KARTHIK MANNEM

Los Angeles, CA
747-347-6409 | mannemkarthik01@gmail.com
linkedin.com/in/karthik-mannem-2008b4225 | github.com/mannemkarthik010| Portfolio

Summary

Machine Learning Engineer who approaches problem-solving as an end-to-end AI pipeline, ingesting complex real-world inputs, performing structured feature engineering, training and fine-tuning deep learning and transformer-based architectures, and deploying scalable inference systems. With 3+ years of experience across predictive modeling, computer vision, and LLM-powered Retrieval-Augmented Generation (RAG) workflows, I focus on evaluation rigor, performance optimization, and production-grade deployment.

Personal Projects

•Preventive Care Recommendation Agent | LLM + RAG (LangChain)

LLM-powered healthcare decision support system deployed publicly.

- Designed and deployed an LLM-powered preventive care recommendation system that analyzes patient age, medical history, and risk factors to generate personalized screening and vaccination guidance aligned with CDC guidelines.
- Architected a Retrieval-Augmented Generation (RAG) pipeline using LangChain, implementing document chunking, embeddings, and vector-based semantic retrieval.
- Applied prompt engineering techniques to improve reasoning accuracy and reduce hallucinated medical outputs.
- **Tech:** Python, LangChain, OpenAI API, Vector Search, Flask, Docker

•SplitWise Pro – AI-Powered Expense Sharing Platform

Full-stack expense management system with OCR-based receipt parsing and cloud deployment.

- Built a full-stack expense sharing platform for group expense tracking, receipt uploads, balance calculation, and optimized settlement generation.
- Integrated Firebase Authentication with a Node.js/Express backend to secure protected routes and synchronize authenticated users with MongoDB Atlas.
- Implemented AWS S3-based receipt storage with Google Vision API and OpenAI-assisted OCR parsing to extract structured expense details.
- **Tech:** Next.js, TypeScript, Node.js, Express, MongoDB Atlas, Firebase Auth, AWS S3, Google Vision API, OpenAI API

•Estimating Product Length using BERT on Amazon Dataset

NLP-based regression model for structured e-commerce prediction.

- Fine-tuned a pre-trained BERT model using product titles, descriptions, and category metadata to predict product length in a large-scale e-commerce dataset.
- Implemented text preprocessing, tokenization, and feature engineering pipelines to improve model convergence and prediction stability.
- Applied log transformation and model ensembling techniques to enhance regression performance and reduce prediction variance.
- **Tech:** Python, BERT, HuggingFace Transformers, Scikit-learn, Pandas

•Fake Currency Detection Using CNN

Computer vision-based counterfeit currency detection system.

- Developed a CNN-based image classification model using TensorFlow and OpenCV to detect counterfeit currency from a dataset of 10,000+ images.
- Implemented image preprocessing and data augmentation techniques to improve robustness and reduce false positives.
- Optimized training and evaluation workflows to enhance generalization across varied lighting and image conditions.
- **Tech:** Python, TensorFlow, OpenCV, CNN, NumPy

Technical Skills

Programming: Python, Java
Machine Learning & Deep Learning: PyTorch, TensorFlow, Scikit-learn, OpenCV
NLP & LLM Systems: Transformers, HuggingFace, LangChain, Prompt Engineering, Retrieval-Augmented Generation (RAG)
Data Processing: Pandas, NumPy, Feature Engineering, Data Preprocessing
Deployment & Tools: Docker, Flask, REST APIs, Git, GitHub
Databases: MongoDB, PostgreSQL, MySQL
Cloud: AWS (EC2, S3)

Education

•California State University, Northridge *Los Angeles, CA*
Master of Science in Computer Science | ***CGPA: 3.92 / 4.0*** Aug 2024 – May 2026

•Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology *Chennai, India*
Bachelor of Technology in Computer Science and Engineering | ***CGPA: 9.21 / 10.00*** Aug 2020 – May 2024

Experience

•AI Model Evaluation Analyst
Handshake AI Sep 2025 – April 2026

- Evaluated 1,000+ large language model (LLM) outputs using structured quality benchmarks to assess response accuracy, coherence, and factual consistency.
- Identified recurring model failure cases and categorized hallucination patterns, contributing to improved prompt refinement workflows and reducing evaluation inconsistencies by 20%.
- Collaborated with cross-functional teams to provide systematic feedback for improving model behavior across multiple domain-specific evaluation tasks.
- Maintained high annotation agreement standards across large datasets, increasing benchmarking reliability and validation consistency.

•Machine Learning Engineer (Freelance)
Self-Employed Aug 2023 – June 2024

- Designed and deployed deep learning models (LSTM, CNN) for 3+ client applications, delivering production-ready inference pipelines.
- Built end-to-end ML workflows including data preprocessing, feature engineering, model training, and cross-validation, improving model training efficiency by 30%.
- Containerized ML services using Docker and exposed REST APIs via Flask, enabling scalable real-time prediction systems with reduced deployment overhead.
- Optimized data pipelines and training configurations, reducing overall training time by 25% and improving inference latency in deployed systems.

Positions of Responsibility

•Board of Directors – Facilities and Operations Specialist
University Student Union, CSUN Aug 2024 – May 2026

- Oversaw campus infrastructure initiatives impacting 38,000+ students and managed a $5M operational budget.
- Collaborated with executive leadership to improve sustainability, accessibility, and space utilization across university facilities.

•Content Lead
Google Developer Student Clubs Aug 2022 – Apr 2024

- Led a team of 8+ members, organizing workshops and publishing technical content for 200+ participants.
- Mentored contributors in Git/GitHub workflows and coordinated cross-campus technical collaborations.

•Student Volunteer
National Service Scheme (NSS) 2021 – 2024

- Organized blood donation drives, campus clean-ups, and environmental awareness campaigns impacting 1000+ community members.
- Led donation drives supporting orphanages and local social programs.

Licenses and Certifications

Artificial Intelligence and Machine Learning (Foundation and Associate) – AWS
Fundamentals of Deep Learning – NVIDIA
Program Essentials in Python Programming (Intermediate Level) – CISCO
Model Validation 2 Expert – Handshake AI

Updated: July 9, 2026