Mudit Rastogi

Gurugram, Haryana, India | +91 8826383466 | ✉ mudit.rastogi.cv@gmail.com

🔗 LinkedIn: Mudit Rastogi

Professional Summary

Experienced AI Lead with 8+ years in building and deploying scalable AI/ML solutions across NLP, Computer Vision, and Generative AI. Proven track record in driving innovation, reducing operational costs, and deploying deep learning systems in production environments. Passionate about cutting-edge AI, including LLM fine-tuning, speech systems, and edge-based intelligence.

Professional Experience

Lead AI Engineer | Bobble AI

Gurugram, India | Feb 2021 – Present

- Led AI initiatives focused on Generative AI, Large Language Models (LLMs), and Vision-based models; reduced ops costs by **76%** and improved delivery speed by **30x**.
- Built **multilingual vector-based content search** using FAISS and HuggingFace Transformers, improving relevance by **71%**.
- Designed and deployed robust AI pipelines with **PyTorch**, **TensorFlow**, **ONNX**, **AWS/GCP**, and **Docker**.
- Developed **Disney-style GAN** models for real-time cartoon generation, handling 200K+ daily requests with 95% faster inference using TensorRT optimizations.
- Engineered speech-to-text and NLP systems, enhancing WER by **17.7% (Hindi)** and **19% (Bengali)** through acoustic and language model tuning.
- Fine-tuned **LLMs (LLaMA 2, Mistral)** on proprietary dialog datasets for content moderation, grammar correction, and response suggestion; deployed using **FastAPI** + **Quantized Models** for edge devices.
- **Awarded the Bobble AI Innovation Award (2024)** for outstanding contributions in Generative AI applications.

Chief Technical Officer | Astute Resolutions

Gurugram, India | Nov 2018 – Jan 2021

- Created edge-based AI video analytics for Smart Cities using **OpenCV**, **YOLO**, and **TensorFlow Lite**, enabling real-time traffic detection.
- Developed a **GIS data processing pipeline** using **GDAL** and **NumPy**, reducing data processing time by **55%**.
- Partnered with GIS providers to automate map annotation and asset detection through instance segmentation models.

Machine Learning Engineer | Attentive AI

New Delhi, India | Feb 2018 – Oct 2018

- Automated geospatial feature extraction (buildings, roads) using **UNet** and **ResNet** variants in **PyTorch** and **TensorFlow**; improved throughput by **63%**.
- Streamlined large-scale satellite image labeling pipelines for civil and urban infrastructure datasets.

Trainee Software Engineer | Voxta Communication Pvt. Ltd

Hyderabad, India | Jun 2017 – Jan 2018

- Built context-sensitive NLP spell-check and phoneme classifiers using **BiLSTM** and **MLP**, focused on Indian languages.
- Developed interactive dashboards with **Flask** for real-time phonetic data inspection.

Education

B.Tech – Computer Science & Engineering
Krishna Institute of Engineering and Technology, Ghaziabad
CGPA: 79.8%

Key Skills

- **Languages & Frameworks:** Python, PyTorch, TensorFlow, Keras, OpenCV, GDAL, NumPy, Pandas
- **Generative AI & LLMs:** GANs, Stable Diffusion, Transformers, PEFT, LoRA, QLoRA, LangChain, HuggingFace
- **Deployment & MLOps:** Docker, Kubernetes, FastAPI, Flask, AWS (SageMaker, EC2), GCP (Vertex AI), Firebase
- **NLP & Speech:** NER, Speech-to-Text, ASR, Translation, WER Optimization, Multilingual Models
- **Vector Search & Embeddings:** FAISS, Pinecone, Sentence Transformers, Chroma
- **Tools:** Git, Jupyter, Postman, TensorBoard, Weights & Biases

---

Patents & Publications

- **Patent:** *System and Method for Personalized Cartoon Image Generation* – US20240046536A1
  🔗 View Patent
- **Publication:** *Low-Light Image Enhancement using Conditional GANs* – IEEE ICSC
  🔗 View Paper