Chaitya Shah

cs4621@columbia.edu | (646) 271-6316 | shahchaitya.com | linkedin.com/in/chaityas | github.com/Chaitya0623

EDUCATION

Columbia University - Fu Foundation School of Engineering and Applied Science New York, NY
Master of Science in Data Science | GPA: 3.83/4 Expected Dec 2026
University of Mumbai - Dwarkadas J. Sanghvi College of Engineering Mumbai, India
Bachelor of Technology in Computer Engineering | GPA: 8.76/10 Dec 2021 - May 2025

EXPERIENCE

SimPPL New York, NY
Founding AI Engineer Aug 2023 - May 2025

- Spearheaded a team of 4 to launch an AI-powered audience analytics MVP, scaling it to a pilot with grants from **Mozilla and Google (exploreCSR)**, helping **New York Public Radio** boost engagement for **20K+ users**.
- Deployed a **full-stack production system** with a Next.js, Flask API's, PostgreSQL and Redis data layers, containerized with Docker and served via Nginx on **AWS**, **processing 1M+ user interactions** to drive real-time content recommendations.
- Engineered an **AI Agent** using LangGraph to analyze **120k+ social media posts** for our **UNICEF-incubated tech venture**.
- Productionized predictive backend services for persona and trend based click forecasting, **boosting CTR by 8%**.

Avrio Energy Berlin, Germany
Data Science Intern Jun 2024 - Jul 2024

- Optimized energy usage across **70+ McDonald's outlets** in 22 countries by engineering a data pipeline to track appliance level consumption and visualize real-time insights via Power BI, **reducing operational costs by 15%**.
- Automated real-time fault detection with Airflow alerts and a **92% accurate XgBoost model**, cutting energy waste by 21%.

Indian Institute of Technology Patna Patna, India
Research Intern May 2023 - Nov 2023

- Designed a fraud detection system on **250k+ Ethereum transactions** using Graph Neural Networks (Temporal GATs) and Directed Random Walks to classify fraudulent nodes based on behavioral patterns, achieving **85% accuracy**.

Infiheal Mumbai, India
Artificial Intelligence Intern Jun 2023 - Sep 2023

- Engineered a **RAG-based mental health chatbot** with LangChain & Haystack, **fine-tuned LLaMA** on 500k+ user chat transcripts to handle sensitive user queries, reducing hallucinations and **increasing user retention by 12%**.
- Deployed scalable ML architecture on AWS with Terraform & Docker, reducing manual setup time by **5+ hrs/week**.

LEADERSHIP

DJS Antariksh Mumbai, India
Coding Head May 2022 - Oct 2024

- **Represented India in 6 International Competitions**, leading a 30-member coding department. Developed autonomous systems for space robotics, finishing 3rd at International Rover Challenge 2024 and 2nd at European Rover Challenge 2023.
- Trained junior members, co-authored system design and technical documentation to support competition qualification.

PROJECTS

Dynamic KV Cache Compression for Reasoning in LLMs

- Applied Compressed Context Memory for chain-of-thought reasoning by training a **token-level classifier** with LoRA adapted LLaMA-3.1-8B, using GPT-4o-mini knowledge distillation to dynamically compress GPU KV caches.
- Achieved **1.60× memory compression** with 32% reasoning accuracy (vs 26% baseline) and <0.4% inference overhead.

AI Agent for Multi-Step Smart Contract Vulnerability Discovery | Research Assistant

- Developed an LLM-agent system leveraging 6 models to detect and repair **sequential exploit vulnerabilities** in smart contracts, modeling complex attack chains and validating on **SCONE-Bench (Anthropic)** and real-world contracts.

SKILLS

Programming & Tools: C/C++, Python, Django, JavaScript, SQL, R, MATLAB, LangChain, LangGraph, FastAPI
Machine Learning: Numpy, Pandas, Matplotlib, Scikit-Learn, Keras, TensorFlow, PyTorch, OpenCV, NLTK, CUDA
Databases, Analytics & Cloud: PostgreSQL, Redis, Power BI, BigQuery, GCP, AWS, Docker, Linux, Nginx