

Meriat Joseph

AI/ML Engineer

 meriatjoseph |  Meriat Joseph |  meriatjoseph@gmail.com |  +91 7902800574

SUMMARY

Results-driven AI/ML Engineer with hands-on experience in developing intelligent applications using Large Language Models (LLMs), **Retrieval-Augmented Generation (RAG)**, **LangChain**, Stable Diffusion, and LangChain. Skilled in Deep Learning, NLP, Computer Vision, and deploying scalable AI solutions with **FastAPI** and Streamlit. Strong in **Prompt Engineering**, Transformers, and MLOps practices. Adept at translating complex theory into practical innovation, committed to continuous learning and impactful problem-solving.

WORK EXPERIENCE

AI Engineer – Nxtgen Cloud Technologies

April 2025 - Present

- Built and deployed end-to-end LLM-powered applications for enterprise use cases, including intelligent assistants, document search systems, and personalized conversational bots.
- Implemented **Retrieval-Augmented Generation (RAG)** pipelines using **LangChain**, **OpenAI embeddings**, and **vector databases (FAISS/Chroma)** for context-aware question answering and semantic search.
- Designed and optimized **prompt engineering** techniques to improve accuracy, relevance, and contextual depth of large language model responses.
- Developed and deployed **RESTful APIs with FastAPI**, serving both machine learning and image generation models in production-grade environments.
- Utilized **MongoDB** and vector databases (Chroma, FAISS) to store embeddings and metadata, enabling efficient similarity search and retrieval at scale.
- Worked extensively on **Stable Diffusion and ControlNet** for conditional image generation, and **SAM-ViT-H** for high-accuracy object segmentation.
- Explored and fine-tuned **Vision-Language Models (VLMs)** such as **Qwen-VL** for multimodal reasoning across text and image inputs.
- Developed and tested LLM workflows leveraging open-source models like **Meta LLaMA 3.1 (70B)** for high-performance natural language understanding and domain-specific reasoning.
- Implemented a **knowledge graph** using **Neo4j** to model and store legal domain data, supporting semantic relationships and graph-based retrieval.
- Performed **supervised fine-tuning (SFT)** of LLMs to generate question-answer pairs tailored to legal and regulatory text understanding.
- Built a multilingual AI chatbot integrating **Whisper Large-v3** for real-time transcription and translation, and a RAG-based LLM to deliver accurate, context-aware query responses.

PROJECTS

English Mastery – AI-Powered English Communication App

[GitHub Repo](#)

Developed an AI-powered mobile application to enhance English proficiency and IELTS preparation. Integrated **LLM**, **RAG**, and **LangChain** to provide personalized, context-aware learning across Reading, Writing, Listening, and Speaking modules. Implemented **Whisper Large-v3** for multilingual speech recognition, **GTTS** for natural text-to-speech, and **Gemini**, **Llama3**, and **GPT-3.5** for intelligent conversational assistance. Utilized **OpenAI embeddings** for semantic search and adaptive content recommendations. Built with **Flutter** (frontend) and **FastAPI** (backend), deployed on **AWS**.

MINI PROJECTS

NASA Jet Engine RUL Prediction

- Developed a Remaining Useful Life (RUL) prediction model using NASA’s C-MAPSS dataset. Applied time-series forecasting, feature engineering, and **Random Forest classifier** modeling to predict engine failure under variable conditions and improve maintenance planning.

BCGX – Customer Churn Analysis

- Built a **Random Forest classifier** achieving 85% accuracy in predicting customer churn. Conducted **Exploratory Data Analysis (EDA)**, feature engineering, and data visualization using **Pandas** and **NumPy** to identify churn drivers and provide actionable business insights.

Interactive Road Accident Analysis Dashboard

- Created an interactive Power BI dashboard with advanced visualizations including **donut charts, pie charts, maps, and stacked columns** to analyze accident trends and deliver actionable insights for stakeholders.

ChatBot – NLP-Powered Conversational Agent

- Developed an NLP-based chatbot using Python with **TF-IDF vectorization**, stemming, and **stop-word** removal for enhanced text understanding and response accuracy.

SKILLS

Programming Language	Python, R, C
Data Visualization	Tableau, Power BI, Matplotlib, Seaborn
Machine Learning & Deep Learning	Supervised & Unsupervised Learning, Neural Networks, LSTM, RNN, CNN, Transformers, Feature Engineering, AutoML, Model Evaluation, Data Preprocessing, EDA
Generative AI & LLMs	LLMs, Generative AI, Prompt Engineering, RAG, LangChain, Stable Diffusion, Google Generative AI, OpenAI API, Hugging Face Transformers
AI Agents & Agentic AI	AI Agents, Agent Workflows, Multi-Agent Collaboration Platforms (MCP), Agentic Reasoning Systems, LangGraph
Natural Language Processing (NLP)	Text Processing, Sentiment Analysis, Named Entity Recognition (NER), Language Modeling, Text Classification
Computer Vision	OpenCV, Image Classification, Object Detection, Face Recognition, Diffusion Models, Image Processing, Image Segmentation
Frameworks & Libraries	TensorFlow, Keras, PyTorch, Hugging Face, Scikit-learn, Pandas, NumPy
Web & Backend Development	FastAPI, Streamlit, RESTful API Development, Streamlit Cloud
Databases	PostgreSQL, MongoDB, Vector Embedding Storage (FAISS, Chroma)
Tools & Platforms	Git, GitHub, Google Colab, VS Code, Jupyter Notebook
Core Competencies	AI Deployment, Prompt Engineering, Analytical Thinking, MLOps (Basic), Problem Solving, Collaboration, Teamwork

EDUCATION

2023 - 2025	Data Science, Brototype, Bengaluru
2020 - 2022	MA Econometrics, Mahatma Gandhi University
2017 - 2020	BSc Mathematics, Mahatma Gandhi University