

Sooraj Kakkanatt Manoj

✉ soorajmanoj@gmail.com ☎ +1 585 540 0455 in Sooraj Kakkanatt Manoj 🌐 soorajmanoj 📄 soorajmanoj.github.io

Skills

Programming & Databases: Python, SQL, R, Java, JavaScript, PostgreSQL, MySQL, Snowflake, MongoDB
Machine Learning & Deep Learning: PyTorch, TensorFlow, Keras, LangChain, CUDA, NLP, HuggingFace
MLOps, Cloud & Deployment: AWS, Azure ML, Git, Docker, Kubernetes, FastAPI, Flask
Data Engineering & Visualization: Apache Spark, Apache Airflow, Hadoop, Tableau, Power BI, Jupyter, QGIS, SPSS

Experience

- Graduate Teaching Assistant - Department of Computer Science** Rochester, NY
Rochester Institute of Technology Aug 2025 – Apr 2026
- Mentored 25 students per semester in advanced server programming, backend system logic, and automation concepts, grading 750+ weekly script and data-pipeline assignments across 2 semesters with code reviews.
 - Held 120+ hours of office hours across 2 semesters, advising 60–90 students on debugging complex programming errors and bridging theoretical concepts with production-ready code.
- Intern: Analysis & Design of Commissions for Gig Workers** Chennai, India
One Integral Technologies Pvt. Ltd. Jan 2022 – Jul 2024
- Designed and optimized performance-based pay models using Digital Twin simulations across 50+ data points, improving worker engagement by 15%.
 - Developed a React.js analytics dashboard that turned performance data from 10+ sources into visualizations covering 500+ gigs, reducing manual analysis time.
 - Partnered with the engineering team to integrate the dashboard with the backend Digital Twin system via MongoDB, feeding a live pipeline for predictive incentive analysis.

Projects

- Bias in AI Travel Narratives: Indic vs. Western LLMs** [🔗](#) May 2026
- Investigated representational bias in AI-generated travel narratives by architecting a modular 5-layer data processing pipeline to ingest and normalize 600,000 user comments.
 - Implemented an automated relevance classifier using the Gemini API to extract valid contextual records, discarding 300,000 irrelevant entries.
 - Orchestrated a fault-tolerant batch generation workflow across 4 Large Language Models (LLaMA32, Gemma2, GPT2, and Sarvam) to produce counterspeech responses addressing culturally biased travel narratives.
- AI-Powered YouTube Q&A Assistant** [🔗](#) Oct 2025
- Built a fully local RAG pipeline for YouTube Q&A using FastAPI, LangChain, and Ollama (llama3 + nomic-embed-text), indexing transcripts in under 2 seconds and answering questions in 8–12 seconds.
 - Implemented session-scoped semantic search with FAISS vector retrieval and per-video chat memory, served through a single-page UI directly from FastAPI.
 - Added a pytest suite (13 tests, 71% code coverage) covering request validation, video-URL parsing, and session handling, with transcript/embedding/LLM calls mocked for CI reliability.
- Dockerized MLOps API with FastAPI** [🔗](#) Sep 2025
- Containerized a TF-IDF + Logistic Regression sentiment classifier behind a FastAPI service, with Pydantic-validated request/response schemas and auto-generated OpenAPI docs.
 - Trained and evaluated the model on a synthetic 800-review dataset, validating generalization on held-out novel sentences.
 - Set up CI via GitHub Actions to run the pytest suite and validate the Docker build on every push/PR.
- Kafka-Based Credit Card Transaction Processing System** [🔗](#) Apr 2025
- Engineered a real-time + batch transaction pipeline on a Kafka-based Lambda Architecture, streaming 400+ transactions through fraud checks (geolocation mismatch + credit-utilization limits) before batch-layer reconciliation.
 - Developed stream-layer fraud validation flagging 2.5% of transactions for distance-based anomalies and credit-limit violations, then reconciled and pushed finalized results (updated balances/scores) back to MySQL.

Certifications

- Six Sigma White Belt – Council for Six Sigma Certification** Jul 2026
Microsoft Azure AI Fundamentals (AI-900) - Coursera / Microsoft May 2025
Diploma in Programming and Data Science, IIT Madras Aug 2022

Education

- Rochester Institute of Technology** May 2026
MS in Data Science
PSG College of Arts and Science May 2023
BSc in Computer Science with Data Analytics