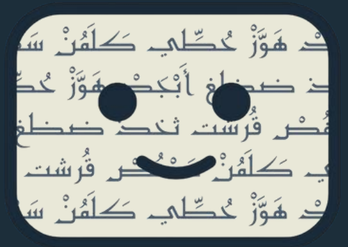

هَوَّزْ حُطِّي كَلَمُنْ سَ
ضظغ أَبْجَدْ هَوَّزْ حُطِّ
قُرِسْتْ ثَخَذْ ضَظَغْ
كَلَمُنْ سَعْفَصْ قُرِشْتْ
هَوَّزْ حُطِّي كَلَمُنْ سَ

Rannan

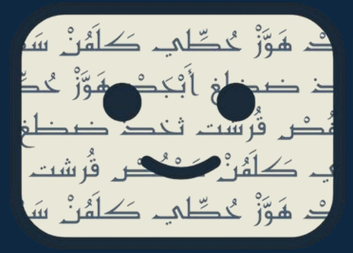
رنّان

SAMSUNG INNOVATION Campus

AI Course

SAMSUNG

The Intelligent Linguistic Advisor



Rannan

رنّان

Contents

01 Rannan Team Member

02 WHY IT MATTER?

03 OUR TARGET USERS

04 OUR BIG IDEA

05 HOW WE BUILT IT?

06 WHAT WE ACCOMPLISHED

07 WHAT NEXT?

08 TAKEAWAYS

Rannan

رنّان

TEAM MEMBER

Maher Alhijile

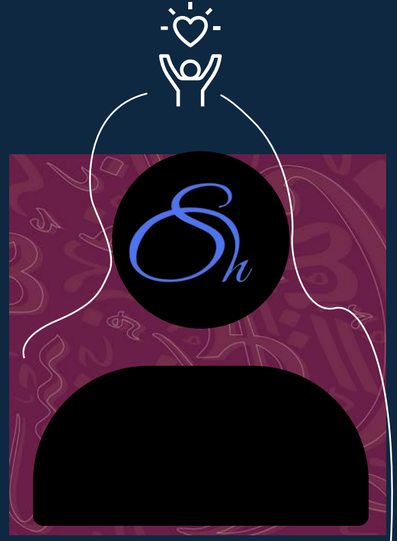
AI Engineer

Abdulrazaq Al-Dawsari

Project Manager , Full stack Dev, Data Engineer & Assistant AI Engineer

Bader Alshamrani

AI Engineer

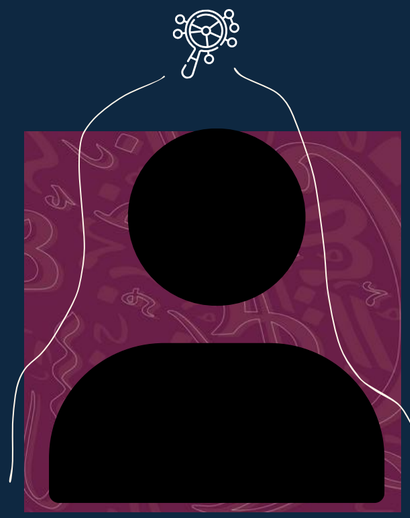


Shahad Alnashwan

Project Documentation & Assistant AI Engineer

Rahaf Faiz

Data Engineer & Assistant AI Engineer



Raniyah Alghamdi

Data Engineer & Assistant AI Engineer

The team works with a unified vision to shape a smarter future for the Arabic language, driven by passion, excellence, and diverse expertise.

WHY IT MATTER?

Arabic lacks a unified, trustworthy source for clear linguistic guidance, leaving millions relying on scattered and inaccurate information. This gap creates confusion, weakens writing quality, and increases mistakes across education, media, and digital communication. As a rich and complex language, Arabic becomes harder to navigate without accessible expert support—yet such support is rarely available. Our solution matters because it bridges this gap, restores clarity, empowers users with instant, reliable answers, and strengthens the presence of the Arabic language in today's digital world.

Our big idea

WHAT WE DO

- "**Rannan**" is an **AI-powered Arabic linguistic advisor** that provides accurate, real-time responses to questions related to grammar, morphology, and spelling.
- It uses **trusted linguistic data** from the **King Salman Global Academy for arabic language** and responds to user queries via a conversational chatbot interface.

WHY IT MATTERS

- The Arabic digital space lacks reliable, accessible linguistic tools.
- Our solution addresses the high error rates in formal writing across education, media, and digital content.
- It empowers users to write correctly with confidence, reducing dependency on outdated or unreliable sources.

HOW IT WORKS

- Rannan combines a Retrieval-Augmented Generation [RAG] model with **Qwen** for both retrieval and generation **[LLM and embeddings]**, using **ChromaDB** as the vector database.
- It retrieves the most relevant linguistic answers from a semantic search over 8,294 curated Q&A records, then generates clear, context-aware responses.
- Built using **FastAPI**, **vector DB**, and **React frontend** for seamless user interaction.

WHAT SETS US APART

- **Authoritative & Contextual**: All answers are grounded in verified data from KSGAAL.
- **Arabic-first AI RAG system**: Specifically tailored for the complexities of Modern Standard Arabic.
- **Scalable & Modular**: Can expand into broader Arabic domains [e.g., rhetoric, poetry, editing].
- Clean UI + instant feedback = frictionless user experience.

Timeline

20/ Nov — Project Initiation & Planning

- Defined the project vision and selected the final idea
- Drafted the action plan
- Assigned roles and responsibilities

29/ Nov — Data Preparation

- Acquired Arabic grammar datasets from KSGAAL
- Cleaned, normalized, and preprocessed text using CAMeL Tools
- Structured Q&A pairs for high-quality retrieval

30/ Nov — RAG Development

- Built the semantic retrieval pipeline
- Configured vector database and embeddings
- Integrated the Qwen language model with retrieval

03/Dec — UI Integration & Deployment

- Developed the backend [FastAPI] and connected it to the RAG pipeline
- Built the frontend using React and implemented the chatbot interface
- Conducted model evaluation, user testing, and applied iterative improvements
- Finalized documentation and prepared the presentation

Rannan
رنّان

HOW WE ACCOMPLISHED?

DATA

8888

Number of Rows Before

8294

Number of Rows After

HOW WE ACCOMPLISHED?

📍 BEFORE CLEANING:

Q: ما نوع الإعلال في الفعل (نأت) ؟ وما وزنه ؟

A: نأت: فيه إعلال بالحذف لمناسبة الجزم وأصله نأتي، وزنه نفع.

لغويات ونحو 4 ألف سؤال,

📍 AFTER CLEANING:

Q: ما نوع الإعلال في الفعل (نأت)؟ وما وزنه؟

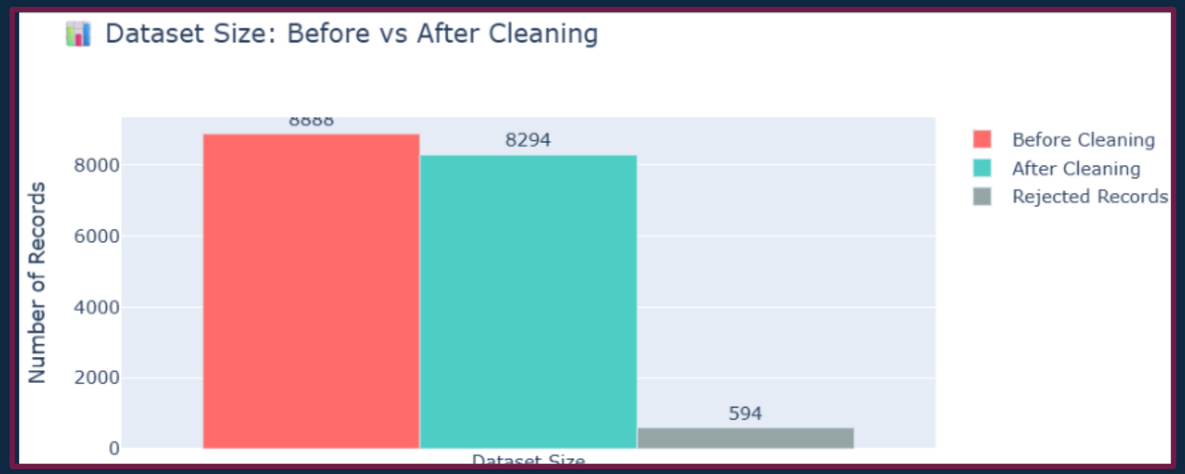
A: نأت:فيه إعلال بالحذف لمناسبة الجزم وأصله نأتي، وزنه نفع.

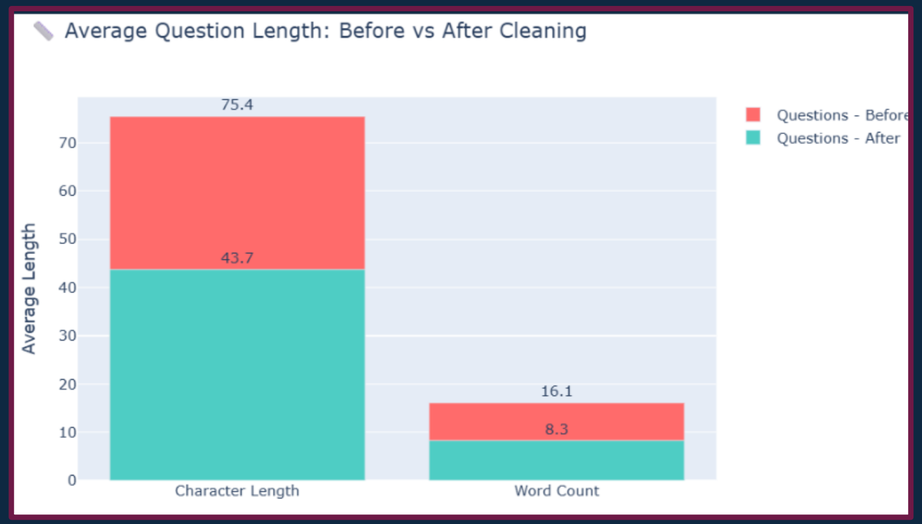
📊 Character Count Change:

Question: 44 → 40 (-4)

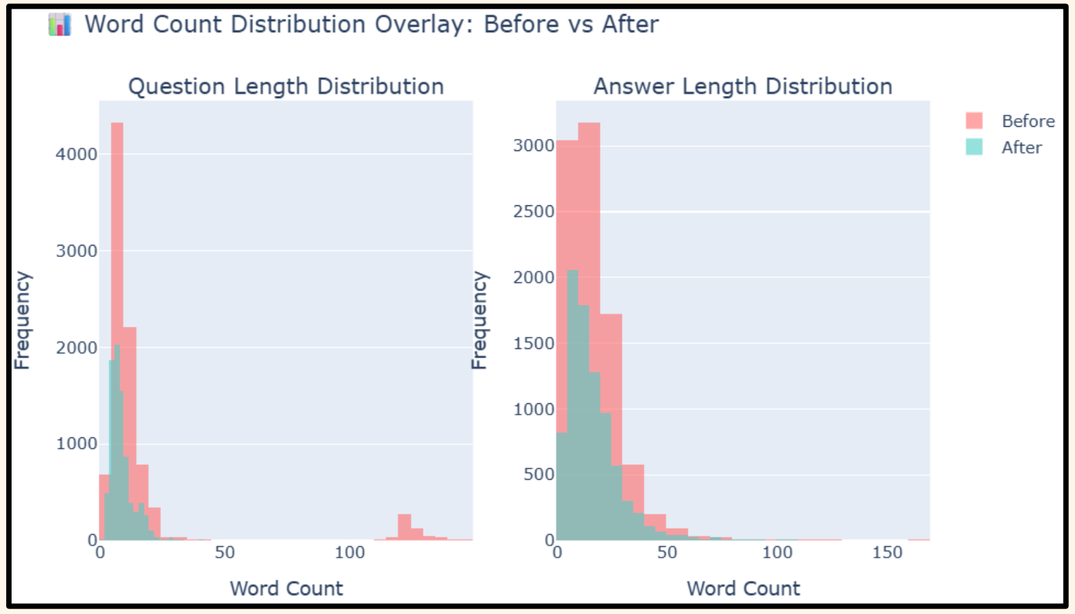
Answer: 84 → 56 (-28)

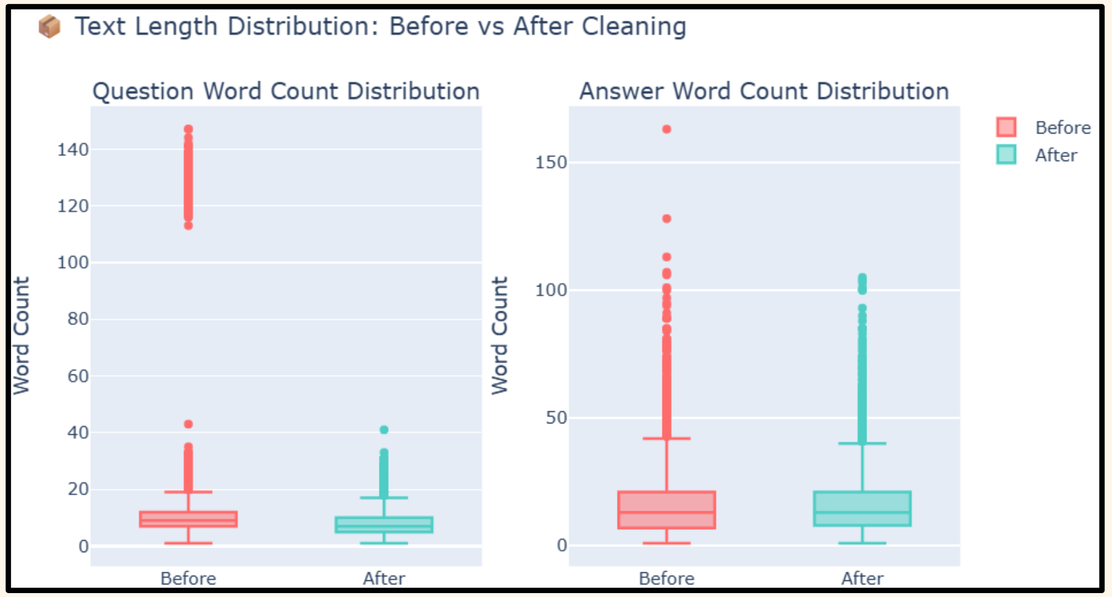
HOW WE ACCOMPLISHED?





HOW WE ACCOMPLISHED?





HOW WE ACCOMPLISHED?

```
================================================================================
📈 Data Quality Improvement Metrics
================================================================================
                  Metric Percentage          Status
     Data Retention Rate     93.32%    ✅ Excellent
Average Text Cleanliness     98.66%      ✅ Improved
  Record Validation Rate     93.32% ✅ High Quality
```

OUR RAG SYSTEM ARCHITECTURE PIPELINE

Dataset Preparation:
Our dataset is stored in clean **JSON** format, where each question and answer is merged into a single well .structured text document

Semantic Embedding Generation:
We generate embeddings using Qwen3 Embedding 8B, producing high-dimensional vectors up to 4096 dimensions for precise ..semantic similarity search

Vector Storage and Indexing:
All vectors and documents are stored in ChromaDB, enabling fast and accurate retrieval of the .most relevant linguistic entries

User Query Embedding and Retrieval:
When a user submits a question, the system generates its embedding, retrieves the closest matching entries, and builds a focused context.

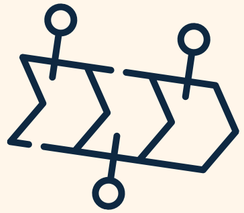
Context-Grounded Answer Generation:
Qwen 32B uses this retrieved context to produce accurate, grounded Arabic answers while avoiding hallucinations.

Reliable and Evidence-Based Response Delivery;
This RAG workflow ensures every response is reliable, authoritative, and fully supported by real linguistic data.

Testimonial

C-R	C-P	A-R	Faithfulness	Correctness	L-Overlap
99%	79%	98%	92%	92%	72%

Testimonial



```
============================================================
EVALUATION METRICS SUMMARY (Scale 0-1)
============================================================

Metric                          Mean            Std
------------------------------------------------------------
context_relevance               0.990           0.090
context_precision               0.787           0.181
answer_relevance                0.980           0.132
faithfulness                    0.925           0.220
correctness                     0.923           0.224
lexical_overlap                 0.723           0.327

============================================================
RETRIEVAL METRICS SUMMARY (Scale 0-1)
============================================================
```

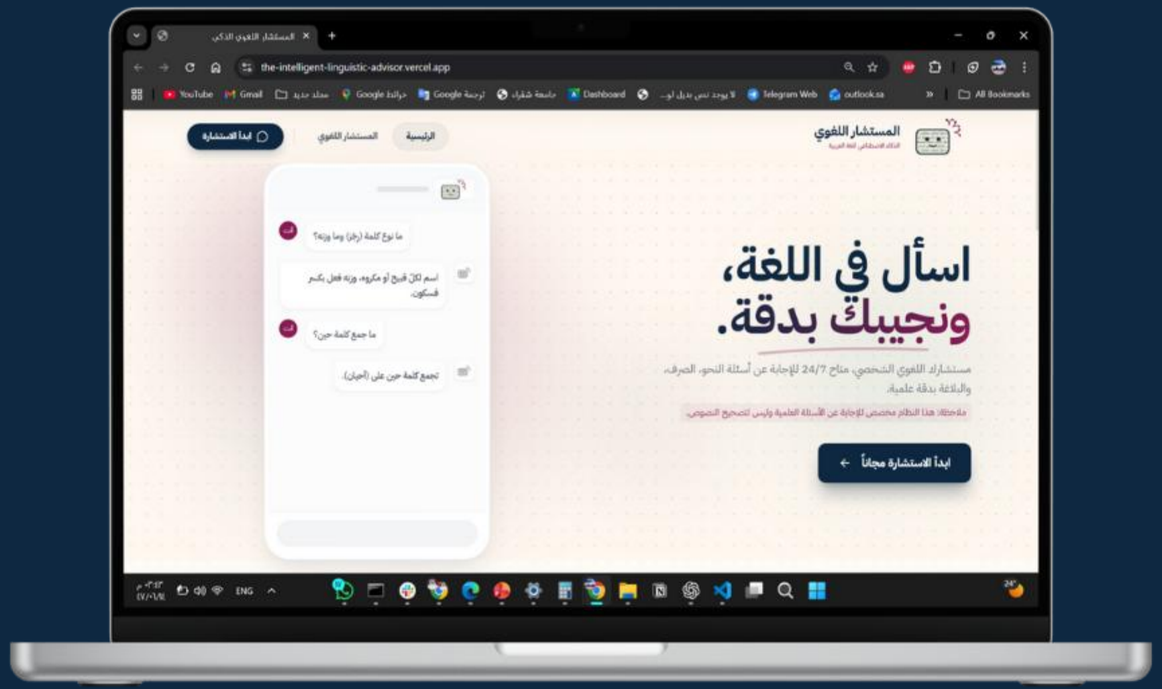
The Website

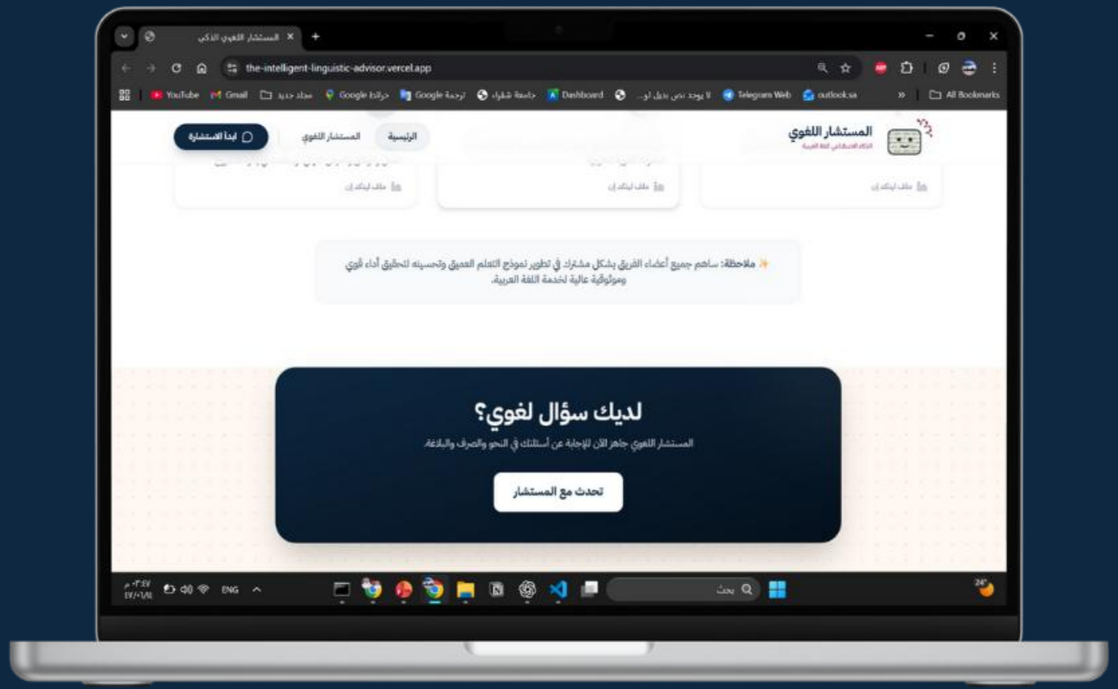
1

2

3

The Website





4

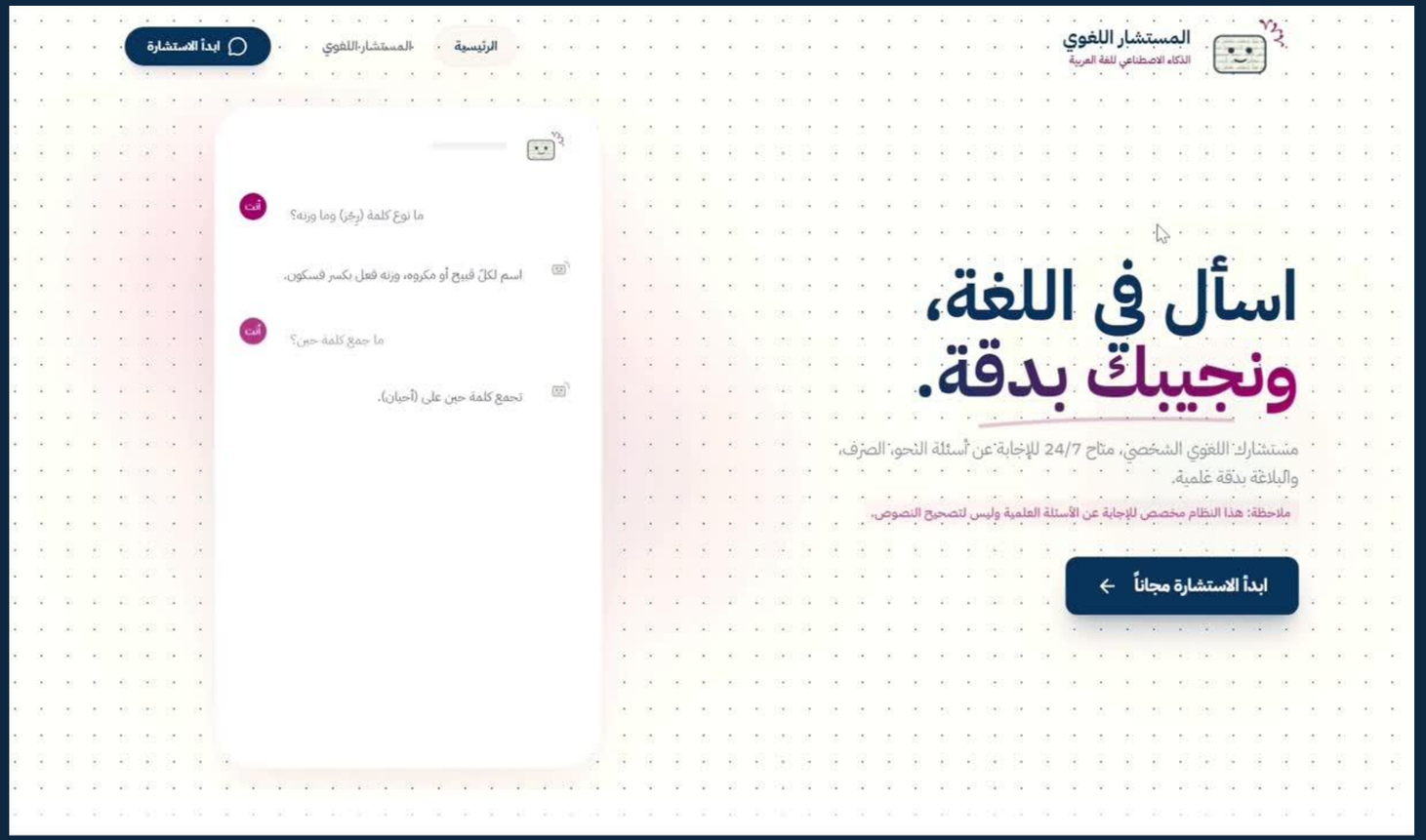
5

Test It Yourself

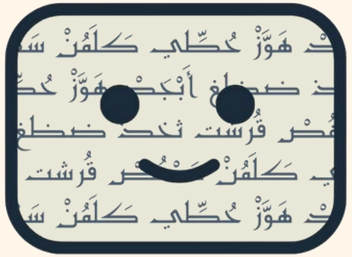
To Continue the improvement of the model with cooperation with KSGAAL, to improve the scope of the model, and to improve its responses to be more accurate.

What is Next?

DEMO

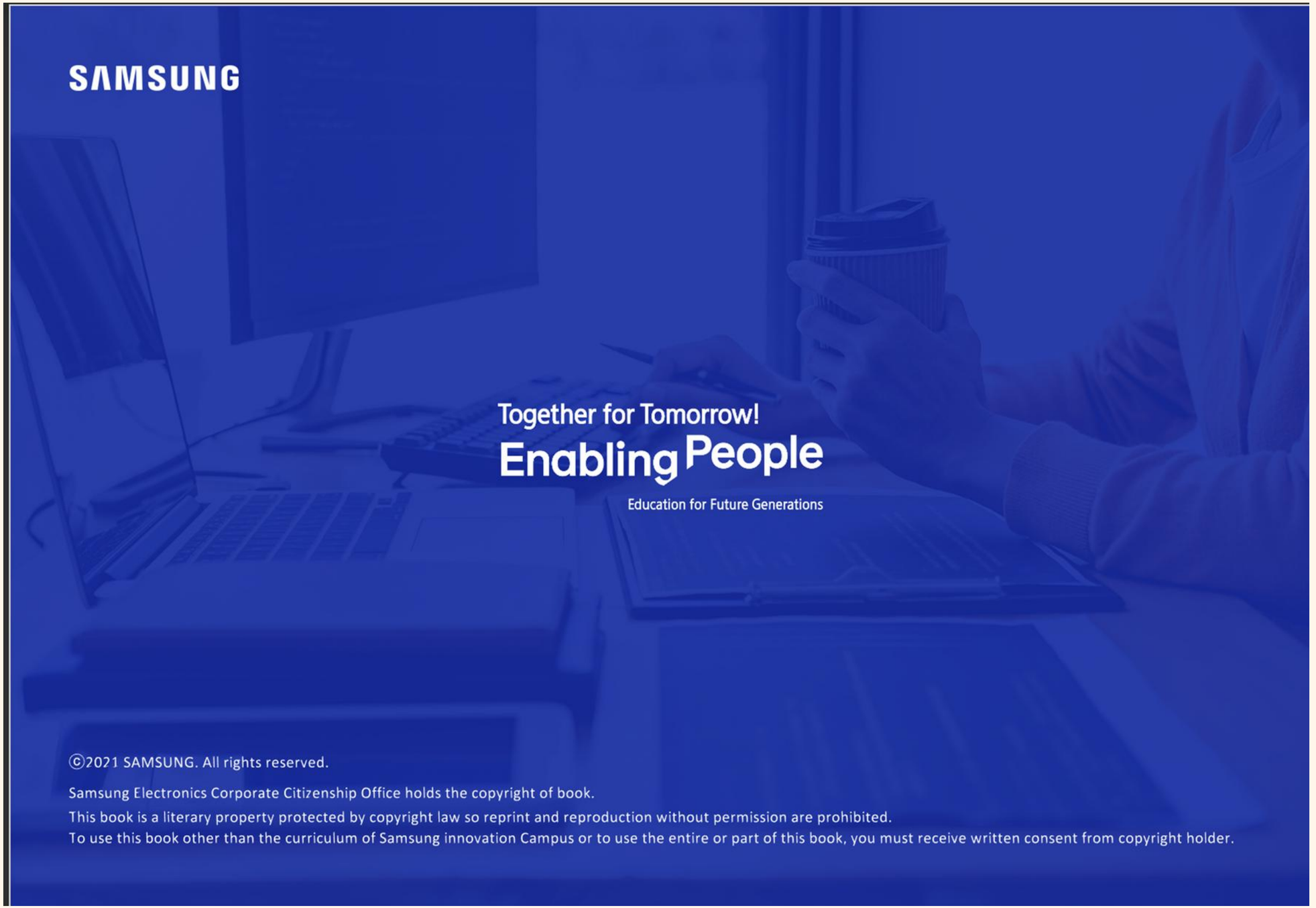

المستشار اللغوي
الذكاء الاصطناعي للغة العربية
الرئيسية
المستشار-اللغوي
ابدأ الاستشارة
ما نوع كلمة (رِجْز) وما وزنه؟
اسم لكلّ قبيح أو مكروه، وزنه فعل بكسر فسكون.
ما جمع كلمة حين؟
تجمع كلمة حين على (أحيان).
اسأل في اللغة،
ونجيبك بدقة.
مستشارك اللغوي الشخصي، متاح 24/7 للإجابة عن أسئلة النحو، الصرف، والبلاغة بدقة علمية.
ملاحظة: هذا النظام مخصص للإجابة عن الأسئلة العلمية وليس لتصحيح النصوص.
ابدأ الاستشارة مجاناً

Thank you!



SAMSUNG

SAMSUNG



Together for Tomorrow!

Enabling People

Education for Future Generations

©2021 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of book.

This book is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this book other than the curriculum of Samsung innovation Campus or to use the entire or part of this book, you must receive written consent from copyright holder.