Dinesh Kothandaraman

West Haven, CT | dineshkothand@gmail.com | +1 475-227-5994 | LinkedIn | Github

Summary

AI/ML Engineer with 6+ years of experience building and deploying scalable ML systems and GenAI solutions. Skilled in LLM fine-tuning, RAG pipelines, and MLOps practices for reliable production delivery.

Technical Skills

Languages: Python, SQL, TypeScript, JavaScript, R
AI/ML & GenAI: LLM, RAG, NLP, Transformers, Attention Mechanism, Agentic AI workflows
Frameworks: PyTorch, TensorFlow, Scikit-learn, FastAPI, LangChain
MLOps & Tools: Docker, Kubernetes, MLflow, DVC, GitHub Actions, Airflow, Kubeflow, MLflow
Cloud & Data: GCP, AWS, Spark, Hadoop, Hive, PostgreSQL, Snowflake

Experience

Agentic AI Developer | Kamali Inc**,** Remote-United States| Sep 2024 – Jun 2025

Built and deployed document-based QA chatbots using LangChain, RAG, and LLMs. Designed multi-agent workflows for intent routing, integrated ASR and SSML for voice interaction, and deployed scalable FastAPI services on AWS using CI/CD and Docker.

Machine Learning Engineer | Cognizant Technology Solutions, India | Jan 2021 – July 2023

Developed ML pipelines for insurance underwriting automation using PySpark and AWS. Trained classification models for risk scoring, improving policy processing time by 35%. Delivered explainable ML solutions via Flask APIs integrated into internal dashboards.

Machine Learning Engineer | Jubliee Information Technology Pvt Ltd, India | July 2017 – Dec 2020

Created recommendation and segmentation models for insurance clients using PySpark and clustering. Automated daily scoring workflows on AWS EMR with Kafka integration, improving targeting and operational efficiency.

Publication

Hybrid CNN-LSTM Model for Jail Activity Recognition | IEEE UEMCON | Oct 2024
https://ieeexplore.ieee.org/abstract/document/10754670

Projects

Document-based QA Chatbot: Built RAG-powered chatbot using LLMs and LangChain for document queries.

Healthcare IoT System: Developed AWS-based patient monitoring pipeline for real-time alerts.

Text Summarization Models: Fine-tuned T5 and BART for abstractive summarization with attention visualization.

Car Logo Detection: Applied transfer learning with CNNs (ResNet, VGG) for high-accuracy logo recognition.

Jail Activity Recognition: Combined CNN-LSTM for real-time activity classification.

Gender & Age Detection: Built face-based deep learning classifier with pretrained models.

Education

MS, University of New Haven, Data Science | Aug 2023 – May 2025 | GPA: 3.89/4.0

BE, Chennai Institute of Technology, Mechatronics Engineering | June 2013 – April 2017 | GPA: 7.35/10.0