

# JAMES GERONIMO

☎ +1 (925) 470-7668 ✉ [jegeronimo@berkeley.edu](mailto:jegeronimo@berkeley.edu) [in linkedin.com/in/james-geronimo](https://www.linkedin.com/in/james-geronimo) [github.com/jegeronimo](https://github.com/jegeronimo)

## EDUCATION

**University of California, Berkeley** Expected Graduation: May 2026  
*Bachelor's in Computer Science, Bachelor's in Data Science* GPA: 3.868/4.0  
**Relevant Coursework:** Data Science, Machine Learning, Artificial Intelligence, Data Engineering\*, AI Agents\*, Inference\*, Data Analytics, Probability Theory, Linear Algebra, Data Structures and Algorithms, Machine Structures, Computer Security  
**Honors:** Dean's List, Honors to Date (4x), Upsilon Pi Epsilon (Computer Science Honors Society)  
**Organizations:** Data Science Society (Vice President, PM), Computer Science Mentors, Berkeley Engineers and Mentors

## EXPERIENCE

**Oracle** September 2025 – Present  
*Data Science Consultant* Berkeley, CA

- Build a full-stack **AI tutor** on **Oracle Cloud Infrastructure (OCI)** (React + Flask/FastAPI) with a RAG pipeline on **Oracle Database** vectors, delivering source-grounded answers and auto-generated MCQs in **<2s P50** latency SLO
- Implement keyword/semantic retrieval using **JSON-relational duality** and **vector data types** for answer relevance
- Construct resilient data layer for **60+ users** with usage stats and conversation logs, enabling insightful dashboards
- Productionize with **CI/CD** (GitHub Actions), **OCI** autoscaling, and observability to sustain **>99%** uptime

**UC Berkeley College of Computing, Data Science, and Society** Jun 2024 – Present  
*Lead Undergraduate Student Instructor* Berkeley, CA

- Teach **1200+ students** in Data 100 ☑, host discussions for **60+ students**, and lead an **infrastructure team of 4**
- Streamline course website ☑ construction with Jekyll development and automate workflows through GitHub Pages

**Amgen** Jun 2025 – Aug 2025  
*Data Science Intern* Thousand Oaks, CA

- Achieve **87%+ accuracy** in text-to-SQL operations by building generative AI and testing frameworks in Databricks
- Optimize Databricks SQL connection and executions by **86%+** in the worst case, reducing operations to **0–80 seconds**
- Deliver a Streamlit application optimized with caching to save **500+ hours** and automate **60,000+ queries** every year
- Leverage large language models (LLMs) to automate campaign analytics across **27 brands** using distributed AI agents

**AmigoAI** Dec 2024 – May 2025  
*Software Engineering Intern* Berkeley, CA

- Devise and build an automated end-to-end testing framework using Playwright to detect regressions in the product's UI
- Engineer a web scraper that emulates user agents, preloads consent cookies, and bypasses JavaScript-rendered content
- Scrape **1000+ data entries** into JSON files to fine-tune custom LLMs for standardized exam question generation
- Construct a cross-format file processor to extract text for a computer vision pipeline that tracks real-time user progress

**UC Berkeley Data Science Modules Team** Feb 2024 – May 2025  
*Software Developer* Berkeley, CA

- Partner with **5 professors** to create **6 educational data science modules** for LS 22 ☑ and El Camino College
- Introduce **500+ students** every semester to fundamental programming, statistics, and data-driven decision-making

## PROJECTS

**Scalable AI Tutor (UberTA)** ☑ | *Python, TypeScript, FastAPI, Azure, React, Next, Tailwind* Jan 2025 – May 2025

- Architect event-driven ingestion pipeline on **asynchronously** chunk/embed documents into **Azure AI Search** vectors
- Build **TypeScript** frontend powered by **FastAPI** backend with **MSSQL** schemas and **RBAC** for multi-tenant courses
- Ship endpoints for chat (conversation history, markdown rendering), document **CRUD**, and admin management separated into services (Edison chat, embedding worker, API) to improve reliability and deployability

## SKILLS

**Languages:** Python, SQL, Java, JavaScript, TypeScript, C, C++, Golang, R, Scheme, x86, RISC-V, HTML, CSS, L<sup>A</sup>T<sub>E</sub>X  
**Data Science/Machine Learning:** Pandas, NumPy, Scikit-Learn, PyTorch, TensorFlow, LangChain, Databricks, Spark, Matplotlib, Seaborn, Tableau, OpenCV, Hugging Face, NLTK, GENSIM, TextBlob, SpaCy, Tesseract-OCR, PyMuPDF/Fitz  
**Software Development:** Git, Docker, React, Next, Tailwind, Jekyll, Azure, REST, FastAPI, Flask, Selenium, Playwright