

VERSO – AI-Powered Arabic Poetry Generation

Final Presentation

The Inspiration Behind VERSO

- **Arabic poetry, a cornerstone of cultural heritage, faces challenges of modern accessibility and understanding.**
- **Inspired by the lack of engaging resources, VERSO aims to simplify poetry learning and creation.**

Features

- **Interactive Creation:** Compose poetry line by line with real-time prosody feedback.
- **Meter Classification:** Ensures poems adhere to traditional rhythms
- **Enhanced Analysis:** Provides linguistic and artistic explanations.

System Implementation Highlights

- **Custom Tokenizer:** Accurately processes Arabic text, respecting diacritics and grammar for poetic analysis.
- **Prosody Analysis Engine:** Matches poetic meters to Al-Bihar with >90% accuracy, validating traditional rules.
- **AI Model:** Fine-tuned Silma-9B for Arabic language tasks using optimized methods like LoRA and quantization.
- **Frontend Interface:** Interactive features for seamless poetry creation and analysis.
- **Generate Poem Explanation** in 3 variations
- **Generate Artworks**

System Implementation Highlights

Load the 8-bit Model

```
1 from transformers import AutoTokenizer, AutoModelForCausalLM, BitsAndBytesConfig
2
3 # Define the quantization configuration for 8-bit
4 quantization_config = BitsAndBytesConfig(
5     load_in_8bit=True, # Enable 8-bit quantization
6     llm_int8_enable_fp32_cpu_offload=True # Offload to CPU if GPU memory is
7     insufficient
8 )
9
10 # Model ID from Hugging Face
11 model_id = "silma-ai/SILMA-9B-Instruct-v1.0"
12
13 # Load the model with 8-bit quantization
14 model = AutoModelForCausalLM.from_pretrained(
15     model_id,
16     quantization_config=quantization_config, # Apply 8-bit quantization
17     device_map="auto" # Automatically map model to available devices
18 )
19
20 # Load the tokenizer
21 tokenizer = AutoTokenizer.from_pretrained(model_id)
22
23 print("Model loaded in 8-bit mode for prompt engineering.")
```

```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as secret in your Google Cloud Platform console. You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
warnings.warn(
config.json: 100% ██████████ 895/895 [00:00<00:00, 63.5kB/s]
model.safetensors.index.json: 100% ██████████ 39.1k/39.1k [00:00<00:00, 2.99MB/s]
Downloading shards: 100% ██████████ 5/5 [07:19<00:00, 82.93s/it]
model-00001-of-00005.safetensors: 100% ██████████ 3.91G/3.91G [01:32<00:00, 41.9MB/s]
model-00002-of-00005.safetensors: 100% ██████████ 3.96G/3.96G [01:34<00:00, 42.5MB/s]
model-00003-of-00005.safetensors: 100% ██████████ 3.96G/3.96G [01:33<00:00, 42.9MB/s]
model-00004-of-00005.safetensors: 100% ██████████ 3.96G/3.96G [01:33<00:00, 42.7MB/s]
model-00005-of-00005.safetensors: 100% ██████████ 2.69G/2.69G [01:03<00:00, 41.5MB/s]
Loading checkpoint shards: 100% ██████████ 5/5 [00:10<00:00, 2.04s/it]
generation_config.json: 100% ██████████ 168/168 [00:00<00:00, 13.3kB/s]
tokenizer_config.json: 100% ██████████ 46.9k/46.9k [00:00<00:00, 3.73MB/s]
tokenizer.model: 100% ██████████ 4.24M/4.24M [00:00<00:00, 18.0MB/s]
tokenizer.json: 100% ██████████ 17.5M/17.5M [00:00<00:00, 43.1MB/s]
special_tokens_map.json: 100% ██████████ 636/636 [00:00<00:00, 51.7kB/s]
Model loaded in 8-bit mode for prompt engineering.
```

Image: System architecture or tokenization process diagram.

System Implementation Highlights

عن فيرسو

توثيق فيرسو

VERSIO

البيت 1

الشطر 1

الشطر 2

البيت 1

وزن القصيدة

اختر البحر:

الطويل

شرح القصيدة

شكيل

شكيل

الكتابة المشكلة: دُكَأَ يُضِيءُ الدُّرُوبَ لَنَا أَمَلًا
الكتابة العروضية: دُكَأُنْ يُضِيءُ دُرُوبًا لَنَا أَمَلُنْ
البحر: فعولن مفاعل فعولن مفاعيلن
نسبة التطابق: 95.45%

الكلمة	ترميز الكلمة	التفعيلة	ترميز التفعيلة	ترميز التطابق	نسبة التطابق
دُكَأُنْ	11010	فعولن	11010	11010	100.00%
يُضِيءُ	11010	مفاعل	11010	11010	100.00%
دُرُوبًا	11010	فعولن	11010	11010	100.00%
لَنَا أَمَلُنْ	1101110	مفاعيلن	1101010	1101#10	85.71%

الكتابة المشكلة: سِلْمَا أَتَى بِالْمَرْأَيَا إِلَى الْعِلْمِ
الكتابة العروضية: سِلْمَا أَتَى بِالْمَرْأَيَا إِلْ لِعِلْمِي
البحر: فعول مفاعلن فعولن مفاعيلن
نسبة التطابق: 77.27%

الكلمة	ترميز الكلمة	التفعيلة	ترميز التفعيلة	ترميز التطابق	نسبة التطابق
سِلْمَا	1010	فعول	1101	1###	25.00%
أَتَى	110100	مفاعلن	110110	1101#0	83.33%
مَرْأَيَا	11110	فعولن	11010	11#10	80.00%
إِلْ لِعِلْمِي	1101010	مفاعيلن	1101010	1101010	00.00%

Image: System architecture or tokenization process diagram.

User Journey

- Enter a poetry idea or theme.
- Customize settings: meter selection, verse count, and style.
- Generate the poem with analysis and optional explanations.
- Save, share, or visualize the poem.

Results

- Generated Arabic poetry aligned with traditional metrics.
- 90%+ accuracy in poetic meter classification.
- Example input: "اكتب شعر عن الذكاء الصناعي سلما بالبحر الطويل من ٦ ابيات"
- Example Output:

سلما أَتَى بِالْمَزَايَا إِلَى الْعِلْمِ
ذَكَاءٌ يُضِيءُ الدُّرُوبَ لَنَا أَمَلًا
يُنَافِسُ فِي سَاحَةِ الْعِلْمِ جَاهِدًا
وَيَبْنِي لَنَا فِي الْمَعَارِفِ مَنْزِلًا
يَرَى فِي طَرِيقِ النُّجُومِ طُمُوحَهُ
وَيَسْعَى إِلَى كُلِّ فَوْزٍ وَمُشْتَهَى

Results

عن فيرسو

توثيق فيرسو

VERS0

نظم القصيدة

اختر البحر:

الطويل

عدد الأبيات: 3

شرح القصيدة

الذكاء الصناعي سلما

↑

حذف

سلما أُنِّي بِالْمَرَايَا إِلَى الْعِلْمِ

حذف

وَيُنَافِسُ فِي سَادَةِ الْعِلْمِ جَاهِدًا

حذف

يَرَى فِي طَرِيقِ التَّجْوِمِ ظُفُوفَهُ

حذف

دُكَاءٌ يُضِيءُ الدُّرُوبَ لَنَا أَمَلًا

حذف

وَيُبَيِّنِي لَنَا فِي الْمَعَارِفِ مَنَزَلًا

حذف

وَيَسْعَى إِلَى كُلِّ مَوْزٍ وَمُسْتَهَى

Results

توثيق فيرسو

VERSO

عن فيرسو

عرض صوري

قصة

الشرح المفصل

الشرح العام

الشرح العام

القصيدة تحتفي بشخصية سلما كرمز للأمل والتقدم في مجال العلم. تُصور ذكاء سلما كضوء يضيء الدروب للآخرين، مانحًا إياهم الإلهام والطموح. تعكس الأبيات اجتهد سلما في السعي نحو التفوق العلمي، ورؤيته لبناء إرث دائم في عالم المعرفة. الرحلة التي يخوضها سلما تتسم بالإصرار والعزيمة، مع تطلع دائم نحو النجوم، التي ترمز إلى الطموحات العالية والأحلام اللامتناهية.

جميع الحقوق محفوظة لفيرسو

Challenges and Solutions

- **Dataset Quality:**
 - **Issues:** Inconsistencies in poetic styles and missing diacritics.
 - **Solution:** Developed Python scripts for cleaning and annotating over 112,000 entries.
- **Model Optimization:**
 - **Issues:** Resource constraints during fine-tuning.
 - **Solution:** Used QLoRA and 8-bit quantization for better performance.
- **Integration:**
 - **Issues:** Ensuring seamless communication between backend and AI models.
 - **Solution:** Modular design and extensive testing resolved compatibility issues.

Testing Outcomes

- **Tokenization:** 95%+ accuracy verified with manual annotations.
- **Poetry Generation:** Evaluated for linguistic coherence and metric alignment.
- **Performance Testing:** Ensured efficient model fine-tuning on limited resources.

Future Enhancements

- Expand datasets for richer training and diverse styles.
- Develop a GUI for wider accessibility and better UX.
- Integrate additional multimedia features like text-to-image transformation.

Business Vision

Target Market:

- Universities, schools, and educational institutions.
- Poets, writers, and literary enthusiasts.
- Students and researchers in Arabic literature.

Core Value:

- Make Arabic poetry creation and learning accessible to everyone.

Revenue Opportunities:

- Subscription-based services for advanced features.
- Licensing for educational institutions.
- Partnerships with cultural organizations.



Thank You