Sanatan Khemariya

📞 7898009986 — ✉ sanatankhemariya@gmail.com — 🔗 linkedin.com/in/sanatank — 🐙 github.com/sanatren

Education

Jaypee University of Engineering and Technology — **August 2022 - Present**
Bachelor of Technology in Computer Science and Engineering — *Guna, India*

IIT Guwahati (Certification Program) — **Jan 2025 - Present**
Credit-Linked Program in Data Science — *Guwahati, India*

Key Courses: Operating Systems, Computer Networks, Data Structures, Algorithms, Machine Learning, Databases

Certifications
– **Kaggle** – Pandas for Data Analysis | **Link**
– **Kaggle** – Generative AI Intensive | **Link**

Skills

Frameworks	FastAPI, LangChain, OpenAI SDK	**OS**	MacOS, Windows
Libraries	PyTorch, TensorFlow, Scikit-learn, cv2, NumPy, Pandas	**Databases**	PostgreSQL, MySQL
Languages	Python, C++, Swift	**Tools**	Xcode, Git, GitHub, Cursor AI, Google Colab, Hugging Face, Vector DB, Exploring vLLM

Projects

Azazel - Supercharging Legacy Models — **LLMs, OpenAI, FAISS, LangChain, Streamlit**
— **Project Link**
– Built an AI toolkit that transforms legacy language models into multimodal assistants capable of text generation, image analysis, code execution, and real-time web search
– Integrated GPT-4o-mini vision with LangChain RAG and FAISS vector storage, improving response accuracy by 25% on complex user queries
– Reduced API costs by 30% through efficient prompt engineering, dynamic model switching, and caching strategies
– Developed an interactive Streamlit interface with multilingual support, speech-to-text input, document chatting, and advanced contextual memory for a seamless user experience

Easy-Notes - Effortless Writing — **FastAPI, TensorFlow, Transformer, Supabase**
— **Project Link**
– Designed a full-stack AI-powered note-taking platform that enhances writing with smart text suggestions and translation
– Trained an LSTM-based next-word prediction model achieving 66% accuracy for an improved writing experience
– Developed and fine-tuned a Transformer-based English-to-Hindi translation model reaching 94.6% translation accuracy
– Optimized PostgreSQL database architecture with Supabase, reducing query response times by 25%
– Engineered RESTful API endpoints with FastAPI, Uvicorn & Jinja2 for seamless frontend-backend integration
– Implemented low-latency ML model serving with average prediction time under 100ms for real-time text suggestions

Legal Document Analyzer — **BART, Transformers, PPO RL, Hugging Face**
— **Project Link**
– Built an NLP system for legal summarization using a fine-tuned BART model and a custom transformer prototype.
– Fine-tuned facebook/bart-large-cnn on legal datasets achieved improved summary coherence for domain-specific text
– Explored a **custom transformer architecture** with Byte Pair Encoding (50k vocab), showcasing in-depth understanding of multi-head attention and encoder-decoder design
– Implemented **PPO-based RL** to optimize ROUGE & BLEU scores, further improving summary quality
– Hosted the fine-tuned model on Hugging Face for easy public access & inference

ChurnMonitor — **ANN, Keras, Scikit-learn, NumPy, Pandas**
— **Project Link**
– Built an AI-powered system to predict customer churn in the banking sector, helping businesses identify at-risk customers
– Engineered a machine learning model achieving 87% accuracy on 50,000+ customer records
– Created an interactive Streamlit dashboard for real-time insights and visualization of key churn factors
– Developed an automated data preprocessing pipeline, reducing data preparation time by 20%

Achievements

– Solved **300+ LeetCode** problems, demonstrating strong problem solving and algorithmic skills
– Led a team to the **National Tata Innovant (Generative AI Hackathon)**, being the only selected team from Madhya Pradesh out of 2600+ teams