

Arsum Nadeem Chaudhary

Greater LA Area | arsumnadeemch16203@gmail.com | 8402056625 | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

EDUCATION

Pomona College, Claremont, CA

Expected Graduation: Jun 2025

Double Major of Bachelors of Arts in Computer Science and Mathematics | 3.8 GPA

Relevant Coursework: Data Structures, Algorithms, Big Data, Managing Complex Systems, Artificial Intelligence, Neural Nets, Computer Languages and Theory, Advanced Linear Algebra, Operations Research, Biostatistics, Probability, Combinatorics, Vector Calculus

SKILLS

Languages Python · SQL · Javascript · Java · C/C++ · R

Cloud & DevOps AWS (EC2, S3) · Docker · Kubernetes · GitHub Actions · Google Cloud Functions · Depot

ML & Frameworks Flask · React · OpenCV · TensorFlow · Pandas · Robotic Process Automation · Playwright

Security & Data Data pipelines (Snowflake, BigQuery) · Logging/Observability (NewRelic, PostHog, PagerDuty) · Security/Compliance (SOC 2)

EXPERIENCE

AI Engineer

July 2025 - Present

[Pyq AI](#)

San Francisco, CA

- Built Docker-containerized **Python microservices**—orchestrated via **Kubernetes** and Depot—on GCP Cloud Functions and AWS S3 to streamline insurance back-office data ingestion and processing
- **Automated ~90%** of policy onboarding and document validation for **GEICO** and **Berkshire Hathaway Guard** using RPA pipelines (Playwright, AI CodeGen) against REST APIs, cutting manual steps and errors

AI Engineering Resident

Sep 2024 - Dec 2024

[Headstarter](#)

Remote

- Engineered end-to-end AI/ML services (Python backends & TypeScript frontends), deployed on AWS Lambda, S3, and Vercel with automated CI/CD via GitHub Actions
- Orchestrated LLM-chaining pipelines and hyperparameter-tuned 10+ models to hit sub-100 ms inference latency and >90% accuracy, integrating real-time monitoring for performance and error tracking
- Collaborated in 7-day agile sprints with ML, Kubernetes, and security engineers—embedding compliance checks and best practices to deliver robust, enterprise-grade solutions

Software Engineering Intern

Sep 2023 - Dec 2023

[Learning Economy Foundation](#)

Washington, DC

- Designed & implemented **ChatGPT plugin** architecture for LearnCard - LEF's novel all-purpose Web 3 education wallet
- Enhanced LearnCard's Boost Feature with **automated** previews using OpenAI function calling empowering 5k+ daily users and reducing processing time by **60%**
- Optimized verification speed and scalability to support real-time **credential access** for rapidly growing user base

Software Engineering Intern

Jun 2023 - Aug 2023

[Inabia Software and Consulting](#)

Redmond, WA

- Developed 3 compliant ETL pipelines ingesting & normalizing 1 TB of health data, boosting processing efficiency by 20%
- Improved text classification accuracy by 15% and reduced translation error by 20% on 500K client inquiries
- Enhanced data governance and reliability, enabling clients to make informed, compliant health-care decisions

SELECTED PROJECTS

Brain Tumor Classification | Open-Source Project - [GitHub](#)

Jan 2025

- Designed a high-throughput data pipeline to ingest and preprocess 1k+ MRI scans, feeding an Xception + custom CNN ensemble for 3-way tumor classification with <200 ms per image latency
- Produced shareable saliency-map reports in Jupyter in <200 ms, demonstrating end-to-end data ingestion / visualization workflows

Twitter Clone Web Application | Personal Project - [GitHub](#)

May 2024

- Architected a cloud-ready Flask/PostgreSQL service with sharded, indexed tables supporting 1M+ records—showcasing data modeling & performance tuning at scale
- Containerized and deployed with Docker, Nginx & Gunicorn on AWS/GCP; implemented CI/CD (GitHub Actions), RBAC-style authentication, full-text search and observability hooks for real-time performance insights

WebOps VM Management | Team Project (5 engineers) - [GitHub](#)

Jan 2024

- Hardened AWS EC2 (TLS/SSL) hosting PostgreSQL/Tomcat & JVM; integrated New Relic & PagerDuty for SLI/SLO alerts
- Orchestrated auto-scaling via AWS ELB and EFS, achieving <1% error rate at 10k RPM; built automated health-check tests to validate failover and maintain <500 ms response under load
- Gained skills in **error diagnosis** from complex input information, 24/7 availability scheduling, and meticulous note keeping

P-Fraud, P-ai, Claremont Colleges AI Club | Team Project - [GitHub](#)

Apr 2023

- Engineered an ETL pipeline to normalize 500K financial records for SVM, Random Forest & XGBoost models—attaining 94% fraud-detection accuracy while ensuring data integrity and reproducibility

LEADERSHIP & ACTIVITIES

Personal Portfolio | Public Landing Page - [GitHub](#)

Feb 2025

- Wrote 11k lines of HTML, CSS, Javascript to showcase my projects, experience, and skills, optimized for UI/UX, SEO and SEM

UCLA Datafest | Team Project (4 data scientists)

Nov 2024

- Led team of 6 to present **policy recommendations** for enhancing customer satisfaction in pro bono case analyzing with R and Pandas around 450k+ data points from the American Bar Association

Coronavirus Twitter Analysis Project | Personal Project - [GitHub](#)

Feb 2024

- Analyzed 1.1b tweets in parallel to identify **sentiment trends** in 10+ languages across 20+ countries to create 4 critical visuals to comprehend pandemic reactions to inform better response for global health communication