Navyatej Tummala

Raleigh, NC | (425) 515-1639 | navyatej31@gmail.com | LinkedIn | GitHub

EDUCATION

North Carolina State University   **Aug 2024 – May 2026**
Master of Computer Science, GPA: 3.97   *Raleigh, NC*
GITAM University   **Aug 2020 – May 2024**
Bachelor of Technology in Computer Science and Engineering, GPA: 7.93   *Visakhapatnam, India*

EXPERIENCE

Cloud Backend Engineer   **May 2025 – Aug 2025**
ARoS Lab, NC State University, Raleigh, NC

- Engineered core features for a cloud web application to manage and label research image datasets, improving dataset readiness for downstream model training.
- Built a serverless backend using **AWS Amplify** and **API Gateway**, delivering **REST APIs** for dataset upload, labeling workflows, and export operations.
- Implemented event-driven ingestion with **S3** and **Lambda** to ingest and preprocess large compressed image archives, improving reliability via trigger-based processing and structured logging.

Software Engineering Intern   **May 2023 – May 2024**
Symbiosis Technologies, Visakhapatnam, India

- Developed a **FastAPI** ingestion service for **50+** sources with schema validation and pagination, reducing manual collection effort by about 40%.
- Built an OpenCV processing pipeline using Celery and Redis workers; persisted outputs to PostgreSQL with retries for resilient background execution.
- Delivered analytics APIs and React + Power BI dashboards, improving insight quality by about 20% and reducing refresh time by about 25%; productionized with Docker, GitHub Actions, pytest/mypy, auth, and rate limiting.

Software Engineering Researcher   **Jan 2022 – May 2023**
GITAM University, Visakhapatnam, India

- Built a Python ETL with Pandas and NumPy to ingest and clean 5k+ news and social posts per day; exposed typed FastAPI endpoints with validation and caching, cutting preprocessing time about 40%.
- Fine-tuned Llama-2 models in PyTorch and Keras with tests and monitoring; shipped a Flask trends dashboard that improved signal precision by about 15% and reduced analysis time to minutes.

PROJECTS

OnCall Copilot | *Python, FAISS/Pinecone, LangChain, Slack Bolt, FastAPI, Docker, Redis*

- Built an AI-powered Slack incident copilot that answers on-call questions from runbooks and postmortems, generating mitigation steps with linked dashboards and log queries.
- Implemented RAG with hybrid retrieval and reranking; deployed a containerized FastAPI service with Redis caching to achieve 89% top 3 source recall and about 780 ms median latency.

GenAI Drug Information Assistant | *Python, BERT, FDA API, Rasa, FastAPI, SQL*

- Built a drug information chatbot using BERT-based NLP plus the FDA API to answer real-time medication and regulatory questions for clinicians and the public.
- Implemented the assistant in Rasa with intent classification, entity recognition, and dialogue management, plus a secure API and refresh pipeline to keep responses current.

TECHNICAL SKILLS

Languages: Python, Java, TypeScript, SQL
Backend: FastAPI, Flask, REST APIs, OAuth2, JWT, OpenAPI
Cloud/DevOps: AWS (S3, Lambda, API Gateway, ECS/Fargate, CloudWatch), Docker
AI/ML: PyTorch, LangChain, scikit-learn, Sentence Transformers, RAG (FAISS/Pinecone)
Databases/Systems: PostgreSQL, MySQL, Redis, SQS, Celery, ETL pipelines
Testing/Quality: pytest, JUnit, mypy, observability (logging/metrics, OpenTelemetry)