

Midhun M

8111931361 • midhunpaapu@gmail.com • linkedin.com/in/midhun-m-d2001 •
<https://my-portfolio-lilac-gamma.vercel.app/>

AI Solutions Engineer Machine Learning Engineer | Applied AI | Backend AI Systems

AI Solutions Engineer with 1.5+ years of experience building and deploying production-grade machine learning systems. Strong expertise in designing scalable AI microservices using FastAPI and Docker, developing multilingual NLP systems, AI agents, and document intelligence pipelines. Proven track record in taking AI products from research to real-world deployment with monitoring, explainability, and automation.

WORK EXPERIENCE

Gritstone Technologies AI/ML Engineer

12/2024 – Present

- Designed and deployed production-grade AI systems, including goal-driven agentic assistants and real-time inference pipelines.
- Built and optimized multimodal solutions combining text, audio, and image data for end-to-end automation.
- Led model evaluation and monitoring workflows using cross-validation, accuracy benchmarking, and latency optimization.
- Automated deployments using FastAPI + Docker, integrated CI/CD pipelines, and published containerized solutions to Docker Hub.

iDatalytics Associate Data Science Intern

05/2024 – 11/2024

- Developed ML models for forecasting and classification tasks, improving decision-making accuracy.
- Collaborated with product teams to integrate models into business-critical workflows, ensuring measurable ROI.
- Conducted feature engineering, data preprocessing, and model performance tuning for production readiness.

EDUCATION

Data Science and AI Generation

Certification

Learnbay • 01/2024 – 11/2024

Bachelors of Technology in Production Engineering

Government Engineering College • GPA: 7.2

Thrissur • 07/2020 – 07/2024

CERTIFICATIONS

HuggingFace NLP Course – Natural Language Processing with Transformers

Deep Learning Specialization – Andrew Ng (Coursera)

MLOps Fundamentals – Model Deployment & Monitoring

Data Science and AI Generation– Learnbay

PROJECTS

AI-based Fraud Detection System

Tech: Python, Scikit-learn, SHAP, MLflow, FastAPI, Docker

Built end-to-end fraud detection pipeline using ensemble ML models.
Implemented SHAP for model explainability to analyze feature importance.
Used MLflow for experiment tracking and model comparison.

AI Agent System – Multi-Agent Platform

Tech: CrewAI, LLMs, FastAPI, Python

Designed multi-agent system using CrewAI with Planner, Executor, and Critic agents.
Implemented task delegation, memory handling, and automated decision pipelines.
Built production-style backend for agent orchestration using FastAPI.
Demonstrated use cases in document analysis and workflow automation.

AI Resume Enhancer agent

Engineered a comprehensive AI Resume Enhancer leveraging Transformer-based models (BERT, Hugging Face) and LangChain frameworks to deliver precise ATS score evaluations and keyword optimization recommendations.
Automated multi-threaded processing pipelines utilizing Python's async capabilities and Docker containerization to expedite resume-job description matching, reducing processing time by 50%.
Applied cross-validation and hyperparameter tuning to optimize NLP classifiers for skill extraction and formatting issue detection, improving prediction accuracy by 25%.

AI-Powered Document Requirement Extractor (Gemini Pro)

Designed a system to parse and extract requirements from unstructured PDFs using Gemini Pro.
Implemented semantic chunking & vector store integration for precise contextual retrieval.
Automated manual review, saving significant processing time for document-heavy workflows.

Real-Time Audio & Multimodal Processing

Designed real-time audio transcription and image-processing pipelines with optimized latency for live use cases.
Implemented multi-threaded processing to handle parallel inference requests at scale.
Delivered scalable solutions for real-time voice-to-text, object detection, and multimodal analytics.

SKILLS

Languages & Tools:: Bash, Git, Linux, Python, SQL

ML / AI:: Avatar-based Interfaces, PyTorch, TensorFlow, Scikit-learn, HuggingFace, Transformers, YOLO, LLMs, CrewAI, Prompt Engineering, LangIndex, LangChain, DPO, Self-Rag

Backend & MLOps:: ChromaDB, PineCone, Vector Embedding,, FastAPI, Docker, MLflow, SHAP, REST APIs, Gradio, CI/CD

Domains:: Explainability (SHAP/LIME), Grafana/Prometheus, Hyperparameter Tuning, NLP, Speech Processing, AI Agents, Computer Vision, Document AI, Model Deployment