Kintur Shah
+1 (682) 307-0426 | shahkintur@gmail.com | LinkedIn | Github | Website

SUMMARY

AI/ML Engineer with over 2 years of experience designing and deploying machine learning models and intelligent systems. Skilled in Python, TensorFlow, and cloud platforms (AWS, GCP, Azure) for scalable, production-ready AI solutions. Adept at integrating models into APIs and automating ML workflows using data engineering best practices.

WORK EXPERIENCE

Resolve Tech Solutions | AI/ML Engineer **Sept-2025 - Present**

- Designed end-to-end AI system architecture integrating Titan embeddings, OpenSearch vector databases, and retrieval-augmented pipelines, enabling semantic search and contextual data enrichment across enterprise sources.
- Engineered an intelligent incident-management framework powered by AWS Bedrock AgentCore and ServiceNow, automating ticket triage and enrichment while ensuring seamless integration across multi-cloud infrastructure.
- Partnered with platform teams to implement observability, latency tracking, and failover for reliable AI service delivery.

Holiday Channel LLC | Data Scientist Intern **June 2025 - Sept 2025**

- Architected LLM-based recommendation pipelines for recipes, gifting, and marketplace personalization leveraging LangChain, vector databases, and RAG workflows.
- Leveraged ML and NLP techniques with GCP Vertex AI to improve personalization, predictive accuracy, and deliver actionable insights through interactive dashboards for business teams.
- Contributed to end-to-end prototyping and deployment of AI/ML analytical solutions to ensure seamless production integration.

UTA Office of Admissions | Data Analyst Intern **Nov 2024 - May 2025**

- Analyzed KPI metrics and performance trends using SQL and Python, providing actionable insights that drove process optimization and informed key business decisions.
- Made interactive dashboards and automated Power BI and Excel reports, visualizing key performance data.
- Collaborated with cross-functional teams to interpret data and enhance performance metrics, contributing to a 15% improvement in operational efficiency.

CSRBOX Foundation | Data Analytics Intern **Feb 2023 - May 2023**

- Optimized and maintained data pipelines in Python and Azure Data Factory to collect, clean, and preprocess large-scale datasets, improving data reliability and analytical consistency.
- Executed Exploratory Data Analysis (EDA) using Pandas and Seaborn; applied time series and predictive modeling with A/B testing, resulting in a 15% increase in forecast accuracy.
- Automated ETL workflows with SQL and PySpark, reducing processing time and enabling quick access to clean, ready data.

PROJECTS

Financial Assistant Chatbot - AI Agent Project [Python, LangChain, GCP, Cloud SQL, Vertex AI, BigQuery]

- Constructed a multi-source data pipeline (structured, PDFs, web-scraped) on GCP for unified financial analytics.
- Modeled document embeddings with Vertex AI and LangChain for intent detection, Natural language to SQL conversion, and transparent data attribution.
- Deployed scalable AI with Cloud SQL (pgvector), Vertex AI, and Gemini LLM, cutting financial lookup time by 30%.

Gemini AI Knowledge Assistant Bot - LLM Project [Python, Gemini API, Streamlit]

- Created an AI-powered conversational assistant using Google Gemini API to generate real-time, context-aware responses through natural language input, leveraging Azure ML Studio for model training, validation, and deployment pipelines.
- Enabled hands-free voice input using the SpeechRecognition module, reducing input friction and increasing accessibility.
- Built with Streamlit, implementing 5+ key features including session-based chat history, reset/download options, and dynamic reruns, while hosting the model on Hugging Face Cloud Spaces for real-time interaction.

MovieFindr - Machine Learning Project [Python, NLP, Hugging Face, Elastic Search, Streamlit]

- Devised a semantic movie recommendation engine using Hugging Face Sentence Transformers and Elasticsearch, improving search relevance by 25% for natural language queries.
- Orchestrated backend workflows for data preprocessing, feature extraction, and model implementation for efficient data pipelines.
- Launched the model through a Streamlit interface, delivering real-time recommendations and a 30% lift in user engagement.

HateScan - Deep Learning Project [Python, NLP, TensorFlow, Keras, BERT, LSTM]

- Implemented a deep learning-based hate speech detection system using LSTM and BERT, achieving 93% accuracy by applying text cleaning, tokenization, and class-balancing techniques.
- Performed end-to-end preprocessing: text cleaning, tokenization, and class balancing for robust model training.
- Applied dropout, early stopping, and hyperparameter tuning within TensorFlow/Keras training pipelines, cutting overfitting by 20%.

SKILLS

Programming Languages: Python, SQL, R, HTML, CSS, JavaScript
Data Analysis & ML Libraries: Pandas, NumPy, Scikit-learn, PySpark, TensorFlow, PyTorch, OpenCV, XGBoost, Scrapy, Statsmodels
LLM & Generative AI Frameworks: OpenAI, Anthophic Claude, Langchain, Hugging Face, Google Gemini, CrewAI, AutoGen
Visualization Tools: Matplotlib, Seaborn, Plotly, Bokeh, Sweetviz, D3.js, Looker, Google Charts, Tableau, Excel, Power BI
Databases: MySQL, PostgreSQL, SQLite, ChromaDB, Neo4j, DynamoDB, Azure SQL Database, Redis
Tools: Git, Github, Supabase, Streamlit, Gradio, Flask, Django, Apache Spark, MLflow, Databricks, Langchain, AWS, MS Azure, GCP

EDUCATION

The University of Texas at Arlington | Master of Science in Computer Science | August 2023 - May 2025 | GPA: 3.83
L.J. Institute of Engineering and Technology | Bachelor of Engineering in Computer Engineering | August 2019 - June 2023 | GPA: 3.70