

# JOEL SUSHANTH MARKAPUDI

Boston, MA • 617-510-2051 • markapudi.j@northeastern.edu • linkedin.com/in/joemjs/ • github.com/mjsushanth

## EDUCATION

### Northeastern University

M.S. in Artificial Intelligence (G.P.A - 3.9)

Relevant Coursework: MLOps, Deep Learning, Computer Vision, Machine Learning.

Boston, MA.

### KL University

Bachelor's in Computer Science

Jan 2024 – Present

Vijayawada, India.

Jun 2016 – Jun 2020

## SKILLS

**Programming & ML/AI:** Python (PyTorch, Pandas, Polars, Scikit-learn, Gensim, OpenCV), Java, SQL.

**MLOps & Cloud:** Docker, Airflow, FastAPI, Streamlit, AWS (Bedrock, S3 Vectors, Lambda), DuckDB.

**Data & BI Tools:** Power BI, Talend, PostgreSQL, Parquet, ETL Architecture, Data Modeling.

## ACADEMIC PROJECTS

### FinSights: LLM RAG Financial Data Intelligence System

Sep 2025 – Dec 2025

- Engineered E2E **RAG system** for analyst-style queries over SEC Form 10-K data with **hybrid retrieval architecture**. Combines structured KPI tables and sentence-level vector search. Implemented fuzzy NLP entity adapters and multi-stage context assembly pipeline, provenance tracking, and citation header generation for exact source traceability.
- Designed retrieval pipeline with **query variant generation** to improve recall, **S3 Vectors** store with metadata filtering, triple retrieval regimes, context neighbor expansion followed by de-duplication and assembly orchestration.
- Deployed **3-tier production system** (ML orchestrator, FastAPI backend, Streamlit frontend) with cost modeling that achieves **\$0.017/query** through S3 Vectors. Developed scaling formulas and example scenarios for cost analysis.
- Built offline evaluation pipeline validating retrieval quality with **automated anchor tests, curated gold sets** tested with metrics (ROUGE-L/BERTScore/BLEURT), and manual grounding analysis achieving 80%+ factual accuracy.

### SocrAItic Circle: Multi-Agent LLM Debate & Evaluation Framework

Jan 2025 – Apr 2025

- Built a multi-agent LLM debate system with structured preparation phases, iterative debate refinement, and adversarial critique mechanisms. Implemented a specialized judge framework, achieving structured LLM argument assessment.
- Developed persistent memory architecture with YAML-based prompt templates, prompt orchestration, auto-tuning, experiment configurations, and local deployment using Ollama.

### Pose Synthesis With Cross-Modal Diffusion Architecture (PyTorch, CLIP)

Jan 2025 – Apr 2025

- Developed cross-modal transformer-diffusion architecture integrating CLIP embeddings for 3D pose generation.
- Implemented specialized cross-attention mechanisms: hierarchical semantic conditioning across UNet resolution levels, with attention specialization for body part detection, spatial relationships, and action understanding.
- Applied biomechanical constraints with bone-length consistency, anatomical validation, and CFG techniques.

### Multi-View Image Analysis for 3D Reconstruction

Sep 2024 – Dec 2024

- Engineered a **pose-guided feature correspondence system** that integrates camera pose estimation with view overlap analysis for multi-view matching across varying baselines and viewpoint changes.
- Implemented an adaptive pipeline with scene-type classification, geometric validation, match quality assessment, and rejection techniques. Optimized with further occlusion handling, track formation, and stability analysis.
- Built GPU-accelerated preprocessing pipeline with LAB color space transformations, CLAHE enhancement, and bilateral filtering using PyTorch. Attained a 0.93 quality score with 56K+ features per image, and 28% spatial coverage.

### Additional Projects:

Jan 2024 – Dec 2024

- Late-interaction Image–Text Model (ViT+CLIP, MaxSim retrieval, LoRA), Protein Structure Prediction (HMM, CRF, BiLSTM, SVM), Art Classification (CNN/SVM-SIFT image classification), DNA Sequence Matching (Suffix Trees/Aho–Corasick genomic sequence search), Stock Trading Platform (Java MVC client–server application).

## EXPERIENCE

### Innova Solutions, Hyderabad

Data Engineer — Business Intelligence

Dec 2019 – Nov 2023

### Accomplishments

- Architected and led development of scalable data solutions for FP&A teams, serving as a lead dev. Automated **ETL in Talend** for boosted data integration, optimized **PL-SQL procedures**, and crafted several **PowerBI data models**.
- Led migrations and upgrades of ETL systems with CI/CD pipelines that reduced deployment time by over 50%. Led cross-functional teams of size 3-6 and mentored new junior developers, boosting team productivity by over 25%.
- Designed **25+ dashboards** encompassing **500+ metrics**, providing comprehensive business insights.
- Partnered with in-company teams to help refine data modeling. Conducted **35+ interviews** to assist recruitment.
- Earned Star Awards and **Innova Idol Award** for exceptional contributions to team projects and efficiency.

## **FP&A and SG&A Reporting**

Mar 2020 – Nov 2023

- Orchestrated multiple workspaces with tailored FP&A reports for diverse team needs including Quarterly Business Reviews (QBRs), Producer Reports, Hiring Metrics, and specialized reports for different business units and divisions.
- Forecasted revenue and hiring metrics, integrating an adaptive awards system aligned with dynamic business rules.
- Standardized revenue trends and eligibility criteria to enhance the accuracy and relevance across multiple reports.
- Developed SG&A reporting models, custom hierarchies to streamline financial data and executive decision-making.
- Automated margin and audit processes, reducing manual Excel-based tasks by over 70 hours and enhancing process reliability. Implemented automated systems for searches, exclusions, remarks, and data verifications.
- Managed **over 11 major projects**, including the Producer Commission Report, Reconciliation Report, Daily Sales and Outstanding Report, and CSM Dashboard, contributing to crucial operational efficiency and data transparency.

## **Operational Reporting**

May 2022 – Nov 2023

- Architected end-to-end data models integrating multi-source financial data and custom-prepared lookups, enhancing data accessibility and accuracy.
- Collaborated with cross-functional teams to align performance indicators with strategic business goals, engineering over 200 high-impact KPIs.
- Managed the project across two phases, optimized code and data loaders to accelerate report generation by 60%. Established in-depth data view for key metrics and detailed presentations, boosting data exploration capabilities by over 80%. Innovated a data integration solution, achieving over **70% accuracy** in correlating disparate system data.