Rishi Guptha Mankala

934-263-3087 | rishiguptha.mankala@gmail.com | linkedin.com/in/rishi-guptha | github.com/rishiguptha

EXPERIENCE

Nevara AI New York, NY
Software Engineer Intern (Artificial Intelligence) *Jun 2025 – Aug 2025*
- Engineered a near real-time sales auditing platform (Next.js, Supabase) for 100+ active users, architecting an asynchronous AWS Lambda pipeline integrating OpenAI/Deepgram APIs that slashed manual review time by 80% via automated AI coaching reports.
- Optimized system throughput by fine-tuning AWS Lambda provisioned concurrency and memory allocation, achieving a 40% reduction in latency overhead to sustain sub-second transcription speeds during concurrent traffic spikes.
- Designed a temporal data synchronization engine (Node JS, HLS.js, React Query) that maps high-frequency audio timestamps against transcript segments, enabling millisecond-level navigation and instant context retrieval for auditors.

Stony Brook University Stony Brook, NY
Data Science Research Assistant *Jan 2025 – May 2025*
- Engineered an end-to-end Python and SQL pipeline to process 20+ years of S&P 500 data, implementing automated feature normalization that boosted training efficiency by 40%.
- Developed LSTM and CNN models (PyTorch) achieving an R^2 of 0.82—reducing error by 18% vs. ARIMA—and deployed a Streamlit dashboard for real-time strategy assessment.

Kimberly-Clark Bengaluru, India
Data Engineer Intern *Jan 2023 – Jul 2023*
- Led the migration of 10,000+ data assets into Collibra, writing validation scripts to ensure 100% data integrity when transitioning from legacy Excel trackers to a centralized, automated enterprise data catalog.
- Engineered automated metadata extraction pipelines using Python and Java to ingest schemas from Data Lakes, Data Marts, and SAP HANA, streamlining discovery for the Analytics team and saving 10+ hours of manual work weekly.
- Deployed custom business workflows using Python for real-time lineage validation to accelerate the stakeholder reporting approval cycle by 35% across global teams.

Matchday AI Hyderabad, India
Software Engineer *Jan 2022 – Dec 2022*
- Built computer-vision pipelines using OpenCV to detect player trajectories and ball positions, achieving high-fidelity event tracking for professional analytics used by coaching staff to optimize in-game tactical strategies.
- Implemented homography transformations to convert dynamic broadcast camera feeds into accurate 2D tactical field projections.
- Created a full-stack data annotation platform using Flask and JavaScript that accelerated the training data labeling process for 6+ ISL teams integrated with Star Sports broadcasts.

PROJECTS

Distributed Spatial-Temporal Clustering Engine | *Python, Pandas, Scikit-Learn, DBSCAN, mpi4py, PCA*
- Implemented distributed clustering on NYC building permit data (3.9M rows, 1991-2025) using DBSCAN and parallel processing with mpi4py and K-means to analyze spatial-temporal patterns and identify construction hotspots at scale.

NutriSmart: Semantic Search Engine | *Python, FastAPI, AWS, Docker, FAISS*
- Deployed a RAG service on AWS EC2 using Sentence Transformers and FAISS to index 1M+ tokens of complex medical papers, enabling sub-second semantic search queries via FastAPI.

Snowflake Data Warehouse Modernization | *dbt, Snowflake, S3, SQL*
- Built an S3-to-Snowflake ELT pipeline on 20M+ Movielens dataset, using dbt incremental models and SCD2 logic to track historical changes while minimizing query costs.

NYC Taxi Insights | *Python, GCP, Airflow, SQL*
- Engineered an Airflow-orchestrated workflow on GCP to process 10M+ records, designing star-schema data marts and integrating BERT feature embeddings for predictive modeling.

Spotify Audio Analytics | *AWS Lambda, Glue, Spark, Docker*
- Designed a serverless architecture using AWS Glue and Spark, implementing optimized partitioning strategies to reduce job runtime and compute costs by 40% through efficient data skew handling.

EDUCATION

Stony Brook University Stony Brook, NY
Masters of Science, Data Science *Jan 2024 – Dec 2025*

Vellore Institute of Technology India
Bachelors, Computer Science *Jun 2019 – May 2023*

TECHNICAL SKILLS

Languages & CS: Python, Java, SQL, TypeScript/JavaScript, R, Git, Linux
Data & Analytics: Data Modeling, ETL Pipelines, Apache Spark, Airflow, dbt, Snowflake, PostgreSQL, BigQuery, Tableau, Power BI
ML & AI: scikit-learn, Pandas, NumPy, PyTorch, TensorFlow, NLP, RAG Architectures, OpenAI API, Vector Databases (FAISS)
Cloud & DevOps: AWS (Lambda, S3, EC2, Glue), GCP, Docker, CI/CD (GitHub Actions), Vercel, Supabase, FastAPI, Next.js, Streamlit, Node JS