Aaron Lee

(650) 382-7190 hylee@ucsb.edu San Jose, CA

Programming Languages: Python, C/C++, C#, Java, Go, Verilog, JS, HTML/CSS, Perl
Frameworks: PyTorch, Keras, Scikit-learn, Apache Spark, AvalancheGo, Spring Boot, Node.js, React
Technologies: Git, Azure, Docker, PostgreSQL, MongoDB, Databricks, ARM, Vivado, Altera, Selenium

EDUCATION

University of California, Santa Barbara — *Santa Barbara, CA* September 2021 - March 2024
B.S. in Computer Engineering (Honors) GPA: 3.78 / 4.0

- **Coursework:** Data Structures & Algorithms, Object-Oriented Design, Machine Learning, Deep Learning, Computer Networking, Computer Architecture, Real-Time Embedded Systems

EXPERIENCE

Software Engineer Intern - Data — *Procter & Gamble, Cincinnati, OH* June 2024 - September 2024

- Designed and maintained scalable ETL pipelines for 20 markets across 5 EU countries, integrating Azure Data Lake with SQL Warehouse using PySpark and Spark SQL in Databricks
- Automated manual transformation processes, saving 100+ hours annually while reducing human error and cutting resource costs and system downtime by 25%
- Optimized pipelines, reducing query complexity, resource usage, and runtime by 35%, leading to faster analytics and improved data-driven decision-making
- Initiated DevOps standards for project documentation and handovers, reducing onboarding time by 50% and improving cross-team collaboration efficiency
- Conducted a branch-wide workshop series on Generative AI applications with 80+ participants, with over 80% reporting a significant increase in confidence when using AI technologies

Computer Science Assistant Teacher — *X-Camp Academy, Santa Clara, CA* June 2021 - January 2023

- Developed 50+ tailored lessons in Python, C++, and Java, improving student coding proficiency
- Coordinated 100+ staff meetings, managed 200+ timely student data records, and initiated 1-on-1 office hours with weekly progress updates and personalized guidance

PROJECTS

Low-Latency On-Disk Database For Blockchain | Overview | GitHub | *Go, Avalanche, Delve* 2024

- Introduced a novel approach leveraging inherent Merkle trie structure for direct on-disk storage, eliminating industry-wide inefficiencies in conventional blockchain storage systems
- Implemented a free list and diff layer to optimize disk space reuse and support version control rollbacks, reducing data fragmentation by 40% and ensuring 99% crash recovery
- Outperformed Google's LevelDB, achieving 44% faster execution times and 32% lower memory usage across core operations, validated against real-world blockchain workloads/benchmarks
- Designed to enhance and scale any system relying on key-value stores for data persistence

Real-Time Facial Recognition Analyzer | GitHub | *Python, OpenCV, NumPy, Pandas, Scikit-Learn* 2023

- Engineered real-time object permanence algorithms with overlapping edge detection to predict foot traffic trends with 85% accuracy, supporting data-driven visit time recommendations
- Built and trained datasets using OpenCV's Haar Cascade Classifiers and Region-Based CNN (R-CNN), improving detection speed by 30% and reducing false positives by 20%
- Adapted R-CNN-based machine learning algorithms to achieve 95% and 74% accuracy for facial detection and recognition, respectively

Parallelized Web Scraper Tool | GitHub | *Python, HTML, CSS* 2023

- Developed a high-speed data extraction tool using BeautifulSoup, capable of pulling and parsing text, images, and keywords from any given website or HTML file
- Scaled tool to recursively search all sub-directories across multiple website inputs, processing 10,000+ URLs and outputting results in a structured format
- Boosted runtime efficiency by implementing multithreading with concurrent.futures, achieving an 18x speedup (36.8 min to 2.1 min for 1,000 URLs)