

Mohammed Abrar Hussain

AI Engineer | LLM Engineer | AI Agent Developer

✉ 11abrarhussain11@gmail.com

🌐 LinkedIn

🐙 GitHub

☎ +91 9959394534

🌐 Portfolio

Professional Summary

AI Engineer who builds production LLM systems, not demos. Built and deployed voice and WhatsApp assistants that cut LLM prompt size by 70% (5,000→1,500 tokens) through backend-driven orchestration, while maintaining sub-2-second response times. Skilled in FastAPI, PostgreSQL, Qdrant, and multi-step agentic workflows for lead qualification and business automation.

Technical Skills

Programming Languages	Python, SQL
LLM Engineering	LLMs, Retrieval-Augmented Generation (RAG), AI Agents, Agentic AI, Prompt Engineering, Function Calling, Structured Outputs
Frameworks	FastAPI, LangChain, LangGraph, CrewAI, AutoGen
Databases	PostgreSQL, Qdrant
AI Stack	Groq Llama, Groq Whisper, ElevenLabs, Google Gemini Embeddings, MiniLM Embeddings
Tools	Docker, Git, GitHub, Linux, REST APIs, WebSockets, Streamlit

Experience

Data Science Intern | Prinston Smart Engineers Jan 2026 – May 2026

- Led a team of five interns in developing a Startup Success Prediction system using a Kaggle dataset containing approximately **900 U.S. startup records**, coordinating project tasks from data preparation to model development.
- Performed data preprocessing, missing value handling, feature engineering, feature scaling, and evaluated multiple machine learning models including Random Forest and Support Vector Machine (SVM), selecting Random Forest based on predictive performance.
- Contributed to the development of a Streamlit-based application that predicts startup success using business attributes such as industry, location, investors, funding rounds, and company profile.

Projects

AI Voice Agent for Lead Qualification & Meeting Scheduling Jul 2026 – Aug 2026

- Designed and developed an AI voice agent using Groq Llama, Groq Whisper, and ElevenLabs to enable natural voice conversations, lead qualification, meeting scheduling, cancellation, and context-aware interactions.
- Engineered a backend-first orchestration architecture integrating FastAPI, PostgreSQL, CRM, calendar, and enterprise workflows, maintaining conversational response times under **2 seconds** while backend-assisted operations typically completed within **3–4 seconds**.
- Reduced LLM prompt size by **70%** (from **5,000** to **1,500** tokens) by moving business logic into the backend and implemented RAG using Qdrant with MiniLM embeddings, session memory, lead scoring, admin analytics, and automated email notifications.

AI-Powered WhatsApp Business Assistant Jun 2026 – Jul 2026

- Developed an AI-powered WhatsApp Business Assistant using Baileys, FastAPI, Groq LLMs, and Retrieval-Augmented Generation (RAG), delivering context-aware responses with typical reply latency under **2 seconds**.
- Integrated Qdrant with Google Gemini Embeddings to index and retrieve information from a knowledge base of **7 company documents**, while implementing four backend workflows for lead qualification, CRM updates, automated email notifications, and knowledge retrieval.
- Built persistent phone number-based conversation management using summarized conversation history and validated the system through **10+ test conversations** involving **5 users**.

Education

Bachelor of Technology (B.Tech.) in Computer Science & Engineering 2022 – 2026

HKBK College of Engineering | Visvesvaraya Technological University (VTU) | CGPA: 8.13

Certifications

- Google AI Professional Certificate — Google Credential ID: 072FTPFQ5F2G
- IBM RAG and Agentic AI Professional Certificate — IBM Credential ID: 7UWVUXOW6YJM

Languages

English	Fluent (Duolingo English Test Score: 130)	Hindi	Professional Working Proficiency
Urdu	Native	Kannada	Basic