SANJAY RAJU

+1 (201) 268-9997 | sanjjayrj@gmail.com | Jersey City, NJ, USA | linkedin.com/in/sanjayraju5/ | sanjjayrj.github.io/

EDUCATION

Stevens Institute of Technology   **September 2023 - May 2025**
Master's, Data Science GPA: 3.7

Amrita Vishwa Vidyapeetham   **June 2019 - August 2023**
Bachelor's, Artificial Intelligence GPA: 8.8

PROFESSIONAL EXPERIENCE

This Just In News   **New York City, NY**
Co-Founder/ AI Engineering   *June, 2025 - Present*

- Designed and built a Gen-AI powered personalized news platform: FastAPI backend with multi-role Docker containers (API, ingest, push, HLS workers) on AWS ECS/FARGATE, ALB, CloudFront CDN, OpenSearch Serverless, EventBridge scheduling, and CloudWatch Container Insights.
- Owned database schema design, migrations, and performance tuning for high-throughput transactional and analytical workload.
- Implemented CI/CD pipelines with Github Actions and AWS ECR for rapid iteration and safe reliability.
- Implemented E2E encrypted messaging using Signal Protocol patterns: X25519 key exchange, AES-256-GCM encryption, HKDF key derivation, forward secrecy via ephemeral keypairs, and PIN-based key backup with PBKDF2-SHA256 100K iterations) stored in iOS Keychain.
- Built hybrid search combining SSE Streaming, PostgreSQL BM25, pgvector HNSW, AWS Bedrock Titan embeddings, and OpenSearch with reciprocal rank fusion-plus agentic AI-powered creator briefings with Claude-generated story recommendations and events timeline aggregation.
- Developed iOS app in SwiftUI/UIKit with Supabase RealTime Broadcast Channels for instant message delivery, typing indicators, presence detection, and LRU video caching for smooth infinite-scroll feeds.

Tata Elxsi   **Bangalore, KA, India**
AI Intern   *January 2023 - June 2023*

- Developed multithreaded C++ systems for high-performance video processing pipelines.
- Integrated multiple ML components into a unified, production-ready workflow.
- Improved pipeline throughput and latency via CUDA acceleration and profiling.
- Performed performance benchmarking and regression testing across model iterations.

PROJECTS & OUTSIDE EXPERIENCE

Lunar Crater Detection and Depth Analysis   *May 2022 - August 2023*
Leader

- Utilized images from Chandrayaan-2 satellite to map out craters on the lunar surface to ensure precision in landing rovers.
- Created a standardized dataset using web scraping and used Google Colab Pro to handle the huge dataset size; analyzed images obtained from the Optical High-Resolution Camera on Chandrayaan-2 satellite, accessed from the ISRO website.
- Provided said dataset for research to other universities and published a paper to a Journal.
- Implemented a geospatial coordinate tracking and depth analysis algorithm to estimate the location and depth of the craters.
- Manually labelled 900+ images of craters and trained a R-CNN and YoloV5 model for object recognition. *Link to paper*

Job Aggregator Conversational AI Agent   *September 2024 - December 2024*
Leader

- Processed user resume and preferences to scrape 1000+ job listings from the last 72 hours.
- Prompt engineered for the GenAI agentic execution to understand user queries to update recommendations.
- Developed a state-of-the-art matching algorithm to rank top 20 jobs for recommendation.
- Interface developed using Flask, GenAI agent able to provide career advice based on resume and recommendations.
- Implemented ranking, feedback loops, and dynamic scraping with real-time response generation.
- *Link to project*

SKILLS

Backend & Systems: Python, C/C++, CUDA, PostgreSQL, OpenSearch, MCP, REST APIs, RAG, Kubernetes, Multithreading
Cloud & DevOps: AWS (ECS, S3, CloudFront), Docker, CI/CD, Git
Data & ML: PyTorch, LightGBM, OpenCV, NLP
Languages: Spanish, Malayalam

EXTRA-CURRICULARS

- **CPE MLK Hackathon 2025**: Mentored students building AI-driven solutions
- **CivicHacks 2025 Hackathon @ Boston University:** Built AI tutor with generated animations and interactive graphs.