SAI VARSHITH NUNE
AI/ML Engineer

7093590846 | saivarshitnune@gmail.com | Hyderabad

CAREER SUMMARY

Experienced ML Engineer with 3+ years in NLP, Generative AI, and Machine Learning. Skilled in deep learning, LLMs, and AI-driven solutions. Passionate about building scalable, intelligent systems to solve complex problems.

TECHNICAL SKILLS

Python | NLP | Deep Learning | Machine Learning | Exploratory Data Analysis | Computer Vision | SQL | Transfer Learning | NumPy | Pandas | Scikit Learn | Keras | PyTorch | Generative AI | Vector Database | LLM's | Lang-Chain | RAG | Rest-API

PROFESSIONAL EXPERIENCE

MACHINE LEARNING ENGINEER | MLAI Digital Private Limited | Hyderabad   Oct 2024 – Present

iPal - Multilingual FAQ Chatbot

- Implemented Language Detection Module using a MODERN BERT model to accurately classify user queries as English, Hindi (Devanagari), or Hinglish (Latin script).
- Developed a Translation Module utilizing the NLLBV model to ensure seamless translation between English and Hindi while preserving formatting for HTML, URLs, and code snippets.
- Enhanced Admin Panel to support multilingual FAQ management, including language toggling, translation uploads, and consistent content synchronization.
- Upgraded System Architecture by integrating language detection, translation, and multilingual data management into the chatbot framework, ensuring smooth cross-language interactions.
- Integrated Firestore Database & External APIs to store bilingual FAQs and leverage third-party services for optimized language processing and response translation

AI-Driven Anomaly Detection & Behavioral Analytics for Banking

- Developed an AI-driven anomaly detection system for verifying government-issued ID proofs using Tesseract OCR and Azure OpenAI, ensuring accurate identity verification.
- Optimized bulk image processing, enabling 5,000–10,000 uploads per hour through multi-threaded API calls for enhanced efficiency.
- Built an advanced user behavior analytics model to extract, categorize, and predict transactions from bank statements using LLM-based prompts.
- Integrated both solutions into a scalable FastAPI framework, improving fraud detection, risk assessment, and financial decision-making.

SENIOR SYSTEMS ENGINEER | Infosys BPM Limited | Hyderabad   Dec 2021 – Oct 2024

AI Agent for Intelligent Job Search

- Built an AI agent to extract job listings from Google Jobs using SerpAPI.
- Processed and structured job data into a tabular format for efficient analysis.
- Integrated GPT for personalized job recommendations based on user prompts.
- Automated job search using LangChain's React Agent.
- Secured API key management for seamless integration with SerpAPI and OpenAI.

RAG Insurance Chatbot

- Developed an RAG-based Insurance Chatbot using OpenAI APIs and LangChain to extract policy details from PDFs, ensuring high faithfulness and answer relevance.
- Implemented semantic chunking for accurate retrieval from 100+ documents, optimizing context recall and relevance.
- Integrated Chroma DB with OpenAI embeddings for efficient semantic search and retrieval.
- Enhanced chatbot performance with prompt engineering, response re-ranking, and retrieval tuning, reducing hallucinations and improving answer quality.
- Built a scalable FastAPI backend for query handling and response generation while monitoring metrics for continuous improvement.

Healthcare Intent Mapper Using BERT
- ➢ Built a BERT-based classifier for categorizing medical symptoms into 25 intents from descriptive text.
- ➢ Leveraged transfer learning with a pre-trained BERT model, adding and fine-tuning a custom classification layer.
- ➢ Achieved 99.40% accuracy on the test set after 20 epochs of fine-tuning.
- ➢ Applied to medical chatbot for accurate symptom identification and intent classification.

CERTIFICATIONS

GOOGLE CLOUD CERTIFIED Professional Machine Learning Engineer   2025 – 2027
Please Click to View Certificate

EDUCATION

BACHELOR'S DEGREE | Electrical & Electronics Engineering | BVRIT   2015 – 2019