ARMAAN GULATI

New York, NY · 510-514-3075 · gulatiarmaan@gmail.com · linkedin.com/in/armaan-gulati · github.com/armaangulati1

Objective: Applied-AI / forward-deployed engineer seeking to build LLM-powered products in healthcare. Biomedical engineer who designs and ships agentic systems end to end on the Claude and OpenAI APIs, from clinical data extraction to full-stack production deployment.

SKILLS

AI / ML: LLM system design & application development (Claude, OpenAI & Google Gemini APIs), agentic systems & orchestration (tool-calling, multi-step router → validator → verifier pipelines), Model Context Protocol (MCP), agent identity & auth validation, RAG (retrieval-augmented generation, reranking, RAGAS), context engineering, structured/constrained generation & schema validation (Pydantic/JSON), evaluation harnesses (precision/recall/F1, LLM-as-judge, CI eval gates), prompt engineering, multimodal AI, vector embeddings & semantic search (Chroma)
Backend & Full-Stack: FastAPI, Streamlit, REST APIs & third-party API/webhook integrations, Docker, CI/CD (GitHub Actions), testing (pytest), observability (Langfuse), Git/GitHub, Poetry, Jupyter; AI-assisted / live coding
Languages: Python (Pandas, NumPy, Matplotlib), Java, SQL, MATLAB
Healthcare: FHIR (resource modeling & validation, live REST API, batch/transaction bundles, Synthea, local HAPI server), clinical NLP / chart abstraction, prior-authorization / RCM workflows, PHI redaction & de-identification
Data & Tools: Data cleaning & preprocessing, statistical analysis, data visualization, Excel (pivots), Figma

EXPERIENCE

AI / Software Developer (Part-Time), *Trust.care* May 2024 – Present

- Architected a full-stack, AI-powered referral platform (**FastAPI, Streamlit, LangChain, Chroma**) connecting healthcare providers with community-based organizations (CBOs) for Medicaid Marketplace referrals; built ground-up as a two-person engineering effort, now processing 5.12M+ outreach messages to a ~955K-member Medicaid cohort across 100+ CBOs.
- Designed the LLM system architecture: orchestrated **Claude, OpenAI, and Google Gemini models via LangChain** over a Chroma vector store, enabling multimodal (text + document/image) semantic search, plus a reply-classification pipeline that analyzed 118K+ inbound member replies by intent, separating high-signal responses from noise.
- Built autonomous, **tool-calling agents with an agent identity & authentication layer**, giving each agent a verifiable identity to securely validate and act on people-initiated operations, addressing a core trust challenge in agent-based systems.
- Drove the platform from concept to production: 19.2% cohort engagement, ~12.8x the published Medicaid SMS benchmark, contributing to 35K+ reported Medicaid enrollments and re-enrollments.

Data Analyst Intern, *Sycamore Informatics* May 2024 – Aug 2024

- Investigated pharmaceutical fraud by consolidating 4-5 raw financial extracts and building chunked pandas pipelines over larger-than-memory datasets, surfacing anomalous activity patterns for internal review.
- Applied statistical methods and data visualization to make fraud signals interpretable for non-technical stakeholders, partnering cross-functionally to strengthen data integrity.

Technical Intern, *Harmony Health* May 2023 – Aug 2023

- Built predictive models (campaign response, no-show risk, enrollment likelihood) from healthcare campaign data; the no-show model was used to prioritize outreach lists, outperforming the prior targeting approach.

PROJECTS

Clinical Prior-Authorization Agent, deployed MCP-powered agentic system 2026

- Built and deployed an end-to-end agentic system automating a clinical prior-authorization workflow: authored two **Model Context Protocol (MCP) servers** (read + action) and an LLM agent that reads structured clinical facts from a **live FHIR server (HAPI/Synthea)**, fuses them with free-text note extraction (with provenance), checks payer policy, and decides (submit / request-more-info / deny-risk), gated by a **human-in-the-loop approval step** and a deterministic guardrail.
- Engineered a rigorous **eval harness** (**93.7% macro-F1 on a held-out set**, per-class precision/recall/F1, trajectory + LLM-as-judge with judge validation, CI regression gate); added PHI redaction, prompt-injection defense, retries/idempotency, and **deployed the MCP server live** (Docker, Fly.io, StreamableHTTP, token auth) (GitHub).

ChartExtractor, agentic clinical data-extraction platform 2026

- Built an **agentic pipeline** (router → extractors → validator → verifier) extracting structured oncology variables from clinical notes into schema-validated JSON; **93.6% macro-F1** against a ≥85% CI deploy gate, with a per-field eval harness, containerized deploy, and a live demo + API (demo · GitHub).

EDUCATION

Washington University in St. Louis May 2026

- Bachelor of Engineering in Biomedical Engineering. Partial scholarship, College of Arts & Sciences.
- WashU's undergraduate Biomedical Engineering program is ranked #6 in the nation (U.S. News 2026); the university ranks #20 among National Universities (McKelvey School of Engineering).
- Relevant coursework: Optimization, Statistics, Engineering Mathematics, Differential Equations, Signal & Image Processing, Quantitative Physiology.
- High school: 4.65 GPA; ranked #38 in California.