REETH JAIN

+91-9302171850 | reeth_j@ch.iitr.ac.in | linkedin.com/in/reeth-jain-rj777 | github.com/reethj-07

EDUCATION

- **Indian Institute of Technology, Roorkee** — Uttarakhand, India
  B.Tech. in Chemical Engineering — *2022 – Present*

INTERNSHIPS

- **Data Science Intern | PocketFM** — May 2025 – July 2025
  - Built scalable pipelines on user behavior logs to surface early churn signals for ML-driven retention analysis.
  - Segmented 2M+ users using clustering to support personalization and churn risk modeling.
  - Developed NLP sentiment analysis on 50K+ reviews to extract product and content signals.
  - Evaluated A/B experiments using CTR, sessions, and impressions to validate model-driven features.

- **Data Analyst Intern | Paragon** — Dec 2024 – Jan 2025
  - Cleaned and standardized multi-source sales data using SQL to improve downstream analytics reliability.
  - Performed EDA to uncover sales trends, customer behavior patterns, and operational bottlenecks.
  - Built optimized SQL queries and KPI tables to support analytics dashboards.
  - Applied clustering on customer purchase behavior to enable targeted segmentation strategies.

PROJECTS

- **Hallucination-Aware Hybrid LLM System | RAG, QLoRA, FAISS, Phi-3 Mini** [GitHub]
  - Built a hallucination-aware RAG pipeline using FAISS dense retrieval with constrained prompt grounding.
  - Fine-tuned Phi-3 Mini (4k) using QLoRA with 4-bit quantization for parameter-efficient adaptation.
  - Enforced context-only generation with explicit refusal when answers lacked retrieved evidence.
  - Compared RAG vs non-RAG inference modes to empirically demonstrate hallucination reduction.
  - Developed a Streamlit dashboard with retrieved-document inspection for explainability.

- **Multi-Modal Emotion-Aware Music Recommendation System | Wav2Vec2, ResNet50** [GitHub]
  - Built a multi-modal music recommender using ResNet50, Wav2Vec2, and DistilRoBERTa for emotion inference.
  - Fine-tuned ResNet50 on RAF-DB and integrated HuggingFace transformers for real-time emotion inference.
  - Mapped detected emotions to Spotify genres using Spotipy for personalized music recommendations.
  - Developed an end-to-end Streamlit app supporting webcam, audio, and text-based inputs.

- **GenAI-Based YouTube & Web Summarizer | Whisper, LangChain, LLaMA-3** [GitHub]
  - Built a Streamlit-based system to summarize YouTube videos and web pages using Whisper and LangChain.
  - Implemented audio extraction, transcription, and summarization pipelines for long-form content.
  - Integrated Groq-hosted LLMs via LangChain for fast, high-quality abstractive summaries.

- **Credit Card Default Prediction | Random Forest, XGBoost, SMOTE** [GitHub]
  - Engineered financial risk features from 25K+ customer records and addressed class imbalance using SMOTE.
  - Trained Logistic Regression, Random Forest, and XGBoost models for default prediction.
  - Optimized Random Forest using GridSearchCV with F2-score to prioritize recall for risk detection.
  - Identified key default drivers and tuned the decision threshold to 0.30 for risk-sensitive decisions.

AWARDS / SCHOLARSHIPS / ACADEMIC ACHIEVEMENTS

Silver Medal in Techshilla 2.0 (Tech-Enhanced AI Interview Platform) | Pre-Finalist, WWT Unravel 2025 | AIR 6328 (JEE Main) | AIR 7306 (JEE Advanced)

TECHNICAL SKILLS

Programming: Python, SQL, C++
ML & Deep Learning: PyTorch, TensorFlow, Scikit-learn
NLP & GenAI: Transformers, LangChain, HuggingFace, FAISS, Whisper, RAG
MLOps & CI/CD: Model Versioning, Experiment Tracking, Pipelines, Docker, GitHub Actions
Backend & Serving: FastAPI, Flask, Streamlit
Data & Tools: Pandas, NumPy, Matplotlib, Seaborn, Git, Jupyter, VS Code, GCP