Mounika Bolla

Gainesville, Florida
(913)-238-8079 mounikabollav@gmail.com Linkedin Github

Summary

Machine Learning Engineer with 4+ years of experience building production AI systems in deep learning, NLP, and computer vision. Expert in deploying scalable ML pipelines and agentic AI systems using PyTorch, TensorFlow, OpenAI API, Google Gemini API, AWS, and Azure, with proven results in model optimization, real-time analytics, and cost reduction across healthcare, cybersecurity, and manufacturing domains.

Education

University of Central Missouri **Aug 2023 – May 2025**
Master of Science in Computer Science – GPA: 3.5/4.0 *Lees Summit, MO*

Professional Experience

JerseySTEM **Aug 2025 – Present**
Data Scientist *Remote*

- Developed an agentic AI pipeline to automatically design customized curricula based on course requirements, reducing design time by 60%, and built ML models achieving 85% accuracy in evaluating curriculum effectiveness using feedback from 500+ students
- Consolidated attendance, assessment, and survey datasets using Python and SQL (reducing data wrangling time by 35%), and engineered interactive Tableau dashboards with FastAPI backend, contributing to ∼10% increase in workshop engagement
- Participated in Agile development cycles with weekly sprint planning and Scrum boards, collaborating cross-functionally to deliver ML pipeline iterations and analytics features

University of Central Missouri **Aug 2024 – May 2025**
Graduate Research Assistant *Lees Summit, MO*

- Engineered an advanced ECG risk-stratification system integrating deep learning with multi-channel signal processing; preprocessed WFDB recordings (MIT-BIH) using Butterworth filtering (0.5–40 Hz) and adaptive R-peak detection for 180-sample heartbeat extraction
- Designed and optimized CNN, LSTM, and attention-based multi-scale architectures with SMOTE and augmentation pipelines, achieving 98.84% accuracy, 0.999 AUC, and 0.98 recall across Normal, Moderate-Risk, and High-Risk classes through systematic hyperparameter tuning and cross-validation, ensuring minimal false negatives critical for clinical diagnosis
- Developed a production-ready clinical prototype with automated risk scoring and real-time ECG visualization using Streamlit, supporting multi-channel analysis and detailed reporting

Attacked.ai **May 2024 – Dec 2024**
AI Product Engineer Intern *London, UK (Remote)*

- Built an agentic AI pipeline to autonomously ingest, deduplicate, and score global cyber-attack reports, powering a real-time intelligence map used by enterprise security teams
- Developed a voice-interactive simulation platform (LiveKit + Django + OpenAI APIs) enabling stakeholders to query attack scenarios and assess business impact via natural language
- Designed the Guard Framework, combining LLMs with knowledge graphs to automate security-control mapping and ensure alignment with NIST and ISO-27001 compliance standards
- Developed NLP classification models (Logistic Regression, Random Forest, fine-tuned BERT) for automated categorization of 50K+ cybersecurity incident reports, improving classification accuracy by 14% over baseline

ESAB India Pvt Ltd **May 2022 – Jul 2023**
Machine Learning Engineer *Chennai, India*

- Developed a computer vision system using MobileNetV3 and EfficientNet-B0 with OpenCV preprocessing, achieving 95%+ accuracy for automated consumable classification and reducing manual inspection time by 70%
- Built a predictive maintenance "Life Meter" using Temporal Convolutional Networks (TCN + attention) to forecast consumable degradation from sensor streams, reducing equipment downtime by 40%
- Designed scalable ML pipelines integrating CV models with manufacturing inventory systems, processing 10K+ consumable transactions daily in production
- Collaborated with cross-functional engineering teams to translate requirements into production-ready deep learning solutions for welding equipment optimization

IIITDM Kancheepuram **Nov 2021 – May 2022**

Research Intern *Chennai, India*

- Engineered a learning-based model for fundus image quality assessment using ResNet50 and EfficientNetB0, optimizing hyperparameters to improve diagnostic accuracy by 5% over previous research, achieving 86.2% accuracy
- Presented research at BMT 2023 Conference, effectively communicating complex ideas to both technical and non-technical audiences

Veyyil Robotics Private Limited **Aug 2021 – Oct 2021**

AI Development Intern *Chennai, India*

- Leveraged the YOLO algorithm for automated fault detection, reducing human intervention by 40% and enhancing operational efficiency
- Conducted comparative analysis of YOLO versions, optimizing detection precision by 25% through IoU-based probability scoring

Key Projects

Intelligent Violence Surveillance System | *FastAPI, TensorFlow, Azure, Next.js, Gemini API* **2024**

- Architected full-stack AI violence detection system using FastAPI, TensorFlow (MobileNetV3Small CNN with 88.8% accuracy), Google Gemini API, and Next.js, deployed on Azure with CI/CD pipeline achieving sub-50ms frame inference
- Designed intelligent three-tier frame extraction algorithm prioritizing violence detection segments and scene changes, generating condensed videos (20-30 frames at 2 FPS) that reduced API costs by 80-90% for long video processing while maintaining analysis quality
- Built production-grade monitoring system with real-time metrics tracking (latencies, model confidence, processing times), OpenTelemetry integration with Azure Monitor, and user feedback collection for continuous model improvement

AlignAI - AI-Powered Resume Tailoring System | *Python, Flask, PostgreSQL, RAG, LangChain* **2025**

- Developed an intelligent resume optimization platform using RAG (Retrieval-Augmented Generation) to automatically tailor resumes to job descriptions, processing PDF/DOCX files with 95%+ accuracy through Python, Flask, and PostgreSQL backend
- Engineered multi-agent system with custom prompt engineering and context-aware chat interface that learns from user feedback and iteratively refines resume content based on job requirements
- Built full-stack web application with secure authentication, real-time chat functionality, resume library management, and RESTful API architecture for seamless personalized document generation

Financial Market Map with Autonomous AI Agents | *Python, LangChain, FinBERT, APIs* **2025**

- Developed autonomous multi-agent system using LangChain and OpenAI API to analyze financial news sentiment and correlate impact on global market indices in real-time, processing 1K+ daily news articles
- Built interactive global map visualization displaying financial news propagation effects across international markets with automated data pipeline and real-time sentiment analysis
- Implemented automated news scraping pipeline with sentiment scoring using fine-tuned FinBERT model, achieving 87% sentiment classification accuracy on financial text

Certifications

AWS Certified AI Practitioner – AWS Certification, Nov 2025

Tata iQ AI-Powered Data Analytics & Strategy – Completed job simulation involving AI-powered data analytics for Financial Services. Conducted EDA using GenAI tools, proposed no-code predictive modeling framework for customer delinquency risk assessment, and designed AI-driven collections strategy incorporating agentic AI, ethical AI principles, and regulatory compliance, 2025

Introduction to Big Data with Spark and Hadoop – Coursera Certification, Sep 2024

Technical Skills

Machine Learning & Deep Learning: TensorFlow, PyTorch, Scikit-learn, XGBoost, LightGBM, Keras, CNN, RNN, LSTM, Transformers, Attention Mechanisms, Transfer Learning, Model Optimization, Hyperparameter Tuning

NLP & Computer Vision: Hugging Face Transformers, BERT, GPT, FinBERT, spaCy, NLTK, OpenCV, YOLO, MobileNet, EfficientNet, ResNet, Image Classification, Object Detection, Sentiment Analysis

MLOps & Cloud: Docker, Kubernetes, MLflow, Apache Airflow, AWS (SageMaker, Lambda, S3, EC2), Azure (Azure ML, Azure Monitor), CI/CD Pipelines, OpenTelemetry

Big Data & Analytics: Apache Spark, Kafka, Hadoop, PySpark, Real-time Stream Processing, Distributed Computing,

Tableau, Data Pipelines, ETL
Programming & Development: Python (Advanced), SQL, R, Java, Shell Scripting, Git, Linux, RESTful APIs, Agile/Scrum
Data & Databases: PostgreSQL, MySQL, MongoDB, Redis, Elasticsearch, FAISS, Pinecone, Vector Databases
Web Development & Frameworks: FastAPI, Flask, Django, React.js, Next.js, Streamlit, LiveKit, Node.js
LLMs & AI Agents: OpenAI API, Google Gemini API, LangChain, LlamaIndex, RAG, Multi-Agent Systems, Agentic AI, Prompt Engineering
ML Specialties: Recommendation Systems, Anomaly Detection, Time Series Forecasting, A/B Testing, Feature Engineering, Predictive Maintenance

Publications

Deep Learning Based Quality Prediction of Retinal Fundus Images – Sep 2023
Published peer-reviewed research on medical image quality assessment using deep learning at BMT 2023 Conference
https://degruyter.com/document/doi/10.1515/cdbme-2023-1177/html