

NAZZAL BIN KAUSAR

AI Engineer | Generative AI Engineer

Email: nazzal5448@gmail.com | [LinkedIn](#) | [GitHub](#) | [Website](#)

PROFESSIONAL SUMMARY

AI Engineer specialising in Generative AI, Large Language Models (LLMs), and AI-powered applications. Experienced in developing, deploying, and optimising AI models using LangChain, TensorFlow, and PyTorch. Strong knowledge of cloud-based AI deployments using Streamlit, Flask, and AWS services. Skilled in GitHub Actions, shell scripting, and automation for AI/ML workflows. Passionate about building scalable, user-centric AI applications that solve real-world problems.

SKILLS

Programming & Frameworks:

- **Languages:** Python
- **AI/ML Frameworks:** TensorFlow, PyTorch, LangChain, LangGraph
- **LLM Technologies:** OpenAI, Llama2, Mistral, Google Gemini Pro, DeepSeek-R1, Qwen2.5, Stable Diffusion
- **Vector Databases:** ChromaDB, Pinecone, FAISS
- **Deployment & Cloud Services:** Streamlit, Flask, Gradio, AWS (EC2, Lambda, CodeBuild), Azure Functions
- **Automation & DevOps:** GitHub Actions, Shell Scripting, CI/CD Pipelines
- **Database Management:** MySQL

Soft Skills:

Analytical Thinking | Problem-Solving | Teamwork | Effective Communication

PROFESSIONAL EXPERIENCE

AI Engineer | Generative AI Specialist

(Freelance / Personal Projects) | 2023 - Present

- Developed & fine-tuned LLM applications using LangChain, LangGraph, and Python, improving response accuracy by 25%.
- Optimised vector database retrieval (ChromaDB, Pinecone, FAISS), increasing AI model efficiency by 30%.
- Deployed AI-powered applications using Streamlit, Flask, and Gradio, ensuring smooth user interaction and scalability.

PROJECTS

Pin8n - An end-to-end Pinterest Blogging automation SaaS

- **Description:** Developed an AI saas that automatically publishes SEO-optimised blogs with images to WordPress sites and also generates Pinterest-ready pins for users based on the content uploaded to their website.
- **Technologies:** OpenAI, n8n, Llama3, FastAPI
- **Impact:** Enabled 1,200+ creators to generate SEO-optimised high-quality content, increasing engagement rates by 133%.

TalkToDocs - RAG Application

- **Description:** Developed an AI-powered Retrieval-Augmented Generation (RAG) system that enables users to query uploaded documents with LLMs for better information retrieval.
- **Technologies:** LangChain, Python, AWS, Google Gemini Pro, Llama2
- **Impact:** Improved user document interaction efficiency by 40%, reducing manual search time.

EDUCATION

BS Software Engineering

Virtual University, Pak | 2025 - ongoing

Relevant Coursework: Data Structures & Algorithms, Systems Design, Development Frameworks, Programming

FSc - Pre Medical

The Fazlehaq College Mardan | 2024

Research Interests

- Efficient AI model development for edge and resource-constrained devices.
- Multilingual AI for low-resource languages (Chinese, Arabic, Urdu, etc.)
- Safe, ethical, and interpretable AI for real-world use cases.