Mokshit Surana

Email LinkedIn GitHub Scholar +1 (331) 269-4898

Education

University of Illinois at Chicago **August 2024 – May 2026**
Masters of Science in Computer Science (Thesis) *GPA: 3.73/4*

University of Mumbai (Thadomal Shahani Engineering College) **August 2020 – June 2024**
Bachelor of Engineering in Information Technology *CGPA: 9.43/10*

Experience

Machine Learning Engineer **February 2023 – June 2024**
SimPPL *Remote*

- Built a scalable cross-platform data pipeline processing **80M+ comments and 500K+ videos** from YouTube and Truth Social, using chunked Pandas processing and batched GPU inference to run large-scale model evaluation under tight memory constraints.
- Engineered an automated LLM evaluation harness scoring LLaMA and Mistral 7B outputs across 6 toxicity dimensions via the Google Perspective API, benchmarking model behavior against human baselines derived from 460K+ messages.
- Developed a labeled taxonomy of 40 adversarial prompt patterns (jailbreak and persuasion techniques) to systematically stress-test open-source LLM safety and reproducibility.
- Conducted the research in collaboration with Prof. Deb Donig (UC Berkeley), contributing methodology and experimental validation.

Graduate Teaching Assistant **January 2025 – May 2026**
University of Illinois at Chicago *Chicago, USA*

- Co-authored a research paper on LLM-generated metadata enrichment for RAG systems, accepted at **IEEE CAI 2026**; led experimental validation and benchmarking.
- Designed and deployed RAG infrastructure for student research teams, Graph-RAG pipelines on Neo4j and production vector stores (ChromaDB, FAISS), plus clinical NLP workflows using ClinicalBERT for ICU readmission prediction.
- Mentored 9+ graduate project groups on production vector search, retrieval evaluation, and multi-agent LLM design patterns.

Technology Intern **June 2023 – July 2023**
Deutsche Bank *Pune, India*

- Cut manual processing time by **96.7%** by replacing a manual workflow with an automated CSV validation engine built on Spring Boot and Oracle DB.
- Built a Spring Boot PDF-generation microservice using iText, applying OOP design principles within an Agile delivery team.

Publications / Research

- **A Systematic Framework for Enterprise Knowledge Retrieval: Leveraging LLM-Generated Metadata to Enhance RAG Systems.** *IEEE Conference on Artificial Intelligence (IEEE CAI)* 2026.
- **Multi-Agent Simulators for Social Networks.** *Multi-Agent Security Workshop, NeurIPS,* 2023.
- **Examining the Implications of Deepfakes for Election Integrity.** *AI for Credible Elections Workshop, AAAI,* 2024.
- **Leveraging CNNs and Ensemble Learning for Automated Disaster Image Classification.** *Springer ICSISCET,* 2023.

Projects

Ecosystem Change Intelligence (ECI) | *LangGraph, LangChain, ChromaDB, pgvector, Next.js* **GitHub**

- Built an end-to-end multi-agent system with 5 specialized agents (Ingestion, Delta Detection, Graph-RAG Builder, Sentinel, Coordinator) orchestrated via a LangGraph supervisor pattern to monitor 10+ feeds and emit evidence-backed action tickets.
- Designed a DeltaRAG retrieval architecture fusing dense vector search with a NetworkX knowledge graph via Reciprocal Rank Fusion, reaching **0.93 MRR** and **88% P@1** on a 100-query benchmark, on par with a tuned RAG baseline while adding graph-grounded, traceable evidence for each ticket; served through a Next.js dashboard.

UIC Labor Docs RAG Chatbot | *Python, Streamlit, Neo4j, LangChain, RAGAS* **GitHub**

- Built a retrieval-augmented QA system over UIC labor documents using a hybrid vector + graph retrieval pipeline (Neo4j knowledge graph and ChromaDB semantic search).
- Implemented a RAGAS evaluation harness measuring faithfulness, answer relevance, and context precision; shipped as a Streamlit web application.

Technical Skills

Languages: Python, Java, SQL, C++
ML / AI: PyTorch, TensorFlow, HuggingFace, Scikit-Learn, Pandas, NumPy
LLM / Agents / RAG: LangChain, LangGraph, Graph-RAG, Multi-Agent Systems, RAGAS
Vector & Graph DBs: ChromaDB, pgvector, FAISS, Qdrant, Pinecone, Neo4j, NetworkX
Cloud & MLOps: GCP (Vertex AI, BigQuery), AWS (Lambda, EC2, S3), Docker, Git
Backend & Frontend: FastAPI, Spring Boot, Next.js, React.js, REST APIs, Streamlit

Leadership / Extracurricular

- **Competitive Programming**: LeetCode Knight Badge (top 5% in contests); 1st place in TSEC Weekly Coding Challenges '23.
- **Hackathons & Clubs**: Won 5 hackathons; Core Team Member at Google Developer Student Club; Student Mentor.