

JOEL SUSHANTH MARKAPUDI

Boston, MA ☎ 617-510-2051 ✉ markapudi.j@northeastern.edu 🌐 [linkedin.com/in/joemjs/](https://www.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. KL University Bachelor's in Computer Science	Boston, MA. Vijayawada, India.
	Jan 2024 – Present Jun 2016 – Jun 2020

SKILLS

Programming & ML/AI:	Python (PyTorch, Pandas, Polars, NumPy, Scikit-learn, OpenCV), SQL, Java; HuggingFace Transformers/Diffusers, XGBoost, ViT; LoRA/PEFT, PyLate, LEANN.
MLOps, LLM & Cloud:	Docker, MLflow, Optuna; FastAPI, Flask, Streamlit; Airflow; ELK Stack. AWS (S3, Lambda, Bedrock, S3 Vectors, ECS); GCP; GitHub Actions CI/CD.
Data & BI:	DuckDB, PostgreSQL, Power BI, Talend; ETL architecture, Data modeling.

ACADEMIC PROJECTS

FinSights: LLM RAG Financial Data Intelligence System	Sep 2025 – Dec 2025
<ul style="list-style-type: none">Built 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, powered by fuzzy entity extraction.Built token-aware embedding pipeline processing 1,850 vectors/min with intelligent batching (96 texts/call, 15K token budget) and real-time cost tracking. Utilizes S3 Vector store with metadata filtering.Structured retrieval pipeline with query variant generation, triple retrieval regimes (filtered/global/union), sentence window expansion, deduplication, and stratified distance-based selection (0.35 cutoff). Includes provenance tracking.Deployed 3-tier production system (ML orchestrator, FastAPI + Streamlit) with cost modeling that achieves \$0.017/query through S3 Vectors. Derived scaling formulas and example scenarios for cost analysis.Achieved 0.82-0.86 BERTScore F1 on business-realistic gold test. Validated retrieval quality through deterministic neighbor tests (98%+ Self@1, 70-92.5% Hit@5). Evaluation includes multi-metric analysis (BLEURT, ROUGE-L).	
Pose Synthesis With Cross-Modal Diffusion Architecture (PyTorch, CLIP)	Jan 2025 – Apr 2025
<ul style="list-style-type: none">Devised Conditioned UNet diffusion architecture processing 23.3K+ training samples with 8-head cross-attention & 4 resolution levels, generating 3D poses from 66-dimensional skeletal representations.Modeled hierarchical cross-attention with first-action syntactic segmentation and cluster-aware sampling, obtaining body-part and action-aware semantic conditioning. Achieved 0.69 final loss through 3-phase progressive training.Enforced biomechanical structure awareness, anatomy validation, forward kinematics constraints, tanh-bounded outputs (1.5× scaling) and CFG guidance scale 7.0. Achieved 70% visual alignment and topological consistency.	
SocrAItic Circle: Multi-Agent LLM Debate & Evaluation Framework	Jan 2025 – Apr 2025
<ul style="list-style-type: none">Created a multi-agent LLM debate system with structured preparation phases, iterative refinement, and adversarial critique. Integrated specialized judge framework achieving 35% improvement in argument coherence scores.Engineered a persistent memory architecture with 25+ YAML-based prompt templates, prompt orchestration, auto-tuning, experiment configurations, and local deployment using Ollama.	
Multi-View Image Analysis for 3D Reconstruction (Computer Vision)	Sep 2024 – Dec 2024
<ul style="list-style-type: none">Engineered a pose-guided feature correspondence system processing 360K+ 2D observations (24MP) across 38 high-resolution views, integrating camera pose estimation and view overlap analysis for multi-view matching.Orchestrated a 6-stage adaptive pipeline integrating scene-type classification, RANSAC geometric validation, and dual-layer occlusion handling. Reduced false matches by 42%. Improved feature track stability to 3.56-frame persistence. Achieved 38.5% 2D-3D correspondence rate generating 33K+ 3D points.Implemented a GPU-accelerated preprocessing pipeline with LAB color space transformations, CLAHE enhancement, and bilateral filtering. Attained a 0.93 quality score with 56K+ features per image, and 28% spatial coverage.	
Late-interaction Image-Text Model: (ViT+CLIP, MaxSim retrieval, LoRA)	Nov 2025 – Dec 2025
Protein Secondary Structure Prediction (HMM, CRF, BiLSTM, SVM)	Sep 2024 – Dec 2024
Stock Trading Platform (Java MVC client-server application).	Jan 2024 – Apr 2024

EXPERIENCE

Innova Solutions, Hyderabad Data Engineer — Business Intelligence	Dec 2019 – Nov 2023
Accomplishments <ul style="list-style-type: none">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%. Guided cross-functional teams of size 3-6 and mentored new junior developers, boosting team productivity by over 25%.	

- Created **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. Deployed automated systems for searches, exclusions, remarks, and data verifications.
- Oversaw and personally executed across **11+ 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

- Developed 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.
- Drove 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.