VS

VENKAT SUNKARA

AI Engineer

Phone

+91 7093902986

Email

Venkat.genai264@gmail.com

Education

Bachelor of Technology in Computer Science and Engineering

Acharya Nagarjuna University

Certifications

Generative AI Engineer

MINDX Solutions

Data Science

MINDX Solutions

Machine Learning

MINDX Solutions

OBJECTIVE

Dynamic AI/ML Engineer with 3 years of experience in developing and deploying machine learning models, specializing in fine-tuning, model optimization, and advanced algorithms. Seeking to transition into a Generative AI Engineer role to leverage expertise in deep learning, RAG techniques, and scalable AI solutions. Eager to contribute to the creation of innovative generative AI applications, driving transformative advancements in the field of artificial intelligence.

EXPERIENCE

GENERATIVE AI ENGINEER, STRAIVE, UK, 2022-Present

Worked as a Generative AI Engineer at Straive, focusing on building intelligent solutions using OpenAI, AWS Bedrock, Azure AI, and Google's Multimodal APIs. Integrated real-time APIs to enable dynamic text generation, multimodal processing, and automated decision-making. Fine-tuned domain-specific models with optimized feature engineering and preprocessing techniques. while leveraging cloud services like AWS, GCP, and Azure to build scalable AI/ML pipelines. Applied Generative AI to solve real-world problems like content generation, intelligent automation, document processing. Collaborated with cross-functional teams to ensure AI systems aligned with business goals and user needs and deploying high-performance AI applications in production environments.

PROJECTS

Multimodal Chatbot: Built a chatbot capable of handling both text and image inputs using OpenAI GPT-4 Vision and integrated image processing. Enabled real-time image-text interactions and deployed on Azure with LangChain and Streamlit interface.
Tech Stack: OpenAI GPT-4 Vision, LangChain, Streamlit, Azure

Medical AI Agent: Developed an intelligent medical assistant capable of answering patient queries, analyzing symptoms, and summarizing medical documents using OpenAI GPT-4 and LangChain. Integrated with clinical data sources and deployed on Azure for secure and scalable access.
Tech Stack: OpenAI GPT-4, LangChain, Azure, Streamlit

Multilingual Q&A Chatbot: Created a document-based chatbot with multilingual support (EN/FR) using Azure OpenAI + Weaviate, deployed on GCP Vertex AI.
Tech: LangChain, Weaviate, Azure OpenAI, GCP

Video Subtitle Semantic Search: Designed a semantic subtitle search engine using BERT and TF-IDF to improve video content relevance.
Tech: BERT, TF-IDF, FastText, Streamlit, Langserver

SKILLS

Programming Languages: Python

Frameworks & Tools: TensorFlow, PyTorch, OpenCV, LangChain, LlamaIndex, Streamlit, Flask, NLTK, SpaCy, Hugging Face Transformers, Unsloth, Crew AI, LangGraph, N8N.

Generative AI Technologies: RAGs, Supervised Fine-Tuning, PEFT, Quantization, LORA, QLORA, Transfer Learning, Zero-Shot Learning, Few-Shot Learning, Prompt Engineering, LLMs, GPT Models, Agentic AI, LLMOps.

Vector Databases: ChromaDB, FAISS, Weaviate.

Text Embedding Techniques: Word2Vec, FastText, ELMo, LLM modules.

Text Classification & Sentiment Analysis: Multinomial Naive Bayes, Decision Trees, Random Forest, SVM, CNN, RNN, LSTM.

Advanced NLP Techniques: Transformer Models (BERT, GPT), Sequence-to-Sequence Models, Attention Mechanisms.

Search & Retrieval: Keyword-based Search, Semantic Search Engines, Document Chunking, Cosine Similarity

Deployment Platforms: AWS Bedrock, GCP, Azure, Hugging Face Spaces, NVIDIA Deep Stream, Triton Inference Server

AI/ML Techniques: Machine Learning, NLP, Deep Learning, Fine-tuning with Custom Data, Vector Embedding, Neural Network Optimization, MLOps.

Computer Vision: Image Processing, Object Detection (YOLO, SSD, Faster-R-CNN), Image Segmentation (U-Net, Mask R-CNN), Human Tracking, Facial Recognition, OCR (Tesseract, EasyOCR), Real-Time Video Processing (OpenCV), Model Optimization (ONNX, TensorRT).

Data Science: Data Preprocessing, Feature Engineering, Statistical Analysis, Model Evaluation, Predictive Analytics, Data Visualization.

Tools & Platforms: OpenCV, Detectron2, MediaPipe, Triton, NVIDIA Jetson, Raspberry Pi

Soft Skills: Analytical Thinking, Teamwork, Effective Collaboration.