CONCEPT

Deploying ALLaM and Arabic Large Language Models in Saudi Organisations: RAG and Fine-Tuning



Deploying ALLaM and Arabic Large Language Models in Saudi Organisations: RAG and Fine-Tuning

Technical depth: Practitioner · **Practical mode:** Lab

Introduction

Saudi organisations serving Arabic-speaking customers and staff need language models that handle Arabic morphology, Saudi and Gulf dialects and formal registers. ALLaM, the Arabic model family developed at SDAIA and now carried forward by HUMAIN, offers a national option alongside multilingual models, but choosing, adapting and hosting it within PDPL and NCA requirements takes evidence rather than assumptions. This coreconcept course equips technical teams to evaluate and deploy ALLaM and other Arabic models, from benchmarking through retrieval pipelines to production controls. Participants build in hands-on labs and leave with an Arabic LLM Deployment Blueprint backed by a working prototype.

Course Objectives

- Compare ALLaM with other Arabic-centric and multilingual language models on quality, licensing, hosting and cost for a defined use case
- Evaluate candidate models with Arabic benchmarks and task-specific test sets before selection
- Build retrieval-augmented generation pipelines on ALLaM that answer accurately from Arabic documents
- Adapt ALLaM to domain language using prompt design and LoRA fine-tuning where the evidence supports it
- Apply PDPL, NCA cloud and SDAIA generative AI requirements alongside OWASP and ISO/IEC 42001 controls to Arabic LLM applications
- Produce an Arabic LLM Deployment Blueprint backed by a tested prototype

Target Audience

- Data scientists and machine learning engineers building language applications
- Solution and software engineers integrating language models into business systems
- AI product owners responsible for Arabic chatbots, assistants and search services
- Data and platform engineers managing model hosting, GPUs and pipelines in KSA
- Information security and data protection specialists reviewing AI applications against PDPL and NCA controls
- Digital service teams preparing Arabic content and knowledge bases for AI use

Course Outline

Day 1: ALLaM and the Arabic LLM Landscape in Saudi Arabia

- Arabic NLP Challenges: Morphology, Dialects, Diglossia and Script Variation
- ALLaM Model Family: ALLaM 7B Open Weights, ALLaM 34B and HUMAIN Chat
- ALLaM Compared with Jais, Falcon-Arabic and Aya
- Deployment Options in KSA: IBM watsonx, Azure AI Foundry, Open Weights and CCC-2:2024 Hosting
- Arabic Use Case Screening Matrix and Current-State Content, Data and Infrastructure Review

Day 2: Architecture and Evaluation Standards

- Transformer Architecture, Tokeniser Fertility and Arabic Context Costs
- Hugging Face Transformers and vLLM Serving Stack
- Retrieval-Augmented Generation Architecture with Arabic Embedding Models
- Arabic Benchmarks: ArabicMMLU, Open Arabic LLM Leaderboard and BALSAM
- AraGen 3C3H Criteria for Generative Quality Evaluation

Day 3: Building Arabic Applications on ALLaM

- ALLaM Prompt Design for Formal Arabic and Saudi Dialect Requests
- Arabic Text Normalisation and Pre-Processing with CAMeL Tools
- Document Chunking and Vector Index Build for Arabic Corpora
- Parameter-Efficient Fine-Tuning of ALLaM 7B with LoRA on Domain Data
- RAG Pipeline Evaluation with RAGAS Faithfulness and Relevance Metrics

Day 4: Security, Risk and Optimisation

- OWASP Top 10 for LLM Applications Controls for Arabic Chatbots
- Arabic Red-Teaming for Hallucination and Cultural Alignment
- PDPL and NDMO Data Classification Controls: Masking, Retention and Access Logging
- Quantisation, Batching and GPU Cost Optimisation
- SDAIA Generative AI Guidelines, ISO/IEC 42001 Controls and Production Model Monitoring

Day 5: Lab Build and the Deployment Blueprint

- Lab Build: ALLaM-Based Arabic Policy Assistant over Internal Documents
- Side-by-Side ALLaM and Multilingual Model Evaluation on the Lab Test Set
- Production Readiness Checklist and Go-Live Criteria
- Arabic LLM Deployment Blueprint Drafting
- Technical Review Board and Blueprint Defence

Skills You Will Gain

- Arabic NLP Engineering
- LLM Model Selection
- Retrieval-Augmented Generation Design
- Parameter-Efficient Fine-Tuning
- LLM Evaluation and Benchmarking
- LLM Application Security
- Inference Cost Optimisation
- Production Model Monitoring

Why Attend This Course

- Leave with an Arabic LLM Deployment Blueprint and a working ALLaM-based prototype built on a representative document set
- Choose between ALLaM and multilingual models on evidence from your own test data rather than leaderboard headlines
- Anticipate the dialect, normalisation, retrieval and data residency problems that degrade Arabic deployments in production
- Work alongside engineers and product owners from Saudi government, banking, telecoms and healthcare facing the same deployment choices

Conclusion

Arabic language models deliver value only when the chosen model, the retrieval design and the controls around them have been tested against the organisation's own content and users. This course moves from ALLaM and the Arabic model landscape in Saudi Arabia, through architecture and benchmarks, to hands-on prompting, retrieval, fine-tuning and evaluation, and then to PDPL, security and cost. The final day brings that work together in a lab prototype and an Arabic LLM Deployment Blueprint, ready for technical and governance review in the participant's organisation.