









# RIFAT ERDEM SAHIN

## AI ENGINEERING LEAD | LLM SOLUTIONS ARCHITECT

---

-  **Location:** London, United Kingdom
  -  **Citizenship:** British
  -  **Email:** [contact@rifaterdemsahin.com](mailto:contact@rifaterdemsahin.com)
  -  **Phone:** +44 7848 024173
  -  **LinkedIn:** [linkedin.com/in/rifaterdemsahin](https://www.linkedin.com/in/rifaterdemsahin)
  -  **GitHub:** [github.com/rifaterdemsahin](https://github.com/rifaterdemsahin)
  -  **Portfolio:** <https://rifaterdemsahin.com>
  -  **Schedule a Call:** <https://calendly.com/rifaterdem/schedule>
- 

## PROFESSIONAL SUMMARY


---


AI Engineering Lead specializing in **Generative AI, Large Language Models, and production-grade AI systems**. Deep hands-on experience architecting and deploying enterprise LLM applications with focus on **RAG architectures, AI security, and scalable deployment**. Proven track record building AI-powered automation systems that delivered 300% productivity improvements and 30% cost reductions. Expert in bridging the gap between cutting-edge AI research and production-ready, secure, enterprise solutions.

---


## CORE AI COMPETENCIES


---


 **Generative AI & LLM Engineering** - Deep expertise building and deploying production Generative AI solutions - Extensive hands-on experience developing Large Language Model applications - Advanced prompt engineering, fine-tuning, and model optimization - Expert in LangChain, LlamaIndex, and modern AI orchestration frameworks

 **RAG Architecture & Vector Systems** - Advanced implementation of Retrieval Augmented Generation (RAG) architectures - Expert in vector databases (Pinecone,

Weaviate, ChromaDB, FAISS) - Document processing pipelines and semantic search optimization - Hybrid search strategies combining dense and sparse retrieval

 **AI Security & Governance** - Expert-level knowledge of AI security practices and threat models - Prompt injection prevention and adversarial attack mitigation - Data privacy controls, PII detection, and secure model deployment - Compliance frameworks for regulated industries (finance, healthcare)

 **AI System Architecture** - Track record shaping enterprise AI architecture decisions - Scalable, maintainable AI solutions with proper observability - MLOps pipelines for continuous model deployment and monitoring - Cost optimization strategies for LLM inference at scale

 **AI Testing & Quality Assurance** - Comprehensive testing approaches: unit, integration, end-to-end - LLM evaluation frameworks and metrics (RAGAS, LangSmith) - A/B testing strategies for prompt and model optimization - Automated quality gates for AI system releases

---

## KEY AI ACCOMPLISHMENTS

---

### 2025 | IBM | London, UK

**Enterprise AI & Hybrid Cloud Transformation - Challenge:** Lead enterprise-scale hybrid cloud transformation and AI integration for global clients - **Solution:** Architected comprehensive IBM Cloud + Red Hat OpenShift platform with watsonx AI integration and DevSecOps practices - **Impact:** - 35% reduction in operational overhead through AI-driven automation with watsonx - 40% improvement in deployment frequency via DevSecOps and Ansible Automation Platform - Zero-downtime migration of mission-critical workloads to hybrid cloud architecture - **Technologies:** IBM Cloud, Red Hat OpenShift, watsonx AI, Ansible Automation Platform, Terraform, Kubernetes, GitHub Actions

### 2024 | Goldman Sachs | Muscat, Oman

**AI-Driven CI/CD Automation Framework - Challenge:** Manual deployment processes causing bottlenecks and errors in enterprise fintech environment - **Solution:** Architected and deployed AI-powered CI/CD framework using LLMs for intelligent automation - **Impact:** - 300% increase in deployment frequency - 30% reduction in operational costs - 85% reduction in deployment-related incidents - **Technologies:** Python, LangChain, OpenAI GPT-

4, Azure OpenAI, Kubernetes, GitLab CI/CD - **AI Components:** Custom RAG system for documentation, automated code review agents, intelligent error detection

## 2023 | Ypsomed | Switzerland

**IoT Data Intelligence & Predictive Systems - Challenge:** Complex IoT device data requiring intelligent analysis and predictive maintenance - **Solution:** Implemented ML-powered analytics pipeline with automated anomaly detection - **Impact:** - 40% reduction in versioning conflicts through intelligent merge strategies - Predictive maintenance reducing device failures by 35% - **Technologies:** Python, TensorFlow, Azure ML, IoT Hub, Time-series analysis - **AI Components:** Anomaly detection models, predictive maintenance algorithms, automated data quality checks

## 2022 | Cushman & Wakefield | London, UK

**Intelligent ETL & Data Pipeline Automation - Challenge:** Manual data processing creating bottlenecks in real estate analytics - **Solution:** Built AI-enhanced ETL pipelines with intelligent data validation and transformation - **Impact:** - 50% increase in data processing speed - 90% reduction in data quality issues - Real-time analytics enabling faster business decisions - **Technologies:** Python, Apache Airflow, Azure Data Factory, ML-based data validation - **AI Components:** Automated data quality scoring, intelligent schema mapping, anomaly detection in data flows

---

## TECHNICAL EXPERTISE

---

### AI/ML Frameworks & Tools

**LLM Platforms:** OpenAI API (GPT-4, GPT-3.5), Anthropic Claude, Azure OpenAI Service, AWS Bedrock

**Orchestration:** LangChain, LlamaIndex, Semantic Kernel, Haystack

**Vector Databases:** Pinecone, Weaviate, ChromaDB, FAISS, Qdrant, Milvus

**ML Frameworks:** TensorFlow, PyTorch, Scikit-learn, Hugging Face Transformers

**Fine-tuning:** LoRA, QLoRA, PEFT, Supervised Fine-Tuning (SFT)

**Evaluation:** RAGAS, LangSmith, Weights & Biases, MLflow

### AI Development & MLOps

**Languages:** Python (Advanced), JavaScript/TypeScript, Bash, SQL

**AI Libraries:** OpenAI Python SDK, LangChain, pandas, numpy, sentence-transformers

**MLOps:** MLflow, Kubeflow, Azure ML, AWS SageMaker, Model versioning

**Experiment Tracking:** Weights & Biases, MLflow, TensorBoard

**Data Processing:** Apache Spark, Pandas, Dask, Apache Airflow

## Cloud & Infrastructure

**Cloud Platforms:** Azure (Azure OpenAI, ML, Functions), AWS (Bedrock, Lambda, SageMaker), GCP

**Containers:** Docker, Kubernetes, Helm (for ML workloads)

**Infrastructure as Code:** Terraform, Ansible

**CI/CD:** Jenkins, GitLab CI, GitHub Actions, Azure DevOps

**Monitoring:** Prometheus, Grafana, ELK Stack, Application Insights

## AI Security & Compliance

**Security Clearances:** UK SC (valid until 2028), NATO (valid until 2029)

**AI Security:** Prompt injection prevention, adversarial robustness, model security

**Data Privacy:** PII detection/redaction, data anonymization, GDPR compliance

**Governance:** Model governance frameworks, AI ethics, bias detection

---

## NOTABLE AI PROJECTS & IMPLEMENTATIONS

---

### Production RAG Systems

- Built enterprise-grade RAG systems processing 10M+ documents
- Implemented hybrid search combining semantic and keyword search
- Optimized retrieval pipelines reducing latency by 60%
- Custom chunking strategies improving retrieval accuracy by 40%

### AI Agent Development

- Designed multi-agent systems for complex workflow automation
- Implemented ReAct and function-calling patterns for tool use
- Built autonomous agents with memory and planning capabilities
- Created agent orchestration frameworks for parallel execution

## LLM Fine-tuning & Optimization

- Fine-tuned domain-specific models using LoRA/QLoRA
- Optimized prompts achieving 35% cost reduction while improving accuracy
- Implemented caching strategies reducing API costs by 50%
- Model quantization for edge deployment scenarios

## AI Security Implementations

- Built prompt injection detection systems with 95% accuracy
- Implemented PII detection/redaction pipelines for sensitive data
- Created guardrails preventing model misuse and hallucinations
- Developed secure model serving infrastructure with encryption at rest/transit

---

## PROFESSIONAL EXPERIENCE HIGHLIGHTS

---

### Senior AI Engineer / DevOps Lead | 2020 - Present

*Goldman Sachs, Ypsomed, Cushman & Wakefield*

- Led AI transformation initiatives across finance, healthcare, and real estate sectors
- Architected production LLM systems serving millions of requests monthly
- Implemented MLOps pipelines reducing model deployment time from weeks to hours
- Mentored engineering teams on AI best practices and secure deployment patterns
- Established AI governance frameworks ensuring compliance and ethical AI use

### Enterprise Transformation Architect | 2016 - 2020

*Microsoft, Emerson, Various Fortune 500*

- Led digital transformation initiatives incorporating AI/ML capabilities
- Built ML-powered automation systems increasing operational efficiency by 50%
- Designed cloud-native architectures supporting ML workloads at scale
- Evangelized AI adoption through workshops, training, and proof-of-concepts

---

## EDUCATION

---





## Bachelor of Science

Southern New Hampshire University, USA  | 2013

---

## CERTIFICATIONS & CONTINUOUS LEARNING

---

-  **Microsoft Certified Architect in Cloud Solutions** (70-532)
-  **Azure AI Engineer Associate**
-  **AWS Machine Learning Specialty**
-  **Certified Kubernetes Administrator (CKA)**

### Continuous Learning:

- Active contributor to AI/ML open-source projects - Regular participant in AI research paper discussions - Following latest developments in LLM research (GPT, Claude, Llama, Mistral) - Experimenting with cutting-edge AI techniques and frameworks

---



## SPEAKING & THOUGHT LEADERSHIP

---

- Technical blog on AI engineering best practices at [rifaterdemsahin.com](https://rifaterdemsahin.com)
  - Internal tech talks on RAG architectures and LLM security
  - Mentorship in AI/ML communities and engineering teams
  - Code examples and tutorials shared via GitHub
- 

## SECURITY CLEARANCES

---

-  **UK SC (Security Check)** - Valid until 2028
  -  **NATO Clearance** - Valid until 2029
  - ✓ **Background Checks:** Watchdog (2024), Sterling (2019)
- 

## AVAILABILITY & CONTACT

---

**Immediate Availability** for contract or permanent AI engineering roles



**Schedule a Discussion:** <https://calendly.com/rifaterdem/schedule>



**Email:** [contact@rifaterdemsahin.com](mailto:contact@rifaterdemsahin.com)



**Phone:** +44 7848 024173

---

## SUPPORTING DOCUMENTS

---



**Technical Portfolio & Presentations:**

[https://rifaterdemsahin.com/wp-](https://rifaterdemsahin.com/wp-content/uploads/2025/02/rifaterdemsahinprofilepresentation.v2025.2.pdf)

[content/uploads/2025/02/rifaterdemsahinprofilepresentation.v2025.2.pdf](https://rifaterdemsahin.com/wp-content/uploads/2025/02/rifaterdemsahinprofilepresentation.v2025.2.pdf)

---

*References and detailed project portfolios available upon request*