Kaushal Jambhale

☎ 7021780426 ✉ kaushal.jambhle@gmail.com | linkedin.com/in/Kaushal_jambhle | github.com/KaushalWalker | Portfolio

SUMMARY

Physics graduate building production-grade agentic AI systems, not just prototypes. Designed and deployed an LLM agent with persistent memory, guardrails, and automated LLM-as-a-Judge evaluation, containerized with Docker and deployed across distributed cloud services. Self-taught into AI/ML Engineering through shipped projects, not coursework alone.

TECHNICAL SKILLS

Agentic AI & LLM: LangChain, Prompt Engineering, RAG, Tool-Calling, Groq, Hugging Face Transformers

Languages & Data: Python, SQL, NumPy, Pandas, Scikit-learn

Vector Search: ChromaDB, Sentence Transformer, Semantic Search

Cloud & Deployment: AWS(EC2,ECR), Docker, FastAPI, REST APIs, MLflow, SQLAlchemy, JWT Authentication, PostgreSQL, Git, Railway

ML/DL: Gradient Boosting, Regression, Classification, TensorFlow, Feature Engineering

PROJECTS

AI Shopping Assistant with Memory, Guardrails & Agent Evaluation | *Python, LangChain, Groq, SQLite, Streamlit, Docker, AWS EC2,ECR* GitHub

- Built an agentic shopping assistant using Langchain tool-calling for multi-step search, recommendations, and checkout workflows on Groq LLMs.
- Added persistent user preferences and session isolation for personalized, budget-aware recommendations across users.
- Cut hallucination risk with an automated **LLM-as-Judge-pipeline** scoring tool-call accuracy and format compliance, backed by input guardrails and integration testing.
- Containerized with **Docker**, pushed to **Amazon ECR**, and deployed on **AWS EC2** with env-based credentials and persistent SQLite storage and configure EC2 security groups to securely expose the app for remote demos.

Telecom Customer Care Chatbot (RAG) | *Python, ChromaDB, Sentence Transformers, Streamlit* GitHub

- Unified fragmented support content into a single retrieval system by building an **end-to-end RAG pipeline** with ChromaDB
- vector search and sentence transformer embeddings, **achieving 100% Recall@3** on curated benchmark query set.
- Increased response trustworthiness via confidence-scored fallback logic by implementing similarity-threshold gating and inline source citations, validated through systematic failure-case analysis.

Fraud Detection System using Machine Learning | *Python, Scikit-learn,MLflow, Streamlit, Power BI* GitHub

- Minimized false negatives in financial identification with **0.96 Recall and 1.00 Precision** by engineering and tuning a Gradient Boosting pipeline on 11K+ transactions.
- Enabled real-time operational monitoring and auditability via full input/prediction/probability logging by deploying a **Streamlit app paired with a 6-visual Power BI dashboard**

AI Interview Pilot — Full-Stack Cloud Deployment | *React, FastAPI, PostgreSQL, SQLAlchemy, JWT, Docker, Vercel, Railway, Neon* GitHub

- Delivered a production-accessible full-stack application, live and publicly reachable, by architecting a distributed **deployment across Vercel, Railway, and Neon PostgreSQL.**
- Secured application access with authenticated, session-scoped API endpoints by implementing **JWT-based authentication** with SQLAlchemy ORM data models.
- Resolved cross-platform integration failures for successful, repeatable production builds by debugging Docker configuration, environment variables, CORS conflicts, and PostgreSQL driver incompatibilities across three independently hosted services.

CERTIFICATIONS

Data Science & AI Certification — Boston Institute of Analytics (BIA), 2026

Data Visualisation: Empowering Business with Effective Insights — Tata, 2025

EDUCATION

Bachelor of Science in Physics — University of Mumbai, 2021 – 2024