



Teaching Arabic Grammar to Children Through Storytelling with the ALLAM Model (Tamakan)

Group members

Ahmed Al-Jaid
Norah Al-Muhaysini
Rawan Al-Ghamdi
Zainab Al-shingiti

09/11/2024

Table of Contents

Introduction	3
Problem	3
Solution	4
Related work	4
Methodology	5
The Tools	7
The Results	7
Evaluation Methods	7
1. Automated Metrics (BERTScore)	7
2. Human Evaluation:	8
Conclusion	10
Reference	10

Introduction

In today's educational landscape, mastering foundational language skills is essential for both academic achievement and personal growth. Yet, many students find themselves struggling with one of the most critical aspects of language: Arabic grammar (إعراب). Adding to this challenge, many teachers face difficulties in presenting Arabic grammar in a simple and engaging manner. Traditional teaching methods often make grammar seem daunting and tedious, leaving students disengaged and uninterested. This gap between instruction and comprehension highlights the urgent need for innovative approaches that can make learning Arabic grammar both enjoyable and effective.

Problem

According to a recent study, about 65% of Arab students suffer from difficulties in Arabic grammar and its practical application, especially in the primary and preparatory stages. This is due to the fact that the traditional teaching methods currently used in learning the Arabic language, such as memorization and routine exercises, are not interactive and boring for children, which reduces the interest of many children in the Arabic language. Studies have shown that more than 70% of students are interested in and interact with interactive teaching methods that include games or story activities, indicating that traditional methods may be ineffective in developing language skills. This is a large gap in education that causes a lack of linguistic proficiency and difficulty in using the Arabic language in a way that qualifies them to engage with society and the requirements of the labor market. [6][7]

Despite the importance of grammar in mastering any language, Arabic grammar, in particular, has become a stumbling block for many students. Several factors contribute to this issue:

- **Traditional Teaching Methods:** Conventional approaches to grammar education are often rigid and repetitive, failing to capture students' interest. Lessons are typically focused on rote memorization of rules rather than practical application in real-life contexts.
- **Lack of Engagement:** Arabic grammar, or إعراب, is perceived as challenging and abstract by many students. This lack of engagement results in students losing motivation and interest in learning, further widening the gap in their language skills.
- **Limited Personalization:** Every student learns at a different pace, and current educational systems rarely cater to individual needs. This one-size-fits-all approach leaves behind students

who may need extra support or different learning techniques, particularly in a subject as nuanced as grammar.

- **Long-Term Impact:** The consequences of these issues are far-reaching. As students progress through school, their weak grammar skills lead to difficulties in reading comprehension, writing, and communication. This, in turn, affects their academic performance and future career prospects.

Solution

To address the challenges of teaching Arabic grammar (إعراب) in an engaging and effective way, Tamakan introduces a comprehensive, interactive learning approach that adapts to each student's needs. The solution is designed to make grammar learning fun, personalized, and continuously adaptive, ensuring students improve their skills as they progress.

The Tamakan project offers an innovative solution that enhances the process of learning grammar in a fun and interactive way through an interactive environment that enables children to learn grammar rules through various stories that suit their interests and linguistic level, and enables them to create their own stories, which develops their imagination and encourages them to express themselves. After reading each part of the story, the child must answer a question related to grammar to move to the next part of the story, which stimulates thinking, enhances a deep understanding of grammar rules, and increases the child's enthusiasm and continuity in learning. Tamakan also provides an exciting competition between students, as they are ranked according to their levels, which increases their motivation to learn and contributes to their mastery of the Arabic language in a distinctive and fun way.

Our project includes these points that enhance and solve the gaps in children's education:

- Providing a fun and interactive platform for students.
- Providing a variety of stories according to the child's interest
- Providing an easy-to-use platform for teachers and parents.
- Offer rewards to the most advanced students.

Related work

Research on Arabic grammar learning shows that combining engaging activities with strong educational content is key to making learning both effective and enjoyable for students. In this section, we highlight related work that examines story generation and grammar learning through interactive, large language models (LLMs) to enhance engagement and understanding.

As demonstrated by researchers in this study [1] explores the potential of Large Language Models (LLMs) for Grammatical Error Correction (GEC) in Arabic, a language with complex grammar and rich morphology that presents unique challenges due to its low-resource status. The research evaluates various models, including Arabic-specific transformers like AraBart and AraT5v2, as well as multilingual models such as mT0 and mT5, all fine-tuned with specific learning rates and batch sizes. Additionally, ChatGPT is used to examine the effectiveness of prompting strategies like few-shot chain-of-thought (CoT) and expert prompting for generating synthetic data.

This paper [2] presents the first study on automatic Arabic story generation using large language models (LLMs), specifically leveraging the powerful Arabic LLM, AraLLaMA. The authors focus on two Arabic dialects—Egyptian and Moroccan—and fine-tune the models on datasets in these dialects. They compare their approach to both comparable and larger models, demonstrating through extensive automated and human evaluations that their models outperform strong baseline models.

In addition to this paper [3], the authors explore methods for automatically generating and simplifying children’s stories, focusing on creating content that is both engaging and accessible by adjusting vocabulary and sentence structures to age-appropriate levels. Utilizing mBERT, they apply two main strategies: the first involves introducing new entities into existing story structures to preserve coherence, albeit with some limitations on creativity; the second uses remodeled story spaces as prompts for large language models (LLMs) to generate complete stories, enhancing creativity but sometimes affecting structural integrity, as noted by human evaluators. Additionally, the Arabic ComVE dataset was used to train a commonsense evaluation model, yielding strong results that highlight the effectiveness of the approach in producing high-quality narratives that balance engagement, coherence, and developmental appropriateness for readers.

The Clear Arabic Grammar [4] is an educational app designed to teach Arabic grammar. It offers a structured approach to grammar through three sections, covering topics from basic to advanced levels, such as nominal sentences and transitive verbs. The app includes interactive exercises, performance evaluations, and progress tracking charts to enhance learning.

Looking at the prior work discussed above, we notice some limitations in the proposed approaches. Mainly, very few have specifically targeted Arabic grammar learning, and none have leveraged large language models (LLMs) to teach Arabic grammar through storytelling tailored for young learners. Our work aims to address this gap by developing an app that uses LLMs to create engaging, educational stories that make Arabic grammar learning accessible and enjoyable for children. Through this approach, we hope to combine the power of LLMs with interactive storytelling to create a unique and effective learning experience in Arabic grammar for young audiences.

Methodology

The methodology follows a structured process, utilizing a mix of assessment, interactive learning, feedback loops, prompt engineering, and AI-based enhancements. The key steps of our methodology include:

1. Initial Evaluation (Pre-Test)

The learning journey begins with a pre-test designed to assess the student's existing knowledge of Arabic grammar. This evaluation consists of various questions covering fundamental grammar rules, allowing us to determine the student's skill level—whether beginner, intermediate, or advanced. This initial assessment is critical in creating a personalized learning path, ensuring that the content provided aligns with the student's current abilities.

2. Personalized Content Delivery

Based on the pre-test results, students are presented with grammar-related stories that match their proficiency level. These stories are designed to integrate grammar questions naturally within the narrative, promoting active engagement. For beginners, the stories focus on simple grammatical rules, while advanced users are given more complex challenges. The system adapts to the student's progress, offering increasingly difficult questions as they advance through the lessons.

3. Interactive Learning with Feedback

As the student progresses through the stories, they are required to answer grammar-related questions to continue. These questions test their understanding of specific grammar rules and concepts. If a student answers a question incorrectly, they are given feedback in real time. This

feedback, powered by the **ALLAM** model, not only explains the correct answer but also offers further clarification, reinforcing the learning process.

4. Adaptive Learning Model with Prompt Engineering (ALLAM)

The ALLAM model plays a key role in enhancing the learning experience. It monitors the student's performance and adapts the difficulty of questions and exercises based on their needs. If a student consistently struggles with a particular grammar rule, ALLAM presents additional exercises, simpler explanations, or revisits earlier concepts to reinforce their understanding. Conversely, if the student excels, the system adjusts the difficulty level, introducing more advanced grammar challenges. Prompt engineering plays a crucial role in delivering personalized and relevant interactions between the student and the system. By carefully crafting the prompts, we ensure that ALLAM generates grammar questions and feedback tailored to each student's learning needs.

5. Learning Through Gamification and Progression

To maintain student interest and motivation, the platform incorporates elements of gamification. Students can unlock new levels, progress through increasingly challenging stories, and earn rewards as they answer grammar questions correctly. This game-like progression encourages consistent engagement and fosters a sense of accomplishment as the student advances through the learning material.

Throughout the learning process as shown in figure 1, students' progress is continuously monitored and evaluated. This allows us to track their growth in grammar proficiency and adjust the content or difficulty level as needed. The platform's adaptive learning system ensures that students receive personalized feedback and exercises at every stage, promoting long-term retention of grammar rules.



Figure 1 Tamakan 's Learning Process

Tamakan workflow consists of four phases as shown below in figure 2. The process begins with user signup and a pre-test to assess the child's level, which will inform the selection of appropriate stories and grammar content. The output from this initial phase feeds into the second phase, which involves Allam. Here, multiple prompts are generated sequentially, starting with story generation, followed by story splitting, and then question generation based on the story. The third phase entails generating an API using Flask, which will facilitate communication between the backend and the frontend. Finally, the workflow culminates in the development of the Tamkeen app using Flutter, providing an engaging platform for children to interact with stories and grammar exercises effectively.

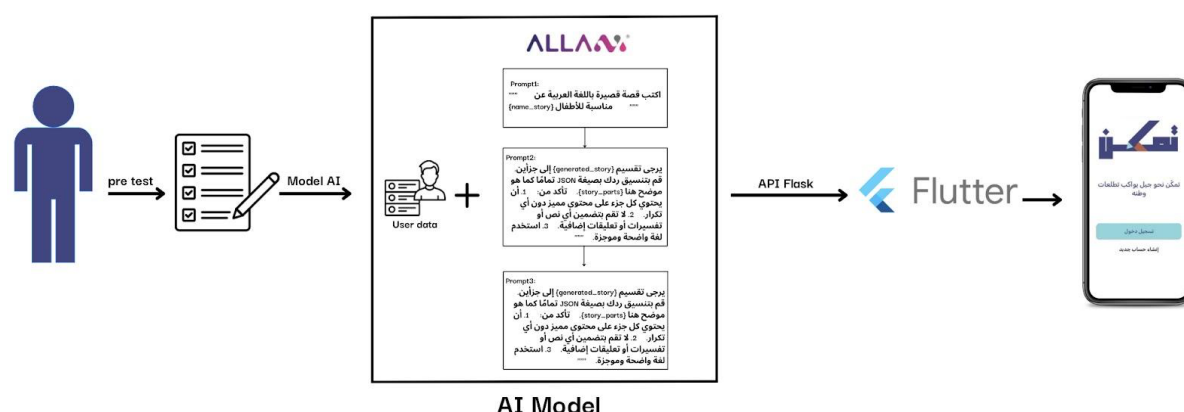


Figure 2 System Architecture

Tools

The implementation of **Tamakan** leverages advanced technology to create a seamless learning experience. The platform uses **Lang-chain**, a large language model, to generate grammar questions, provide feedback, and evaluate students' performance. **Python** serves as the primary programming language. Additionally, **Flutter** and **UIzard** are used to build an interactive, mobile-friendly model, ensuring that **Tamakan** is accessible on various devices.

Results

The evaluation of our methodology for **Tamakn** is based on a combination of automated metrics and human feedback to assess the effectiveness of the learning model in teaching Arabic grammar through interactive stories. This section outlines the criteria, tools, and results of our evaluation process.

Evaluation Methods

We used two primary methods for evaluating the model including Automated Metrics (BERTScore) and Human Evaluation:

1. Automated Metrics (BERTScore)

We employed BERTScore to measure the semantic similarity between the model's generated content and ideal reference responses. To enable a robust evaluation, we created a synthetic dataset using GPT-based generation techniques, designed to closely mirror the types of stories and grammar-related questions our model produces. This synthetic dataset served as a valuable benchmark, enabling consistent, large-scale comparisons across generated outputs. BERTScore provides three core metrics:

- **Precision:** Evaluates how much of the generated content is semantically aligned with the reference.
- **Recall:** Measures how well the generated responses capture the intended meaning in the reference text.
- **F1 Score:** Combines precision and recall to give an overall measure of the model's alignment with reference responses [5].

The results show in the below figure 3, that while the model is effective in covering the necessary content (high recall), there's an opportunity to improve precision across all tasks, especially in story generation and splitting. Enhancing precision could lead to a better balance and higher F1 scores, resulting in outputs that are both accurate and relevant across different content types.

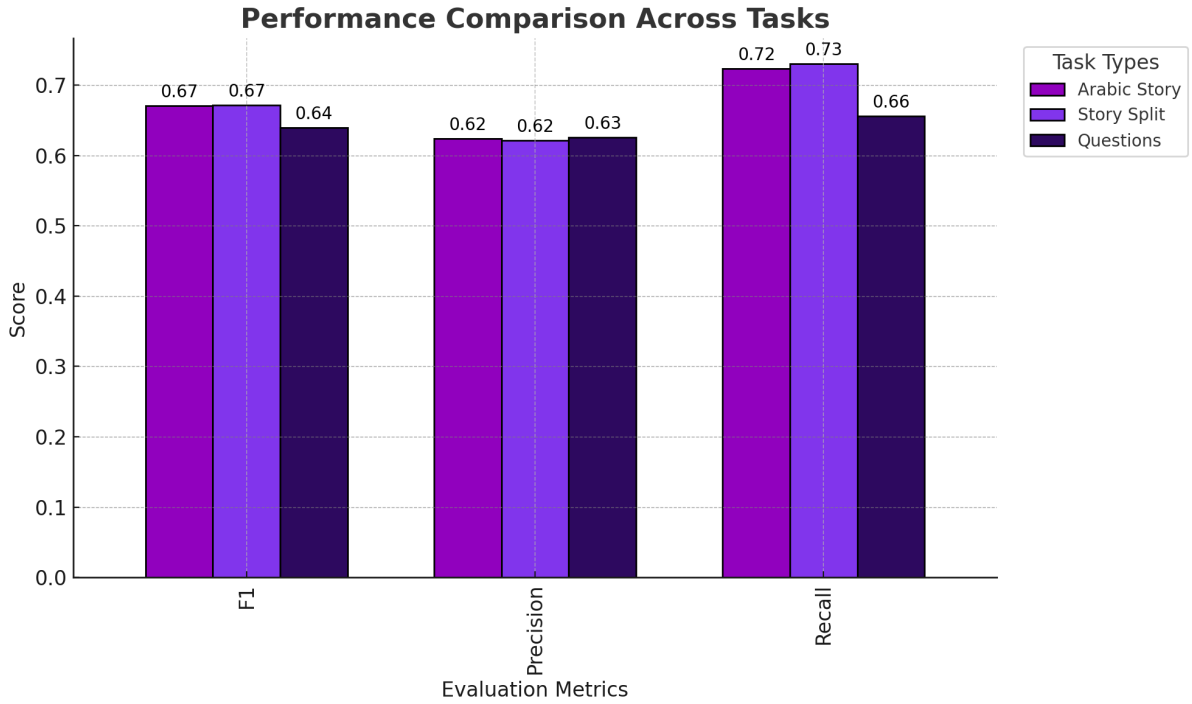


Figure 3 Performance Comparison For Tamakan' tasks

2. Human Evaluation:

Alongside automated metrics, To further validate the model's performance, we conducted a human evaluation on a random sample data from the generative content. This sample was reviewed by human evaluators using four key qualitative criteria essential for effective language learning content: Fluency, Coherence, Creativity, and Relevance [6]. The example below illustrates the evaluation process in action as shown in table 1:

Input Prompt: "Tell a story about friendship for children in Arabic."

Generated Story:

في يوم من الأيام، كان هناك صديقان يدعى أحمد وعلي. كانا يعيشان في قرية صغيرة جميلة. كان أحمد شخصًا ذكيًا وطموحًا يحب الدراسة والتعلم. أما علي فكان لطيفًا ومحبوبًا من الجميع. كانا يتميزان بأخلاق عالية وصفات رائعة. في يوم من الأيام، قرر أحمد وعلي أن يتوجها إلى المدينة المجاورة لزيارة معلم تاريخي. كانا متحمسين لاستكشاف المكان والتعرف على تاريخ المنطقة. خلال الرحلة، تعرفا على أصدقاء جدد وتعلما الكثير عن الثقافات المختلفة.

Criterion	Evaluation	Explanation
Fluency	4.5/5	The story is mostly grammatically correct and smooth, with no awkward phrasing. Minor improvements in wording could enhance readability further, but overall, the language flows naturally and is clear for young learners.
Coherence	4.7/5	The events follow a logical sequence, from Child's curiosity to his learning experience. Each part connects well, making it easy for the reader to understand the story's progression.
Creativity	4.2/5	The story is engaging and encourages curiosity, but it could include more details or scenarios to make child's journey more imaginative.
Relevance	4.8/5	The story aligns closely with the prompt, focusing directly on the child learning

Table 1 Human Evaluation criterias

In this evaluation, the size and quality of our dataset play a key role in understanding the model's performance. With only 100 synthetic records, the dataset gives us a limited snapshot and may not fully capture the variety found in real-world scenarios, which is important for educational content. Expanding to a larger dataset of 1000+ records would allow the model to learn from a wider range of story themes, language structures, and grammatical patterns, helping it to generalize better. However, creating such a large dataset would require considerable resources and time.

Given this limitation, combining BERTScore with human evaluation provides a comprehensive assessment of the model's performance. This approach helps us evaluate how well the model meets educational goals, particularly in terms of clarity, accuracy, and engagement—qualities that are often difficult to gauge through synthetic data alone. Human feedback, paired with BERTScore, offers a well-rounded view of the model's strengths and highlights areas for improvement in generating content that is both engaging and grammatically precise. This dual evaluation approach enables a reliable understanding of the Allam model's current capabilities and sets a foundation for more extensive evaluations with larger datasets in the future

Conclusion

Learning Arabic grammar is crucial for students' academic success, but traditional teaching methods often make it feel boring and hard to understand for children. Many students struggle because the lessons are not engaging, and a one-size-fits-all approach doesn't meet everyone's needs. **Tamakan** aims to change this by offering interactive solution that adapts to each student's learning pace. Using stories powered by large language models (LLMs)(ALLAM), **Tamakan** teaches grammar in a way that make it natural and enjoyable. This approach helps children build strong grammar skills over time, making learning both effective and enjoyable, and setting them up for greater success in language and communication.

Reference

- [1] S. Y. Kwon, G. Bhatia, E. M. B. Nagoudi, and M. Abdul-Mageed, "Beyond english: Evaluating llms for Arabic grammatical error correction," arXiv.org, <https://arxiv.org/abs/2312.08400> (accessed Nov. 5, 2024).
- [2] A. O. El-Shangiti, F. Alwajih, and M. Abdul-Mageed, "Arabic Automatic Story Generation with Large Language Models," Jul. 10, 2024, *arXiv*: arXiv:2407.07551. Accessed: Nov. 05, 2024. [Online]. Available: <http://arxiv.org/abs/2407.07551>
- [3] A. Alhussain and A. M. Azmi, "Beyond Event-Centric Narratives: Advancing Arabic Story Generation with Large Language Models and Beam Search," *Mathematics*, vol. 12, no. 10, p. 1548, May 2024, doi: 10.3390/math12101548.
- [4] Amer, I. (2018) *Clear arabic grammar*, App Store. Available at: <https://apps.apple.com/sa/app/clear-arabic-grammar/id1355833585> (Accessed: 08 November 2024).
- [5] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTScore: Evaluating Text Generation with BERT," Feb. 24, 2020, *arXiv*: arXiv:1904.09675. Accessed: Nov. 05, 2024. [Online]. Available: <http://arxiv.org/abs/1904.09675>
- [6] R. Awasthi *et al.*, "HumanELY: Human evaluation of LLM yield, using a novel web-based evaluation tool," Dec. 27, 2023. doi: 10.1101/2023.12.22.23300458.
- [7] Ministry of Education. (2022). Report on the difficulties of learning the Arabic language.
- [8] International Association for the Advancement of Education. (2021). The Impact of Interactive Methods on Education.