

Umeshkumar Ghaskata

+1 (716) 580-0523 | umeshghaskata2001@gmail.com | [LinkedIn](#) | [GitHub](#) | [LeetCode](#) | USA (Willing to relocate)


University at Buffalo, SUNY (MS in Computer Science)

May 2026

Indian Institute of Technology, Bombay (B.Tech in Aerospace Engineering)

May 2024

PROFESSIONAL EXPERIENCE

DRONES Lab (NASA AIST-21 Project) — ML Engineer [GitHub](#)  USA [Aug'2025-Jan'2026]

- Designed, trained, evaluated, and deployed **DL models** (PCA+MLP, FPGA-based architectures) for hyperspectral image classification improving land-cover recognition accuracy from 82.1% to 92.2%.
- Reduced **edge inference latency** from **2.7s to 32ms** using optimized FPGA Model for real-time output.
- Built cross-platform benchmarking framework across **RTX 4060, Jetson Orin, and Xavier** to quantify deployment trade-offs and guide edge hardware selection.

Roswell Park Research Corporation — Full Stack Developer *Buffalo, USA* [Jan'2026-May'2026]

- Fine-tuned the Gemma 3-1B model on AWS EC2 GPU using QLoRA adapter and **Hugging Face PEFT and deployed to UB server** to build a personalized conversational AI.
- Developed cross-platform mobile applications using **React Native 0.78**, implementing multi-screen navigation, daily check-ins, push notifications, and local storage support for both **iOS and Android**.
- Engineered **CI/CD pipelines** reducing deployment time by **40%**, and optimized **Redux state management** to reduce redundant API calls by **25%**.

Scitara Corporation — Software Developer Intern *Mumbai, India* [Dec'2022-Feb'2023]

- Reduced **manual testing time by 30%** implementing virtual server responses to **mock RESTful API responses**, streamlining testing by JSON responses for GET-POST to **waiving credential** requirement

PROJECTS

Multimodal AI Platform with RAG, Streaming [GitHub](#)  [Jan'2026-Mar'2026]

Transformers, LangChain, FAISS, CLIP, Kafka, PySpark, MLflow, FastAPI, Docker, AWS/GCP

- Built a **multimodal AI system** integrating **RAG pipelines**, **Agentic Workflow**, real-time streaming (Kafka), and big data processing (PySpark) for context-aware reasoning across text, documents, and images
- Designed an **end-to-end retrieval and inference pipeline** using CLIP embeddings, FAISS vector search, and LLMs with re-ranking, implementing **MLflow-based** model lifecycle management with automated retraining triggered by data and model drift

Deepface-Similarity-Engine | Image Processing [GitHub](#)  [Jul'2025-Sep'2025]

TensorFlow, Keras, OpenCV, MTCNN, VGGFace (ResNet50), NumPy, Streamlit

- Built a facial recognition system using **MTCNN** for face detection and **VGGFace (ResNet50)** for feature extraction, generating 2048-d embeddings for identity classification
- Developed an end-to-end pipeline with **TensorFlow** and **NumPy**, and deployed an interactive interface using **Streamlit** for real-time image upload and inference

GenAI Full Stack Interview System | AI-Powered Web Application [GitHub](#)  [Jan'2026-Mar'2026]

React.js, Node.js, Express.js, MongoDB Atlas, Google Gemini, JWT, Zod, Puppeteer

- Developed an AI-powered interview preparation platform using **React.js**, **Node.js**, **MongoDB Atlas**, and **Google Gemini**, enabling resume-JD matching, automated skill-gap analysis, ATS-friendly resume generation, and personalized technical interview roadmaps.
- Engineered a production-grade backend with **structured LLM outputs (Zod schema validation)**, JWT authentication, token blacklisting, PDF parsing, and dynamic interview question generation.

TECHNICAL SKILLS SUMMARY

Languages	Python, React-Native, C++, JS, SQL
ML/AI Framework	PyTorch, NLTK, Scipy, TensorFlow, Scikit-learn, Transformers, LangChain, LlamaIndex, RAG, LLM Fine-Tuning, Multi-modal Learning, OpenCV, PySpark, SFT, XGBoost, Hugging face, AXLearn
MLOps/Deployment	Automation, Docker, CI/CD, FastAPI, Kafka, Model Deployment, AWS, GCP, Algorithm, CMOS, CCNA
Database	Redis, FAISS, PostgreSQL, MySQL, Firebase, REST APIs, Distributed Systems, Streaming Pipelines
Certifications	Generative AI Leader(Google), Cloud Computing

SCHOLASTIC ACHIVEMENT

- Ranked in the **top 0.7% of 1.3+ million candidates** in national Joint Entrance Examination [2020]