



Arabic NLP and Speech AI Engineering

Course: Text Analytics, ASR and Voice Applications



Arabic NLP and Speech AI Engineering Course: Text Analytics, ASR and Voice Applications

Introduction:

The Arabic NLP and Speech AI Engineering Course on text analytics, ASR and voice applications is a 5-day course for data scientists, NLP and speech engineers that ends with an Arabic Text and Speech Analytics Solution Plan for a case organisation. Organisations hold large volumes of Arabic documents, posts and recorded calls, yet generic tools mishandle morphology, dialect variation and missing diacritics, so analytics and transcripts are unreliable. Nominees already write Python and train machine learning models at work and learn through hands-on labs on Arabic corpora and audio. CoreConcept Training Center delivers this Arabic NLP and speech engineering course.

Course Objectives:

- Build Arabic text pre-processing pipelines covering orthographic normalisation, tokenisation, morphological analysis and diacritisation
- Train and evaluate dialect identification, sentiment analysis and named entity recognition models on Arabic text
- Select and apply Arabic word and sentence embeddings for semantic search, clustering and topic discovery
- Fine-tune and evaluate Arabic automatic speech recognition models using word error rate and character error rate
- Assemble text-to-speech and speech analytics pipelines for call recordings and public-facing voice services
- Produce an Arabic Text and Speech Analytics Solution Plan with evaluation targets and a deployment design

Target Audience:

- Data science staff responsible for building text classification and analytics models on Arabic content
 - Machine learning engineers responsible for training, packaging and serving language and audio models
 - Speech technology staff responsible for transcription, voice interfaces and audio data pipelines
 - Analytics staff responsible for turning Arabic customer feedback and call recordings into operational insight
 - Technical leads responsible for selecting Arabic language tooling and setting model quality targets
-

Course Outline:

Day 1: Arabic Language Structure and Text Processing Foundations

- Arabic Script, Orthographic Variation and Normalisation Rules
- Diglossia Between Modern Standard Arabic and Spoken Dialects
- CAMEL Tools Word Tokenisation and Text Cleaning Utilities
- Root and Pattern Morphology with Clitic Segmentation
- Current-State Audit of Arabic Text and Audio Assets

Day 2: Arabic NLP Model Architectures and Evaluation Methods

- Morphological Analysis and Disambiguation with CAMEL Tools
- Automatic Diacritisation and Dediacritisation for Downstream Tasks
- Subword Tokenisers and Pretrained Arabic Transformer Encoders
- Precision, Recall and F1 Scoring for Arabic Labelling Tasks
- Annotation Guidelines and Inter-Annotator Agreement for Arabic Corpora

Day 3: Arabic Text Analytics Labs

- Dialect Identification Classifier Training on Regional Text Samples
- Aspect-Level Sentiment Analysis for Arabic Customer Feedback
- Named Entity Recognition for Arabic Persons, Places and Organisations
- Arabic Sentence Embeddings for Semantic Search and Clustering
- Topic Discovery Across Arabic Complaints and Service Requests

Day 4: Arabic Speech Recognition, Synthesis and Production Risks

- Acoustic Features, CTC and Attention Models for Speech Recognition
- Whisper Transcription and Language Prefix Fine-Tuning for Arabic Audio
- Word Error Rate and Character Error Rate Diagnostics
- Arabic Text-to-Speech Front End with Diacritised Input
- Code-Switching, Noisy Telephony Audio and Dialect Drift Controls

Day 5: Lab Build and Arabic Text and Speech Analytics Solution Plan

- Lab Build Call Recording Transcription and Sentiment Pipeline
 - Lab Build Government Service Request Entity Extraction Pipeline
 - Lab Benchmark Report Against Word Error Rate and F1 Targets
 - Model Serving, Monitoring and Retraining Design for Language Pipelines
 - Arabic Text and Speech Analytics Solution Plan Review Panel
-

Skills You Will Gain:

- Arabic Text Normalisation
- Morphological Analysis
- Dialect Identification
- Sentiment and Entity Extraction
- Arabic Embedding Selection
- Speech Recognition Fine-Tuning
- Speech Error Analysis
- Text-to-Speech Pipeline Design

Why Attend This Course:

- Deliver an Arabic Text and Speech Analytics Solution Plan to the data science lead and the service owner, backed by benchmark results from the labs
- Choose between off-the-shelf and fine-tuned Arabic speech and text models on measured error rates rather than vendor claims
- Prevent costly rework caused by unnormalised text, ignored dialects and noisy telephony audio before models reach production
- Coach colleagues in the normalisation, annotation and error analysis routines practised during the hands-on labs

Conclusion:

Back at work, the participant hands the Arabic Text and Speech Analytics Solution Plan to the data science lead and the owner of the service it supports. They use it to approve which text and voice use cases move to build, which models to fine-tune and which quality targets each must reach before release. After the first pipeline runs on live data, the unit should compare production word error rate and F1 scores with the lab benchmarks and record where dialect coverage, audio quality or annotation needed correction.

Frequently Asked Questions FAQ:

What should participants know before an Arabic NLP and speech AI engineering course?

Participants should write Python comfortably, understand supervised machine learning and have trained at least one classification model. Reading Arabic is needed for error analysis. Prior speech processing experience is not required, because acoustic features and recognition models are introduced before the speech labs.

How does an Arabic NLP and speech AI engineering course differ from a course on deploying large language models?

This course builds task-specific Arabic text and speech models: tokenisation, morphology, dialect identification, sentiment, entities, embeddings, speech recognition and synthesis. A large language model course focuses on prompting, retrieval-augmented generation and fine-tuning general generative models rather than measured text analytics and transcription pipelines.

Why does Arabic NLP and speech AI need dialect-aware processing?

Spoken and informal written Arabic differ from Modern Standard Arabic in vocabulary, morphology and pronunciation, so models trained on formal text lose accuracy on calls and social posts. Dialect identification and dialect-specific evaluation sets reveal where a model fails and where extra training data is needed.

What do participants take back from the Arabic NLP and speech AI engineering course?

Participants take back an Arabic Text and Speech Analytics Solution Plan for a case organisation, working lab pipelines for transcription, sentiment and entity extraction, and a benchmark report showing word error rate and F1 results against agreed targets.
