on validation data

## Leadership & Extracurriculars

**ACM AI at UCLA** — April 2025 – Present
*Workshop Officer & AI Scholarship Foundation Co-Director* — *Los Angeles, CA*

## 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