Jenil Shah

323-212-1218 | jenilashah1201@gmail.com | linkedin.com/in/jenilshah11/ | github.com/Jenil311 | Graduating in May 2026

Education

California State University, Los Angeles — Aug 2024 – May 2026
Master of Science in Computer Science ***GPA: 4.0/4.0*** — *Los Angeles, CA, USA*

Charusat University — Jan 2020 – May 2024
Bachelor of Technology in Computer Science and Engineering ***GPA: 3.9/4.0*** — *Changa, GJ, India*

Work Experience

Full Stack Engineer Co-op — May 2025 – **Present**
Drive Health — *Gilbert, AZ, USA*

- Architected a **RAG-based retrieval system** with Pinecone and FAISS to eliminate AI hallucinations, increasing patient-education accuracy by **20%**.
- Refactored inference and retrieval pipelines **to reduce latency from several minutes to 200ms**, supporting real-time medical Q&A at scale with 38% faster response times.
- Transformed patient feedback collection through an **automated Twilio-based LiveKit platform with Prometheus and Grafana analytics**, cutting review time by **45%**.

Research Assistant — Sep 2024 – Apr 2025
California State University, Los Angeles — *Los Angeles, CA, USA*

- Analyzed **5,000+ student records** to identify at-risk populations by developing predictive models in Python (scikit-learn), reaching **85%** accuracy in forecasting dropout risk and facilitating early intervention strategies.
- Addressed low survey participation by designing an optimized data collection framework across **12 departments**, boosting response rates by **39%** through strategic timing and follow-up protocols.
- Bridged research-to-policy gap by translating complex statistical findings into actionable recommendations for faculty, presenting to **20+ stakeholders** and securing **$50K** in retention program funding.

Data Science Intern — Jan 2023 – Jun 2024
TechXi — *Vadodara, GJ, India*

- Developed a **multi-speaker speech recognition system** using deep learning and contextual language models to isolate overlapping voices in noisy environments, delivering **85%** accuracy in separating and transcribing concurrent speech.
- Cut transcription error rates by **30%** through integration of n-gram language models and streamlined audio preprocessing pipelines, enhancing speech-to-text accuracy and system stability.
- Implemented spectral subtraction and MFCC-based denoising combined with **post-processing NLP techniques** to improve transcription quality, sustaining **90%** recognition accuracy in high-noise conditions.

AI-ML Team Lead — Jul 2021 – Jul 2022
Charusat Development Club — *Rajkot, GJ, India*

- Developed a **real-time face-mask detection system** using TensorFlow/Keras and OpenCV to enforce COVID-19 safety compliance during campus events, achieving **95%+ detection accuracy**.
- Trained a **custom CNN model** on curated masked/unmasked datasets with data augmentation and lighting-robust preprocessing, improving robustness across **3+ real-world lighting conditions**.
- Deployed the system at **5+ live club gatherings** to **automate mask-compliance checks**, reducing manual monitoring effort by approximately **40%**.

Projects

Financial Risk Prediction | *Python, Kafka, Spark, FastAPI, Grafana*

- Built a fraud detection solution to identify suspicious transactions using Gradient Boosting and Autoencoder models, reaching **94% accuracy** and lowering false positives by **28%**.
- Enabled real-time fraud detection by deploying Kafka/Spark pipelines and FastAPI models, accelerating processing speed by **35%** with Grafana monitoring dashboards.

Retail Demand Forecasting System | *Python, XGBoost, LSTM, AWS, Airflow*

- Lowered inventory costs using an LSTM and XGBoost forecasting framework on **10M+ transactions**, delivering **92% accuracy** and slashing overstock by **25%**.
- Optimized pricing strategy by creating a dynamic pricing engine with Random Forest and Reinforcement Learning, increasing margins by **12%** while maintaining conversion rates.

Context-Aware Conversational AI System | *PyTorch, Transformers, Redis, FastAPI*

- Enhanced multi-turn conversation quality by **38%** through development of context-aware dialogue platform with transformers and semantic retrieval, outperforming baseline GPT models.
- Scaled architecture from prototype to **10K+ concurrent users** by engineering Redis-cached microservice with async FastAPI, slashing latency from **2s to 500ms**.

Technical Skills

Programming Languages: Python, SQL, R, C++, JavaScript, Scala, Shell, Java
ML/DL Frameworks: PyTorch, TensorFlow, scikit-learn, Keras, XGBoost, LightGBM, Transformers (HuggingFace), CatBoost
Data Science: Pandas, NumPy, Matplotlib, Seaborn, Plotly, OpenCV, SciPy, Statsmodels, ggplot2
NLP & RAG: NLTK, SpaCy, LangChain, FAISS, Pinecone, Sentence-Transformers, Weaviate, Haystack, Anthropic
MLOps & Infrastructure: Docker, Kubernetes, MLflow, Airflow, AWS (S3, Lambda, Glue, SageMaker), Redis, Kafka, Spark, Terraform
Deployment & Monitoring: FastAPI, Flask, Django, Prometheus, Grafana, GitHub Actions, Streamlit, Datadog