Sami Halabieh

925-964-7401 | samisam@berkeley.edu | samisam.me

Education

University of California Berkeley — Berkeley, CA

Bachelor of Arts in Computer Science, Data Science, and Applied Mathematics — *Aug. 2022 – May 2026*

• GPA: 4.0/4.0

Experience

Undergraduate Researcher — January 2025 – Present

Honda Research Institute

- Conducting AGI research in collaboration with UC Berkeley to enhance large language model performance in pattern recognition and problem-solving for the Abstraction and Reasoning Corpus (ARC-AGI).
- Working with a team of four UC Berkeley students to create and train a Test-Time Training Neural Network in a domain-specific language.
- Verifying solutions with Z3 (an SMT solver), resulting in accurate predictions for over 200 ARC puzzles.

Software Engineering (Machine Learning) Intern — Jan 2025 – Present

California Water Resources Control Board — *Berkeley, CA*

- Building an interactive UI (React) with a Node.js/Express backend to handle routing, leveraging NLP and LangChain (LLMs) to interpret user inputs, generate search queries, and retrieve literature on chemical contaminants in wastewater.
- Integrating REST APIs (CrossRef, Springer) and scraping automation tools (Selenium) to retrieve and process academic articles, utilizing PDF extraction tools and storing key information in Firebase and vector databases (Pinecone) in the backend.
- Cross-comparing machine learning models to predict wastewater contaminant levels under extreme climate conditions, achieving 96 percent validation accuracy.

Data Science and Automation Intern — February 2023 - August 2024

Topcon — *Livermore, CA*

- Successfully implemented a CNN model written in Python (Sci-Kit-learn, Pandas, Numpy) and deployed it on a DigitalOcean workflow that accurately predicted part failures during quality control testing. Saving the company over $200,000 in one fiscal year by providing live metrics and actionable insights through an internal dashboard.
- Leveraged Agile methodologies with PostgreSQL, DAX, and Python to iteratively develop interactive Power BI/Tableau dashboards, automate data processing, and conduct statistical modeling for actionable insights.

Projects & Outside Experience

Grid Solver | *Python, PyTorch, Numpy, Pandas, Matplotlib*

- Developed a dual-headed CNN with a shared convolutional backbone.
- Multi-task architecture involves pixel-wise classification with discrete sequence prediction to enhance model generalization.
- Optimized training through custom learning rate schedulers, early stopping, successfully deployed the solution on AWS EC2 for scalability.

Guitar Hero | *Java, Maven, StdAudio, JUnit, Git*

- Developed a Java musical instrument simulator using the Karplus-Strong algorithm to mimic a plucked guitar string.
- Implemented a double-ended queue as a ring buffer that uses noise initialization, iterative averaging, and a decay factor to capture harmonic nuances.
- Integrated an interactive GUI with real-time audio playback (StdAudio).

Technical Skills

Languages: Java, Python, C++, SQL (Postgres), JavaScript, HTML/CSS, GoLang, LaTeX
Technologies: React, Node.js, Flask, JUnit, LangChain, Three.js, Rest API, CrossRef, Springer, Selenium, Firebase, Pinecone, Tensorflow, Pandas, Kubernetes
Developer Tools: Git, Docker, Unix, Azure, VS Code, Visual Studio, PyCharm, IntelliJ, Eclipse
Courses: Deep Learning for Computer Vision, Optimization Models, Data Structures and Algorithms, Object-Oriented Programming, Probability Theory, Abstract Linear Algebra, Putnam