

Fahad Shaikh

📍 New York City

✉ fs12516n@pace.edu

☎ +1 (329) 204 9984

🔗 Portfolio

in LinkedIn

🐙 GitHub

Education

Pace University — MS in Data Science — GPA 4.0

NYC, Sept 2024 – Jan 2026

Relevant Coursework: Deep Learning, Machine Learning, Algorithms for Data Science, Reinforcement Learning, Natural Language Processing (NLP), Data Mining, Business Data Science, Probability Statistics, Big Data Analytics, Data Visualization.

Experience

Data Scientist *Endeavor Group (IMG Media)*

NYC, Nov 2023 – Aug 2024

- Led the end-to-end transformation of People Analytics for a 500-member HR team across Endeavor Group subsidiaries including UFC, WWE, IMG, and WME, by migrating legacy reporting into Visier and centralizing key workforce metrics for strategic alignment.
- Engineered end-to-end NLP pipelines using Transformer-based models for semantic analysis of resumes and exit surveys, extracting actionable insights to inform HR strategy.
- Automated the machine learning lifecycle from data ingestion to model deployment and monitoring using Azure Data Factory and Azure ML.

Sr Data Consultant *Mu Sigma*

Bengaluru, Apr 2019 – Jun 2022

- Partnered with accounting and sales teams at a prominent financial services firm to develop and deploy advanced Power BI dashboards, enabling effective management of 1 billion dollars in business operations.
- Orchestrated LSTM-based NLP pipelines and BERT embeddings to extract insights from millions of Microsoft Viva support interactions, enabling precise customer segmentation and AI-driven feature development.
- Designed and implemented advanced anomaly detection systems for Citigroup using machine learning algorithms (Isolation Forest, One-Class SVM), enhancing fraud detection accuracy.
- Constructed and operationalized time-series forecasting models (ARIMA, Prophet) to predict key business metrics, improving strategic planning and resource allocation.
- Architected scalable cloud infrastructure on AWS (EMR, S3, Redshift) to support distributed training of machine learning models on terabyte-scale datasets.

Trainee *KPIT*

Pune, Aug 2018 – Mar 2019

- Contributed to the development of an intelligent adaptive cruise control system, implementing core control logic in Embedded C for automotive AI applications.
- Managed technical requirements for automotive ECUs, focusing on functional safety standards for embedded AI systems.

Projects

GPT Shakespeare (Karpathy-inspired)

[GitHub Link](#) 

- Implemented a character-level GPT model from scratch in PyTorch, inspired by Andrej Karpathy, to generate Shakespearean text and demonstrate core transformer mechanics.
- Visualized token-level attention and trained the model end-to-end, highlighting interpretability and creative text synthesis.

Earnings Call Summarizer

[GitHub Link](#) 

- Developed an AI pipeline using the **all-MiniLM-L6-v2** sentence-transformer and FAISS to generate concise, citation-aware summaries of earnings call transcripts.
- Enabled real-time, context-aware financial insights by integrating semantic search and retrieval-augmented generation (RAG) for large-scale document processing.

YouTube Video Summarizer

[GitHub Link](#) 

- Built an AI pipeline that generates concise, context-aware summaries of YouTube videos using transcript analysis, visual cues, and multimodal language models (TinyLlama, sentence-transformers).
- Integrated RAG techniques for real-time, citation-aware summarization and semantic retrieval at scale.

Skills

Programming, ML & AI: Python, C++, SQL, Scikit-learn, TensorFlow, Keras, PyTorch, Pandas, NumPy, OpenCV, LangChain, ANN, CNN, RNN, Transformers, Autoencoders

Cloud, Databases & DevOps: Azure, AWS (S3, Redshift, EMR), Oracle, Snowflake, Databricks, Docker, Kubernetes, Git, CI/CD, JIRA

Visualization: Power BI, Tableau, Matplotlib, Plotly, ChromaDB

Achievements and Certifications

- **Impact Award, LSEG:** Redesigned revenue-risk model, resulting in \$400K annual savings.
- **Spot Awards, Mu Sigma:** Delivered critical solutions for Microsoft and Citigroup, recognized for exceptional contributions.
- **Participation:** Smart India Hackathon, 2017.
- DeepLearning.AI TensorFlow Developer
- Hugging Face NLP Specialization
- Google Advanced Data Analytics
- AWS Certified Machine Learning Foundational