

AADI MALAVIYA

520-470-5572 | aadim@ucla.edu | www.linkedin.com/in/aadi-mala | github.com/Aad-im | US Citizen

EDUCATION

University of California, Los Angeles

September 2023 – June 2026

BS in Computer Science and Mathematics

Coursework: Data Structures and Algorithms, Computer Systems, Database Systems, Machine Learning, Deep Learning, Artificial Intelligence, Linear Algebra, Optimization, Probability and Statistics, Applied Numerical Methods

EXPERIENCE

Ryniant (AI Drug-Synthesis Platform)

May 2026 – July 2026

Founding Engineer

Phoenix, AZ

- Built the core engine that figures out how to synthesize a target drug molecule, cutting planning time from hours to minutes.
- Replaced LLM-guessed reactions with automated PubChem and FAISS verification, ensuring 100% of results are literature-verified.
- Built a Claude (LangGraph) agent that proposes and fact-checks novel routes, expanding discovery by 25% with zero validity loss.
- Shipped the full product on Cloud Run (React/TypeScript, FastAPI, Redis, PostgreSQL) with a 7-day cache, cutting latency to near-zero.
- Built an internal admin interface with Matplotlib visualizations, giving the team visibility into job status and route rankings.
- Used Celery task queues to process long-running route-planning jobs asynchronously, deployed on Google Cloud for reliability.

UCLA NLP Lab (LLM Interpretability)

September 2025 – May 2026

Machine Learning Researcher

Los Angeles, CA

- Ran probing and layer-ablation experiments across GPT-2, Llama, Pythia, and Gemma, narrowing reasoning signal to 6 of 32 layers.
- Applied activation patching and causal tracing to isolate 3 attention heads causally responsible for hallucinated versus grounded outputs.
- Trained sparse autoencoders on model activations to extract 16 interpretable features tied to reasoning and factual grounding.
- Designed steering vector interventions on those features, cutting the hallucination rate by 30% across controlled benchmark tests.

Gender Capital Management (Wealth Management Firm)

June 2025 – September 2025

Data Analyst Intern

Los Angeles, CA

- Automated portfolio reporting for clients with a Python pipeline (Schwab API, Pandas, Plotly), cutting report time by 3+ hours weekly.
- Built a Schwab scraping and RMD engine for 200+ accounts by age, balance, and IRA type, saving ~40 hours per season.
- Built an automated email system for 200 clients' required withdrawals, cutting turnaround to minutes and saving ~25 hours per cycle.

Chelle AI (LLM-Based Learning Platform)

January 2025 – June 2025

Machine Learning Engineer

Los Angeles, CA

- Designed a rubric-based framework (LangChain, OpenAI, Gemini APIs) to score LLM-generated study guides on clarity and accuracy.
- Used the rubric to compare prompting strategies, identifying top configurations and cutting hallucination rate by 25%.
- Built automated evaluation loops in the FastAPI backend to re-score prompts on every change, lifting engagement scores 18%.

Sentiment Systems (Health Clinic Startup)

August 2022 – August 2023

Backend Engineer

Phoenix, AZ

- Designed a Django dashboard with role-based access for clinical staff, cutting reporting time by 30% via Plotly charts.
- Built an NLP sentiment pipeline (spaCy, NLTK, PyTorch) to classify feedback and surface topic-level trends across clinics.
- Engineered ETL pipelines (Pandas, SQL) and REST endpoints, then shipped a Dockerized AWS deployment for clinic onboarding.

TECHNICAL SKILLS

Technologies: Python, PyTorch, Pandas, Django, PostgreSQL, Scikit-learn

AI / LLM: vLLM, LoRA, LangChain, LangGraph, Hugging Face Transformers, NNSight, RAG

Cloud & DevOps: AWS S3, Docker, GitHub Actions, Google Cloud Run, Redis, CI/CD