
**t**: +44(0)7399 575 082 | **e:** candidates@frs-online.com | **w**: https://FRS-online.com

Role: Architect (A.I. / Artificial Intelligence focussed)
Location: Pune or Bengaluru - India
Experience: 12+ years
Type: Permanent / Full-time
Employer: Multi-National Consultancy
Salary: Depending on Experience
Job Ref: 5165

## Job Summary:

- This will be a high-visibility R&D initiative in both financial research and data intelligence.
- We are seeking an experienced, hands-on AI Architect (12+ years) to lead technical delivery and client engagement for Agentic AI platforms.
- The role owns end-to-end design, productionisation and operationalisation of LLM-based solutions (RAG, fine-tuning, model serving) on Azure, and will act as the technical face to the customer while leading a cross-functional delivery team.

## What Success Looks like:

- Deliver production Agentic AI features that meet SLA targets for latency, throughput and reliability.
- Reduce time-to-value for LLM integrations via repeatable patterns, MLOps pipelines and reusable components.
- Maintain model governance, explainability and security posture appropriate for sensitive data domains.

## Key Responsibilities:

- Architect and build Agentic AI systems: orchestrate agents, action executors, retrieval layers, and feedback loops.
- Design and implement LLM solutions (RAG, retrieval chains, prompt engineering, fine-tuning/LoRA/P-tuning) for production use.
- Own model deployment, serving and scaling on Azure (Azure AI, Azure ML, AKS, container registries) and hybrid setups.
- Build MLOps & ModelOps pipelines: CI/CD for models and services, automated testing, monitoring, drift detection and rollbacks.
- Lead data pipelines for retrieval: vector stores, semantic search, indexing, embeddings, data privacy & access controls.
- Implement model explainability, confidence scoring, adversarial protections and prompt security (prompt injection mitigation).
  Define and enforce model governance: versioning, reproducibility, lineage, audit trails and compliance.
- Collaborate with product, data engineering, security, DevOps and UX to ensure integrated delivery and acceptance.


**t**: +44(0)7399 575 082 | **e:** candidates@frs-online.com | **w**: https://FRS-online.com

- Mentor and upskill engineers; conduct technical reviews and pair programming; recruit when required.
- Act as primary technical contact for clients: present designs, lead architecture reviews, and support RFP/interview processes.

## Required Skills & Experience:

- 12+ years in software engineering/AI with demonstrable, hands-on production experience.
- Deep experience with LLMs and RAG architectures: retrieval design, vector DBs (Pinecone/Weaviate/Milvus), embeddings.
- Proven track record in fine-tuning/customising LLMs (LoRA, full-fine tune, instruction tuning) and prompt engineering.
- Strong Azure AI stack experience: Azure OpenAI/GPT, Azure ML, AKS, Azure Functions, KeyVault, Data Factory/Synapse.
- Expertise in Agentic frameworks and orchestration (LangChain, LangGraph, custom agent frameworks).
- Production MLOps/ModelOps: CI/CD for models, model registry, automated testing, monitoring (Prometheus/Grafana/ELK), drift detection.
- Backend engineering: Python, FastAPI, microservices, Docker, Kubernetes, gRPC/REST, event-driven architectures.
- Data engineering basics: SQL/NoSQL, ETL, schema design, data lineage and data privacy controls.
- Security & compliance: secrets management, access controls, vulnerability remediation, data encryption in transit & at rest.

## Good-to-Have Skills:

- Experience with hybrid/multi-cloud deployments and avoiding provider lock-in.
- Familiarity with LangGraph, agentic safety patterns, and adversarial robustness testing.
- Prior exposure to financial data or private markets / regulated data handling.
- Experience with model explainability tools (SHAP, LIME, integrated gradients) and bias/fairness testing.
- Familiarity with Terraform/ARM for infra as code and GitOps workflows.

## Are You or Anyone You Know Interested?

Please express interest by **email to Sanjay using sanjay@frs-online.com** adding the following **Job reference 5165** in the subject-line after providing **required responses for points below**:

1. CV & Contact Details (email, telephone with current city):
2. Availability (earliest start date or current notice period):
3. Current CTC:
4. Expected CTC: