

Harshal Bhambhani

f20221809@hyderabad.bits-pilani.ac.in +91 8561972727 Hyderabad, Telangana
https://www.linkedin.com/in/harshal-bhambhani-hhbb24 github.com/Harshaaalll

My Portfolio

Professional Summary

Data Scientist designing high-impact GenAI and NLP solutions to accelerate business growth and enhance customer experiences. Proven expertise in structuring LLMs, prompt engineering, and scalable analytics to solve complex financial challenges. Driven to leverage advanced data analytics to optimize operations across lending, payments, and insurance.

Skills

Languages & Databases: Python, SQL, Relational Databases

AI/ML & NLP: LLM Integration, Prompt Engineering, LangChain, Transformer Models, SBERT, Semantic Search, Vector Databases (FAISS), Scikit-Learn, Large Language Models

Data & Analytics: NumPy, Pandas, ETL Pipelines, Data Cleaning, Power BI

Cloud & MLOps: AWS S3, Docker, Kubernetes, GitHub

Experience

Data Science Intern - Standard Chartered Global Business Services Pvt. Ltd. Jul 2025 – Dec 2025

- Engineered Python NLP pipelines processing 100+ page agreements, cutting manual audit time by 90% for business users.
- Built high-recall classifiers achieving 98.6% recall on rare LMA clauses despite 90%+ irrelevant text inputs.
- Deployed SBERT vector search using cosine similarity to retrieve gold-standard legal compliance sample language.
- Implemented confidence thresholding models filtering 95%+ false positives, delivering audit-grade analytical precision.
- Optimized ETL pipelines with 250-token sliding windows and 50-token overlaps, yielding zero data loss.

Data Science Intern - DG Liger Consulting Jun 2024 – Jul 2024

- Constructed RAG chatbots using LangChain loaders to extract intelligence from proprietary PDF documents.
- Generated semantic embeddings via all-MiniLM-L6-v2 and configured FAISS vector databases for rapid similarity search.
- Integrated StableLM Zephyr 3B LLM to ingest retrieved context and formulate accurate, grounded AI responses.

Technical Projects

AutoTransFlow - Multilingual Document Translation - Document AI Aug 2025 – Nov 2025

- Launched doc-layout-yolo mapping text boxes over 0.25 confidence, preserving rigid layouts.
- Orchestrated multilingual translation pipelines using Facebook NLLB-200 transformers hosted on internal servers.
- Extracted elements and re-rendered translated text into matching page coordinates to finalize business reports.

Market Analysis Using LLMs - Financial Market Analytics Jun 2024 – Jul 2024

- Scraped 500+ articles via BeautifulSoup and built unstructured data workflows using LangChain text chunking.
- Summarized financial text via 2-pass BART 4-beam search and deployed RoBERTa sentiment models at 67% accuracy.
- Balanced datasets via SMOTE on 55:46 split and deployed Python Flask APIs serving real-time actionable analytics.

Multilingual Speech Recognition Pipeline - ASR, NLP Aug 2024 – Nov 2024

- Transcribed low-resource Urdu audio using zero-shot OpenAI Whisper models, tracking continuous Word Error Rate.
- Processed raw audio with -20dBFS loudness normalization and spectral subtraction, cutting noise by 10dB.
- Corrected transcription errors via IndicBERT masked language models overwriting incorrect phonemes past 40% threshold.

Medical Bill OCR and Fraud Detection Datathon Nov 2025

- Designed PaddleOCR pipelines with CLAHE and 4x DPI scaling achieving 90%+ accuracy extracting hospital bill data.
- Formulated Qwen 7B LLM workflows via prompt engineering parsing structured JSON from fragments.
- Executed 4 automatic anomaly checks via IQR algorithms and SHA-256 hashing to flag duplicate claims accurately.

Education

B.E. Electrical and Electronics with a Minor in Computer Science BITS Pilani, Hyderabad(2022-2026)

Relevant Coursework: Machine Learning, Database Systems, Artificial Intelligence, OOP, Operating Systems

CBSE: Class 12 - 92.2% — Class 10 - 91.2%

Extracurricular

Double-Sport State Representative — Basketball & Cricket Captain

Gold Medalist, Captain — Intra-BITS Volleyball Tournament (RIXA)

Department of Security & Hospitality, Core Member

- Led a team of 20+ managing venue logistics, crowd access, and safety protocols for campus events.