Ayraf Raihan

Manipal Institute of Technology, Bengaluru
+91-996-700-8090 ayrafr8@gmail.com LinkedIn Profile GitHub Profile

Education

Manipal Institute of Technology **Sep. 2022 – May 2026**
Bachelor of Technology in Computer Science, Artificial Intelligence *GPA: 8.5/10*

Experience

IBM **December 2024 – Present**
Winter Intern *Bengaluru*
- Designed and implemented an advanced answer sheet recognition system to process diverse handwritten layouts, leveraging IBM's internal libraries for efficient data ingestion and analysis.
- Enhanced text extraction quality by **62%** through implementation of custom chunking techniques and optimization of the processing pipeline.
- Scaling the solution to industry-level requirements, ensuring robust performance across varied answer sheet formats and layouts.

Intel Technologies Pvt. Ltd. **March 2024 – September 2024**
Project Intern *Bengaluru-Hybrid*
- Developed 3+ AI-powered internal tools that improved data processing efficiency by **40%** and enhanced model performance metrics.
- Executed advanced research on **LLM fine-tuning** techniques and **Graph Retrieval-Augmented Generation (RAG)** to optimize data retrieval strategies.

IEEE EPICS **September 2023 – September 2024**
IEEE Research Intern *Bengaluru*
- Led a team of 20+ members in designing a cutting-edge multi-terrain autonomous rover using **LiDARs, sensors, edge computing, and machine learning**.
- Managed cross-functional collaboration and technical development cycles for innovative robotics project.

Projects

BrainWave | *Python, Llama-7B, DeepGram, Groq Cloud, LangChain, Streamlit* **January 2025**
- Engineered a NotebookLM-inspired learning platform leveraging **Groq's LPU** inference for real-time podcast content processing and summarization.
- Implemented **DeepGram** for high-fidelity audio transcription with **89%** accuracy, while utilizing Groq to reduce inference latency by **65%**.
- Developed an optimized content processing pipeline achieving less than **500ms** end-to-end latency for transcription, summarization, and key point extraction.

Microsoft Phi-2 SLM Fine-Tuning for Financial Sentiment Analysis | *QLoRA, Phi-2, Kaggle* **August 2024**
- Fine-tuned **Microsoft Phi-2 Small Language Model** using **QLoRA**, improving financial news sentiment classification accuracy by **52%**.
- Leveraged **Kaggle** for comprehensive dataset acquisition and advanced model optimization techniques.
- Demonstrated expertise in machine learning by enhancing sentiment classification across positive, negative, and neutral domains.

Chatbot for Mental Health Support | *Python, ChromaDB, LangChain, LLM, Hugging Face, Streamlit* **February 2024**
- Created a personalized mental health support chatbot utilizing **Retrieval-Augmented Generation (RAG)** to provide empathetic, contextually-relevant emotional support.
- Integrated **ChromaDB** for sophisticated context management and fine-tuned the model using **Hugging Face** best practices.
- Deployed production-ready application on Hugging Face with a user-friendly **Streamlit** interface.

Text2SQL Conversion Tool | *Python, Hugging Face, Streamlit, MySQL, Generative AI, Ollama, Docker, AWS* **February 2024**
- Developed an advanced natural language to SQL conversion solution using **SQL-Coder7B** and **Retrieval-Augmented Generation (RAG)** architecture.
- Integrated **MySQL** for robust backend functionality and demonstrated cloud engineering skills by deploying using **Docker** on **AWS**.

Technical Skills

Languages: Python (Advanced), Java, C, C++, SQL
Technologies/Frameworks: Linux, GitHub, MongoDB, ChromaDB, Firebase, Keras, OpenCV, TensorFlow, LangChain, Figma, Neo4j, CUDA, LLMs, Hugging Face, Kaggle, Groq, Notion

Leadership / Extracurricular

Tech Solstice **April 2023**
Tech Lead Ops *MIT, BLR*
- Orchestrated event management for MIT BLR's 2023 Tech Fest, coordinating 50+ volunteers to successfully execute events with 1000+ participant engagement.
- Implemented strategic planning that enhanced event operational efficiency and participant experience.

CodeX **February 2023 – September 2023**
AI Research Lead *MIT, BLR*
- Led the Research department of MIT's largest coding club, driving multiple successful technical events and research initiatives.
- Fostered innovation and technical skill development among peer researchers, expanding the club's technical capabilities.