Kanan Vyas

vyaskanan98@gmail.com — +91-8780772018 — github.com/KananVyas
— https://www.linkedin.com/in/kanan-vyas27/ — https://medium.com/@kananvyas

SUMMARY

AI Engineer with 6+ years of experience building scalable production-ready ML systems with a strong focus in Computer Vision, NLP and Large language model (LLM) applications.

TECHNICAL SKILLS

- **Languages**: Python, C++
- **Deep Learning**: PyTorch, TensorFlow, OpenCV, DeepStream, ONNX, Transformers
- **LLMs**: HuggingFace, LangGraph, LangChain, LangFuse, Prompt Engineering, RAG
- **MLOps**: Docker, FastAPI, Flask, MLflow, REST APIs, AWS, Azure, GCP, Git, CI/CD, Kubernetes, Distributed Systems
- **Optimization**: TensorRT, OPENVINO, Nvidia Triton

PROFESSIONAL EXPERIENCE

Machine Learning Engineer, Reliance Jio, Bangalore   **Aug 2022 – Present**

- Developed smart AI agents for cab and grocery automation using **LangGraph** and **FastAPI**. Designed an Agent Evaluation Framework using **LangFuse** to cater to 15k requests / day.
- Conducted research on entropy-based selection algorithms to improve reasoning capabilities in small-scale **LLMs**.
- Designed and trained invoice information extraction models using transformer-based multi-model architectures, including **LayoutLMv3**.
- Improved invoice field extraction accuracy from **55% to 80%** by training LayoutLMv3, reducing human verification by 40%.
- Pre-Training of **LayoutLMv3** model using **BERT**-based strategies such as masked image modeling.
- Finetuned **CLIP** for text-based visual search systems for fashion clothing search. Improved recall@1-match from 25% to 56%.
- Developed a video analytics-based worker safety solution for Reliance Jamangar plant, focusing on PPE Compliance and Fire/Smoke/Person detection using **YOLO** and **ResNet**-based architectures.
- Designed end-to-end scalable pipelines using **NVIDIA DeepStream**, capable of processing 200+ cameras with 25 FPS.
- Achieved 93 Cr/year cost savings for scalability up to 10K cameras, also the project received - **The Gulf Energy Award - 2024**

Machine Learning Engineer, Zoop.One, Pune   **Jun 2021 – Jul 2022**

- Developed a framework to check for Active Facial Liveness using **HRNet** based on facial pose estimation.
- Data Curation for facial spoofing activity from mobile-phones, light reflections mainly dedicated to Indian faces.
- Trained a **ResNet** based classification model for facial spoofing recognition. Improved the accuracy from 64% to 89%.
- Created synthetic data pipelines for Indian ID cards (**Aadhar**, **PAN**, **Passport**).
- Developed and integrated the REST APIs with CRUD operation and served 5k+ requests/day throughput to various clients across India.

Deep Learning Engineer, Eagle Eye Networks, Bangalore   **Jun 2019 – Jun 2021**

- Research and optimization of Object Classification, Detection, and Segmentation models, including **HRNet**, **U-Net**, and custom classifiers for recognition of faces, people, heads and actions.
- Developed end-to-end multi-container **Docker** pipelines and deployed them on edge devices for real-time inference.
- Performed model optimization using **Intel OpenVINO** and **NVIDIA TensorRT** to enhance inference speed and efficiency on edge hardware.

- Led multiple proof-of-concepts (POCs) and product demonstrations that showcase the deployment and performance of edge AI solutions.

ML Intern, Eagle Eye Networks, Bangalore   **Jan 2019 – Apr 2019**
- Built text and license plate recognition pipeline.
- Developed facial expression analysis for retail.

EDUCATION

B.Tech in Information and Communication Technology   Ahmedabad University
2015 – 2019   CGPA: 3.15 / 4.33

ACHIEVEMENTS

- Technical blogs on Medium that gained more than 2K+ Reads.
- Finalist - TechGig Geek Goddesses 2019 (Rank 114 All India)
- Top 8 - HackNight2.0 by Microsoft
- Winner - Ingenious Hackathon 2017 for ColdBox
- Top 50 - Sayaji Startup Summit 2018
- Founder - Women in Tech Community (Clique Organization)
- Webmaster - IEEE Student Branch, Ahmedabad University