KANDKKEKAAR SAI TEJ

+916300871001 | saitej13sai@gmail.com | Hyderabad, India | linkedin.com/in/kandkkekaar-sai-tej-728689239

Summary

AI & Machine Learning Engineer specializing in Generative AI, Large Language Models (LLMs), and production-grade AI systems. Experienced in building Retrieval-Augmented Generation (RAG) solutions, domain-specific LLMs, and AI agents for enterprise and healthcare use cases. Strong background in Python, NLP, deep learning, and cloud-native AI deployments across Azure and GCP. Proven ability to deliver scalable, secure, and compliant AI solutions through cross-functional collaboration.

Skills

- **Programming:** Python, SQL
- **AI & Machine Learning:** Machine Learning, Deep Learning, NLP, Generative AI, LLMs, RAG, Vector Search
- **LLM & GenAI Tools:** FAISS, SentenceTransformers, LangChain, Hugging Face, Prompt Engineering, GPT-4o
- **Frameworks:** PyTorch, TensorFlow, Scikit-learn
- **Cloud & MLOps:** Microsoft Azure, Azure OpenAI Service, Google Cloud Platform (GCP), Docker, FastAPI
- **Data & Analytics:** ETL Pipelines, Data Validation, Streamlit
- **Visualization & Analytics:** Tableau, Power BI, Matplotlib, Seaborn
- **Collaboration:** Cross-Functional Coordination, Stakeholder Communication

Experience

Behavior Education Services Team | Remote
AI Specialist | 07/2025 - Present

- Built and deployed a production-ready RAG AI assistant using FAISS, SentenceTransformers, and GPT-4o for semantic document search
- Improved AI response relevance by 40% through optimized embeddings, chunking, and retrieval strategies
- Developed AI agents for healthcare document summarization and clinical insights, reducing manual review time by 70%
- Implemented HIPAA-compliant AI systems with data encryption, role-based access control, and audit logging

Outlier AI | Remote
Software Engineer – AI Data Training | 07/2024 - 06/2025

- Built and automated large-scale AI data training pipelines, reducing processing time by 30% and speeding up model iterations
- Applied supervised learning and data-centric optimization techniques to improve model accuracy, consistency, and reliability
- Developed ML systems for content classification, anomaly detection, and dataset enrichment to support high-quality training data
- Designed feature engineering and data validation pipelines to ensure stable and reliable ML deployments

Epsilon Pi
Machine Learning Specialist | 02/2022 - 05/2024

- Contributed to an AI-powered lung cancer detection system using CNNs and RNNs, improving diagnostic accuracy by 22% and reducing false positives by 18%
- Designed and optimized end-to-end ML pipelines for medical imaging and clinical data using PyTorch and TensorFlow
- Improved model precision by 12% through advanced preprocessing, feature extraction, and hyperparameter tuning

Certificates

AWS Academy Graduate - Machine Learning Foundations, AWS Academy Graduate - AWS Academy Cloud Foundations, Python Programming Essentials - Cisco Networking Academy, Salesforce Developer Virtual Internship - SmartInternz, Salesforce Administrator Virtual Internship - SmartInternz

Education

Vignana Bharathi Institute of Technology | Hyderabad, India

Bachelor of Technology in Computer Science and Data Science

Projects

- **AI Product Manager Assistant**
  Created a feature prioritization assistant using LangChain for agent orchestration, FastAPI for API development, and Docker for deployment—boosting decision-making efficiency for product teams.
- **AI Email Processor for Fashion Retail**
  Built a GPT-4o-powered email classification and response system using Retrieval-Augmented Generation (RAG), improving customer support accuracy and efficiency by 25%.
- **Personalized News Feed Agent**
  Engineered a recommendation engine using Sentence-BERT embeddings and simulated user interactions to deliver adaptive news feeds, increasing user engagement.
- **Voice-Controlled AI with Facial Emotion Recognition**
  Developed a Python-based voice assistant integrated with OpenCV for facial emotion detection, achieving 85% accuracy and enabling emotionally responsive interactions.
- **Customer Churn Prediction with ANN**
  Built a TensorFlow-based Artificial Neural Network to predict customer churn, reducing prediction error by 40% and enabling proactive retention strategies.
- **Weather Forecasting via Time Series Models**
  Applied advanced time series models to enhance forecast accuracy by 35%, leveraging temporal pattern recognition for more reliable weather predictions.