

KAYVAN ZAHIRI

kzahiri@dons.usfca.edu <https://www.linkedin.com/in/kayvan-zahiri/> <https://github.com/Kayvan-Zahiri/>

AWS- and GCP-certified AI Engineer and Software Developer with full-stack, GenAI, and LLM fine-tuning experience.

Experience

AI Engineer

Asurion

Remote

January 2026 - Present

AI Engineer

Spotly Jobs

Remote

October 2025 - January 2026

- Built optimized ATS pipelines, increasing job coverage 5%+ and processing hundreds of postings per run via dlt/dbt.
- Created 7 Dagster assets/sensors to automate ETL workflows/long-run monitors, reducing manual intervention 50%+.
- Increased ingestion throughput by optimizing scraping logic and API handling, improving processing speed by 30–40%.

AI Engineer

Outlier AI

Remote

May 2024 - September 2025

- Trained LLMs using Reinforcement Learning from Human Feedback, improving instruction-following accuracy by 20%.
- Evaluated 1M+ tokens across multiple datasets, surfacing 300+ instruction-following errors.
- Refactored 3+ AI models to generate more precise and succinct output.

Software Developer

University of San Francisco

San Francisco, CA

September 2024 - May 2025

- Lead developer on IDEA team software development, delivering 3 full-cycle applications.
- Created 3 business/entrepreneurship GenAI Python apps, adopted by 50+ students and professors in coursework.
- Conducted research on 10+ existing AI tools and technologies.

Software Developer

Founders Network

San Francisco, CA

August 2024 - December 2024

- Built Slack site integrations into an existing codebase of over 10,000 lines of code.
- Delivered 8 sprint cycles in an Agile environment with a 95% on-time delivery rate.
- Worked with the CEO and a team of professional developers as an intern.

Software Developer

University of San Francisco

San Francisco, CA

May 2024 - August 2024

- Created and deployed 2 generative AI Thinkable apps, one of which was featured on the site's front page.
- Lead developer on a full-stack PeopleMuseum project app and built a codebase of over 5,000 lines of code.
- Used React JS to create front-end UI, achieving load times under 2 seconds with optimized API connections.
- Worked on a team with 5+ graduate students and professors, delivering 100% of milestones on schedule.

Projects

Creator of the PeopleCodeOpenAI Python library

- Allows for easy access and use of OpenAI's tools and API access: <https://www.peoplecode.ai/>
- Contains text-to-speech, voice recognition, RAG, and more
- Being used in intro courses at the University of San Francisco

Developed a full-stack search engine

- Full-stack search engine with inverted index (Java, 1910 lines Javadoc, 1746 SLOC)
- Backend: Java 21, Apache OpenNLP, TF-IDF ranking, Log4j2, multithreading (locks + queues)
- Web crawler: sockets, HTTP, regex, reentrant locks
- Web server: Eclipse Jetty + servlets; Frontend: HTML + Bootstrap
- Collaboration: GitHub PRs, Eclipse IDE, JUnit testing, Maven

Technical Skills

Languages: Python (PySpark, PyTorch, Pandas, NumPy, Matplotlib, Plotly), Java, C, SQL

AI/ML: Generative AI, LLM fine-tuning, Prompt Engineering, Model Evaluation, RAG

Tools: Git, GitHub, Bash, Docker, ETL Pipelines, Distributed Computing, AWS, GCP, MongoDB

Education

University of San Francisco

M.S. in Data Science and Artificial Intelligence

San Francisco, CA

July 2025 - July 2026 (Expected graduation)

University of San Francisco

B.S. in Computer Science

San Francisco, CA

August 2021 - May 2025