YASH KUMAR

LinkedIn
yss107.github.io
jangidyash590@gmail.com
GitHub

Experience

***Alignerr* (Remote)**
Machine Learning Engineer — *Jan 2025 – May 2025*

- Designed and automated **Python-based data validation and quality assurance pipelines** for large-scale datasets (10k+ samples/week), ensuring schema compliance and reliability for **LLM fine-tuning and evaluation workflows**.
- Built and evaluated **Retrieval-Augmented Generation (RAG) pipelines**, designing adversarial and edge-case test sets to identify retrieval failure modes, resulting in an estimated **20% reduction in hallucinations**.
- Supported **production LLM workflows** by performing high-precision data annotation and QA using **CVAT and Labelbox**, maintaining **98%+ ground-truth accuracy** for supervised and preference-based learning.
- Analyzed **inter-annotator agreement (Fleiss' Kappa)** and collaborated with senior ML engineers to refine labeling guidelines, improving dataset consistency by **~15%** and downstream model stability.
- Worked cross-functionally with ML and infrastructure teams to ensure **datasets, schemas, and evaluation metrics** aligned with **production inference requirements**.

***Outlier AI* (Remote)**
AI Trainee — *Aug 2024 – Feb 2025*

- Optimized **SQL-based data extraction and ETL pipelines**, reducing processing latency by **65%** and accelerating iterative **NLP model training and evaluation cycles**.
- Curated and engineered **50,000+ multilingual samples** for **RLHF (Reinforcement Learning from Human Feedback)** pipelines, directly supporting instruction-tuned and preference-optimized LLMs.
- Developed **automated data cleaning and validation scripts** using Pandas and NumPy to detect statistical anomalies, improving dataset robustness and model generalization.
- Defined and enforced **strict JSON schemas** for ML data outputs in collaboration with core engineering teams, ensuring compatibility with **downstream training, evaluation, and inference systems**.
- Assisted in improving **data reliability and pipeline scalability** for production-facing ML systems.

Projects

Math Problem Solver with Real-time OCR — *Aug 2025 – Present*

- **Engineered** a real-time Optical Character Recognition (OCR) pipeline by integrating **EasyOCR** and **Tesseract**, achieving **90%+ accuracy** on complex mathematical expressions.
- **Reduced** inference latency to **<200ms** by implementing aggressive image preprocessing (adaptive thresholding, grayscaling) and deploying the optimized model via **Docker**.
- **Architected** a scalable **Django** backend integrated with **WebRTC**, enabling live, low-latency video stream processing for instant solution generation.
- Deployed OCR inference pipeline using Docker on AWS EC2, exposing REST APIs for real-time prediction.

Smartshop — Personalized Grocery Recommendation System — *Jan 2023 – May 2023*

- **Developed** a collaborative filtering model (SVD & Matrix Factorization) tailored for retail transaction data, achieving a **78% Top-N accuracy rate**.
- **Optimized** real-time inference endpoints by implementing **Redis** caching and a **FastAPI** wrapper, supporting a simulated load of **10,000 DAUs** with **<50ms latency**.
- **Conducted** rigorous hyperparameter tuning and benchmarking against NMF and Neural Network baselines to select the most performant model architecture for production.

Education

Bachelor of Science in Information Technology
Lovely Professional University, Punjab June 2021 – July 2024

Skills

- **Languages:** Python (Expert), SQL, C/C++, Java, R, Node.JS, HTML.
- **Machine Learning & NLP:** PyTorch, Transformers (Hugging Face), LangChain, RAG, LLM Fine-tuning, Scikit-learn.
- **MLOps & Cloud:** Docker, AWS (EC2, S3), Git/GitHub Actions, Linux.
- **Data Engineering:** Pandas, NumPy, Vector Databases, MongoDB.

Certifications

- LangChain & Vector Databases in Production – Activeloop (Jun 2024)
- Supervised Machine Learning: Regression and Classification – DeepLearning.AI (July 2023)
- Exploratory Data Analysis for Machine Learning – IBM (Aug 2023)
- Geoprocessing Using Python – Indian Space Research Organization (ISRO) (Aug 2023)