

Kabyik Kayal

📍 Kolkata, West Bengal, India ✉ mail@kabyik.dev 🔗 Kabyik.dev in Kabyik-kayal 🎧 Kabyik-Kayal

Projects

UnArxiv

[Github](#) 

- Engineered a memory-efficient fine-tuning pipeline for Qwen 2.5-3B-Instruct on Intel Arc A750 (8GB VRAM), utilizing Intel Extension for PyTorch (IPEX) and LoRA adapters (Rank 2) to overcome XPU hardware constraints while maintaining bfloat16 precision.
- Architected a Knowledge Distillation pipeline using the Groq API (Kimi K2) to synthesize a custom instruction-tuning dataset, enabling a 3B parameter "Student" model to achieve +12.5% ROUGE-1 score improvement over the base model.
- Built a fault-tolerant training orchestration system with subprocess isolation and automatic checkpoint resumption, solving critical VRAM fragmentation issues inherent to the Intel XPU stack during long-running training sessions.
- Deployed the fine-tuned model via a FastAPI backend with Server-Sent Events (SSE) for real-time streaming, reducing output complexity by 2 grade levels (Flesch-Kincaid) to make scientific literature accessible to a 7th-grade audience.

Supa Voice Cloner

[Github](#)  — [Live](#) 

- Developed a multilingual voice cloning app using cutting-edge XTTS v2 for natural, emotion-rich speaker adaptation in 8+ languages with as minimum as 3 secs of voice sample.
- Designed an intuitive, fast UI for real-time speech generation, with support for custom voice uploads and regional dubbing.
- Deployed on Hugging Face Spaces for instant access with GPU acceleration, ensuring scale and global accessibility.
- Integrated post-processing pipeline with audio normalization and download links.

Anime Recommender System

[Github](#)  — [Live](#) 

- Trained a Tensorflow Keras Model with a mean squared error of 0.089 incorporating live experiment tracking with CometML for Recommending Animes with a Hybrid System of both User-based and Content-based filtering.
- Data versioning with DVC and GCP Bucket was performed to ensure 0% data loss.
- Built a Flask app with an attractive UI with HTML, CSS and JavaScript. Then Docker Containerized the app and deployed to Google Kubernetes Engine ensuring 100% uptime and scalability.
- Built a Jenkins Pipeline to automate the complete process from data ingestion to deployment on cloud saving more than 70% of time and manual effort.

Technical Skills

Languages: Python, SQL, GO

AI/ML: Scikit-learn, Pandas, MLflow, PyTorch, TensorFlow, LangChain, HuggingFace, Transformers

MLOps: Git, DVC, Docker, ZenML, Jenkins, GCP, AWS, Circle CI, Prometheus, Grafana, Kubernetes, FastAPI

Data Visualization: Power BI, Tableau, Matplotlib, Plotly, Excel

Education

Indian Institute of Technology Madras (IITM)

Jun 2024 – Present

BS in Data Science and Applications

- Currently Pursuing Diploma in Programming
- **Coursework:** DBMS, DSA, App Dev, Python, Statistics, Mathematics