Aman Jaiswal

AI/ML Engineer

aman2107jaiswal@gmail.com | +44 7776364344 | London, UK | linkedin.com/in/amanjaiswal2107

github.com/Aman-2107

PROFILE

AI / ML Engineer with a Distinction in MSc Artificial Intelligence from Queen Mary University of London. Strong hands-on experience in building end-to-end AI systems using Large Language Models (LLMs), Natural Language Processing (NLP), and Retrieval-Augmented Generation (RAG). Skilled in Python, PyTorch, FastAPI, PostgreSQL (pgvector), and full-stack AI prototyping. Seeking a junior AI / ML engineer role in London focused on applied machine learning and production-oriented AI systems.

SKILLS

- **Programming Languages:** Python, SQL, JavaScript, TypeScript.
- **AI / ML:** Machine Learning, NLP, Large Language Models (LLMs), Fine-Tuning, Embeddings.
- **LLM Systems:** Retrieval-Augmented Generation (RAG), Prompt Engineering, LLM Orchestration.
- **Frameworks & Libraries:** PyTorch, Hugging Face Transformers, sentence-transformers.
- **Backend & Data:** FastAPI, PostgreSQL, pgvector, SQLAlchemy, JSONB.
- **Frontend:** React, Tailwind CSS.
- **Tools:** Git, Docker, Linux.

PROJECTS

Anamnesis – AI Patient Intake & Clinical Memory System ✐

- Developed an end-to-end AI system that conducts structured conversational patient intake and automatically generates validated clinical summaries from free-text dialogue.
- Designed a multi-stage intake workflow using a large language model (LLaMA-3.3-70B via Groq) to extract chief complaints, symptom attributes, medications, allergies, and safety red flags.
- Implemented a Retrieval-Augmented Generation (RAG) pipeline using sentence-transformer embeddings and PostgreSQL with pgvector to enable evidence-grounded, natural language question answering over patient history.
- Engineered a FastAPI backend with structured JSONB storage and a React + TypeScript frontend separating patient and clinician workflows.

Calma – Offline Multi-Modal AI Emotional Support System ✐

- Developed a fully offline AI emotional support system to ensure user privacy, supporting real-time text and voice interaction.
- Fine-tuned a RoBERTa-based model for multi-label emotion classification, achieving an F1-Macro score of 0.56 using custom Focal Loss and layer-wise learning rate decay.
- Integrated LLaMA-based response generation with Whisper (ASR) and F5-TTS (TTS) in a multi-threaded architecture for low-latency interaction.

Legal Search Engine with BM25 Ranking ✐

- Built a Python-based desktop application for legal document retrieval using the Lex-GLUE benchmark.
- Implemented BM25Okapi ranking for relevance-based search and designed a responsive Tkinter GUI using multithreading to maintain UI performance.
- Engineered automated schema normalization to resolve inconsistencies across heterogeneous legal datasets.

EDUCATION

Queen Mary University of London   Sep 2024 – Sep 2025
MSc in Artificial Intelligence
Score: Distinction.
Key Modules: Machine Learning, Artificial Intelligence, Applied Statistics, Natural Language Processing, Neural Networks and Deep Learning.

CHRIST (Deemed to be University)   Jun 2021 – Apr 2024
Bachelor of Computer Applications

CERTIFICATES & LEADERSHIP

- Certifications: Large Language Models: Text Classification for NLP using BERT (LinkedIn); Natural Language Processing Fundamentals and Artificial Intelligence and Machine Learning Fundamentals(Infosys Springboard); Using MySQL Database with PHP (Coursera).
- Leadership: Served as Event Head of INTERFACE 2024, a National Level Undergraduate IT Fest.