

JAMES GERONIMO

☎ +1 (925) 470-7668 ✉ jegeronimo@berkeley.edu 🔗 linkedin.com/in/james-geronimo 🐙 github.com/jegeronimo

EDUCATION

University of California, Berkeley

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, Inference*, Data Engineering*, Probability Theory, Linear Algebra, Data Analytics, Data Structures, Efficient Algorithms, Machine Structures, Computer Security, CS/AI Education

Honors: Dean's List (College of Letters & Science), Honors to Date, Upsilon Pi Epsilon (Computer Science Honors Society)

Organizations: Data Science Society (Vice President, Project Manager), Computer Science Mentors, Open Project, Intramural Sports

SKILLS

Languages: Python, SQL, R, Java, JavaScript, TypeScript, C, C++, Golang, Scheme, Ruby, x86, RISC-V, HTML, CSS, LaTeX

Data Science/Machine Learning: Pandas, NumPy, Scikit-Learn, PyTorch, TensorFlow, LangChain, Databricks, Spark, Matplotlib,

Seaborn, Tableau, Power BI, OpenCV, Hugging Face, NLTK, GENSIM, TextBlob, SpaCy, Pix2Text, Tesseract-OCR, PyMuPDF/Fitz

Software Development: Git, Docker, React, Next, Tailwind, Jekyll, Azure, Heroku, FastAPI, Flask, Selenium, Playwright, OAuth 2.0

EXPERIENCE

Amgen

Jun 2025 – Present

Data Science Intern

Remote

- Use large language models (LLMs) and generative AI to create a tool that automates internal processes related to commercial sales
- Deploy an interface that incorporates **10+** data sources providing a comprehensive outlook delivered through text and speech

AmigoAI

Dec 2024 – May 2025

Machine Learning Intern

Remote

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

UC Berkeley College of Computing, Data Science, and Society

Jun 2024 – May 2025

Lead Undergraduate Student Instructor (uGSI)

Berkeley, CA

- Teach **1200+** students in the upper-division course Data 100 ☞, covering data processing, statistical/probabilistic foundations, exploratory data analysis/visualization, feature engineering, dimensionality reduction, predictive modeling, and optimization
- Streamline course website ☞ navigation by developing in Jekyll and automating deployment workflows through GitHub Pages
- Lead discussion sections to **40+** students, host office hours, debug DataHub issues, and maintain general course infrastructure

UC Berkeley Data Science Modules Team

Feb 2024 – May 2025

Software Developer

Berkeley, CA

- Partner with four El Camino College professors to create educational data science modules teaching probability distributions, data visualization, classification, statistical modeling, and business analytics under the Data Science Modules ☞ team
- Design and construct two data science modules deployed through Jupyter for the undergraduate interdisciplinary course LS 22 ☞
- Introduce **500+** students every semester to fundamental programming practices, statistics, and data-driven decision-making

University of California, San Francisco

Feb 2025 – Apr 2025

Research Assistant

Remote

- Evaluate LLM robustness via conversational red-teaming and multi-turn jailbreaking techniques under the Abbasi Lab ☞
- Build an agentic LangChain framework to automate multi-turn LLM conversations and LLM-as-a-judge evaluations with GPT-4o
- Customize and apply **10+** prompting goals and target strings from the MedSafetyBench ☞ dataset for attacker and target LLMs

PROJECTS

Scalable AI Tutor System (UberTA) ☞ | Python, TypeScript; FastAPI, Azure, React, Next, Tailwind

Jan 2025 – May 2025

- Engineered a scalable backend with an event-driven Retrieval-Augmented Generation (RAG) pipeline for asynchronous processing
- Automated document processing using a custom embedding pipeline to eliminate manual maintenance and standardize workflow
- Developed an end-to-end AI tutor with user authentication, student chat interface, and instructor portal hosted on Azure

Optimizing Kickstarter Campaigns ☞ | Python; Scikit-Learn, NLTK, TextBlob

Nov 2024 – Dec 2024

- Analyzed **370,000+** Kickstarter campaigns using TF-IDF vectorization, sentiment analysis, and classification modeling
- Obtained **70.1%** cross-validation accuracy with a fine-tuned Random Forest classifier, further evaluating performances through F1-scores and confusion matrices against Logistic Regression, Decision Tree, and Neural Network models

GCN and LLM Directory ☞ | Python, JavaScript; PyTorch, Flask, Heroku, React

Feb 2024 – Apr 2024

- Created an AI-assisted directory for **100+** members with a Python/Flask backend, React frontend, and Heroku deployment
- Employed a Graph Convolutional Network (GCN), achieving a **0.72** ROC AUC score for link classification accuracy
- Incorporated Google OAuth 2.0, fetched from an AirTable database, and generated contextualized responses from GPT-3.5 Turbo