King Saud University
College of Computer and Information Sciences
Department of Computer Science

Arabic Poetry Generation using Large Language Models

Prepared by:
Mohammed Alageel
Mohammed Aldawsari
Hatim Alhomid
Msaad Alameel

Supervised by:
Prof. Mohamed Maher Ben Ismail

Outline

- Motivation
- Problem Statement
- Background
- Literature Review
- Proposed Solution
- Dataset
- Experimental Setup
- Evaluation
- Results
- Conclusion

Motivation

- **Arabic poetry:** A cornerstone of Arab cultural heritage, celebrated for its aesthetic beauty, emotional depth, and distinctive stylistic conventions.

- **AI Advancements:** Recent progress in Machine Learning (ML) and Natural Language Processing (NLP) presents novel opportunities to support and explore the Arabic poetry domain.

- **The Challenge:** Generating authentic Arabic poetry that adheres to complex metrical (Buhur) and rhyming (Qafiyah) rules while conveying profound meaning remains a significant scientific challenge.

Motivation

- **LLM Potential:** Large Language Models (LLMs) offer promising avenues by leveraging vast linguistic data and sophisticated architectures.

- **Project Aim:** This research investigates the capabilities of state-of-the-art LLMs for generating high-quality, structurally sound Arabic poetry.

Problem Statement

- The intricate structure of Arabic poetry (Buhur, Arud, Qafiyah) poses a unique challenge for Artificial Intelligence (AI).
- Current state-of-the-art Large Language Models (LLMs) struggle to replicate the traditional poetic forms accurately.
- The difficulty lies in teaching AI to adhere to these forms while maintaining the linguistic beauty and cultural authenticity of classical Arabic poetry.

أنا مشيت وطريقي طويل   والشمس تطلع وتغيب قليل

أحلم كثير وأطير فوق   يمكن أوصل للسماء بس ذوق

Goals and Objectives

- **Fine-Tune LLMs:**

  Adapt models using specialized datasets like Ashaar.

- **Adhere to Traditional Rules:**

  Develop models that generate poetry respecting Buhur, Arud, and Qafiyah.

- **Maintain Authenticity:**

  Ensure generated poetry preserves linguistic and cultural depth.

- **Bridge Tradition and AI:**

  Contribute to preserving Arabic poetry via modern technologies.

- **Evaluate Performance:**

  Assess model accuracy in following rules and maintaining authenticity.

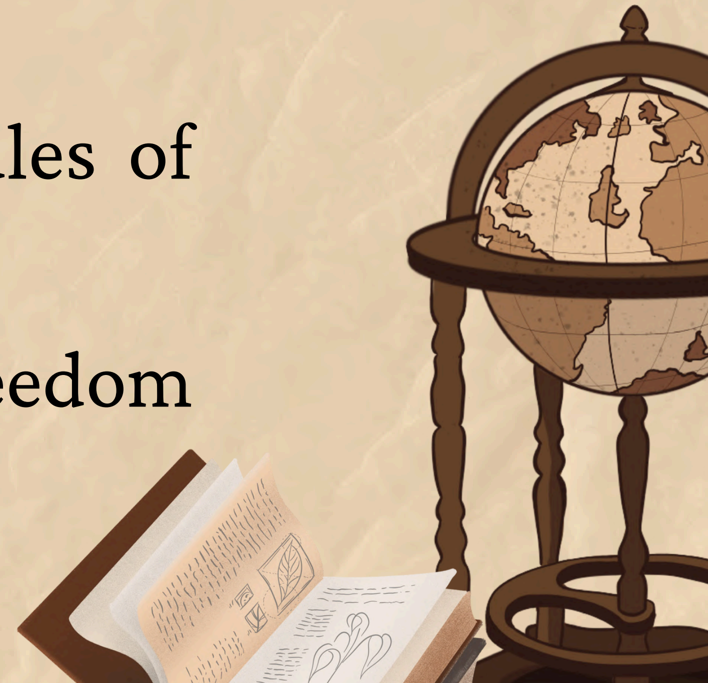
Background

Arabic Poetic Structures

- **Definition:** Intentionally crafted, structured speech, characterized by specific rhyme and meter.

- **Significance:** Deeply embedded in Arab culture and identity; Saudi Arabia's 2023 "Year of Arabic Poetry" highlights its value.

- **Forms:**
  - **Vertical Poetry (Al-shi'r al-'amudi):** Traditional, adheres to strict rules of meter and rhyme.
  - **Free Verse Poetry (Al-shi'r al-hurr):** Modern, offers more structural freedom while often retaining rhythmic qualities.



Background

Arabic Poetic Structures

وَمَنْ لَمْ يَذَقْ مُرَّ التَّعَلِّمِ سَاعَةَ   تَجَرَّعَ ذَلَّ الْجَهْلِ طَولَ حَيَاتِهِ

وَمَنْ فَاتَهُ التَّعْلِيمُ وَقْتَ شَبَابِهِ   فَكَبِّرْ عَلَيْهِ أَرْبَعًا لِوَفَاتِهِ

An example of Vertical Poetry

يا تونس الخضراء .. جئتك عاشقاً
وعلى جبيني وردة وكتاب

إني الدمشقي الذي احترف الهوى
فاخضوضرت بغنائه الأعشاب

An example of Free Verse Poetry

Background

Arabic Poetic Structures

- **Core Elements:**
  - **Meter (Al-Bahr):** The rhythmic framework of a poem (16 classical meters).
  - **Taf’eela:** Fundamental rhythmic units composing the meter.
  - **Wazn:** The overall metrical pattern and structure.
  - **Rhyme (Al-Qafiyah):** The concluding sound pattern of verses, vital for musicality and coherence.

Background

AI & LLMs for Poetic Generation

- **Artificial Intelligence (AI):** Focus on data-driven AI, where models learn patterns from extensive datasets.
- **Machine Learning (ML) & Deep Learning (DL):** Subfields of AI enabling systems to learn from data, using neural networks for complex tasks.
- **Natural Language Processing (NLP):** Equips machines to understand, interpret, and generate human language.

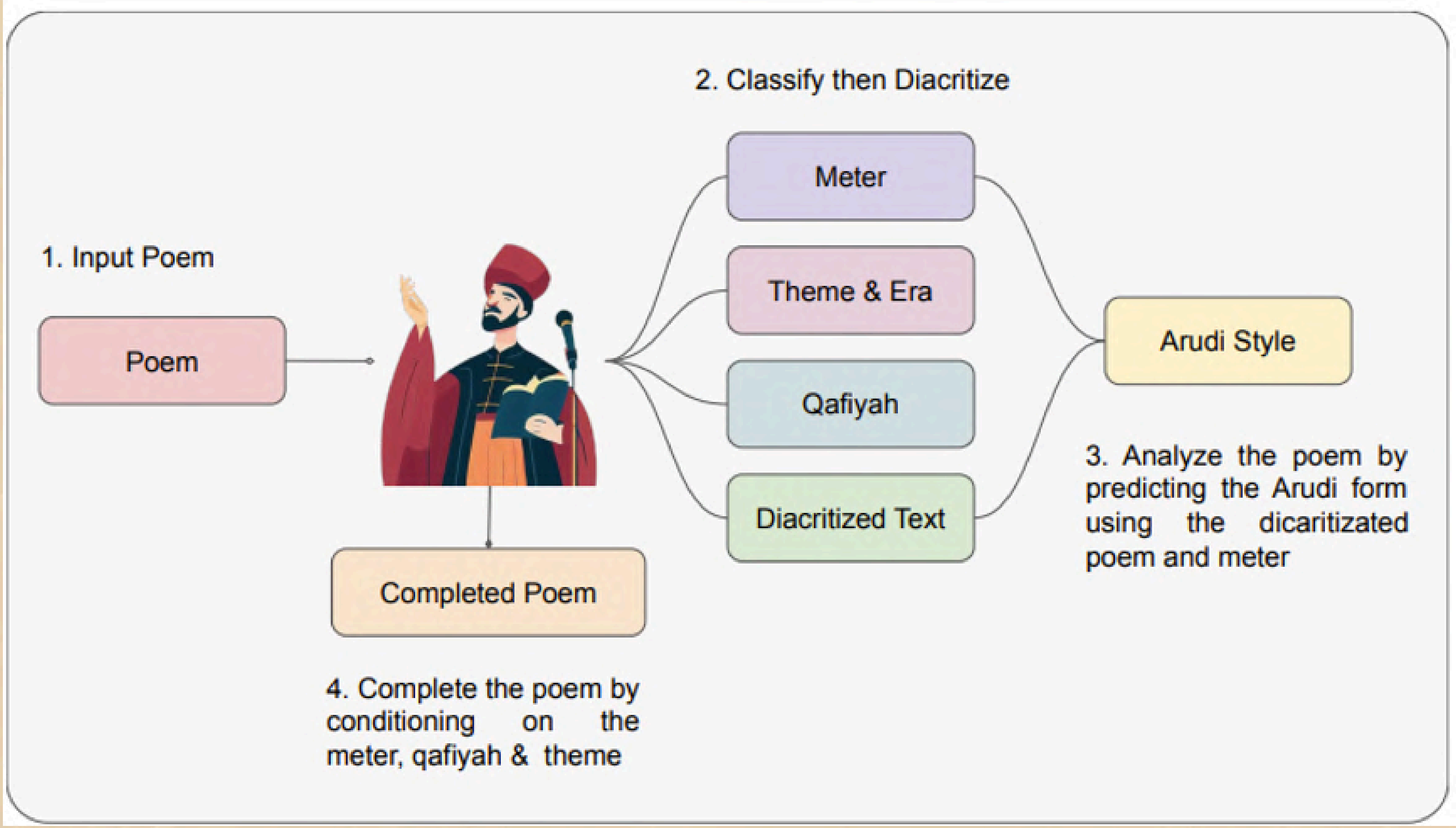
Background

AI & LLMs for Poetic Generation

- **Large Language Models (LLMs):**
  - Pre-trained on massive text corpora, then fine-tuned for specific applications.
  - Transformer Architecture: Key innovation using self-attention for parallel processing and capturing long-range dependencies.
- **Application to Poetry:** LLMs learn grammar, style, and, with specialized training, poetic structures like meter and rhyme.

Literature Review

- **Ashaar (2023):**
  - Integrates deep learning for Arabic poetry analysis and generation.
  - Focuses on meter classification, theme analysis, era recognition, and automatic diacritization.



Literature Review

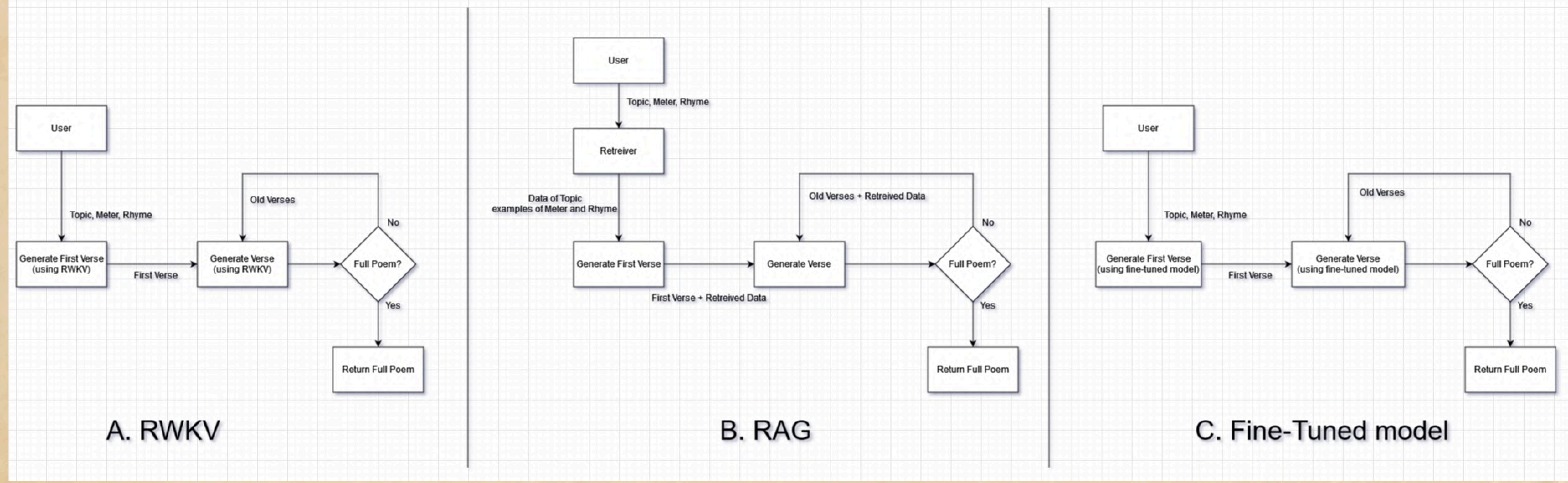
- **Transformer Models (AraGPT2, GPT-J, BERTShared):**
  - Demonstrated improved fluency and structural adherence compared to earlier models.

- **General Trends:**
  - Shift from RNNs/LSTMs to Transformer based models.
  - Challenges persist in consistently mastering complex poetic rules and cultural nuances.

Proposed Solution

- Our research explored several advanced AI methodologies for Arabic poetry generation:
  - **Receptance Weighted Key Value (RWKV)**
  - **Retrieval-Augmented Generation (RAG)**
  - **Fine-Tuning Large Language Models (LLMs) - LLaMA, Qwen, ALLaM**
- Each approach was evaluated for its potential to address the unique structural and semantic demands of Arabic poetry.

Proposed Solutions

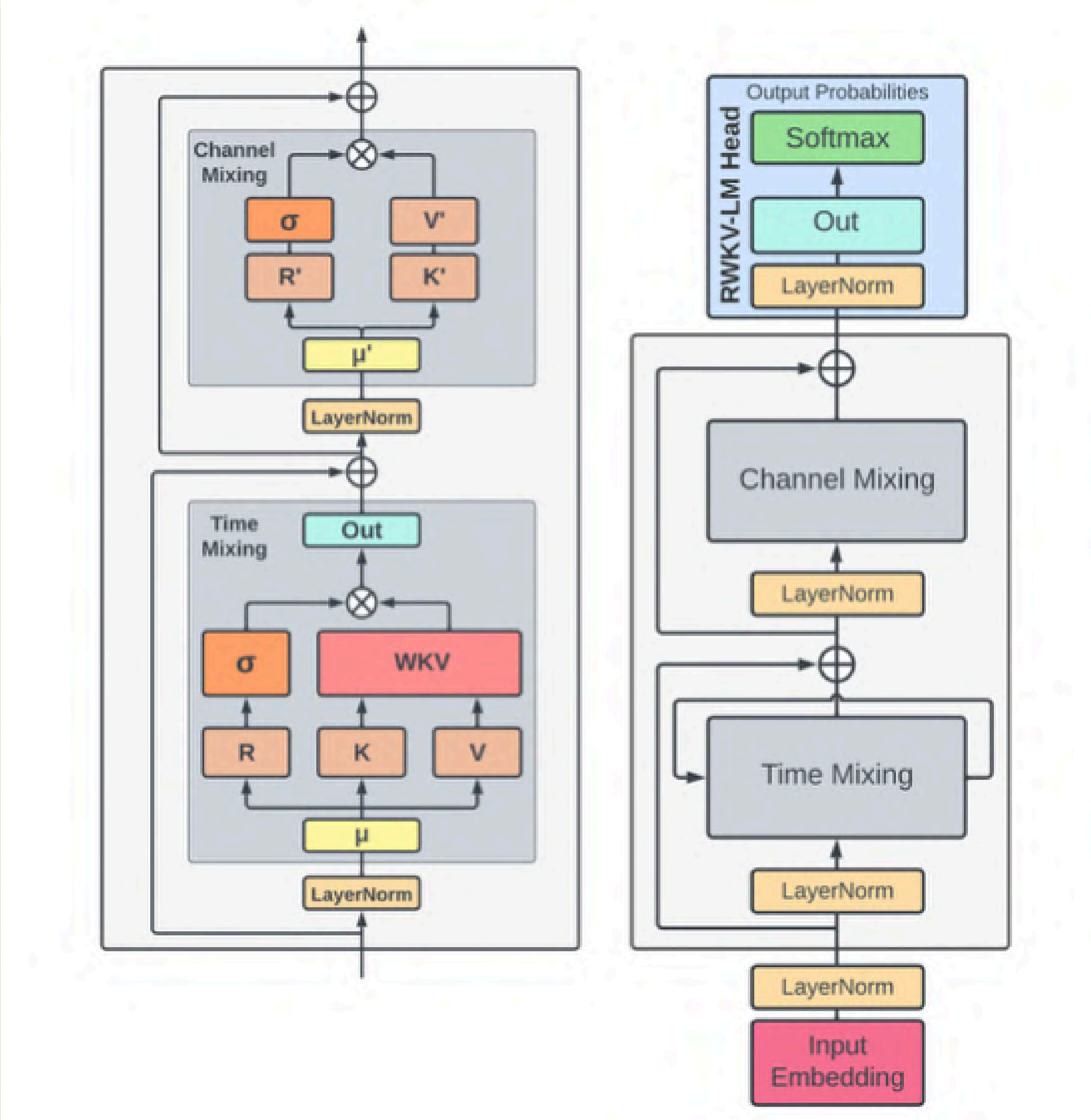


Receptance Weighted Key Value (RWKV)

Combines RNN efficiency with Transformer scalability, theoretically promising for sequential tasks like poetry generation.

Application to Arabic Poetry

Its architecture is well-suited for maintaining long-term dependencies crucial for consistent meter and rhyme.

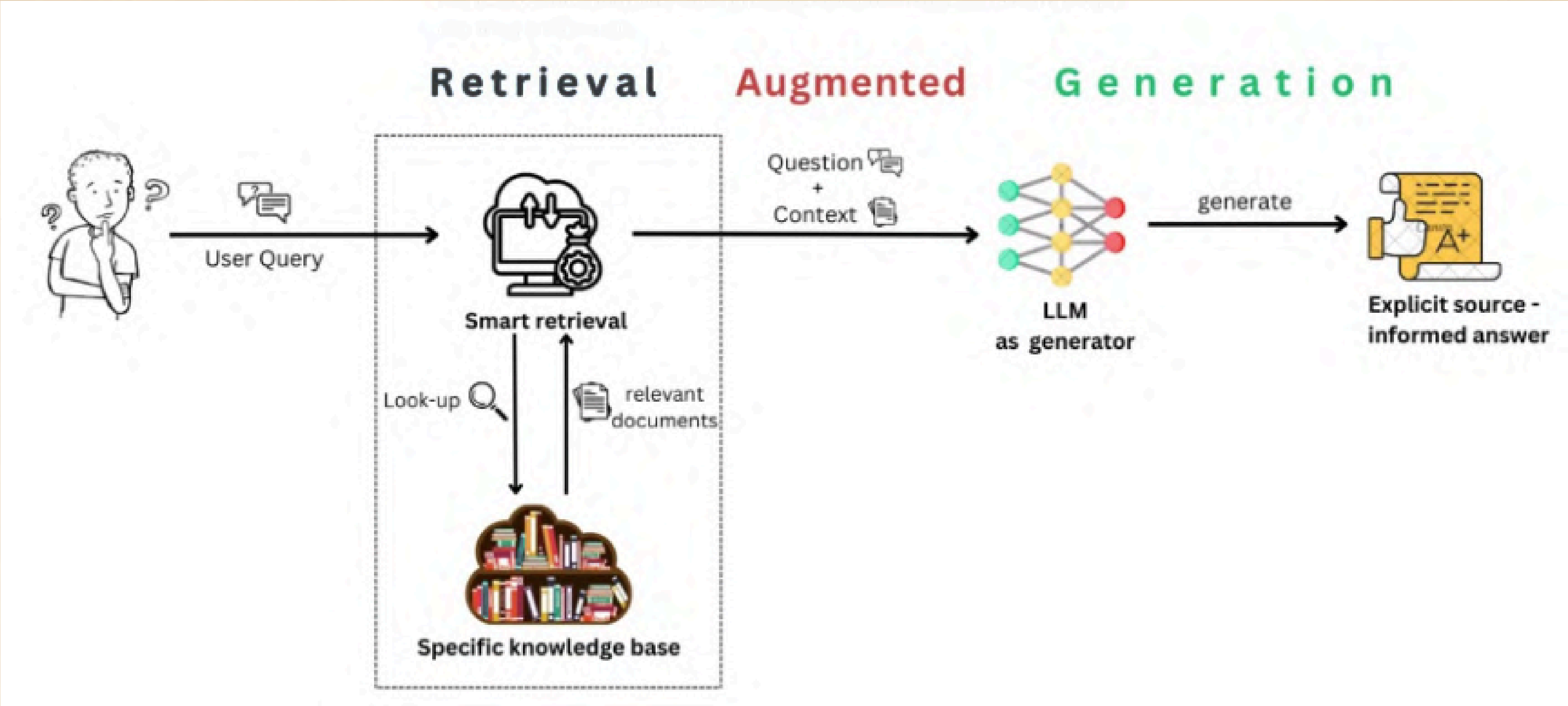


Retrieval-Augmented Generation (RAG)

Designed to enhance LLM outputs by integrating external knowledge dynamically.

Hypothesized Benefit for Poetry

Expected to improve contextual relevance, thematic consistency, and adherence to specific poetic forms by retrieving similar examples.



Fine-Tuning LLMs (LoRA)

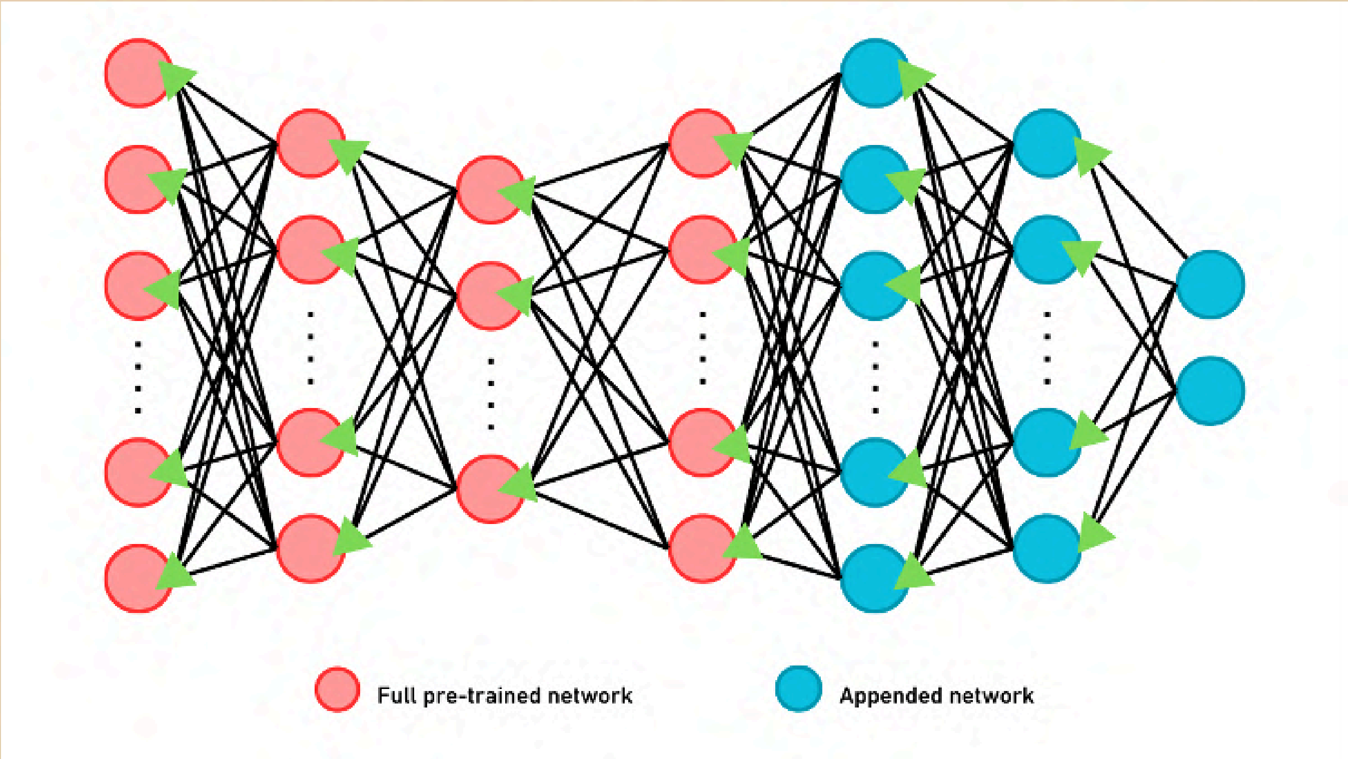
Adapting pre-trained models to the specific domain of Arabic poetry.

Low-Rank Adaptation (LoRA)

Selected due to computational resource constraints. LoRA allows for efficient fine-tuning by injecting trainable low-rank matrices into existing LLM layers while freezing most original weights.

Models

LLaMA, Qwen, ALLaM



Dataset

We have used The Ashaar Dataset is the largest publicly available collection of Arabic poetry, designed to support research in analysis and generation of Arabic poems.

Key Statistics:

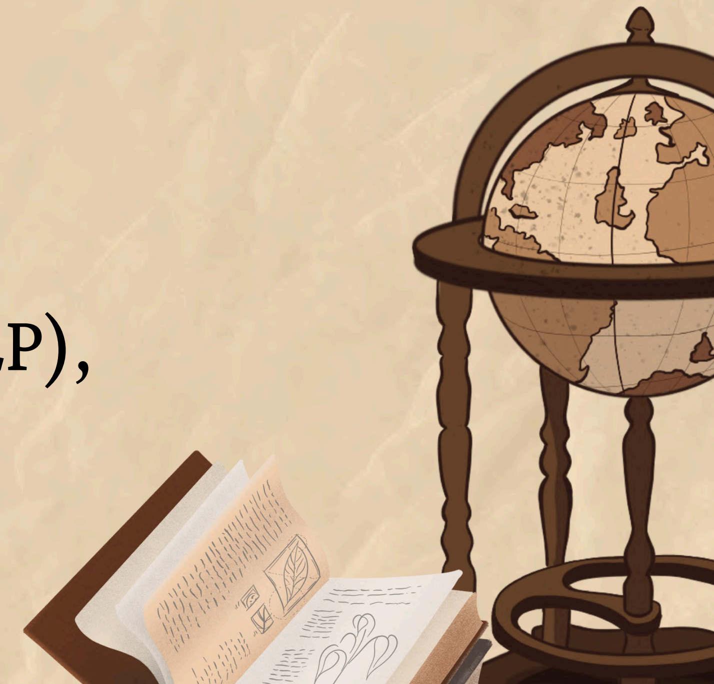
- Number of Poems: 254,630
- Number of Verses (Baits): 3,857,429
- Number of Poets Represented: 7,167

Sources Compiled from six reputable Arabic poetry websites, including:

www.aldiwan.net, www.poetry.dctabudhabi.ae, www.poetsgate.com , www.aldiwanalarabi.com, www.adab.com, www.diwany.org

Purpose:

- Facilitates research and development in Arabic Natural Language Processing (NLP), particularly in poetry analysis and generation



Dataset

poem title	poem meter	poem verses	poem theme	poem url	poet name	poet description	poet url	poet era	poet location	poem description	poem language type
string · lengths 0 – 99	string · classes 101 values	sequence · lengths 0 – 11.6k	string · classes 18 values	string · lengths 36 – 346	string · lengths 1 – 44	string · classes 840 values	string · lengths 38 – 93	string · classes 14 values	string · classes 22 values	list · lengths 1 – 290	string · classes 5 values
[illegible]	بحر الخفيف	[illegible]	قصيدة دينية	https://www.aldiwan.net/poem16182.html	الأمير منجك باشا	منجك بن محمد بن منجك بن أبي بكر بن عبد القادر...	https://www.aldiwan.net/cat-poet-alamir-mnjk-pasha	العصر العثماني	null	null	null
[illegible]	بحر مجزوء الرمل	[illegible]	قصيدة دينية	https://www.aldiwan.net/poem16183.html	الأمير منجك باشا	منجك بن محمد بن منجك بن أبي بكر بن عبد القادر...	https://www.aldiwan.net/cat-poet-alamir-mnjk-pasha	العصر العثماني	null	null	null
[illegible]	بحر البسيط	[illegible]	[illegible]	https://www.aldiwan.net/poem16184.html	الأمير منجك باشا	منجك بن محمد بن منجك بن أبي بكر بن عبد القادر...	https://www.aldiwan.net/cat-poet-alamir-mnjk-pasha	العصر العثماني	null	null	null
[illegible]	بحر الكامل	[illegible]	قصيدة عامه	https://www.aldiwan.net/poem16185.html	الأمير منجك باشا	منجك بن محمد بن منجك بن أبي بكر بن عبد القادر...	https://www.aldiwan.net/cat-poet-alamir-mnjk-pasha	العصر العثماني	null	null	null
[illegible]	بحر الوافر	[illegible]	قصيدة عامه	https://www.aldiwan.net/poem16186.html	الأمير منجك باشا	منجك بن محمد بن منجك بن أبي بكر بن عبد القادر...	https://www.aldiwan.net/cat-poet-alamir-mnjk-pasha	العصر العثماني	null	null	null
[illegible]	بحر الكامل	[illegible]	قصيدة عامه	https://www.aldiwan.net/poem16187.html	الأمير منجك باشا	منجك بن محمد بن منجك بن أبي بكر بن عبد القادر...	https://www.aldiwan.net/cat-poet-alamir-mnjk-pasha	العصر العثماني	null	null	null
[illegible]	بحر الطويل	[illegible]	قصيدة عامه	https://www.aldiwan.net/poem16188.html	الأمير منجك باشا	منجك بن محمد بن منجك بن أبي بكر بن عبد القادر...	https://www.aldiwan.net/cat-poet-alamir-mnjk-pasha	العصر العثماني	null	null	null
[illegible]	بحر البسيط	[illegible]	[illegible]	https://www.aldiwan.net/poem16189.html	الأمير منجك باشا	منجك بن محمد بن منجك بن أبي بكر بن عبد القادر...	https://www.aldiwan.net/cat-poet-alamir-mnjk-pasha	العصر العثماني	null	null	null
[illegible]	بحر الكامل	[illegible]	قصيدة مدح	https://www.aldiwan.net/poem16191.html	الأمير منجك باشا	منجك بن محمد بن منجك بن أبي بكر بن عبد القادر...	https://www.aldiwan.net/cat-poet-alamir-mnjk-pasha	العصر العثماني	null	null	null
[illegible]	بحر الكامل	[illegible]	[illegible]	https://www.aldiwan.net/poem16192.html	الأمير منجك باشا	منجك بن محمد بن منجك بن أبي بكر بن عبد القادر...	https://www.aldiwan.net/cat-poet-alamir-mnjk-pasha	العصر العثماني	null	null	null
[illegible]	بحر الخفيف	[illegible]	قصيدة قصيرة	https://www.aldiwan.net/poem16193.html	الأمير منجك باشا	منجك بن محمد بن منجك بن أبي بكر بن عبد القادر...	https://www.aldiwan.net/cat-poet-alamir-mnjk-pasha	العصر العثماني	null	null	null
[illegible]	بحر البسيط	[illegible]	قصيدة قصيرة	https://www.aldiwan.net/poem16194.html	الأمير منجك باشا	منجك بن محمد بن منجك بن أبي بكر بن عبد القادر...	https://www.aldiwan.net/cat-poet-alamir-mnjk-pasha	العصر العثماني	null	null	null

Dataset Preprocessing

Dataset Overview

- Loaded Arabic poems from two parquet files using the datasets library.
- Includes prompts and poetic completions.

Meter Selection

- After filtering the data based on the selected meters.
- The chosen meters are: الخفيف, الطويل, الكامل, and البسيط.` `

Cultural Accuracy

- Ensure qafiah respects Arabic poetic conventions, such as monorhyme in qasida.
- Supports the following meters

Text Normalization

- Retained diacritics for rhythm and sound.
- Standardized Arabic letters for consistency (e.g., 'ا' → 'أ').
- Removed non-Arabic symbols but kept poetic punctuation and line breaks.

INPUT
Load Parquet Files
TEXT PREPARATION
Normalize Arabic Text
Clean Data
FEATURE EXTRACTION
Extract Qafiah
Tag Poetry Meter
DATA SELECTION
Filter Poems
OUTPUT PREPARATION
Format for Model Input
Save Clean Dataset

Experiment

Receptance Weighted Key Value (RWKV)

Challenges

- Absence of Pre-trained Arabic RWKV Models.
- Computational Constraints for Training from Scratch.

Outcome

Due to these practical limitations, direct experimentation with RWKV was not feasible for this study. However, it remains a theoretically interesting avenue for future research with greater resources.

Retrieval-Augmented Generation (RAG)

Experimental Application

RAG was applied prior to fine-tuning with a selection of base models:

1. Qwen
2. LLaMA
3. Silma
4. ALLaM

Outcome

- The application of RAG in this pre-fine-tuning stage did not yield noticeable improvements in the quality or structural coherence of the generated Arabic poetry for the tested models.
- The added complexity of the retrieval mechanism did not translate to a discernible enhancement in poetic output in this specific experimental context.

Fine-Tuning LLMs (Qwen & LLaMA)

Methodology

LoRA (Low-Rank Adaptation) was employed for efficient fine-tuning on the Ashaar dataset.

Models

- Qwen (variants < 10B parameters)
- LLaMA (variants < 10B parameters)

Outcome

- Despite initial fine-tuning efforts, both Qwen and LLaMA variants showed limited improvement and slow convergence for the task of generating structurally sound Arabic poetry.
- Performance did not reach a satisfactory level to warrant further extensive training within the project's resource and time constraints.

Fine-Tuning LLMs: (Silma)

Training Details

- Conducted for two epochs.
- Utilized a Google Colab Pro A100 GPU instance.
- Total training time: Approximately 38 hours.

Model

Silma (approx. 9 Billion parameters)

- Promising candidate due to its architecture and reported capabilities.

Outcome

- Silma demonstrated a positive response to fine-tuning, showing noticeable improvements in generating text that attempted to follow Arabic poetic conventions.
- Generated outputs (e.g., as evaluated by initial qualitative checks and later by expert human evaluation using tools like Gemini for quick assessment) indicated better structural adherence compared to the baseline pre-trained model.

Fine-Tuning LLMs: (Silma)

مَا بَيْنَنَا سِرٌّ لا يُقَالُ مَا. – هو لاَّ سِرٌّ مِنْ حُبِّنَا مُقْسَمَا

فَحَياتِكَ يَا ذَا الحُسَابِ وَدَمْعَتِي – ولِقَايَ بَعْدَ جَفَائِنَا مَا أَقْسَمَا

وَلَئِنْ شَكَوْتُ فَمَا شَكَوتُ بِلاَئِي – ولكنَّني أَدْعُو العَلِيَّ لِيَدْعَمَا

لا تَطْلُبُوا مِنْ بَعْدِ هَذَا وِدَادِي – فَالشَّيْبُ قَدْ شَيَّبَ الوِشَاحَ المُقَمَّما

وَشَهِدَ لِي مِنْ بَعْدِ ذَاكَ بِرُشْدِي – مَالِي مِنَ الأَيَّامِ حِينَ تَبَدَّمَا

وَعَدَتْ بِلَيْلٍ قَدْ تَجَلَّى ضَوْءُهُ – وَلَهِىً بِحَبِيبٍ كَانَ مِنْ قَبْلُ أَعْصَمَا

فَالشَّمْسُ بَدَّلَتْهُ بِبَدْرِ الدُّجَى – وَالضَّبَابُ قَدْ شَطَّحَتْهُ فَتَجَلَّمَ

Fine-Tuning LLMs: (ALLaM)

Training Details

- Conducted for two epochs.
- Utilized a Google Colab Pro A100 GPU instance.
- Total training time: Approximately 27 hours.

Model

ALLaM(approx. 7 Billion parameters)

- An Arabic-centric model, making it a strong candidate for fine-tuning on Arabic poetry.

Outcome

- ALLaM also responded well to LoRA fine-tuning.
- The model learned to generate poetry that showed characteristics of the target meters and rhyme schemes present in the training data.

Fine-Tuning LLMs: (ALLaM)

أمي شعاع الهدى في ظلمتي أبداً
فحيثما سرت في ليلي وجدت هدى

وفي كل حين كلما نظرت إلى
محياكِ يا أماه أبصرت به الهدى

Baseline Models for Comparison

To provide a benchmark for the performance of our fine-tuned models (Silma and ALLaM), established proprietary LLMs were utilized.

Chosen Baseline Models:

- ChatGPT (OpenAI): Selected for its widely recognized advanced natural language understanding and generation capabilities.
- Claude (Anthropic): Chosen for its strong performance in coherent text generation and reasoning.

Purpose:

- These models, while not specifically fine-tuned on our Ashaar dataset, represent the state-of-the-art in general text generation.
- Comparing our fine-tuned models against them helps contextualize their performance in terms of both fluency and adherence to specific poetic structures.
- These models were primarily used for the final human evaluation phase.

Evaluation

Evaluation Prompts:

Standardized prompts were used across models, specifying theme (e.g., "شكر لجامعة الملك سعود"), Meter (الطويل), and Rhyme letter (e.g., "د").

Evaluation Approach (Quantitative)

Type-Token Ratio (TTR): Measures lexical diversity.

Evaluation Approach (Domain Experts)

Poems were anonymized for impartiality.

Criteria (1-5 scale):

1. Poetic Meaning (المعنى): Coherence, expressiveness, depth.
2. Metrical Accuracy (الوزن): Adherence to the specified Bahr.
3. Rhyme Consistency (القافية): Correctness and consistency of end-rhyme.

Results

ALLaM

شكراً لجامعة سعود درب المجد   طريق التقدم يا سعود لك استمد
من العلم قد أحييت أرضك بالندى   ومن العلم قد أحييت أرضك بالجدد

Silma

جامعة سعود في نجد، فخرها يزداد   في الشرق والغرب، تظل رمز الجداد
ما من مكان في العالم إلا وهي حماة   كالنجم في السماء، يضيء بالنجاد
فصاحتها تعانق الأفق، في كل بلاد   ومدائحها تتردد، بكل لسان وفؤاد
منها العلم ينبع، ويمتد في كل مداد   في نجد، جامعة سعود، فخر الأمجاد

Results

ChatGPT

سعيت للعلا والمجد يا خير مقصد   وسرت بعلم الناس تزْهين بالرشد
رفعت منار الحق فوق ذرا العلا   وغرست في الأرواح نورًا من السعد
إذا ذكر التعليم كنت أولًا   وصوتك بين الخلق أبهى من الغد
لك الحمد يا دار العلوم ومهدها   تفيضين علمًا في السهول وفي النجد

Claude

أيا صرح علمٍ قد سمت بفضلكم   وأشرقت نورًا في ربوع البوادي
تنير عقولًا بالمعارف والنُهى   وتبني رجالًا للغد المتجدد
بك العلم يزهو والثقافة تزدهي   وتسمو بك الآمال نحو السؤدد
فكم أنجبت أجيال علم مضيئة   وكم أطلقت أفكار فكر مسدد
ستبقى منارًا شامخًا في بلادنا   يُشِع بمجد راسخ متجدد
فدُمت لأوطان العروبة مفخرًا   ودام عطاؤك الزاكي على مر الأبد

Results: Lexical Diversity - Type-Token Ratio (TTR)

Type-Token Ratio (TTR): Measures lexical diversity.

TTR = (Number of Unique Words (Types) / Total Number of Words (Tokens)) × 100

Interpretation:

- Claude and ChatGPT (general-purpose, larger models) exhibited higher TTR.
- Fine-tuned models (ALLaM, Silma) showed comparatively lower TTR, potentially due to a stronger focus on adhering to strict structural constraints (meter, rhyme) which can limit lexical choices.

Source	Average TTR (%)
Silma	81.63
ALLaM	78.26
ChatGPT	93.48
Claude	94.83

Results: Expert Human Evaluation Scores

Key Insights:

- Claude excelled in Poetic Meaning. ChatGPT strong in Rhyme Consistency.
- ALLaM and Silma showed efforts in structure; however, metrical accuracy was a notable challenge, especially for Silma.
- Mastering all aspects of Arabic prosody remains difficult for current models.

Criterion	Silma	ALLaM	ChatGPT	Claude
Poetic Meaning (المعنى)	2.0	2.5	3.25	4.0
Metrical Accuracy (الوزن)	1.0	2.0	2.75	2.75
Rhyme Consistency (القافية)	2.25	2.5	3.25	2.0

Conclusion

This research investigated AI techniques (RWKV, RAG, Fine-Tuning) for traditional Arabic poetry generation using the Ashaar dataset.

- RWKV: While theoretically promising, practical implementation was hindered by the lack of pre-trained Arabic models and computational limitations for training from scratch.
- RAG: Did not demonstrate significant improvements for this specific task and dataset.
- Fine-Tuning (LoRA):
  - Initial explorations with LLaMA 3 and Qwen showed limited suitability within resource constraints.
  - ALLaM and Silma, when fine-tuned using LoRA, exhibited encouraging capabilities in generating poetry that attempted adherence to specified meter and rhyme.
- Proprietary models (ChatGPT, Claude) served as valuable benchmarks, often excelling in fluency but facing challenges with strict poetic forms.
- The generation of consistently high-quality, structurally sound, and semantically rich Arabic poetry remains an ongoing research endeavor.

Future Work

- Full Fine-Tuning: If substantial computational resources become available, conducting full fine-tuning (as opposed to LoRA) on models like ALLaM or Silma is highly recommended and expected to yield superior performance.
- Explore RWKV Further: Revisit RWKV if Arabic pre-trained versions emerge or if resources permit training from scratch.
- Advanced Diacritization: Integrate robust Arabic diacritization models to enhance phonetic accuracy and rhythmic adherence.
- Hybrid Systems: Combine LLMs with symbolic rule-based systems or dedicated meter-checking algorithms.
- Reinforcement Learning from Human Feedback (RLHF): Utilize expert evaluations to refine models towards producing more artistically and culturally resonant poetry.
- Dataset Expansion & Diversification: Incorporate a wider range of poetic styles, historical periods, and Arabic dialects.


Thank You

King Saud University
College of Computer and Information Sciences
Department of Computer Science

Arabic Poetry Generation using Large Language Models

CSC 497

Evaluation

Evaluation Prompts:

Standardized prompts were used across models specifying theme (e.g., "شكر لجامعة الملك سعود")
Meter (الطويل)
Rhyme letter (e.g., "د").

Results

ALLaM

شكراً لجامعة سعود درب المجد   طريق التقدم يا سعود لك استمد
من العلم قد أحييت أرضك بالندى   ومن العلم قد أحييت أرضك بالجدد

Silma

جامعة سعود في نجد، فخرها يزداد   في الشرق والغرب، تظل رمز الجداد
ما من مكان في العالم إلا وهي حماة   كالنجم في السماء، يضيء بالنجاد
فصاحتها تعانق الأفق، في كل بلاد   ومدائحها تتردد، بكل لسان وفؤاد
منها العلم ينبع، ويمتد في كل مداد   في نجد، جامعة سعود، فخر الأمجاد