William Francis Arnold

willarnold@berkeley.edu | swaglu.com | github.com/Aphoh | linkedin.com/in/~arnold

Education

Korea Advanced Institute of Science and Technology (KAIST) *September 2023 – Current*
MS, Kim Jaechul School of AI, MLILAB under Professor Eunho Yang
Research on Transformer Inference & Training Efficiency

University of California, Berkeley *August 2017 – May 2021*
BS, Electrical Engineering and Computer Science Major, Math Minor — Highest Honors
Graduate Coursework in High Performance Computing, Theoretical Physics — University Medal Nominee

Projects

Sequence Level Experts for On-Device Infilling - Master's Thesis
Developed and trained hardware-aware MoE modifications for on-device inference by routing sequences rather than tokens. Allows expert offloading to flash, reducing memory usage by 6x while maintaining dense model inference throughput and TTFT latency. Demonstrated model performance through robust training experiments on TPUs using JAX and a custom data pipeline.

WRIST - Torch Auto-Parallelism Search through Simulation
Used `torch.fx.Graph` traces, network topology analysis, and accelerator performance models to efficiently search millions of parallelism configurations for optimal hardware utilization. Achieved 0.96 spearman correlation predicting 20B transformer training performance on 4x4 A100 clusters.

Experience

Google — Student Researcher, Gemini Applied Research *April 2025 — Current*
- Developed a method for statistically efficient measurement of math/coding LLM benchmarks, pending preprint.

Polygon Labs — Senior Engineer, Applied Research *November 2022 — August 2023*
- Redesigned cryptographic commitment schemes, achieving 16x throughput via algorithmic and system-level optimizations.

Eluvio — Software Engineer *Summer 2020, October 2021 – November 2022*
- Built a distributed key management system in Intel SGX/AMD MET securing $100M+ worth of Fortune 500 company media assets.

UC Berkeley — Machine Learning Researcher & Instructor *August 2018 — May 2021*
- Conducted research in Prof. Spanos lab on Reinforcement Learning for intelligent control of energy systems in large buildings, leading to a 3 week energy usage experiment at the National University of Singapore.
- Taught discrete math and probability theory (CS70), receiving 4.97/5.0 on student evaluations (above the department mean of 4.4/5).
- Held over 200 office hours, proctored, wrote questions for, and graded 12 exams.

Procore Technologies — Machine Learning Intern *June 2019 — August 2019*
- Built a image tagging interface using image embedding models and Gaussian clustering accelerate manual tagging of domain-specific latent classes in a construction image dataset, resulting in a 8x larger dataset, increasing accuracy by 40% on edge classes.

Activate Inc., at Google — Technical Lead *February 2016 — August 2018*
- Used machine learning to identify high-demand concepts around 50+ locations in San Francisco for a Fortune 500 company.
- Executed $250k contract building models to predict how Google employees choose on-campus cafeterias to reduce food waste.

Gradescope — Android Developer *September 2014 — July 2016*

Publications

- Factored Agents: Decoupling In-Context Learning and Memorization for Robust Tool Use. In review at *COLM 2025*.
- Chatbot Arena Estimate: towards a generalized performance benchmark for LLM capabilities. In *NAACL 2025*.
- DisruptionBench: Two Advancements in Machine Learning Driven Disruption Prediction. In *Journal of Fusion Energy 2025*.
- Back-to-Basics Revisited: Benchmarking an Expanded Set of RLHF Algorithms. In *NeurIPS SFLLM 2024*.
- Continuous Convolutional Neural Networks for Disruption Prediction in Nuclear Fusion Plasmas. In *NeurIPS CCAI 2023* (1st author).
- Adapting Surprise Minimizing Reinforcement Learning Techniques for Transactive Control. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems 2021* (1st author).
- Pricing in Prosumer Aggregations using Reinforcement Learning. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems 2021*.

Activities

Cal Sailing Team – Varsity Skipper and Fleetmaster *August 2018 — May 2021*