

Hamza Aziz
AI Developer @Global InfoVentures Pvt. Ltd.

hamzaaziz822@gmail.com | +91 8447644188
Sector-23, Sanjay Nagar, Ghaziabad-201002 (U. P.)
LinkedIn | GitHub | Portfolio | Medium



PROFESSIONAL SUMMARY

AI Engineer with **2 years of experience** specializing in **Computer Vision** and **Generative AI**. Proven track record of deploying high-accuracy facial recognition systems and developing conversational assistants leveraging **Rasa, Gemini LLM**, and voice-to-SQL pipelines. Skilled in GPU-accelerated inference (**NVIDIA DGX A100, Jetson Orin Nano, RTX 4090**). Currently pursuing **Executive M. Tech. in AI/ML from BITS Pilani (WILP)** while leading production-grade projects – demonstrating a deep commitment to applied AI innovation and excellence.

SKILLS

- **Programming Languages:** Python, Bash, SQL
- **AI/ML Frameworks:** TensorFlow, PyTorch, InsightFace, YOLO, Scikit-learn, OpenCV, NumPy, Pandas
- **Generative & Conversational AI:** Rasa, Gemini 2.0, Large Language Models (LLMs), NLP, Prompt Engineering
- **Model Optimization & Inference:** TensorRT, ONNX, DeepStream SDK
- **Deployment & DevOps:** Docker, NVIDIA Jetson (Nano, Orin Nano), DGX A100, RTX 4090, Linux, Git
- **Databases & Storage:** Milvus (Vector DB), PostgreSQL, Redis
- **Streaming & Messaging Systems:** Confluent Kafka, Flask-SocketIO
- **Other Tools:** SpeechRecognition, Tesseract OCR, Neo4j, VSCode, PyCharm, Postman

WORK EXPERIENCE

- Assistant Engineer – AI** *Global InfoVentures Pvt. Ltd. | August 2023 – Present*
- Led development of 4 commercial-grade **Facial Recognition Attendance Systems** on **DGX A100, RTX 4090, Jetson Orin Nano**, achieving **>99.99% accuracy**
 - Developed **G6 Voice Assistant**, converting natural language voice queries into precise SQL queries via **Rasa, Gemini 2.0, and Milvus Vector DB**; handled ambiguous queries with advanced intent disambiguation, achieving **near-100% intent classification accuracy**
 - Mitigated domain shift by introducing a two-stage vector database (mobile + CCTV), boosting recognition hit-rate from **85% to 99%** in overhead-angle surveillance scenarios
 - Engineered high-throughput inference pipeline with batched preprocessing and Milvus (GPU-indexed) similarity search, boosting face recognition throughput while maintaining **<20ms latency** per batch
 - Integrated **PostgreSQL** for recognition logs and designed **API sync** with legacy **SIM ERP Portal** for centralized attendance marking and real-time **SMS alert system** for parents
- AI Intern** *Global InfoVentures Pvt. Ltd. | February 2023 – July 2023*
- Contributed as **core developer and internal trainer** for AI/ML training batches across CSE, AIML, and Data Science students, delivering sessions on:
 - Traditional ML algorithms (Logistic Regression, SVM, KNN)
 - CNN Architectures (VGG, ResNet, InceptionNet)
 - Hands-on projects in Computer Vision using YOLOv8

EDUCATION

- Birla Institute of Technology and Science (BITS), Pilani – WILP** 2025 – 2027
Master of Technology (Executive)
- Artificial Intelligence and Machine Learning
- Dr. A. P. J. Abdul Kalam Technical University, ABES Institute of Technology**
Ghaziabad, UP
- Bachelor of Technology** 2019 – 2023
- Computer Science and Engineering

PROJECTS

G6 Voice Assistant – Natural Language to SQL Converter

Global InfoVentures Pvt. Ltd.

Built a robust voice-driven assistant capable of interpreting user speech into SQL queries across multiple institutional modules.

- Integrated **SpeechRecognition**, **Rasa**, and **Gemini 2.0** to convert user voice input into structured SQL queries, enabling seamless access to ERP data from modules like Admission, Employee, Fee, and Library.
- Achieved **~100% intent classification accuracy** using Rasa by training with 50+ voice query variations per module. Implemented **secondary intent resolution** via contextual follow-ups for ambiguous queries (e.g., "Hamza Aziz" as student or employee)
- Provided Gemini with **field-level schema mappings** per module and injected **Milvus Vector DB-based field suggestions** to reduce hallucinations and improve SQL accuracy
- Designed post-processing and summarization logic to convert SQL output into **human-readable, voice-based responses**
- Solved heterogeneity in field values (e.g., department/ designation name variations) using **vector-based field and value disambiguation**, improving precision in SQL generation by **over 40%**.
- Planned future versions with **fine-tuned open-source LLMs** deployed on **NVIDIA DGX A100** to reduce Gemini cost and support **cross-module queries without Rasa**

Facial Recognition Attendance Systems

Global InfoVentures Pvt. Ltd.

Developed and led deployment of 4 AI-powered attendance systems used by educational institutes and enterprises:

- **CCTV-Based Face Attendance System**
 - Deployed on **NVIDIA DGX A100**, handled **20+ live CCTV streams** using `scrfd_10g_gnkps` and `w600k_r50` models (TensorRT Optimized)
 - Used **Milvus (vector DB)** with dual dataset staging (mobile + CCTV) to mitigate domain shift, achieving **>99.99% accuracy**
 - Integrated with PostgreSQL & legacy **SIM Portal ERP** via API
- **PTZ-Classroom Attendance System**
 - Used **rotating PTZ cameras** to periodically capture classroom snapshots, process through the face recognition pipeline, and auto-mark attendance. Helped eliminate roll calls and reduced proxy attendance by **25%**
- **Mobile App-Based Attendance System**
 - Created a low-cost variation allowing faculty to upload classroom photos via mobile app, achieving scalability without PTZ hardware
- **Edge-Based Office Attendance System**
 - Deployed on **Jetson Orin Nano** using **DeepStream SDK**, Raspberry Pi AI Camera, and proximity sensors for real-time recognition. Replaced legacy biometric system with contactless AI-driven attendance tracking

CERTIFICATES

- Generative AI with Large Language Models | Coursera - 2024
- Building Video AI Applications at the Edge on Jetson Nano | NVIDIA Deep Learning Institute - 2024
- Getting Started with AI on Jetson Nano | NVIDIA Deep Learning Institute - 2024
- Artificial Intelligence and Machine Learning | Global InfoVentures Pvt. Ltd. - 2023
- Getting Started with Deep Learning | NVIDIA Deep Learning Institute – 2022