Lakshya Kaushik

lakshya.kaushik811@gmail.com | +91 9456007616 | github.com/lakshya811 | linkedin.com/in/lakshya-kaushik

Summary

AI & Machine Learning Engineer with hands-on experience in designing, training, and deploying production-grade deep learning models and LLM-powered applications. Proven ability to drive significant operational improvements, achieving up to a 92% accuracy boost in data extraction and a 70% reduction in manual workflow effort. Expertise in the end-to-end MLOps lifecycle using PyTorch, TensorFlow, Docker, and cloud platforms including GCP and Azure.

Professional Experience

- **Geeky Bee AI Pvt Ltd** — Pune, India
  AI Developer — *Jan 2025 – Present*
  - Enhanced a critical OCR pipeline for Arabic documents by integrating custom TensorFlow and OpenCV preprocessing models, boosting text extraction accuracy from 78% to 92%.
  - Spearheaded the automation of web data extraction workflows using Playwright and Scrapy, decreasing manual data entry and validation effort by 70% and accelerating dataset creation for LLM pre-training.
  - Architected a Retrieval-Augmented Generation (RAG) QA system indexing 5,000+ technical documents; reduced engineer query resolution time by an estimated 30% using FAISS for sub-second vector search.
  - Engineered and deployed a production-grade sentiment analysis service using FastAPI and Docker, processing over 10,000 customer reviews per month with 85% accuracy, providing key insights for product strategy.
  - Optimized a real-time object detection model (YOLOv12) on an NVIDIA Jetson device, achieving a stable 25 FPS for edge deployment by leveraging PyTorch performance tuning techniques.
  - Collaborated on establishing MLOps best practices, implementing CI/CD pipelines with Git and experiment tracking with MLflow to improve model reproducibility and deployment velocity.

- **Symbiosis Center for Applied AI** — Pune, India
  Research Intern — *Jul 2024 – Dec 2024*
  - Led a research initiative on multimodal fusion for disease prediction, integrating retinal images and clinical data for over 5,000 patient samples, which increased model AUC from 0.82 to 0.90 (an 8% relative improvement).
  - Benchmarked early, late, and hybrid fusion architectures, providing a data-driven recommendation to proceed with a hybrid model for deployment due to its superior performance and computational efficiency.
  - Re-engineered data loading and augmentation pipelines in PyTorch, resulting in a 40% improvement in model training throughput and enabling faster experimentation cycles.

- **Anyology Speciality Services Pvt Ltd** — Remote
  IT Consultant Intern — *Aug 2023 – Nov 2023*
  - Led a team of three to migrate and host the company's WordPress infrastructure on Google Cloud Platform (GCP), resulting in an annual hosting cost reduction of over $1,000.
  - Researched and presented cloud migration strategies to senior leadership, contributing directly to the firm's long-term IT infrastructure roadmap.

Technical Skills

Languages: Python, SQL

ML/DL Frameworks: PyTorch, TensorFlow, Scikit-learn, Keras, XGBoost, Hugging Face Transformers

LLM & Generative AI: RAG, Fine-tuning, Prompt Engineering, LangChain, OpenAI API, Vector Databases (FAISS, Pinecone)

MLOps & Cloud: Docker, Kubernetes, MLflow, CI/CD (GitHub Actions), GCP, Azure, AWS (Basic)

Databases & Data Processing: Pandas, NumPy, OpenCV, Spark, FastAPI, Flask, Scrapy

Projects

- **IntelliShots - News Summarization App:** Developed a full-stack application leveraging T5 and GPT models for abstractive news summarization. Engineered a RAG pipeline to provide sources for summarized content, and built a responsive frontend with React.js and Node.js, handling 100+ daily active users during testing.
- **End-to-End Sentiment Analysis Microservice:** Architected and deployed a BERT-based sentiment classification API using FastAPI and Docker. Established a full CI/CD pipeline with GitHub Actions for automated testing and deployment, achieving 85% accuracy on a custom dataset of 5,000 product reviews.
- **AgNOR Detection in Medical Images:** Designed and trained a custom CNN in PyTorch for detecting AgNOR protein clusters in histopathology images, a key marker for cancer diagnosis. Implemented advanced OpenCV preprocessing techniques that improved model F1-score by 12%.
- **Scalable Web Scraper for LLM Datasets:** Constructed a robust Scrapy pipeline to extract and structure data from over 50 book-selling websites. Integrated with Pandas and a SQL database to create a clean, 100,000+ entry dataset used in downstream LLM pre-training workflow.

Education

- **Symbiosis Institute of Technology** — Pune, India
  B.Tech in Artificial Intelligence and Machine Learning — *Sep 2021 – Jun 2025*
- **Technische Hochschule Ingolstadt** — Ingolstadt, Germany
  Exchange Semester, Computer Science and Artificial Intelligence — *Mar 2024 – Sep 2024*
  - Collaborated with faculty on advanced projects in NLP, recommender systems, and computer vision.
  - Gained proficiency in Prolog and advanced AI methodologies, expanding technical expertise beyond the standard curriculum.