Zaheer Khan

AI Engineer

+91-8450868623 | zaheerkhan00457@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

Indore, India

Summary

AI Engineer skilled in developing intelligent systems using Python, Scikit-learn, and PyTorch, with hands-on experience in fine-tuning large language models (LLMs) and building end-to-end Generative AI solutions. Proficient in RAG pipelines, multimodal agents, and classical ML and deep learning techniques. Adept at deploying real-world applications via Streamlit and optimizing models for performance and task-specific outcomes.

Projects

Multimodal AI Agent – Video Summarizer [GitHub](#)

- Streamlit app powered by Gemini-2.0-Flash-Exp for understanding and summarizing videos.
- Integrated DuckDuckGo web search and PostgreSQL vector store for enhanced responses.
- Multimodal input support and real-time summarization.

Finance Assistant [GitHub](#)

- Built using Phi Agents and Groq's LLaMA 3.3-70B for financial analysis.
- Integrated Finance & Web agents for real-time data and intelligent responses.
- Modular agent architecture with contextual awareness.

RAG Chatbot with Chat History [GitHub](#)

- Retrieval-Augmented Generation chatbot using LangChain, FAISS, and NVIDIA embeddings.
- Added semantic search, chunking, chat memory & persistent history.
- Deployed using Streamlit for clean user experience.

Insurance Premium Predictor [GitHub](#)

- Built with XGBoost; predicts premiums for different age groups with tailored models.
- Preprocessing pipeline to handle input dynamically.
- Web UI powered by Streamlit.

Transformer from Scratch – Attention is All You Need [GitHub](#)

- Implemented full Transformer model (encoder + decoder) in PyTorch based on the original research paper.
- Developed scaled dot-product attention, multi-head attention, positional encoding, and masking logic.
- Trained on a sequence-to-sequence dataset with modular architecture and complete training pipeline.

Education

Sushila Devi Bansal College of Engineering – B.Tech in Computer Science & Engineering – Expected Aug 2025

Certifications

- Master Machine Learning For Data Science – Codebasics
- Mathematics – Basics to Advanced for Data Science and GenAI – Udemy
- Complete Data Science, Machine Learning, DL, NLP Bootcamp – Udemy
- Complete Generative AI Course With Langchain and Huggingface – Udemy

---

Skills

- Languages: C, C++, Python, SQL
- Libraries: Scikit-learn, TensorFlow, PyTorch, Pandas, NumPy, Matplotlib, Seaborn, Hugging Face
- Frameworks: Streamlit, LangChain, google-adk, CrewAI
- Tools: Git, Jupyter, VS Code, Cursor, Postman, MySQL