YASH PAREEK

github.com/pareek-ml • yash.pareek@usi.ch

Professional Experience

Junior Machine Learning Engineer — Bengaluru, IN
Perimattic — Jan. 2023–Aug 2024

- Led a cross-functional team to fine-tune LLAMA 2 on enterprise data, achieving a +70% reduction in response times.
- Directed the deployment of a RAG-LLM chatbot across 10 airports, utilizing advanced NLP techniques and prompt engineering to achieve 80% response accuracy.
- Implemented A/B testing for model validation, ensuring robustness and reducing response latency by 15% in production.
- Engineered an on-premises LLM with biometric authentication, integrating Kubernetes for scalable deployment and Docker for containerization.
- Leveraged model versioning with MLflow to manage updates, enhancing security and scalability for datasets exceeding 15GB.
- Delivered a Machine Learning pipeline using NeuralProphet and TimescaleDB for real-time hourly forecasting, including data preprocessing, model training, and dashboard visualization.

Machine Learning Developer — Jaipur, IN
Freelance — Mar. 2022–Jul. 2023

- Engineered a YOLO-based License Plate Recognition system with Bayesian Optimization to fine-tune hyperparameters, reducing processing time from 45 to 15 seconds.
- Deployed a transformer-based NLP model for automated summarization of sales and customer service call data, enhancing data analysis efficiency and decision-making processes.
- Architected a Face Recognition Service on AWS, with custom Liveness Detection model on SageMaker, resulting in a 30% faster inference time, without compromising precision.
- Created a daily news summarizer using Transformers, achieving a 5x compression
- Increased monthly sales by 30% by deploying a sentiment analysis model on Amazon product reviews, identifying key customer preferences and optimizing product pricing

Projects

- Architected real-time voice-driven chat interface for hands-free conversations with large language models using Fine Tuned Whisper and Coqui TTS.
- Enhanced Vicuna, Llama-13B LLMs performance by applying QLORA and PEFT techniques on a carefully curated dataset, resulting in more accurate and contextually relevant outputs.

Education

MSc Artificial Intelligence — Sept. 2024–Present
USI Lugano, Switzerland

BTech. Computer Science — Aug. 2018–Jul. 2022
Rajasthan Technical University — GPA: 7.84/10.0

Skills

- Python, DVC, Git, CI/CD - MLOps, TensorFlow, PyTorch, HuggingFace, NumPy
- Statistical Analysis, Data Processing, Data Visualization
- English: Full Professional Fluency