

# Henil Gajjar

Boston, MA • +1 (857) 492 8741 • gajjar.he@northeastern.edu • [LinkedIn](#) • [Github](#)

## Education

**Master's in Data Science - Northeastern University** (Boston, USA)

August 2023 - Present

**Related Coursework:** Large Language Models, Supervised and Unsupervised Machine Learning, Data Mining

GPA: 4.0/4.0

Head Teaching Assistant for CS2810 (Data Modelling) and DS3000 (Foundations of Data Science)

**Bachelor's in Electronics and Communication Engineering with Minor's in CSE - Nirma University** (Ahmedabad, India)

June 2023

**Related Coursework:** Database Management, Applied Statistics, Machine Learning, Computer Vision

GPA: 3.9/4.0

## Skills

### Languages

Python, C++, R, Dart, MATLAB

### Frameworks

Langchain, Hugging Face, Transformers, NLKT, TensorFlow, PyTorch, Flask, Keras, Pandas

### (Vector) Databases

Faiss, ChromaDB, DataStax AstraDB, MySQL, MongoDB, Google Firebase, Power Bi, Tableau

### Deployment Operations

Github Actions, AWS Sagemaker, AWS Bedrock, Docker, Streamlit Cloud, Langsmith, Langserve

### Langchain Techniques and LLMs

Document Loaders, Text Splitters, Tools, Agents, LCEL Chain, RAG, GroqCloud, OpenAI, Ollama

## Experience

**Head of AI/ML - Hyperlab Sportech Pvt. Ltd.** (Gandhinagar, India)

January 2022 - August 2023

- Engineered a ML-driven mobile application for an athlete training device 'Helios' resulting in over 5k+ downloads and 4.7-star rating on App store and Play Store within first week of the launch.
- Created HyperAI, an AI-driven insights tool using AWS Lambda, Bedrock, and AstraDB, to analyze athlete performance history and provide detailed weekly insights and performance-related answers.
- Employed LSTM model for timeout drills utilizing past athlete trainings, pivotal in securing \$25M valued investment on Shark Tank.
- Optimized MongoDB data schema post ETL, achieving 30% reduction in query response time through refined data modeling and indexing.

**Student Researcher - Nirma University** (Ahmedabad, India)

May 2021 - December 2021

- Curated a dataset of a Li-ion battery pack encompassing over 150k rows, extracting actionable insights for data-driven strategies.
- Utilized Random Forest regression model to predict optimal temperature during charging/discharging, improving battery health by 12%.
- Employed clustering techniques to understand current rates and temperature of the battery pack over 1300 charge-discharge cycles.

## Projects

**YouTube Video Blog Generator using GPT-4o and CrewAI** [\[Github\]](#)

July 2024

- Implemented a sequential task execution process with memory and caching features to automate content creation from YouTube videos.
- Utilized the YoutubeChannelSearchTool and specialized agents to extract and transcribe video content for comprehensive blog posts.
- Configured agents with specific roles and goals, optimizing the task performance and ensuring high-quality output with minimal manual intervention.

**End-to-End SQL Database Chatbot using Llama3** [\[Github\]](#) [\[Website\]](#)

June 2024

- Implemented integration with both SQLite and MySQL databases, providing flexible and robust data access.
- Created SQL Database Toolkit with Llama3 zero-shot question answering, enhancing the chatbot's ability to handle complex SQL queries accurately.
- Utilized Streamlit for the user interface, allowing users to query databases and retrieve responses in real-time.

**End-to-End RAG Document Q&A with Conversational History** [\[Github\]](#) [\[Website\]](#)

May 2024

- Developed an end-to-end Multidocument Chatbot using ChatGroq and Gemma2 to enable natural language interactions with PDFs.
- Implemented efficient document retrieval using FAISS VectorStore DB and recursive text splitting for high-quality contextual responses.
- Created a user-friendly Streamlit interface for real-time PDF content analysis and dynamic chat session management.
- Utilized stateful chat history to enhance the Q&A experience, maintaining context throughout user interactions.

**Fantasy Team Recommendation for IPL 2024** [\[Github\]](#)

Feb 2024 - April 2024

- Fine-Tuned Gemma & Llama-2 using cricket dataset resulting in 10% improvement in ROUGE score for cricket-specific text generation tasks.
- Established algorithm to extract structured data from unstructured IPL historical data including player stats and match scorecards.
- Leveraged Prompt engineering (2-shot) to enhance prediction accuracy resulting in 85% accurate team prediction for the IPL matches.

## Patent and Selected Publications

**Patent:** Steering System for Autonomous Solar Electric Vehicle. IP India, 363614-001, Issued July 12, 2022

[Patent](#)

**Publications:** A Comprehensive Study on Lane Detecting Autonomous Car using Computer Vision (Elsevier ESA), June 2023,

[Article](#)

A Comparative Analysis of Various Deep-Learning Models for Noise Suppression (EAI endorsed Publication), July 2023

[Article](#)