

HUZEFA JAMBUGHODA

huzefa.jambughoda@gmail.com | [Website](#) | [GitHub](#) | [LinkedIn](#) | +971-58-59-60-011 | Dubai, UAE

AI Engineer with 5.5 years of experience bridging ambiguous business problems and production-grade AI. Experienced in architecting Agentic workflows, RAG pipelines and scalable ML infrastructure.

SKILLS

Languages & Backend: Python, SQL, C++, Javascript, FastAPI, Pydantic, PySQL

ML/DL: PyTorch, Scikit-Learn, Time-Series Forecasting, Computer Vision

GenAI: LangChain, LangGraph, LiteLLM, Anthropic SDK, RAG, Qdrant, PostgreSQL pgvector

LLM: Finetuning, HuggingFace, QLoRA/PEFT, LLM-as-a-Judge, Guardrails

Infra & MLOps: MLFlow, Docker, Kubernetes, GCP Vertex AI, AWS Lambda, AWS SageMaker

EDUCATION

Indian Institute of Technology, Madras | B.Sc. in Data Science & Programming

2020–23 | CGPA: 8.80

Charotar University of Science and Technology | B.Tech. in Information Technology

2016–20 | CGPA: 9.28

EXPERIENCE

ACDC

Dubai, UAE

Founder & Lead AI Engineer

Sep 2023 – Present

- Built a multi-step ERP automation agent using **LangGraph**, custom MCP server for ERP integration, hybrid retrieval with **Qdrant** and **MemO** episodic memory - cutting manual data entry time by 80%.
- Reduced token costs by 33% by building a centralized **LLM Gateway** on top of **LiteLLM** with middleware like semantic **caching** (Redis) and dynamic **model routing** based on task complexity.
- Served agent endpoints via **FastAPI** with Pydantic-validated schemas, containerized with Docker and deployed on Kubernetes. Monitored token usage, latency and per-tenant cost attribution using **MLflow** and **LangSmith**.
- Improved answer relevance from 88% to 94% by fine-tuning Kimi K2 with **QLoRA/PEFT** on UAE construction data, evaluated on a domain-specific benchmark with automated F1 and LLM-as-a-Judge relevance metrics.
- Led B2B client discovery and handling, translating ambiguous operational needs into concrete technical specs.

Google (via Infocusp)

Ahmedabad, India

Senior Machine Learning Engineer

Dec 2022 – Aug 2023

- Reduced gene edit turnaround by 30% and accelerated research by reducing genomic features from 50k to ~100 using Monte Carlo Dropout and Gradient Boosting.
- Accelerated research iteration speed by building an end-to-end ML pipeline that is fully YAML-configured and supports custom NNs in **PyTorch** and all **Scikit-Learn** models with hyper-parameter tuning and automated Docker deployments on **VertexAI**.
- Integrated Monte Carlo Dropout & Variational Dropout to reduce the effects of network sparsity.
- Built 7 **data pipelines** using Pydantic and Pandas to transform raw DNA sequences to trainable features.
- Developed a Computer Vision pipeline on Vertex AI with crowd-sourced labelling to automate physical trait collection from live footage, eliminating hours of manual field observation.

Infocusp

Ahmedabad, India

Machine Learning Engineer

Feb 2021 – Dec 2022

- Autodesk: Boosted predictive accuracy from 74% to 87% for chemical dosage using **LSTM**, **LightGBM** and **XGBoost**, reducing raw-material usage and saving operational energy at water treatment plants.
- Bryte: Engineered serverless, event-driven **AWS Lambda** pipelines to preprocess live IoT smart bed data, shifting from 8-hour batch windows to real-time availability.
- Legal Sifter: **Fine-tuned BERT** architectures to optimize embedding strategies for contract clause classification, boosting parsing accuracy by 18%.

Schneider Electric

Bangalore, India

SE Intern

Feb 2020 – May 2020

- Analyzed energy consumption data from active deployments to identify usage patterns and operational anomalies.
- Researched ML approaches for real-time phishing detection, evaluating **feature engineering** strategies on URL structure and webpage contents.

Enlighten Infosystems

Vadodara, India

SE Intern

Apr 2019 – May 2019

- Built an interactive **geospatial visualization** tool in Python and Bokeh.
- Compared evolutionary and swarm algorithms for **feature selection** from scratch on genetic data - Artificial Bee Colony, Ant Colony Optimization and Genetic Algorithm.

Bhaskaracharya Institute of Space Applications and Geo-Informatics
SE Intern

Gandhinagar, India
Apr 2018 - May 2018

- Automated GIS workflows and deployed them as 3 Python **plugins** for on-premise QGIS application.
- Reduced manual trimming work of ~5mins per spatial feature to seconds utilizing convex hull optimization.

PROJECTS

Leo2 - Multi-Tenant Compliance QA

- Finetuned Kimi K2 using QLoRA/PEFT for construction domain, orchestrated via ADK with a RAG layer for hybrid BM25 + Qdrant retrieval over proprietary documents maintaining per-tenant data isolation.
- Stack: Kimi K2, QLoRA PEFT, ADK, RAG, Qdrant, FastAPI.

YT Ads Compliance RAG System

- Cut ad review time by 80% by building a LangGraph pipeline that ingests YouTube videos, extracts multimodal content using Azure Video Indexer, retrieves compliance information with Azure AI Search and inspects compliance with an OpenAI model on Microsoft Azure Foundry.
- Stack: RAG, LangGraph, FastAPI, Azure Video Indexer, Azure AI Search, Microsoft Azure Foundry.

Mandli - Multi-agent Content Creation

- Specialist agents (research, writing, SEO skills) operating autonomously in CrewAI flows. Uses FireCrawl MCP for web ingestion and Qdrant vector database for deduplication.
- Stack: CrewAI, LangGraph, GLM-5.2 (orchestrator), MiniMax M3 (research), Claude Sonnet (writing), Kimi K2.6 (SEO), FireCrawl MCP, Qdrant, FastAPI, Astro JS (frontend), Streamlit (dashboard).

OpenAccounts

openaccounts.vercel.app

- Fully client-side, minimal, double-entry accounting app with statement ingestion using dual-stage doc parsing.
- Stack: React, Tailwind, IndexedDB, Google Drive, OAuth2, PyMuPDF, PPDocLayout v3, GLM-OCR.

CERTIFICATIONS & ACHIEVEMENTS

Building with Claude API - Anthropic Academy	Certified
AI Agent - HuggingFace	Certified
AI Engineer with Foundry and AIP - Palantir Learn	Certified
National Creativity & Aptitude Test - IFEHE India	National #9
Programming, Data Structures & Algorithms using Python - NPTEL	Top 1%
Codechef	huzefa_j - 5 star