

Daniyal Khan

E-mail: daniyalkhan131@gmail.com Phone number: 7300718950

Linkedin: <https://www.linkedin.com/in/daniyal-khan-1389081a5/> Github: <https://github.com/daniyalkhan131>

PROFILE

AI Research Engineer with 1 year of experience in designing and deploying machine learning systems across speech processing, NLP, and generative AI. Skilled in developing end-to-end pipelines that turn unstructured data into useful insights. Experienced in building scalable, cloud-based solutions with a focus on model reliability, data quality, and real-world results. Comfortable working at the intersection of research and production in fast-paced, product-focused teams.

SKILLS

- **Languages & Frameworks:** Python, SQL, PyTorch, TensorFlow
 - **Machine Learning & AI:** Data Analysis, Statistical Modeling, Classical ML, Deep Learning, NLP, Computer Vision, Generative AI
 - **Experimentation & Tracking:** DVC, ClearML, Weights & Biases
 - **MLOps:** Docker, Kubernetes, AWS, CI/CD, Flask, FastAPI, BentoML, Qdrant
 - **Tools & Platforms:** Git, REST APIs
-

EXPERIENCE

AI Research Engineer, Yoyo (Stealth Startup) — Bengaluru

Aug 2024 – Present

Extract customer–salesperson interactions in 8-hour multi-speaker audio captured via wearable devices in retail environments.

Voice Fingerprinting Pipeline (VFP)

- Built a VFP to extract customer interactions using **denoising** (Resemble AI), **VAD** (Silero, PyTorch), **speaker diarization** (Pyannote), **embedding** generation (ReDimNet), and custom clustering with statistical refinements.
- Integrated **voice fingerprint tracking** and optimized **Qdrant** vectorDB retrieval. Improved detection of partial conversations with **greeting detection**, **voice stitching**, **lower-level clustering**, and **contextual padding**.
- Achieved **~60% precision** (complete/incomplete interactions) and **~49% recall** (single-customer cases).

RAG-Based Event Detection Pipeline (EDP)

- Designed an event-centric pipeline inspired by **RAG-style retrieval** and **HyDE-based multi-query prompting**, identifying customer cuts using event signals.
- Automated classifier training leveraging historical brand data, LLMs (**Gemini**, **Qwen** embeddings), and clustering (**RAPTOR**-inspired) for event labeling; and trained a **KNN** classifier for new event categorization.
- Developed an inference pipeline that classifies new VAD transcript into events, merges them into coherent conversations via temporal proximity and key-event sequencing.
- Achieved **+59% improvement in customer coverage** and **66% reduction in personal speech**.

Impact:

- Combined **VFP + EDP pipelines with LLM-based filtering**, boosted **recall to 73%** and **precision to 89%**.
 - Processed **100,000+ hours of production audio**, **scaling from 2 to 8 clients** across **2,000+ retail stores** in India.
 - Transformed a research prototype into a fully deployed, **high-impact product**, **driving measurable business growth**.
-

PROJECTS

Diabetic Retinopathy Diagnosis using CNN with Interpretability

- Preprocessing like resizing, normalisation done, CLAHE and Gaussian filter applied, did Augmentation
- Transfer learning (InceptionV3) used, did hyper-parameter tuning and got SOTA kappa score 0.82
- Employed Grad-CAM and LIME techniques for explaining predictions

Multimodal Sentiment Analysis

- Audio and text modalities used. Used pre-trained embeddings for text and melSpectrogram technique for audio.
 - Transformer used for encoder (extract features from audio, text) and decoder (tensor fusion)
-

EDUCATION

- Masters in Computer Science and Engineering
Jamia Millia Islamia (2022-24)
9.22 CGPA

ACHIEVEMENTS

- Gate: 97.33 percentile