

# THEJESH MALLIDI

+1 517-730-4554 | thejesh.mallidi18@gmail.com  
[github.com/thejesh-m](https://github.com/thejesh-m) | [linkedin.com/in/thejeshm](https://linkedin.com/in/thejeshm) | [www.thejesh.in](http://www.thejesh.in) | Irving, TX

## EDUCATION

<b>Michigan State University, USA</b> Master of Science in Data Science	08/2023 - 04/2025 GPA: 3.95/4.0
<b>Madanapalle Institute of Technology &amp; Science, India</b> Bachelor of Technology in Computer Science	06/2017- 07/2021 GPA: 9.23/10.0

## EXPERIENCE

<b>Quantiphi Analytics</b>	Irving, TX, USA
<b>Machine Learning Engineer</b>	06/2025 - Present
<ul style="list-style-type: none"><li>Engineered an <b>LLM-driven document ingestion pipeline</b> for a leading telecom company processing <b>2M+</b> <b>contract documents</b>, extracting unstructured legal text into structured Neo4j graph format while preserving full document hierarchy.</li><li>Implemented <b>Agentic Contract Assist chatbot (GraphRAG)</b> over the <b>Neo4j graph + document store</b>, improving query relevance by <b>~45%</b> and reducing retrieval latency from <b>seconds to sub-second responses</b>.</li><li>Developed and deployed CI/CD workflows on Red Hat OpenShift and GCP Cloud Run, enabling zero-downtime updates and Integrated BigQuery and GCS for scalable contract metadata storage.</li></ul>	
<b>Michigan State University</b>	East Lansing, MI, USA
<b>Graduate Research Assistant - Applied AI</b>	01/2024 – 05/2025
<ul style="list-style-type: none"><li>Pretrained <b>domain-adapted BERT (MLM)</b> with domain-aware masking (boosted mask probability for curated terminology) on 1M abstracts, delivering <b>+15% average accuracy</b> on downstream tasks.</li><li>Built a <b>domain-specific knowledge graph</b> with over <b>2 million entities</b> and <b>1.4 million relationships</b> to accelerate hypothesis generation in plant science, leveraging <b>Named Entity Recognition (NER)</b> and <b>Relation Extraction</b>.</li><li>Fine-tuned <b>LLMs</b>, including <b>BERT</b> and <b>LLaMA</b> variants and created a <b>custom BERT-based multi-relation extraction model</b> for information extraction, achieving <b>94%</b> accuracy in entity extraction and <b>84%</b> in relation identification.</li><li>Developed a <b>conversational research assistant</b> with <b>GraphRAG + agentic workflows (LangGraph)</b> to reason over the KG and literature.</li><li>Led large-scale topic analysis of <b>700k evolutionary biology research abstracts</b>, uncovering temporal trends, research shifts, and literature gaps to inform future studies; leveraged <b>BERT-based topic modeling</b> and <b>Agentic AI</b> to extract ~150 high-quality topics.</li></ul>	

<b>Quantiphi Analytics</b>	Bangalore, India
<b>Machine Learning Engineer</b>	05/2021 - 07/2023
<ul style="list-style-type: none"><li>Automated <b>document processing pipelines</b> for Banking, Financial Services, and Insurance (<b>BFSI</b>) workflows using <b>computer vision</b> and <b>NLP-based models</b>, which reduced manual review time by <b>75%</b> and achieved <b>90%</b> data extraction accuracy.</li><li>Delivered <b>10+ Proof of Concepts (PoCs)</b> on <b>document classification &amp; extraction, object detection, image classification, and semantic segmentation</b> in collaboration with cross-functional teams of frontend developers and business stakeholders.</li><li>Developed <b>insurance-focused predictive models</b> for risk assessment and claim outcome prediction using Logistic Regression, Decision Trees, Bagging, Random Forest, and Gradient Boosting Machines (GBM).</li><li>Built and optimized real-time inference pipelines for Nvidia edge devices using <b>DeepStream</b> with <b>custom C++ plugins</b>, achieving <b>27 FPS</b> for latency-sensitive applications.</li><li>Orchestrated the deployment of <b>20+ machine learning models</b> into production environments, leveraging <b>CI/CD pipelines</b> across <b>AWS SageMaker, GCP Vertex AI, and Kubernetes</b>, decreasing model deployment time by <b>30%</b>.</li></ul>	

## SKILLS

- Core Expertise:** Data Science, Machine Learning, Deep Learning, Predictive Modeling, Natural Language Processing (NLP), Computer Vision, Large Language Models (LLMs), Generative AI & Agentic AI, RAG, Graph ML, MultiModal AI, LLM Orchestration
- Programming Languages:** Python, R, C++, C, Bash
- AI/ML Frameworks & Libraries:** PyTorch, TensorFlow, TensorRT, ONNX, LangGraph, LangChain, LlamaIndex, AutoGen, Vector databases, Hugging Face Transformers, OpenCV, NumPy, Pandas
- Big Data & Streaming:** Hadoop Spark, Apache Kafka, Apache Beam
- Databases:** MySQL, PostgreSQL, Neo4j Graph DB, Elastic Search
- Cloud & MLOps Tools:** GCP, AWS, Heroku, Docker, Git, Flask, FastAPI, CI/CD Pipelines, Jenkins, MLFlow, KubeFlow