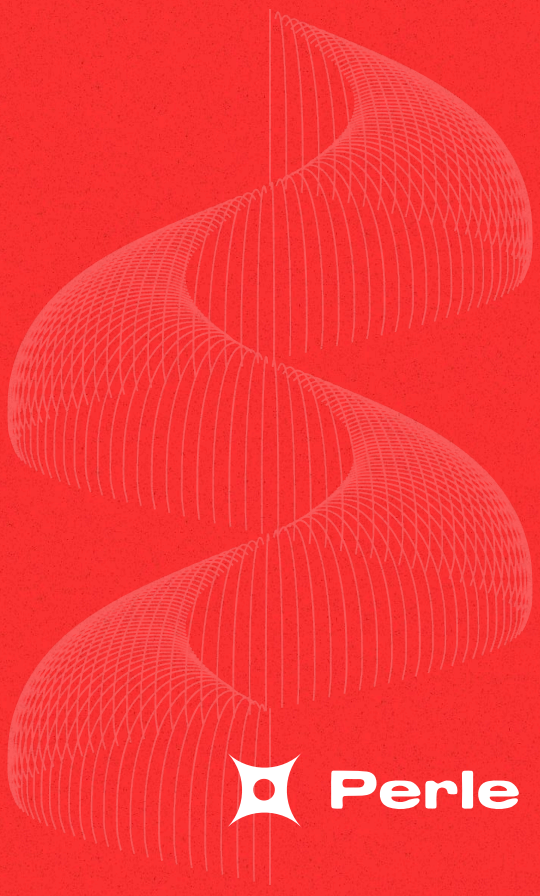
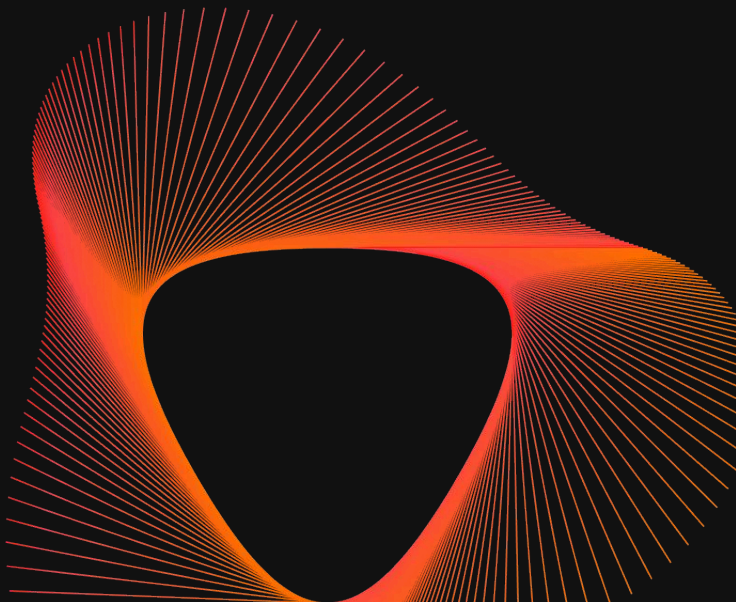
Breaking Language Barriers: AI Technologies in Arabic Legal Document Analysis

Sajjad Abdoli, Founding AI Scientist, Perle



Perle

Index

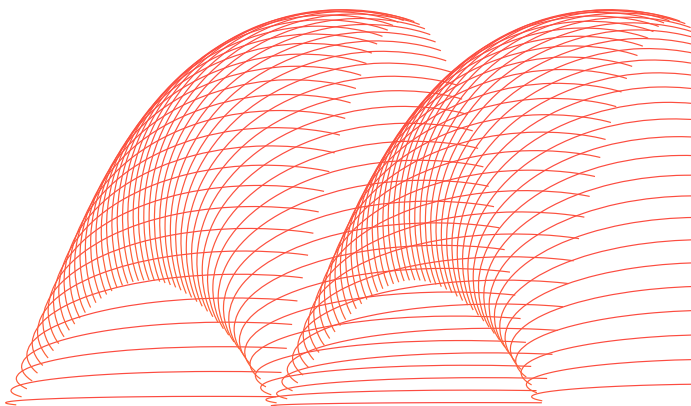


Introduction

As AI systems increasingly support high-stakes enterprise workflows, their application to complex domains like legal document analysis reveals critical gaps-especially in languages beyond English. Languages such as Arabic and Mandarin introduce structural, linguistic, and cultural complexities that challenge large language models (LLMs). These include right-to-left script, rich morphology, nuanced legal terminology, and significant variation across dialects-all of which can degrade model performance and reliability.

In legal contexts, these limitations aren't just theoretical. They can lead to tangible downstream risks-from misinterpreted contract clauses to incorrect legal reasoning and regulatory exposure.

This paper explores the current limitations of LLMs on Arabic legal tasks, outlines our system design, and introduces the Arabic Legal AI Preview-an experts-in-the-loop application built to demonstrate real-world use cases for legal professionals.



The Challenges of Applying LLMs to Arabic Legal Language

Arabic poses several critical challenges for AI systems originally trained on high-resourced languages like English:

01

Atypical Sentence Patterns

Arabic outputs may not always align with sentence structures typically used in other languages.

02

Limited Inter-Sentence Coherence

Transitions between sentences are sometimes abrupt, reducing overall contextual flow.

03

Translation Artifacts

Many AI-generated responses feel like direct translations from English, lacking cultural and linguistic authenticity.

04

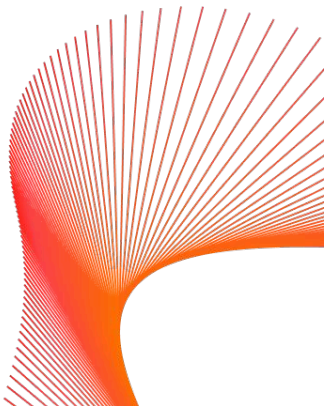
Safaty Calibration

Early testing showed harmful content generation rates in Arabic as high as **89–90%**, necessitating targeted safety interventions [1].

05

Dataset Representation Imbalance

Arabic remains severely underrepresented in LLM training corpora— **73%** of instruction-tuning datasets are English-centric [1].

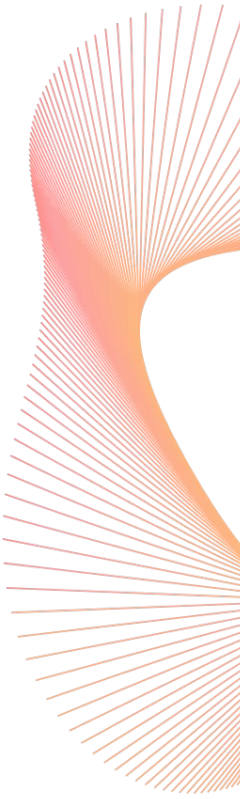


Model Advancement: C4AI Command R7B Arabic

The release of Cohere's C4AI Command R7B Arabic-a 7B-parameter, open-source multilingual model optimized for both Arabic and English-represents a notable advancement in Arabic NLP [2].

The model demonstrates strong capabilities in:

- **Instruction following**
- **Long-context coherence**
- **Retrieval-augmented generation (RAG)**
- **Capturing cultural and linguistic nuance**

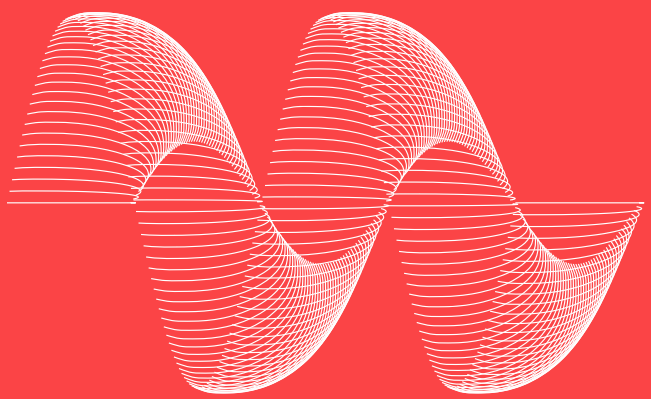


While this architecture reflects progress in multilingual LLM design, real-world deployment-particularly in high-stakes domains like law-demands more than general-purpose modeling. Task-specific fine-tuning, high-quality supervised data, and experts-in-the-loop workflows remain essential to achieving reliable performance in production environments.

Introducing the Perle Arabic Legal AI Demo

Perle is advancing the state of legal AI through domain-specific data pipelines, expert-in-the-loop validation, and LLM orchestration optimized for Arabic-language legal tasks.

Our Arabic Legal AI Demo provides a production-grade preview of our approach to document intelligence in Arabic, built atop frontier model architectures and proprietary annotation workflows.



Key Capabilities

1. Contract Summarization

Our contract summarization capability delivers a comprehensive, structured overview of legal documents by combining prompt-engineered large language models (LLMs) with rules-based post-processing techniques. Key features include:

Clause Detection
Automatically identifies and extracts critical contractual clauses such as obligations, financial terms, termination conditions, confidentiality, warranties, indemnities, and compliance requirements.

Section Classification and Canonical Labeling
Accurately categorizes sections and clauses using a standardized taxonomy, enabling consistent labeling across diverse contract types and formats.

Hierarchical Output Formatting
Produces structured outputs that reflect the logical hierarchy of the contract, facilitating seamless integration with downstream contract lifecycle management (CLM) systems and legal analytics tools.

Document Summary
Generates a detailed, human-readable summary highlighting the contract's essential elements, including:

- Identification of main parties and the overall purpose of the agreement.
- Precise extraction and interpretation of key obligations and responsibilities for each party.
- Comprehensive capture of financial terms, including payment schedules, amounts, and penalties.
- Analysis of quality standards, performance metrics, and implementation timelines.
- Overview of dispute resolution mechanisms, termination rights, and renewal clauses.
- Detection of any special terms, contingencies, or regulatory provisions unique to the contract.

Cross-Referencing and Consistency Checks
Flags inconsistencies or contradictions across clauses to enhance contract review quality and risk assessment.

2. AI-Powered Q&A in Arabic

Our AI-powered question-and-answer system enables users to interact with complex legal documents in Arabic through natural language queries.

This feature combines semantic search, contextual reasoning, and language-specific tuning to support accurate and efficient information retrieval.

Arabic Query Understanding

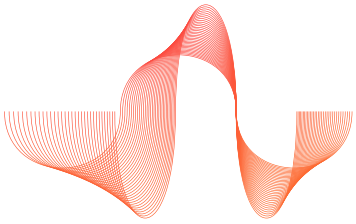
Natively interprets queries in Modern Standard Arabic and dialects, with support for legal phrasing and nuanced linguistic structures.

Semantic Retrieval + LLM Reasoning

Uses dense vector search to retrieve relevant passages and LLMs fine-tuned for Arabic to generate grounded, context-aware answers with source references.

Handles Complex Legal Queries

Deconstructs compound or ambiguous questions, infers implied intent, and provides clear, accurate responses-even in legally intricate cases.



3. Question Similarity Graph

Our interactive question similarity graph enables users and analysts to explore patterns in query behavior, understand user intent, and identify coverage gaps.

It leverages semantic embeddings and real-time visualization to reveal relationships between questions across large datasets.

Node-Based Query Graph

Displays each user query as a node derived from dense embeddings, allowing intuitive visualization of the semantic landscape.

Semantic Similarity Edges

Connects related queries using edge weights based on semantic similarity, helping surface related intents and repeated information needs.

Dynamic Clustering

Automatically groups queries into thematic clusters, making it easier to detect trends, emerging topics, and under-addressed areas.

Expert-Guided Model Development for Arabic Legal AI

Deploying AI in legal contexts-especially in Arabic-requires more than off-the-shelf LLMs. It demands an end-to-end system built with precision, relevance, and continuous oversight. At Perle, we've developed an experts-in-the-Loop (EITL) framework to ground every step of model development in domain expertise. This section outlines our approach.

Native Expert Benchmarking

Our evaluation pipeline incorporates the judgment of native Arabic-speaking legal professionals and linguists through a structured multi-stage process:

- **Layered Review**: All outputs are reviewed across multiple tiers combining legal and linguistic expertise.
- **Agreement Metrics**: IInter-annotator agreement (IAA) is quantified using Cohen's and custom scoring rubrics to ensure evaluation consistency.
- **Feedback Integration**: Structured expert insights are continuously fed back into model fine-tuning and prompt optimization.

Purpose-Built Dataset Creation

A high-performing legal AI system must be trained on data that mirrors the linguistic, structural, and doctrinal nuances of the field.

- **Cross-Jurisdictional Coverage**: Our dataset spans civil, commercial, and administrative legal documents across Arabic-speaking countries.
- **Custom Annotation Schema**: Tailored for legal NLP tasks such as:
  - Legal entity extraction
  - Clause classification
  - Argument role identification
  - Cross reference solution
- **Language-Specific Design**: We account for dialectal variation, Arabic syntax and morphology, and culturally specific legal terminology.

Robust QA Evaluation Criteria

Formatting: Paragraph structure and Arabic-specific layout.

Spelling/Grammar: Diacritics, terminology, and syntax.

Instruction Following: Adherence to user queries.

Verbosity: Balance between conciseness and completeness.

Truthfulness: Legal and factual correctness.

Missing Parts: Identification of content omissions.

Overall Quality: Relevance and usability in workflows.

Foundational Model Assessment

Rather than building on assumptions, we rigorously evaluate which models can serve as the most effective base for Arabic legal AI.

Comparative Benchmarking

Comparative Benchmarking: We test GPT-4, LLaMA variants, and Aya, focusing on legal-specific tasks like summarization, clause retrieval, and QA.

Performance on Arabic Legal Tasks

Performance on Arabic Legal Tasks: We compare model outputs for fluency, accuracy, and legal relevance in non-English legal corpora.

Custom Pretraining:

Custom Pretraining: We are actively exploring the development of a proprietary legal foundational model using our annotated dataset, tailored for MENA jurisdictions.

Join Us in Transforming Legal AI

At Perle, we are advancing the state of legal AI-especially for underrepresented languages such as Arabic-through a combination of domain expertise, rigorous annotation standards, and responsible AI methodologies.

Our platform integrates experts-in-the-loop feedback, domain-specific datasets, and model evaluation frameworks to deliver solutions that prioritize both performance and trustworthiness.

Explore the future of Arabic legal NLP by testing our demo-experience how precise, scalable AI can transform your legal workflows.

Contact us

References

[1] Üstün, Ahmet, et al. "Aya model: An instruction finetuned open-access multilingual language model." arXiv preprint arXiv:2402.07827 (2024).

[2] Cohere team, C4AI Command R7B Arabic model