

# 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

Master of Science in Data Science

New York, NY

Aug 2025 - Dec 2026

### University of Mumbai - Dwarkadas J. Sanghvi College of Engineering

Bachelor of Technology in Computer Engineering | GPA: 8.76/10

Mumbai, India

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.
- Engineered an **AI Agent** using LangGraph for our **UNICEF-incubated tech venture** to analyze 120k+ social media posts from Reddit, Telegram and YouTube delivering actionable insights on trending topics to stakeholders.
- Designed a predictive model that forecasts user clicks based on personas & trending topics, **boosting CTR by 8%**.
- Deployed full-stack AI pipelines on AWS, **processing 1M+ user interactions** to drive real-time content recommendations.

### 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 20%.

### 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

### Columbia University | Teaching Assistant - Natural Language Processing (COMS 4705)

Sep 2025 - Present

- Mentored 8 graduate project teams (172 students), guiding research on LLM fine-tuning, RAG, and model interpretability.
- Led weekly office hours and assisted with assignments and grading for transformer architectures, pretraining, and RL.

### DJS Antariksh | 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.

## PUBLICATIONS

- Chaitya Shah, et al. "Can Social Media Platforms Transcend Political Labels? An Analysis of Neutral Conservations on Truth Social". International AAAI Conference on Web and Social Media 2024. <https://doi.org/10.36190/2024.17>
- C Shah, et al. "Sentence Restructuring with User-Controlled Difficulty using NLP". International Conference on Computing, Communication and Networking Technologies 2023, IEEE. <https://doi.org/10.1109/ICCCNT56998.2023.10307165>

## PROJECTS

### GenAI-Powered Financial Advisory Platform

- Analyzed 100K+ financial data points to generate personalized savings insights; fine-tuned GPT-Turbo and integrated a FAISS vector database for document retrieval, **improving recommendation accuracy by 16%**.

### Economic Policy Impact Analyzer

- Built an end-to-end ML system to assess the effect of Indian government policies on stock volatility by processing 10k+ news articles using NLP and an LSTM model, achieving **87% accuracy** in identifying policy-driven market movements.
- Estimated a statistically significant causal impact of 0.15 on volatility ( $p = 0.04$ ) using **Causal Machine Learning**.

## SKILLS

**Programming & Web:** C/C++, Python, Java, SQL, R, MATLAB, JavaScript, React, Typescript, Django, Flask, FastAPI

**Machine Learning:** Numpy, Pandas, Matplotlib, Scikit-Learn, Keras, LangChain, TensorFlow, PyTorch, OpenCV, NLTK

**Databases, Analytics & Cloud:** PostgreSQL, Redis, Power BI, BigQuery, GCP, AWS, Docker, Linux, Nginx