Mehul Mathur

872-218-2302 | mmath34@uic.edu | linkedin.com/in/mehul | github.com/Mehul1604

Education

University of Illinois at Chicago — Chicago, IL
Masters of Science in Computer Science — *Aug. 2024 – May 2026*

International Institute of Information Technology — Hyderabad, India
Bachelor of Technology in Computer Science — *Aug. 2019 – April 2023*

Relevant Coursework: *Data Science, Statistical Methods in AI, Principles of Concurrent Programming, Compiler Design*

Technical Skills

Languages: Scala, Java, Python, SQL, C/C++, JavaScript, Linux Shell Scripting
Frameworks/Database: Spring Boot, Flask, FastAPI, Express, Akka, React, SQLite, MongoDB, MySQL, PostgreSQL, NoSQL
Libraries/Tools: Pandas, PyTorch, Tensorflow, Matplotlib, Airflow, Kafka, Git, BitBucket, CircleCI, Kubernetes, AWS, GCP
Skills: API development, Microservices, Scalability, Performance optimization, Leadership, Time management

Experience

Paypal — May 2025 – August 2025
Software Engineer Intern

- Built an AI-powered Slackbot in **Python** that automated query resolution in support channels by leveraging **LangGraph multi-agent** workflows for log investigation and **Milvus vector search** for duplicate detection, reducing response time from 2–3 hours to under 5 minutes and saving engineers 5–10 hours weekly by preventing unnecessary JIRA tickets.
- Developed an **MCP** server in **Spring Boot (Java)**, integrated with **Claude Code** and **GitHub Copilot**, enabling engineers to investigate live issues directly within their workspace and reducing root cause analysis time by 40%.

Endowus — July 2023 – June 2024
Software Development Engineer

- Optimized a **Scala**-based trade file processing backend using **Akka** entities for reactive task handling, **cluster sharding** for parallelism, and **Kafka** for idempotency—reducing processing time from 10 to 4 minutes and boosting throughput by 60%.
- Designed an anomaly detection system using **Spring Boot**, connecting to **PostgreSQL** via **JPA** to query inconsistent transactions and sending real-time Slack alerts—eliminating manual checks and detecting 50+ critical discrepancies monthly.
- Developed an **Airflow** DAG in **Python** to automate daily trade file generation at non-uniform times, using XComs state management for dynamic scheduling—unifying scattered workflows to reduce clutter and cutting maintenance time by 40%.

Zevi.AI — March 2022 – August 2022
Artificial Intelligence Engineer Intern

- Engineered a search engine by fine-tuning an LLM in **PyTorch** on augmented product data stored in **AWS S3** for efficient scaling and deploying it on **EC2** using **Transformers** for a lightweight model with inference time less than 50 milliseconds.
- Crafted a phonetic-based spellcheck algorithm in **Python** using **NLTK** and **Fuzzy** matching, reducing unrecognized multilingual queries to under 2% and enhancing product discovery for regional name variations.

Projects

Multi-Agent Natural-Language Analytics | *LangGraph, OpenAI, Python, Flask, PostgreSQL*

- Built a multi-agent analytics visualizer that turned plain-English questions into verified SQL on Postgres and automatically generated Plotly graphs and PDF reports via coordinated tools and agents—cutting time-to-insight from hours to minutes.
- Productionized a Flask web app with a chat UI and REST endpoints to run the end-to-end multi-agent workflow-persisting artifacts and delivering reliable results, reducing multi-tool handoffs to a single request.

Personalized WhatsApp Chatbot using LSTM | *LSTM, NLP, PyTorch, Text Generation*

- Engineered an LSTM-based chatbot trained on personal WhatsApp group chats, preprocessing conversation history to learn unique texting styles, slang, and informal patterns.
- Achieved realistic, personalized responses while optimizing for low-resource, local training, eliminating the need for expensive cloud computing and large-scale transformer models.

RAG Builder — Distributed Semantic Search | *Scala, Hadoop, AWS, S3, EC2, Map/Reduce*

- Deployed REST RAG API on AWS EC2 (Scala/Tapir, Lucene-sharded indices, Ollama) performing parallel fan-out/fan-in search across distributed shards to deliver LLM-grounded answers from 1,000 research papers via top-K retrieval.
- Built a Hadoop MapReduce pipeline on AWS EMR to compute token stats, embeddings, and nearest neighbors in parallel, cutting brute-force $O(n^2)$ runs from days to hours (95%+ speedup).

Achievements/Roles

Teaching Assistant in Courses: Data and Applications, Statistical Methods in AI, Computer Graphics
Awards: Deans List, Merits List, College Hackathon 2022 Runner Up
Publications: "TweetBoost: Influence of Social Media on NFT Valuation", FinWeb, March 2022,
Co-WIN: Really Winning? Analysing Inequity in India's Vaccination Response