**Arjun Sharma**  
📍 Bangalore, India | 📧 arjun.sharma.ai@gmail.com | 📞 +91 98765 43210  
🔗 [LinkedIn: linkedin.com/in/arjunsharma-ai](https://linkedin.com/in/arjunsharma-ai) | 🖥️ [GitHub: github.com/arjunsharma-ai](https://github.com/arjunsharma-ai)

**PROFESSIONAL SUMMARY**

Results-driven **Generative AI Engineer** with **5+ years** of experience in **building, fine-tuning, and deploying LLMs** for various real-world applications. Expertise in **RAG (Retrieval-Augmented Generation), model optimization, prompt engineering, and AI-powered search engines**. Proven ability to integrate **NLP, deep learning, and vector databases** into scalable AI applications. Passionate about **AI-driven automation and intelligent assistants**.

**TECHNICAL SKILLS**

✅ **Languages:** Python, SQL, Bash  
✅ **Frameworks & Libraries:** PyTorch, TensorFlow, Hugging Face Transformers, LangChain  
✅ **LLMs & Gen AI Models:** GPT-4, Llama, Mistral, T5, BART, BLOOM  
✅ **AI Architectures:** RAG, MoE, RLHF, Agentic AI, Fine-Tuning, LoRA, PEFT  
✅ **Vector Search & Databases:** FAISS, Pinecone, Weaviate, ChromaDB, Milvus  
✅ **Big Data & Streaming:** Apache Kafka, Spark, Flink, Databricks  
✅ **Cloud & DevOps:** AWS (S3, Lambda, EC2, Bedrock), Azure OpenAI, Docker, Kubernetes  
✅ **MLOps & Deployment:** MLflow, FastAPI, TensorRT, Triton Inference Server

**WORK EXPERIENCE**

**Senior Generative AI Engineer**

XYZ AI Solutions, Bangalore | **2022 - Present**

* Led the development of a **multi-modal RAG-based AI search engine** integrating **LLMs, FAISS, and Elasticsearch**, improving document retrieval by **40%**.
* Developed an **AI-powered chatbot with real-time RAG** using **Flink, Kafka, and OpenAI APIs**, handling **100K+ daily interactions**.
* Fine-tuned **T5/BART for document summarization**, reducing legal document review time by **60%**.
* Built a **Finance AI Copilot** for real-time fraud detection and market analysis, utilizing **Databricks, Kafka, and MongoDB**.
* Improved inference performance by **3x** using **TensorRT and ONNX for LLM model optimization**.

**AI/ML Engineer**

ABC Tech Pvt. Ltd., Hyderabad | **2019 - 2022**

* Designed a **real-time social media sentiment analysis** pipeline using **Flink, Kafka, and Elasticsearch**, enhancing brand monitoring capabilities.
* Developed **AI-based threat detection** for **cybersecurity**, identifying anomalies using fine-tuned BERT and GPT-based models.
* Created a **customer behavior analysis** system leveraging **NiFi, Spark, and Power BI** for an e-commerce giant.
* Worked on **Hadoop/Spark-based predictive analytics** for supply chain optimization, reducing delays by **25%**.

**PROJECTS**

**1. AI Personal Assistant (LLM + RAG)**

* Built an AI-powered **personal assistant** integrating **LangChain, OpenAI, and vector search** to answer user queries in real-time.
* Achieved **98% accuracy** in providing relevant responses by implementing **hybrid search (dense + sparse retrieval)**.

**2. Generative AI for Cybersecurity**

* Developed an **AI-powered cybersecurity tool** that detects **zero-day threats** using **fine-tuned transformers** and **graph neural networks (GNNs)**.
* Reduced false positives in anomaly detection by **35%** using **self-supervised learning techniques**.

**EDUCATION**

**Bachelor of Technology (B.Tech) in Computer Science**  
Indian Institute of Technology (IIT) Delhi | **2015 - 2019**