ANKUSH PATIL

+91-9822308926 | ankpatil1203@gmail.com | Website | linkedin | Github

Machine Learning & AI Engineer specializing in deep learning, NLP, computer vision, and LLM systems, with hands-on experience building production-grade RAG pipelines, fine-tuned transformer models, and scalable ML services.

Technical Skills

Core Domains: Machine Learning, Deep Learning, NLP, Computer Vision, LLMs, RAG, Model Evaluation.
Models & Techniques: Transformers, CNNs, Fine-tuning (LoRA), Embeddings, Sentence Transformers, Attention Mechanisms.
Frameworks & Libraries: PyTorch, Hugging Face, FAISS, scikit-learn, OpenCV, NumPy, Pandas.
MLOps & Deployment: Docker, FastAPI, GitHub Actions, MLflow, Model Serving, Logging & Monitoring.
Languages: Python, SQL, C++.

Academic Details

Year	Degree / Board	Institute	Marks/GPA
2021	B.E. Civil Engineering	K.K. Wagh I.E.E.R., Nashik	CGPA: 7.88
2017	HSC (Science)	K.T.H.M College, Nashik	73%

Work Experience

Site Engineer **Feb 2023 – Jan 2026**
Charwak Engineers & Developers, Nashik *Onsite*

- Managed construction workflows ensuring adherence to specifications and safety standards under tight execution constraints.
- Developed analytical thinking and structured problem-solving skills in high-pressure, real-world environments.

Projects

Legal RAG System using LLMs (EUR-Lex) [CODE] *Nov 2025 – Jan 2026*

- **Tech:** PyTorch, FAISS, SentenceTransformers, FastAPI, Docker, Hugging Face Hub
- Implemented FAISS-based dense retrieval with sentence embeddings over 57K+ EUR-Lex documents, reducing query latency and improving answer relevance in production RAG pipelines with FastAPI-based inference service.
- Designed comprehensive evaluation framework (Recall@K, Precision@K, MRR, latency analysis) and deployed Dockerized system with stateless API design, structured logging, and artifact hosting on Hugging Face Hub.

Wikipedia RAG + LLaMA Finetuning [CODE] *Sep 2025 – Nov 2025*

- **Tech:** PyTorch, LLaMA, LoRA, RAG, FAISS, SentenceTransformers
- Built production-ready RAG system combining FAISS vector retrieval with LoRA-finetuned LLaMA, achieving measurable improvements in factual QA accuracy and answer grounding.
- Demonstrated consistent training loss reduction and improved coherence through systematic evaluation and iterative model refinement.

Transformer English→Hindi Translation (From Scratch) [CODE] *Aug 2025 – Sep 2025*

- **Tech:** PyTorch, Custom Attention, Positional Encoding, Masking, Beam Search
- Implemented complete Transformer architecture from scratch with custom attention mechanisms and beam search, achieving BLEU 49.76 and deployed on Hugging Face for reproducibility.

Implementations from Scratch [CODE] *Aug 2024 – Jul 2025*

- **Tech:** NumPy, PyTorch, Custom Model Implementations
- ML: Linear/Logistic Regression, Naive Bayes, Oblique Trees, GMM, PCA, L1-SVM, Bagging, XGBoost.
- DL: Neural Networks, CNNs, RNNs, LSTMs, Embeddings, Word2Vec.

Plant Disease Detection (ResNet18) [CODE] *Apr 2025 – Jun 2025*

- **Tech:** PyTorch, ResNet18, OneCycleLR, Grad-CAM
- Built ResNet18 classification pipeline with Grad-CAM explainability across 38 disease classes, enabling interpretable predictions for agricultural applications.

Amazon Reviews NLP Pipeline [CODE] *Jul 2025 – Aug 2025*

- **Tech:** Python, TF-IDF, Naive Bayes, BART, Hugging Face Transformers
- Developed large-scale NLP pipeline processing 3M+ reviews with sentiment analysis, aspect extraction, and BART-based summarization.

Other ML Projects [CODE] *Aug 2024 – Mar 2025*

- **Applied Projects:** Credit card fraud detection (Isolation Forest, PCA, UMAP, GMM, DBSCAN), Student Performance Prediction (PCA, SVM, Gradient Boosting, Stacking), Spam SMS Classification (TF-IDF, Naive Bayes).

Leadership & Activities

Football Departmental Team Vice Captain: Led the departmental football team at K.K. Wagh I.E.E.R., Nashik during B.E. degree.