Kunal Pai

408-620-2339 | pai.kunal05@gmail.com | linkedin.com/in/kunpai | github.com/kunpai | kunpai.space

Education

M.S., Computer Science, University of California, Davis (GPA: **4.0/4.0**) Expected: June 2025

B.S., Computer Science & Engineering, University of California, Davis (GPA: **3.8/4.0**) June 2023

Work Experience

Graduate Student Researcher, *DArchR Lab @ University of California, Davis* Jun 2023 - Present

- Leading a team to develop a superconductor-based graph accelerator in the gem5 simulator.
- Leading a team to deliver full-system, cycle-level simulation models for cryogenic semiconductors and superconductors in gem5.
- Collaborating with a team to develop an autotuning methodology with 90% correlation between gem5 simulation results and hardware profiling metrics.
- Mentoring 5 undergraduate students in the Davis Computer Architecture Lab to prepare them for graduate research.

Student Researcher, *DECAL Lab @ University of California, Davis* Sept 2022 - Present

- Developing a dataset for pairwise code-documentation alignment from open-source Python projects, to enable future research in software maintenance.
- Created a pipeline for measuring calibration and correctness of large language models for code repair, using Defects4J.
- Assisted in validating efficacy of semantic augmentation of language model prompts for code summarization using precision and recall metrics like ROUGE and METEOR.

Teaching Assistant, *University of California, Davis* Sep 2023 - Dec 2023

- Assisted 180 students in a senior-level Probability & Statistical Modeling class.

Undergraduate Researcher, *DArchR Lab @ University of California, Davis* Jun 2022 - Jun 2023

- Implemented a model of the HiFive Unmatched RISC-V board in gem5, achieving 85% accuracy with hardware profiling metrics.
- Authored a poster on the validation of hardware and simpoints with gem5, presented at the gem5 workshop at ISCA 2023.
- Co-authored tutorials on perf compilation for RISC-V and documentation for the Standard Library in gem5.

Software Developer Intern & Tech Lead, *humanID, Davis* Jan 2022 - Jun 2022

- Completed 10 projects with global teams, including:
  - Documentation of a Discord bot that combats spam and fake users.
  - Django-based web application for permission management for 100 users.

Technical Product Marketing Intern, *SiTime Corp., Santa Clara* Jul 2021 - Sep 2021

- Presented strategy to improve distributor margin management and earned profits by $250,000.
- Conducted a market survey on optical transceivers used in AI networking, to identify customers for MEMS timing chips.
- Created Visio diagrams for the product requirements document (PRD) of a timing chip.

Academic Projects

Automated Frameworks of Semantic Augmentation to Improve MWP Solving Apr 2024 - Jun 2024

Machine Learning Project *Python, NLP, Prompt Engineering*

- Improved PaLM 2 LLM prompting accuracy on math word problems (MWPs) by 10% and TinyLlama fine-tuning LM accuracy by 60% through a one-shot digit-level semantics framework.
- Introduced a novel demonstration selection model to improve accuracy of LLMs. Model used BLEU scores and Levenshtein distance to identify the most similar equations for one-shot examples.

Effects of Toxicity on Disengagement in Open-Source Projects Jan 2024 - Mar 2024

Software Engineering Project *Python, GitHub mining, scikit-learn*

- Found a strong correlation ($R^2 = 0.76$) between high developer engagement in FAANG projects with larger codebases and lower levels of toxicity, offering actionable insights for community management.
- Quantified toxic behavior using sentiment analysis and mining corporate and non-profit repositories, revealing that toxicity disproportionately impacts new developers, affecting them up to 1.3x more than experienced ones.

Behavior of Spectre in Different Branch Predictors in gem5 Oct 2023 - Dec 2023

Computer Architecture Project *Python, C++, gem5, Docker*

- Demonstrated up to a 55% reduction in susceptibility to speculative execution attacks by validating design enhancements like longer training periods and minimizing biased branches for Spectre-resistant branch predictors.
- Investigated the vulnerability of x86-based in-order and out-of-order processors to Spectre V1 attacks, revealing a strong correlation between branch predictor training periods and attack effectiveness.

gem5 Vision Jan 2023 - Jun 2023
Framework *Next.js, Python, MongoDB, JSON Schema*

- Boosted resource discovery speed by 20x with optimized search functionality across 1,200+ resources.
- Enabled faster retrieval of resources across 20+ categories by introducing categorization and semantic versioning.
- Enhanced accessibility for 500+ industry and academic user by integrating local/remote JSON files and MongoDB with gem5.

Publications

CoDocBench: A Dataset for Code-Documentation Alignment in Software Maintenance Pai, K., Devanbu, P. & Ahmed, T.
Mining Software Repositories (MSR) 2025: Data and Tool Showcase Track

Calibration and Correctness of Language Models for Code Spiess, C., Gros, D., Pai, K., et. al.
International Conference on Software Engineering (ICSE) 2025

Potential and Limitation of High-Frequency Cores and Caches Pai, K., Nand, A. & Lowe-Power, J.
ModSim 2024: Workshop on Modeling & Simulation of Systems and Applications

Automatic Semantic Augmentation of Language Model Prompts (for Code Summarization) Ahmed, T., Pai, K., et. al.
International Conference on Software Engineering (ICSE) 2024

gem5 Vision Shah, P., Pai, K., et. al.
gem5 Workshop at International Symposium on Computer Architecture (ISCA) 2023

Validating Hardware and SimPoints with gem5: A RISC-V Board Case Study Pai, K., Qiu, Z., & Lowe-Power, J.
gem5 Workshop at International Symposium on Computer Architecture (ISCA) 2023

Technical Skills

Languages: Python, C++, C, Java, JavaScript

Frameworks: React, Next.js, TensorFlow, PyTorch, Django, Flask, scikit-learn, pandas, NumPy, Matplotlib

Tools & Technologies: Git, Docker, MongoDB, gem5, Unix/Linux, LaTeX