**Rahul Sharma**  
Email: rahul.sharma.ai@gmail.com | Phone: +91-9876543210  
LinkedIn: linkedin.com/in/rahulsharma-ai | GitHub: github.com/rahulsharma-ai  
Location: Bangalore, India

**SUMMARY**

Results-driven AI/ML Engineer with 12+ years of experience in designing, developing, and deploying AI-powered solutions, specializing in Generative AI, Large Language Models (LLMs), and AI-driven search engines. Expertise in fine-tuning, Retrieval-Augmented Generation (RAG), multi-modal AI, and productionizing AI systems. Proficient in cloud platforms, vector databases, and real-time AI pipelines. Passionate about AI-driven innovation and scalable architectures.

**SKILLS & EXPERTISE**

* **Generative AI & LLMs:** OpenAI, Anthropic Claude, LLaMA, Mistral, Fine-tuning (LLaMA, BART, T5)
* **NLP & ML:** Transformers, Hugging Face, LangChain, Sentence Transformers, T5, BART, GPT Models
* **Retrieval-Augmented Generation (RAG):** FAISS, Pinecone, Weaviate, ChromaDB, Milvus
* **AI Orchestration & Agents:** LlamaIndex, LangChain Agents, AutoGPT, BabyAGI
* **Big Data & Streaming:** Kafka, Apache Flink, Spark Streaming, Apache NiFi, Databricks
* **Cloud & DevOps:** AWS (S3, Lambda, SageMaker, Bedrock), Azure AI, GCP AI, Kubernetes, Docker
* **Databases:** MongoDB, PostgreSQL, Elasticsearch, Redis, Neo4j, Vector Search (Pinecone, FAISS)
* **AI for Cybersecurity:** Threat detection, AI-driven security analytics, anomaly detection
* **Visualization & Reporting:** Tableau, Power BI, Streamlit, Dash

**WORK EXPERIENCE**

**Lead AI Engineer | TCS (Tata Consultancy Services), Bangalore | Jan 2021 - Present**

* Architected and deployed **RAG-based AI-powered Search Engine** leveraging **LlamaIndex, FAISS, Pinecone, and LangChain** to enhance document retrieval accuracy.
* Developed a **Streaming LLM-based chatbot** using **OpenAI, LangChain, Kafka, and MongoDB**, optimizing real-time response generation for customer support.
* Implemented **AI-powered Financial Copilot** integrating **market analysis and fraud detection** using **Kafka, Spark Streaming, and MongoDB**.
* Designed **Multi-modal Generative AI** solutions integrating **text, images, and documents** for business intelligence and insights.
* Led a **fine-tuned BART/T5-based document summarization system**, enhancing enterprise knowledge extraction.

**Senior AI/ML Engineer | Infosys, Hyderabad | Jul 2017 - Dec 2020**

* Developed a **social media sentiment analysis** pipeline using **Apache Flink, Confluent Kafka, and Elasticsearch** for brand reputation monitoring.
* Built a **real-time fraud detection system** leveraging **Confluent Kafka, Spark Streaming, and MongoDB** to detect anomalies in financial transactions.
* Created a **customer behavior analytics** solution integrating **Kafka, NiFi, Spark, and Tableau**, providing actionable insights for e-commerce businesses.
* Led the adoption of **AI-powered cybersecurity threat detection** using deep learning models for log analysis and anomaly detection.

**Machine Learning Engineer | Wipro, Pune | Jun 2012 - Jun 2017**

* Developed **predictive analytics models** for **supply chain optimization** using **Hadoop, Spark, and MongoDB**.
* Implemented **AI-driven recommendation engines** for e-commerce platforms using **collaborative filtering and deep learning-based models**.
* Designed **AI-assisted personal assistant** integrating **LLMs, RAG, and multi-agent workflows** for enterprise productivity.

**EDUCATION**

**M.Tech in Artificial Intelligence & Machine Learning**  
Indian Institute of Technology (IIT) Bombay | 2012

**B.Tech in Computer Science & Engineering**  
National Institute of Technology (NIT) Trichy | 2010

**CERTIFICATIONS**

* AWS Certified Machine Learning – Specialty
* Google Professional Machine Learning Engineer
* Microsoft Azure AI Engineer Associate
* Generative AI Specialization – DeepLearning.AI

**PROJECTS & RESEARCH**

* **AI-powered Search Engine with RAG** – Built a scalable **multi-modal search** system using **FAISS, Pinecone, and LangChain** for document retrieval.
* **Finance AI Copilot** – Developed an AI-driven assistant for **real-time fraud detection and market analysis**, leveraging **Kafka, Spark Streaming, and MongoDB**.
* **AI for Cybersecurity** – Implemented a **threat detection system** using **deep learning-based anomaly detection** for cybersecurity applications.
* **Social Media Sentiment Analysis** – Built a real-time NLP-based sentiment analysis pipeline using **Apache Flink, Kafka, and Elasticsearch**.
* **Document Summarization System** – Fine-tuned **BART/T5** for summarizing large documents and legal reports.

**TOOLS & TECHNOLOGIES**

* **Programming:** Python, Scala, Java, SQL
* **ML/DL Frameworks:** TensorFlow, PyTorch, Hugging Face Transformers
* **Big Data:** Hadoop, Spark, Flink, Databricks
* **Cloud Services:** AWS, Azure, GCP
* **Vector Databases:** FAISS, Pinecone, Weaviate, ChromaDB

**PUBLICATIONS & CONTRIBUTIONS**

* **"Enhancing Retrieval-Augmented Generation with Vector Search"** – Published research on AI-driven search optimization.
* **Open-source contributions** to **LangChain, LlamaIndex, and Hugging Face repositories**.
* Speaker at **AI conferences** on Generative AI, RAG, and Cybersecurity AI.

**ACHIEVEMENTS**

* Received **TCS Innovator Award** for pioneering AI-powered search solutions.
* Recognized as **Top 10 AI Engineers in India (2024)** by Analytics India Magazine.
* Developed an **AI-based assistant used by Fortune 500 companies** for document automation.