SURYA TEJ MAKINENI

Wichita, KS, USA makinenisuryatej@gmail.com +1 (316)-223-8927 in/makinenisuryatej

SUMMARY

Results-driven AI/ML Engineer with expertise in machine learning, deep learning, and natural language processing (NLP). Proficient in developing and deploying LLM-powered applications, RAG-based retrieval systems, and predictive analytics models to drive data-driven decision-making. Skilled in Python, TensorFlow, PyTorch, LangChain, Prompt Engineering, Vector Databases and AWS, with experience in MLOps, model optimization, and scalable cloud deployments. Passionate about building production-ready AI solutions that solve complex business challenges and enhance developer productivity.

EXPERIENCE

Machine Learning Engineer

Seva Bharathi Organization **February 2021 - May 2022, India**

- Developed and deployed a machine learning model to analyze user behavior, increasing engagement by 30% through personalized content recommendations.
- Optimized data pipelines for real-time analytics, reducing query processing time by 40% and improving decision-making efficiency.
- Designed and implemented an NLP-based chatbot using Python, TensorFlow, and Flask, automating user inquiries and reducing response time by 50%.
- Built predictive models for donation forecasting using scikit-learn, enhancing funding allocation accuracy by 20%.
- Automated model training and validation using TensorFlow and MLflow, ensuring seamless deployment and monitoring in a production environment.
- Technologies: Python, TensorFlow, scikit-learn, Flask, SQL, Git, Selenium, Postman, Agile Methodologies.

Machine Learning Engineer

Internship Studio **July 2020 - September 2020, India**

- Developed a machine learning model that efficiently forecasted with 95% accuracy whether a customer would opt for a personal loan in the future, leveraging their personal information.
- Executed comprehensive Exploratory Data Analysis on a dataset of over 5,000 customer records, involving data cleansing and transformation of 10+ features.
- Ensured clear communication of insights for business understanding, enabling data-driven decision-making and strategic planning.
- Collaborated with cross-functional teams to understand business requirements and translate them into actionable machine learning solutions.
- Technologies: Python, Scikit-learn, Pandas, NumPy, Matplotlib, Git.

PROJECT

Medical Chatbot

Wichita State University

- Developed and deployed an AI-powered conversational medical chatbot using LangChain and LLaMA, applying sophisticated prompt engineering techniques to generate accurate medical responses.
- Implemented a Retrieval-Augmented Generation (RAG) pipeline, integrating real-time knowledge retrieval to improve response relevance by 40% and reduce misinformation.
- Curated and structured a comprehensive knowledge base from the Gale Encyclopedia, using vector embeddings to ensure responses adhered to medical accuracy standards.
- Optimized vector search with Pinecone, increasing information retrieval speed by 60%, reducing response latency from 1.5s to 600ms.
- Designed and built a Flask-based web application, increasing user engagement by 50% through an intuitive and accessible interface for patients.
- Technologies: Python, LangChain, LLaMA, RAG, Pinecone, Vector Databases, Prompt Engineering, Flask, NLP, AI Chatbots.

AI Powered MCQs Generator

Wichita State University

- Developed and fine-tuned a GPT-3.5-turbo-powered AI model using LangChain and custom prompt templates, automating the generation of high-quality, domain-specific multiple-choice questions (MCQs) with 95% accuracy.
- Engineered prompt optimization techniques to improve result consistency and relevance, including few-shot examples and structured output formatting.
- Created an interactive user interface with Streamlit, enabling real-time feedback for seamless MCQ generation and validation.
- Deployed the application on AWS, optimizing cloud infrastructure to handle large-scale MCQ processing, ensuring 99.9% uptime and scalable performance.
- Integrated real-time data processing and model monitoring, improving response times by 40% while maintaining high system efficiency.
- Technologies: Python, GPT-3.5-turbo, LangChain, Prompt Engineering, Streamlit, AWS, NLP, Machine Learning.

EDUCATION

Master's in Computer Science
Wichita State University (WSU) • Wichita,KS • August 2022 - May 2024
• GPA: 3.97/4.

B.Tech in Computer Science and Engineering
P V P Siddhartha Institute of Technology • Vijayawada, India • July 2018 - May 2022
• GPA: 9.43/10.

CERTIFICATIONS

AWS Certified Cloud Practitioner
May 2024

Graduate Certificate in Computational Data Science
April 2024

SKILLS

Languages: Python, Java, C++, JavaScript, SQL
Frameworks/Tools: PyTorch, TensorFlow, Hugging Face Transformers, LangChain, Streamlit, React, Git, Vector Databases, Pinecone, FastAPI
Concepts: Machine Learning, Deep Learning, Natural Language Processing, Prompt Engineering, Retrieval Augmented Generation (RAG), LLM Fine-tuning, Transformer Models, Computer Vision, Generative AI
Cloud/DevOps: AWS, AWS Lambda, AWS Elastic Container Registry, Docker, CI/CD, AWS SageMaker