SURYA VENKATA ROHIT MOGANTI

Buffalo, NY | +1 (716) 612-0309 | rohitsurya7393@gmail.com | GitHub | LinkedIn

SUMMARY

Hands-on Data & ML Engineer with experience building AI document pipelines, agentic healthcare assistants, and RAG-based search tools using AWS, LangChain, and Transformer models. Driven by real-world problem solving, I specialize in scalable, modular systems that run efficiently both on cloud and local infrastructure.

EDUCATION

State University of New York, Buffalo, Master's in Computer Science and Engineering — Jan 2024 - *May 2025*
Relevant Coursework: Deep/Reinforcement Learning, Computer Vision, MLOps, LLM Security
GITAM University, Hyderabad, India, B. Tech in Computer Science and Engineering — *June 2018 - May 2022*
Relevant Coursework: Machine Learning, Algorithm Design, Web Development, Operating Systems

SKILLS

Programming: Python, SQL, JavaScript, Bash
AI/ML: YOLOv8, Detectron2, Vision Transformers, HuggingFace Transformers, PyTorch
Agentic & LLM Systems: LangChain, RAG, FAISS, Tesseract OCR
Data Engineering: Airflow, DBT, PL/SQL, ODI 12C, Data Curation, Data Lineage, Dimensional Modeling, SCD Type 2, Data Quality
Cloud & DevOps: AWS (Lambda, S3, Textract, SageMaker), Airflow, Spark, Hadoop, Docker, Kafka, Kubernetes, Snowflake
Additional: Experience with MLOps, CI/CD pipelines (Git, GitHub Actions), and strong foundation in algorithms and Data Structures

PROFESSIONAL EXPERIENCE

Software Engineer
TEKsystems Global Services – Hyderabad, India — *Aug 2022 - Dec 2023*

- Led end-to-end development of an AI document pipeline using AWS Textract, SageMaker, S3, and DynamoDB boosting text extraction accuracy by 35%.
- Integrated CloudWatch and SNS-based monitoring with custom alerts and metrics for ingestion workflows, reducing pipeline recovery time by 75% and preventing unnoticed failures during off-hours.
- Implemented batch ingestion via S3 bulk uploads and Lambda triggers, boosting document processing speed by 60%.
- Individually managed 50+ ELT workflows in ODI 12C, processing 5M+ records daily across facts and dimensions in Dev, Test, and Prod for a HIPAA-compliant healthcare analytics system.
- Documented data lineage and curated datasets across multiple workflows using ODI 12C and PL/SQL to support compliance, traceability, and reporting requirements.
- Applied data warehousing techniques (SCD Type 2, surrogate keys, temporal tables) to ensure HIPAA-compliant patient history tracking.
- Collaborated with analysts and product stakeholders to understand business needs and deliver data models and curated datasets supporting key reporting pipelines.

Software Engineer Intern
TEKsystems Global Services – Hyderabad, India — *Feb 2022 - July 2022*

- Built production-grade ETL pipelines in Snowflake using SQL, data build tool, and transformation scripts to power analytics dashboards, optimizing query latency for 1M+ row datasets used in reporting.
- Migrated over 500GB of transactional and historical data from Oracle and SQL Server to AWS and Snowflake, reducing storage costs by 30% and simplifying cross-platform maintenance.
- Ensured data integrity and schema compatibility across migrated sources, documenting column-level lineage for key analytics tables post-migration.

PROJECTS

AI Health Companion – Agentic AI for Personalized Healthcare

- Implemented prompt chaining and generic routing across multiple tools and memory modules to allow context-sensitive actions and multi-step reasoning within the agent.
- Developed an autonomous health agent to plan meals, detect symptoms, and suggest real-time actions based on user inputs using LangChain, FAISS, and local vector databases.
- Integrated tool usage (food analysis, symptom lookup) with memory and self-feedback for personalized interaction.
- Designed a modular local-first architecture with LLMs and OCR, enabling offline use for chronic care support.

Road Damage Detection System

- Built an end-to-end road damage detection system achieving 91% precision and 85% recall using YOLOv8.
- Converted XML annotations to COCO/YOLO, applied augmentation, cleaning, and balanced class distributions.
- Performed ablation studies on augmentation techniques (brightness, blur, contrast) to identify optimal strategies for handling poor lighting and occlusion in road scenarios.
- Benchmarked against Faster R-CNN, with YOLOv8 delivering superior speed and generalization accuracy.