

## Soutenance du projet

# Moteur de Recherche Sémantique et Syntaxique sur le Saint Coran utilisant Elasticsearch et le RAG

Soutenu par:

Fatima Zahra BELHAMRA

Oumaima DGUIGOU

Zaïneb EL HASSANI

Safaa LHAKKAOUI

Ilham TIRACH

Jury :

Pr. Noureddine KERZAZI

# PLAN

---

- 01 **Introduction**
- 02 **Partie 1 : Recherche Syntaxique avec Elasticsearch**
- 03 **Partie 2 : Recherche Sémantique avec le Système RAG**
- 04 **Conclusion**



## Introduction:

Le **Saint Coran** est une source de sagesse et d'enseignement, et un accès efficace à son contenu est essentiel pour les chercheurs, les étudiants et les croyants.



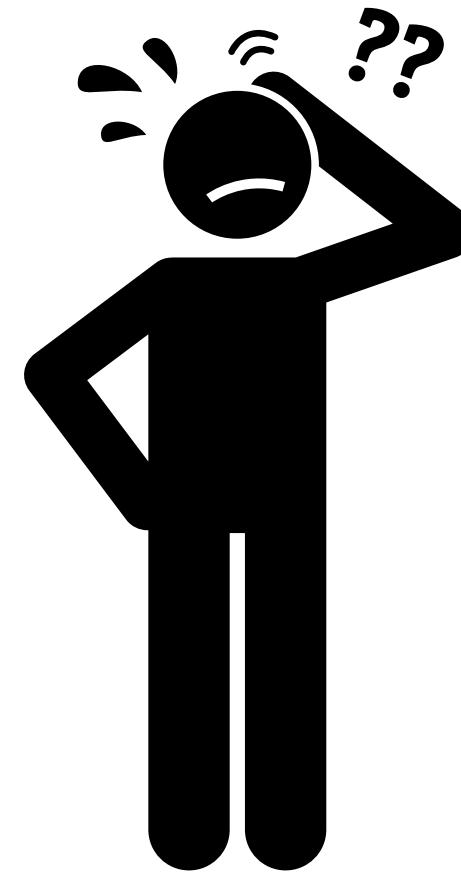
Cependant, rechercher des informations précises dans un texte aussi vaste et complexe pose des défis importants, notamment en termes de syntaxe et de sémantique.





## Problématique:

Comment concevoir un **moteur de recherche** capable de répondre de manière précise et pertinente aux requêtes sur le Saint Coran, en surmontant les limites des approches traditionnelles face aux défis de la **syntaxe** et de la **sémantique** ?

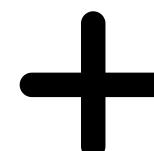




## Solution Proposée:

Nous proposons une solution hybride combinant deux approches complémentaires :

- Elasticsearch, pour une recherche syntaxique performante
- RAG (Retrieval-Augmented Generation), pour une recherche sémantique avancée et contextuelle.



**RAG**

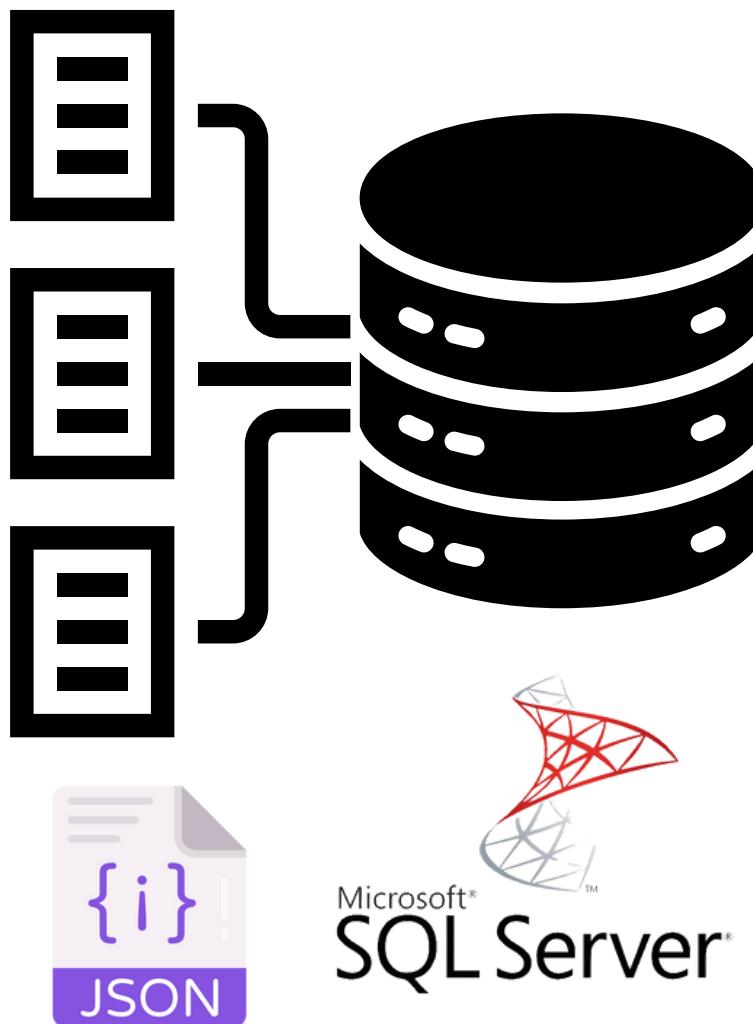


**Une Solution Hybride  
pour la Recherche  
Avancée**

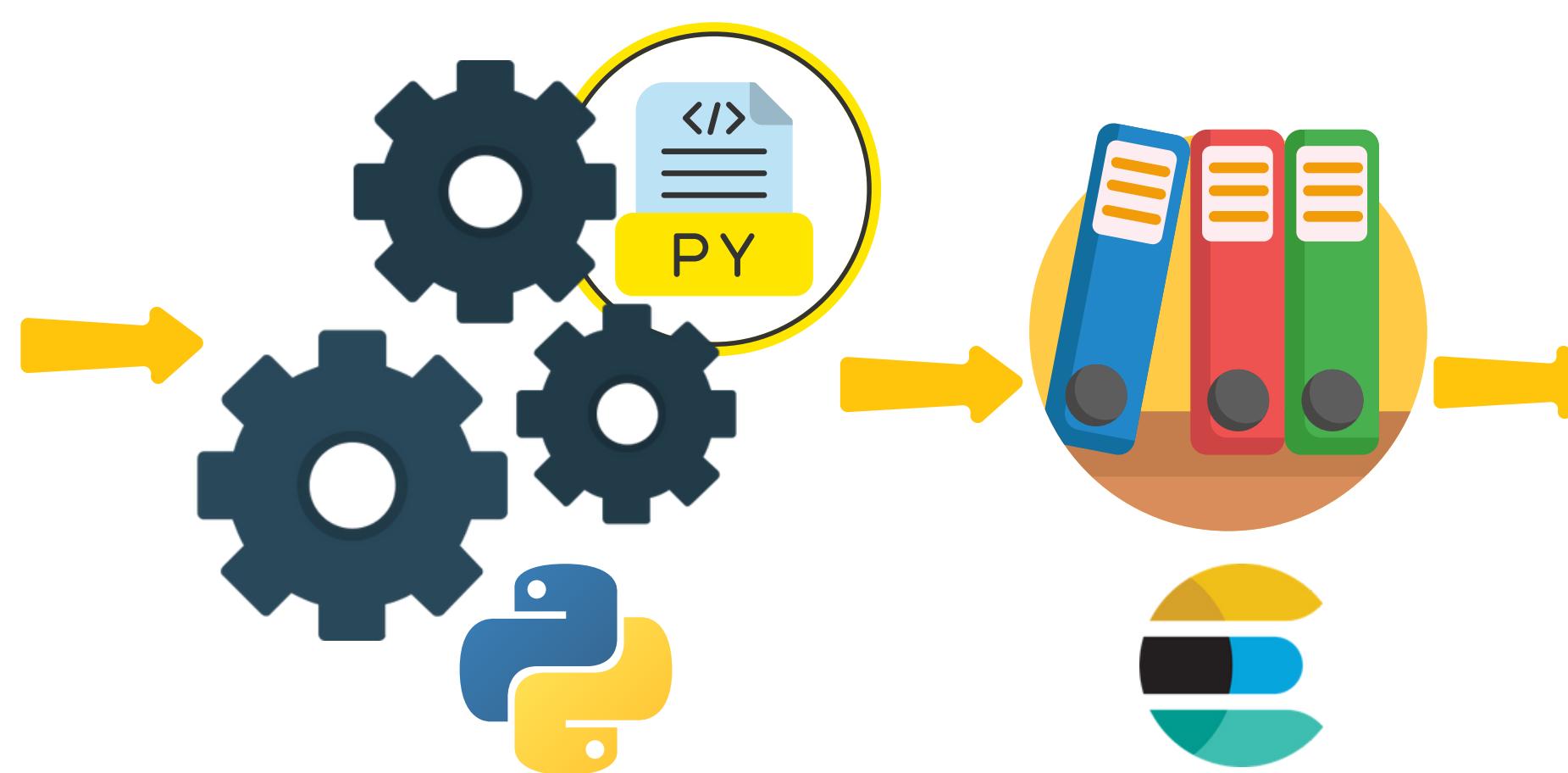


# Architecture de la Solution Proposée:

Source des données :



Traitement et Nettoyage :



Indexation :



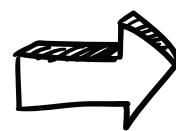
Déploiement :



Vue d'ensemble de l'architecture technique

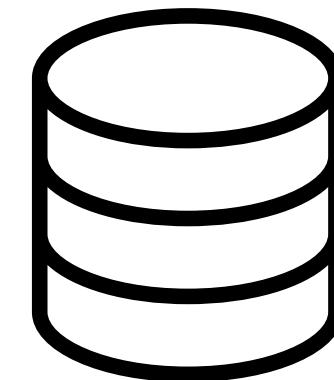


## Les sources de données à utiliser pour faire le preprocessing:



### Fichiers JSON :

- surahs.json : Noms et informations des sourates.
- verses.json : Versets complets du Coran.
- quran\_no\_chakl.json : Versets sans diacritiques.



### Base de données SQL :

- Table tafsir1 : Tafsirs associés aux versets.



### Fichier CSV :

- hadiths.csv : Textes de Hadiths avec sources et chapitres.



## Prétraitement des Données du Coran:

- ➡ **Chargement des Données** : Importation des fichiers JSON.
- ➡ **Connexion SQL** : Récupération des tafsirs depuis la base de données.
- ➡ **Structure des Données** : Construction d'une liste unifiée contenant :
  - Indices des versets
  - Noms des sourates
  - Textes des versets (avec et sans diacritiques)
  - Tafsirs associés
- ➡ **Sauvegarde** : Enregistrement dans **quran.json**.





## Prétraitement des Données des Hadiths:

- ➡ **Chargement des Données** : Lecture de hadiths.csv.
- ➡ **Nettoyage** :
  - Suppression des colonnes inutiles (id, hadith\_id, chain\_indx).
  - Remplacement des noms des sources en anglais par leur version arabe.
- ➡ **Suppression des Caractères Indésirables** : Élimination des caractères HTML et Unicode.
- ➡ **Gestion des Doubloons** : Identification et suppression des doublons dans text\_en.
- ➡ **Réorganisation** : Sélection des colonnes nécessaires.

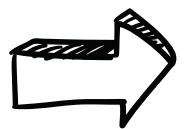


## Résultats de preprocessing :



### Format CSV vers JSON :

- Fonction `csv_to_json` pour convertir le fichier nettoyé **hadith\_processed.csv** en **json\_hadith.json**.



### Facilité de stockage et d'intégration



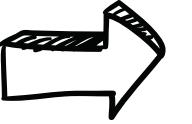
## Mapping et Création d'Index:

→ **Section settings : Cette section contient les paramètres de configuration, tels que :**

- Les paramètres de base :
  - number\_of\_shards : Définit le nombre de partitions logiques pour l'index.
  - number\_of\_replicas : Spécifie le nombre de copies de chaque shard pour assurer la redondance.
- Analyse linguistique :
  - Filtres : Définissent des étapes spécifiques comme la suppression des mots vides (stopwords), la normalisation, le lemmatisation (stemming) et la création de n-grams.
  - Analyseurs : Assemblages de filtres réutilisables pour analyser les textes en fonction des besoins.



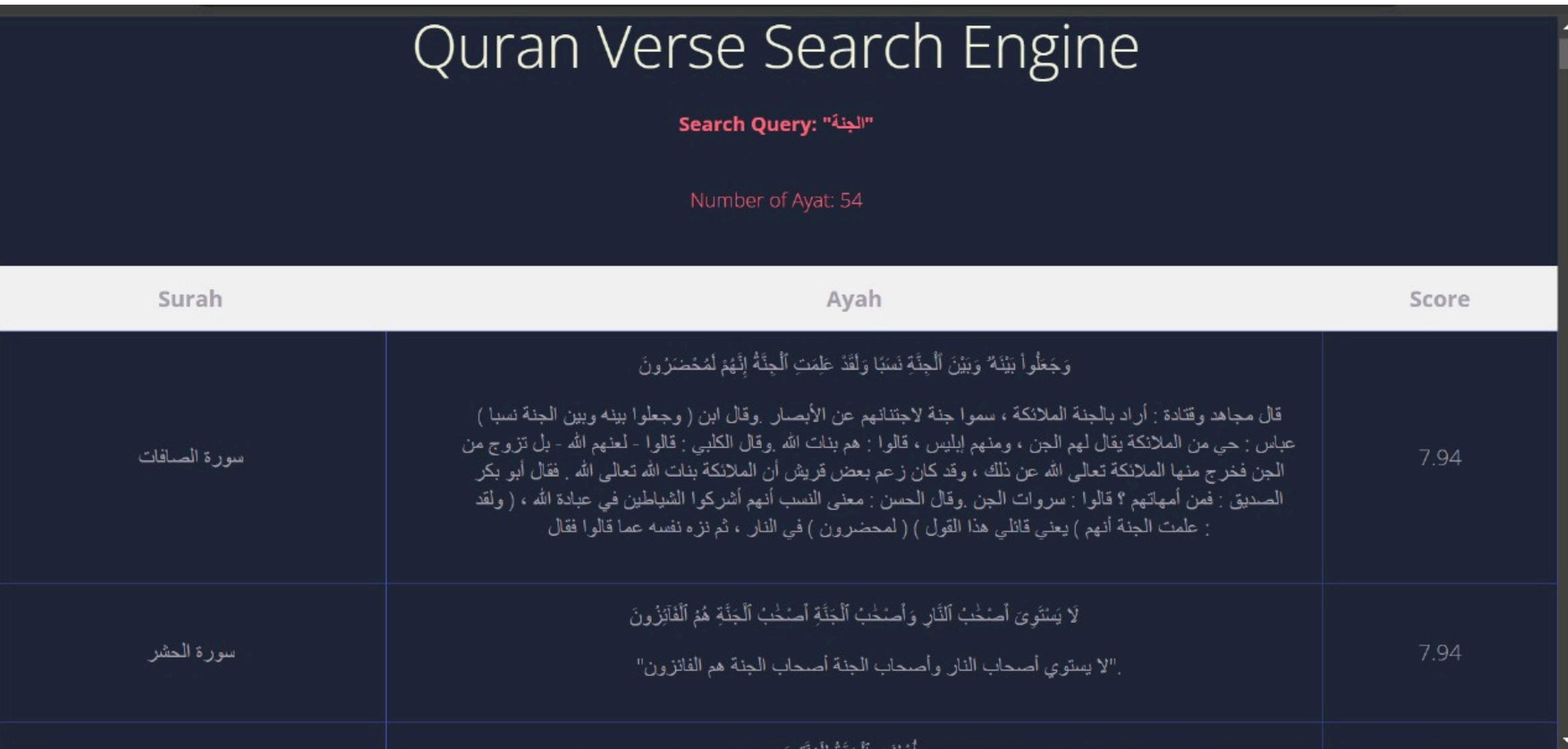
## Mapping et Création d'Index:

 **Section mappings : Cette section définit la structure des données en spécifiant :**

- Les types des champs : (text, integer, keyword, etc.).
- L'utilisation des analyseurs linguistiques : Les champs textuels peuvent être analysés avec les analyseurs définis dans settings.

# Déploiement:

## Une recherche sur "AL-jannah" Dans le Quran



The screenshot shows a search interface for the Quran. The search query "الجنة" has returned 54 results. The results are presented in a table with columns: Surah, Ayah, and Score.

Surah	Ayah	Score
سورة الصافات	وَجَعَلُوا بَيْنَهُ وَبَيْنَ الْجَنَّةِ نَسْبًا وَلَقَدْ عَلِمْتِ الْجَنَّةَ إِنَّهُمْ لِمُخْضَرُونَ قال مجاهد وقتادة : أراد بالجنة الملائكة ، سموا جنة لاجتنانهم عن الأ بصار . وقال ابن ( وجعلوا بينه وبين الجنة نسبا ) عباس : هي من الملائكة يقال لهم الجن ، ومنهم إيليس ، قالوا : هم بنات الله . وقال الكابي : قالوا - لعنهم الله - بل تزوج من الجن فخرج منها الملائكة تعالى الله عن ذلك ، وقد كان زعم بعض قريش أن الملائكة بنات الله تعالى الله . فقال أبو بكر الصديق : فمن أمهاةهم ؟ قالوا : سروات الجن . وقال الحسن : معنى النسب أنهم أشروا الشياطين في عبادة الله ، ( ولقد علمت الجنة إنهم ) يعني قاتلي هذا القول ( لمحضرورون ) في النار ، ثم نزه نفسه عما قالوا فقال	7.94
سورة الحشر	لَا يَسْتُوْ أَصْحَابُ النَّارِ وَأَصْحَابُ الْجَنَّةِ أَصْحَابُ الْجَنَّةِ هُمُ الْفَائِزُونَ ".لا يستوي أصحاب النار وأصحاب الجنة أصحاب الجنة هم الفائزون"	7.94



# Déploiement:

## Une recherche sur "AL-maal"

### Dans le Quran

Quran Verse Search Engine

Search Query: "المال"

Number of Ayat: 4

Surah	Ayah	Score
سورة الفجر	<p style="text-align: right;">وَتُحِبُّونَ الْمَالَ حَتَّىٰ حَمَّا</p> <p>وتحبون المال حباً جماً، أي كثراً، يعني: تحبون جمع المال وتولعون به، يقال: جم الماء في الحوض، إذا كثر واجتمع</p>	11.12
سورة الكهف	<p style="text-align: right;">الْمَالَ وَالْبَنِينَ زِينَةُ الْحَيَاةِ الدُّنْيَا وَالْبَيْتُ الصَّلَاحُ خَيْرٌ عِنْدَ رَبِّكَ تَوَابًا وَخَيْرٌ أَمْلًا</p> <p>التي يفتخر بها عتبة وأصحابه الأغنياء (زينة الحياة الدنيا) ليست من زاد الآخرة. قال علي بن أبي طالب (المال والبنون) رضي الله عنه : المال والبنون حرث الدنيا والأعمال الصالحة حرث الآخرة وقد يجمعها الله لأقوام . (والباقيات الصالحات) اختلفوا فيها فقال ابن عباس وعكرمة ومجاهد : هي قول : سبحان الله والحمد لله ولا إله إلا الله والله أكبر . وقد رويانا أن النبي صلى الله عليه وسلم قال : "أفضل الكلام أربع كلمات : سبحان الله والحمد لله ولا إله إلا الله والله أكبر ". أخبرنا أبو القاسم عبد الله بن محمد الحنفي أباينا أبو بكر محمد بن الحسن الحيري أخبرنا أبو جعفر عبد الله بن إسماعيل الهاشمي أباينا أحمد بن عبد الجبار العطاردي حدثنا أبو معاوية عن أبي صالح عن أبي هريرة رضي الله عنه قال : قال رسول الله صلى الله عليه وسلم : " لأن أقول سبحان الله والحمد لله ولا إله إلا الله والله أكبر أحب إلى مما طاعت عليه الشمس ". أخبرنا</p>	9.23



## Déploiement:

### Une recherche sur "AL-Hayae" Dans le Hadith

Hadith Search Engine

Search Query: "الحياء"

Number of Hadiths: 26

Chapter	Hadith	Score
صحيح البخاري	<p>حدثنا آدم، حدثنا شعبة، عن أبي السوار العدوبي، قال سمعت عمران بن حصين، قال قال النبي صلى الله عليه وسلم الحباء لا يأتي إلا بخير فقال بشير بن كعب مكتوب في الحكمة إن من الحباء وقار، وإن من الحباء سكينة فقال له عمران أحدثك عن رسول الله صلى الله عليه وسلم وتحذثني عن صحيفتك</p> <p>كتاب الأدب - Good Manners and Form (Al-Adab)</p>	13.07
سنن الترمذى	<p>حدثنا ابن أبي عمر، وأحمد بن منيع، - المعنى واحد قالا حدثنا سفيان بن عيينة، عن الزهرى، عن سالم، عن أبيه، أن رسول الله صلى الله عليه وسلم مر برجل وهو يعظ أخاه في الحياة، فقال رسول الله صلى الله عليه وسلم الحباء من الإيمان قال أحمد بن منيع في حديثه إن النبي صلى الله عليه وسلم سمع رجلا يعظ أخيه في الحياة قال هذا حديث حسن صحيح وفي الباب عن أبي هريرة وأبي بكر وأبي أمامة</p> <p>كتاب الإيمان عن رسول الله صلى الله عليه وسلم - The Book on Faith</p>	12.76



## Déploiement:

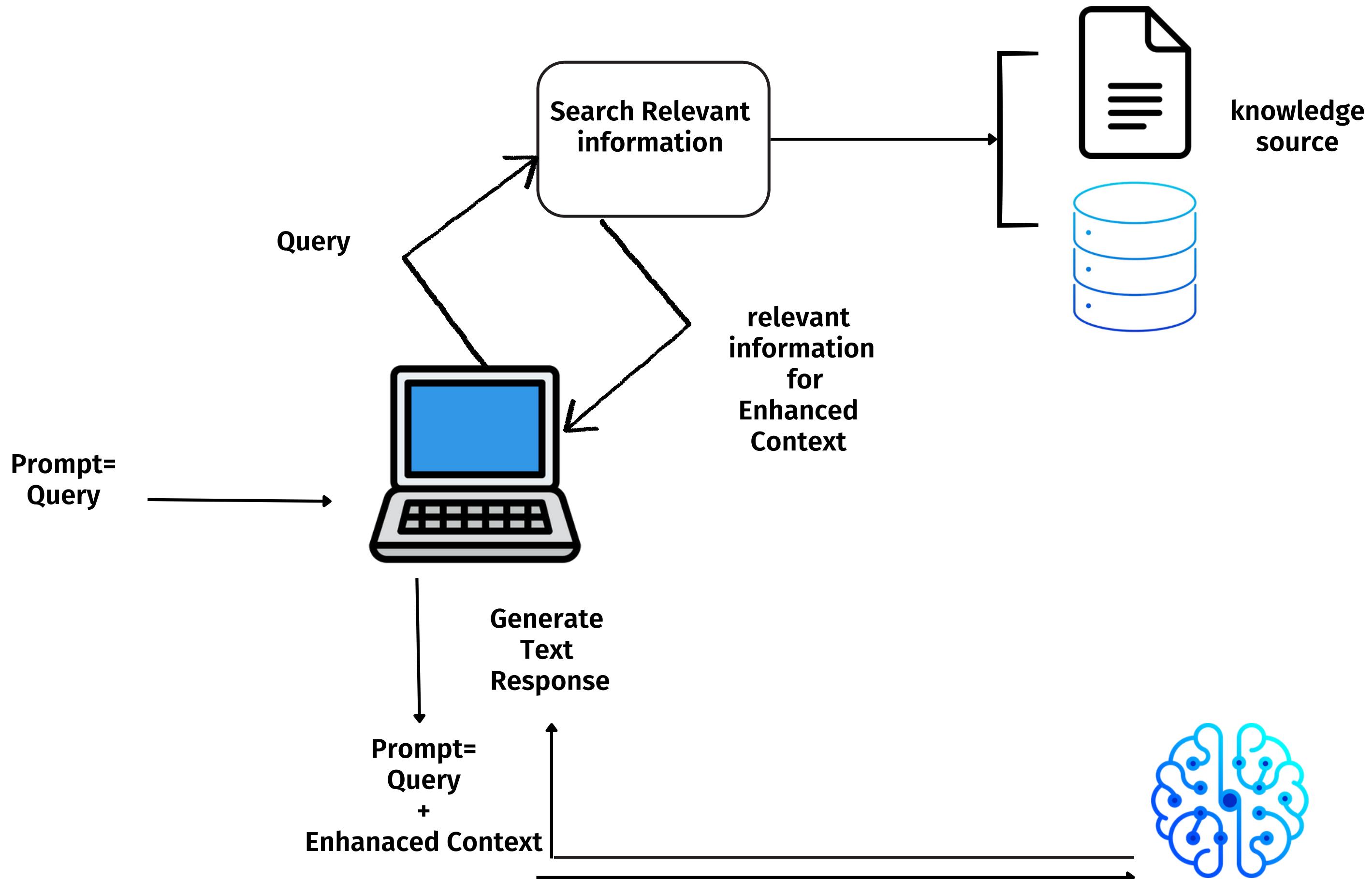
### Une recherche sur "AL-Iman" Dans le Hadith

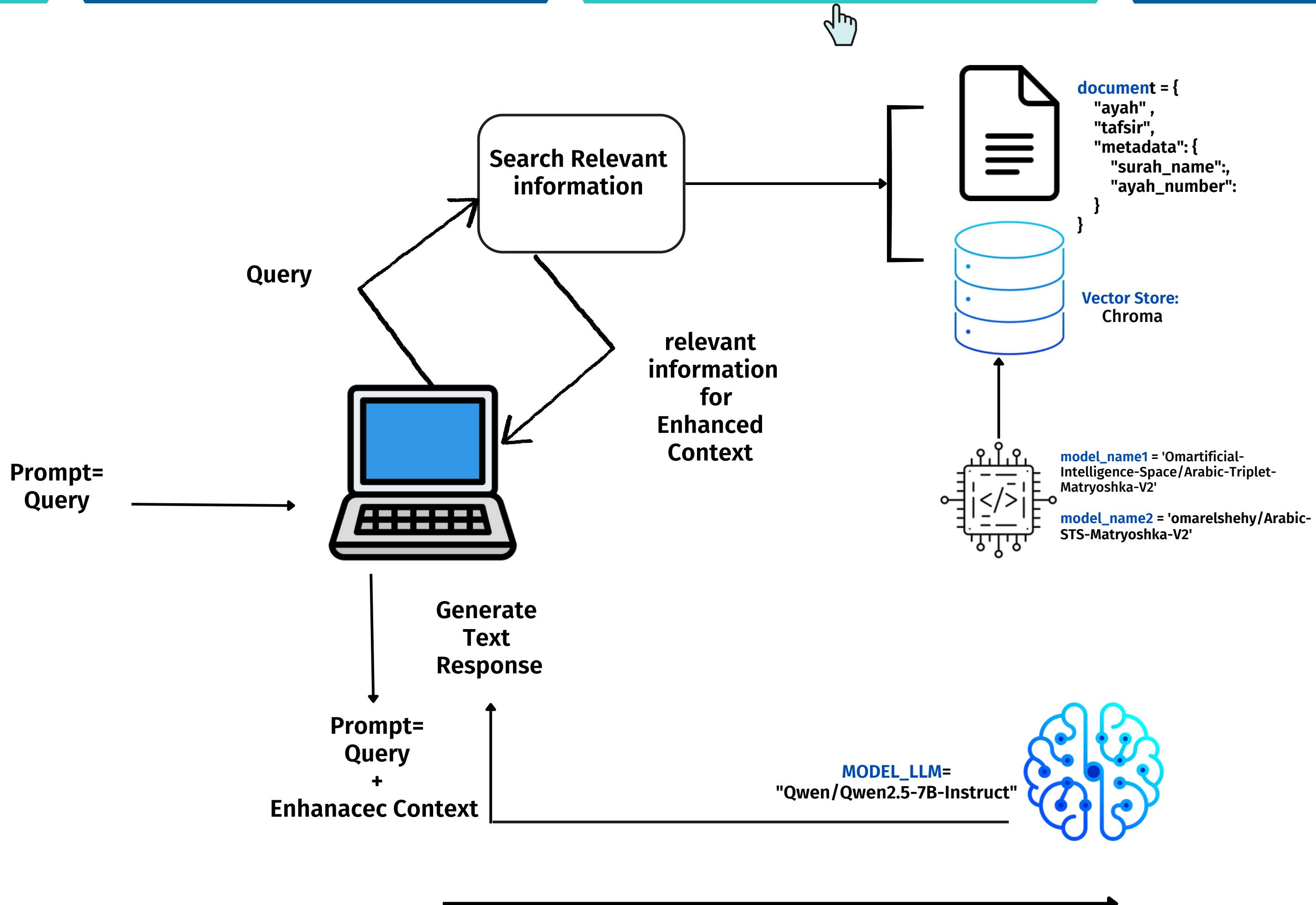
Search Query: "الإيمان"

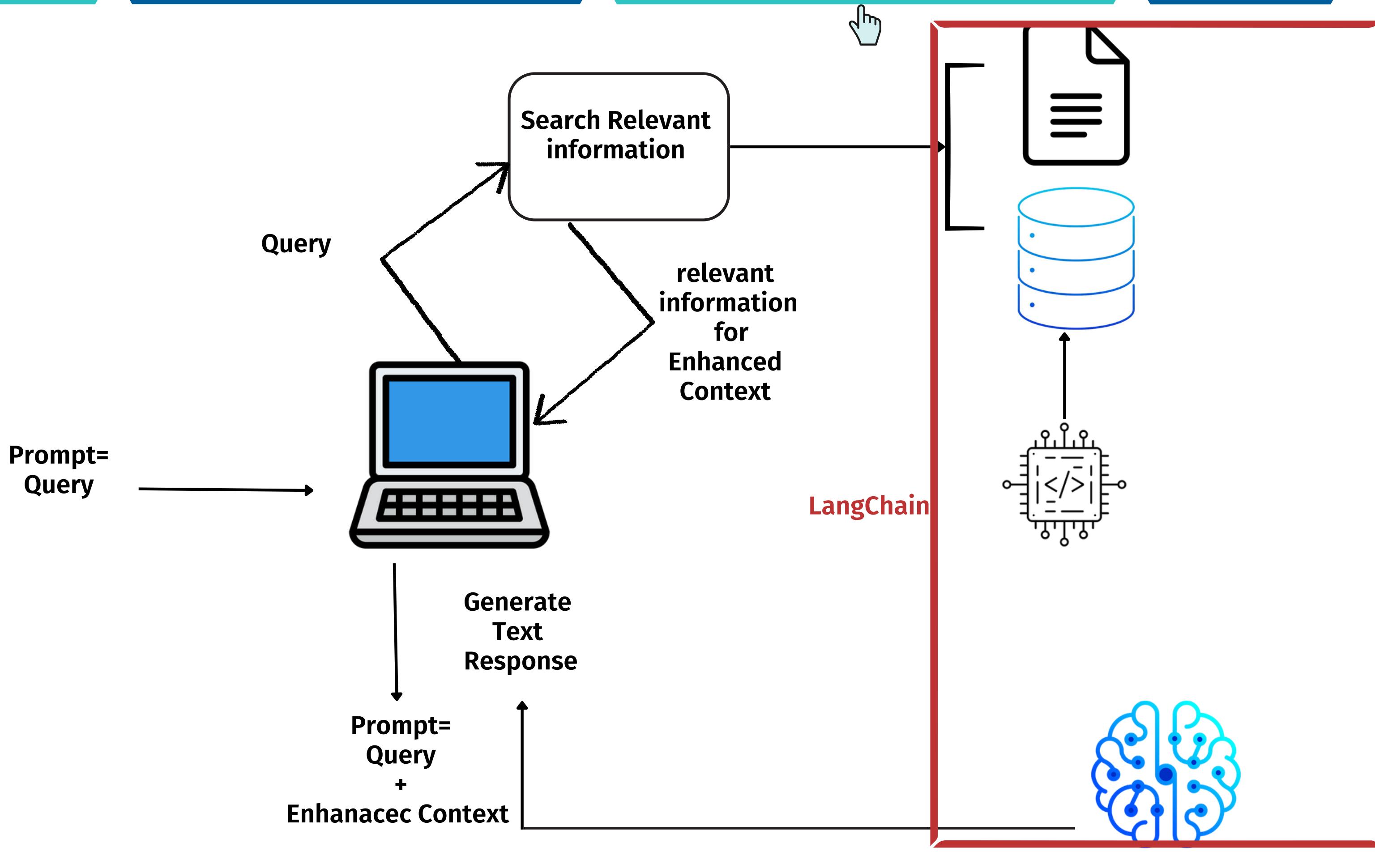
Number of Hadiths: 158

Chapter	Hadith	Score
سنن الترمذى	<p>حدثنا عباس بن محمد الدورى، حدثنا عبد الله بن يزيد المفرى، حدثنا سعيد بن أبي أبوب، عن أبي مرحوم عبد الرحيم بن ميمون، عن سهل بن معاذ بن أنس الجهنى، عن أبيه، أن رسول الله صلى الله عليه وسلم قال من ترك اللباس تواضعاً لله وهو يقدر عليه دعاه الله يوم القيمة على رءوس الخلانق حتى يخربه من أى حل الإيمان شاء يلبسها هذا حديث حسن ومعنى قوله حل الإيمان يعني ما يعطى أهل الإيمان من حل الجنـة</p> <p>كتاب صفة القيمة - Chapters on the description of the Day of Judgement, Ar-Riqaq, and Al-Wara' والرقائق والورع عن رسول الله صلى الله</p>	9.51
صحیح مسلم	<p>حدثنا عبد الله بن سعيد، وعبد بن حميد، قالا حدثنا أبو عامر العقدى، حدثنا سليمان بن بلال، عن عبد الله بن دينار، عن أبي صالح، عن أبي هريرة، عن النبي صلى الله عليه وسلم قال الإيمان بضع وسبعين شعبة والحياة شعبة من الإيمان</p> <p>كتاب الإيمان - The Book of Faith</p>	9.31
سنن النسائي	<p>أخبرنا محمد بن عبد الله بن المبارك، قال حدثنا أبو عامر، قال حدثنا سليمان، - وهو ابن بلال - عن عبد الله بن دينار، عن أبي صالح، عن أبي هريرة، عن النبي صلى الله عليه وسلم قال الإيمان بضع وسبعين شعبة والحياة شعبة من الإيمان</p>	9.31

# RAG et Pre-processing









## Preprocesing

## Input



quran.json



tafseer.json

Bagaoui Tafsir >> Tafsir Al-Muyassar

The screenshot shows a GitHub repository page for 'QuranTafseer-ar-json'. The repository is public and has 1 branch and 0 tags. The commit history shows the following activity:

- 00AhmedMokhtar00 Update README.md (de2c207 · 5 years ago)
- azkar-doaa-data (Add files via upload · 5 years ago)
- quran-data (Add files via upload · 5 years ago)
- README.md (Update README.md · 5 years ago)
- tafsir.json (Create tafsir.json · 5 years ago)



## Preprocessing

## Transformation

```
import json

def merge_json_files(file1_data, file2_data):
    """
    Fusionne deux fichiers JSON en remplaçant les champs "tafsir_aya" du second fichier
    avec les champs "text" du premier fichier, basés sur les correspondances entre
    "number" et "sura_index", et "aya" et "aya_index_sura".

    :param file1_data: Liste de dictionnaires contenant les champs "number", "aya" et "text".
    :param file2_data: Liste de dictionnaires contenant les champs "sura_index", "aya_index_sura", et "tafsir_aya".
    :return: Liste mise à jour de file2_data avec les modifications.
    """

    # Création d'un dictionnaire pour un accès rapide aux textes du premier fichier
    mapping = {(item["number"], item["aya"]): item["text"] for item in file1_data}

    # Parcours du deuxième fichier et remplacement du champ "tafsir_aya"
    for entry in file2_data:
        key = (entry["sura_index"], entry["aya_index_sura"])
        if key in mapping:
            entry["tafsir_aya"] = mapping[key]

    return file2_data
```



## Preprocesing

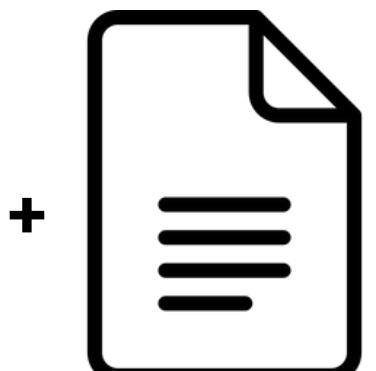
output

```
[  
  {  
    "aya_index_quran": 1,  
    "aya_index_sura": 1,  
    "sura_name": "سورة الفاتحة",  
    "sura_index": 1,  
    "total_verses": 7,  
    "aya_chakl": "بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ",  
    "aya_no_chakl": "الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ",  
    "tafsir_aya": "الْخَلْقُ, (الرَّجِيم) بِالْمُؤْمِنِينَ, وَهُمَا اسْمَانُ مِنْ أَسْمَائِهِ تَعَالَى, يَتَضَمَّنُ إِثْبَاتَ صَفَةِ الرَّحْمَةِ لَهُ تَعَالَى كَمَا يُلْيِقُ بِجَلَلِهِ"  
  },  
  {  
    "aya_index_quran": 2,  
    "aya_index_sura": 2,  
    "sura_name": "سورة الفاتحة",  
    "sura_index": 1,  
    "total_verses": 7,  
    "aya_chakl": "الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ",  
    "aya_no_chakl": "الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ",  
    "tafsir_aya": "الْخَلْقُ, الْقَائِمُ بِأَمْرِهِ, الْمَرْبُو لِجَمِيعِ خَلْقِهِ بِنَعْمَهُ, وَأَوْلِيَّاهُ بِالْإِيمَانِ وَالْعَمَلِ الصَّالِحِ (الْخَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ)"  
  },  
]
```

→output\_tafseer.json



يَا أَيُّهَا الَّذِينَ آمَنُوا إِذَا تَدَاءَنْتُم بِدِينِ إِلَّا أَجَلٌ مُسَعٌ  
فَاصْنُبُوهُ وَلَا يَكُتبَ بَيْنَكُمْ كَاتِبٌ بِالْعَدْلِ وَلَا يَأْبَ  
كَاتِبٌ أَنْ يَكُتبَ كَمَا عَلِمَ اللَّهُ فَلَيَكُتبْ وَلَيُمْلِلِ  
إِلَذِي عَلَيْهِ الْحَقُّ وَلَيَسْتَقِي اللَّهُ رَبُّهُ وَلَا يَبْخَسْ مِنْهُ شَيْئًا  
فَإِنْ كَانَ الَّذِي عَلَيْهِ الْحَقُّ سَفِيهًا أَوْ ضَعِيفًا أَوْ لَا يَسْتَطِيعُ  
أَنْ يُمْلِلْ هُوَ فَلَيُمْلِلْ وَلِيُهُ بِالْعَدْلِ وَاسْتَشِهِدُوا شَهِيدَيْنِ  
مِنْ رِجَالِكُمْ فَإِنْ لَمْ يَكُونَا رَجُلَيْنِ فَرَجُلٌ وَامْرَأَتَيْنِ  
مِمَّنْ تَرْضَوْنَ مِنْ أَشْهَدَاهُ أَنْ تَضْلِلَ إِحْدَاهُمَا فَتُذَكِّرَ  
إِحْدَاهُمَا الْأُخْرَى وَلَا يَأْبَ أَشْهَدَاهُ إِذَا مَا دُعُوا وَلَا تَسْعُوا  
أَنْ تَكْتُبُوهُ صَغِيرًا أَوْ كَبِيرًا إِلَّا أَجْلَهُ ذَلِكُمْ أَقْسَطُ  
عِنْدَ اللَّهِ وَأَقْوَمُ لِشَهَدَةِ وَأَدْبَنْ أَلَا تَرْتَابُوا إِلَّا أَنْ تَكُونَ  
تِجْرَةٌ حَاضِرَةٌ تُدِيرُونَهَا بَيْنَكُمْ فَلَيَسْ عَلَيْكُمْ جُنَاحٌ  
إِلَّا تَكْتُبُوهَا وَأَشْهِدُوهَا إِذَا تَبَايعُتُمْ وَلَا يُضَارَ كَاتِبٌ  
وَلَا شَهِيدٌ وَإِنْ تَفْعَلُوا فَإِنَّهُ فُسُوقٌ بِكُمْ وَاتَّشَأْ  
اللَّهُ وَيَعْلَمُكُمْ أَلَّا اللَّهُ وَاللَّهُ يَكُلِّ شَيْءٍ عَلَيْكُمْ



tafseer.json

docs	full_tokens	full_token_count	truncated_tokens	truncated_token_count
288	السور، #، :، سورة، البقرة، [[CLS]]... الآية: يا أيها السورة: سورة البقرة ...الذين آمنوا	527	السور، #، :، سورة، البقرة، [[CLS]]... الآي، #، ...	512

# Embeddings



# Semantic search & Embeddings models

Traditional Keyword  
Based Search



VS

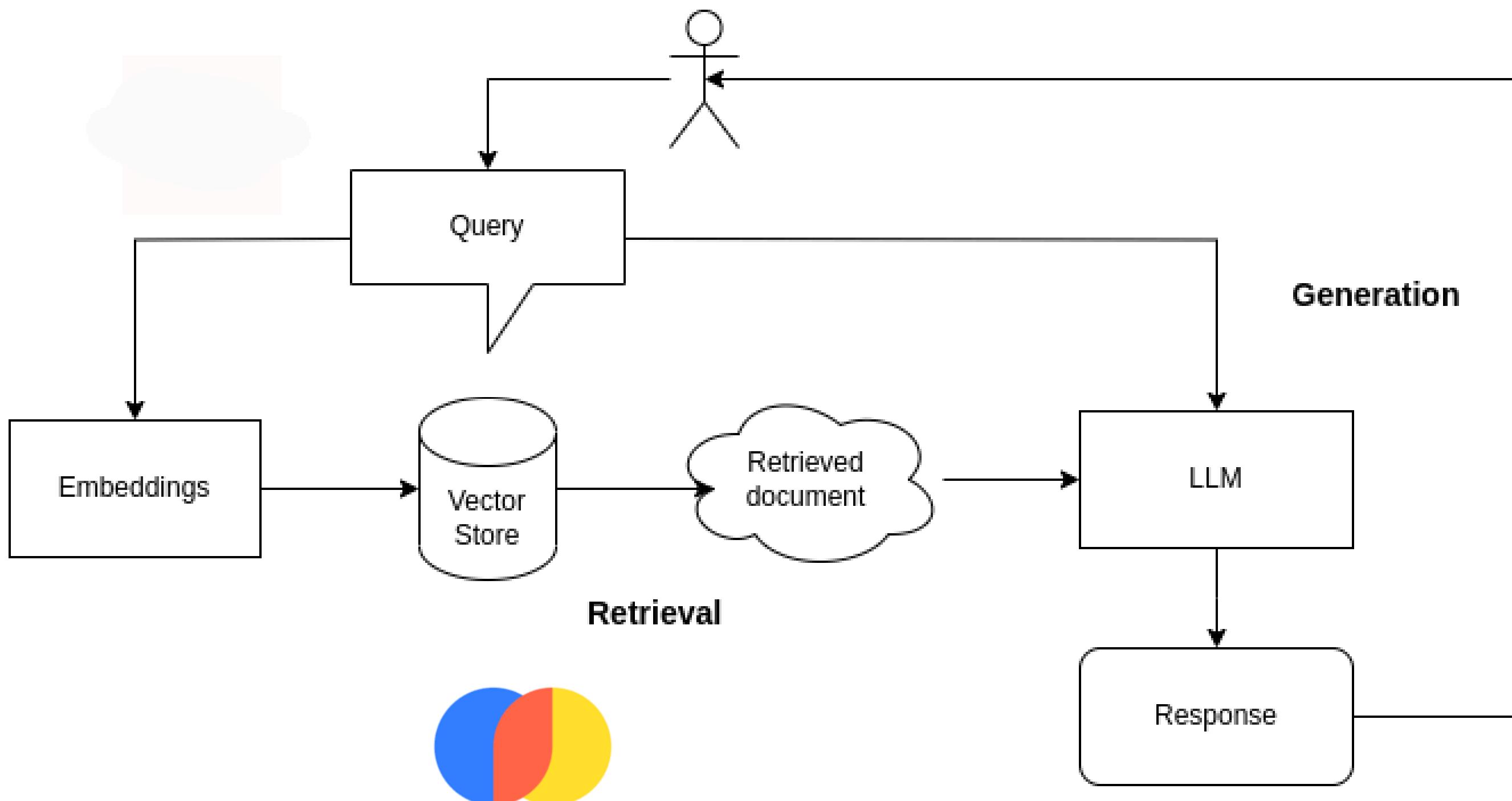
Semantic Search

- Query's Intent
- Context
- Semantics





# Semantic search & Embeddings models : Architecture du RAG



**Chroma**



# Semantic search & Embeddings models - MTEB Leaderboard

Overall   Bitext Mining   Classification   Clustering   Pair Classification   Reranking   Retrieval   **STS**   Summarization   MultilabelClassification  
Retrieval w/Instructions

Semantic Textual Similarity is the task of determining how similar two texts are.

English   Chinese   French   Polish   Russian   **Other**

**STS Other leaderboard** 🍀

- Metric:** Spearman correlation based on the model's similarity metric (usually cosine)
- Languages:** Arabic, Chinese, Dutch, English, French, German, Italian, Korean, Polish, Russian, Spanish (Only language combos not included in the other tabs)

Rank	Model	Model Size (Million Parameters)	Memory Usage (GB, fp32)	Embedding Dimensions	Max Tokens	Average	STS17 (ar-ar)	STS17 (en-en)	STS17 (en-de)	STS17 (es-es)
280	<a href="#">Arabic-STS-Matryoshka-V2</a>	135	0.5	768	512		86.06	13.34		
308	<a href="#">silma-embeddding-sts-v0.1</a>	135	0.5	768	512		85.59	47.7	31.86	9.1
130	<a href="#">Arabic-Triplet-Matryoshka-V2</a>	135	0.5	768	512		85.31			



## Embeddings models : Arabic-Triplet-Matryoshka-V2

- Is a **sentence-transformers** model finetuned from **aubmindlab/bert-base-arabertv02**.
- Génère des vecteurs denses de **768 dimensions**.
- Applications : semantic textual similarity (STS), semantic search, text classification, clustering, ...
- Entraîné sur 1 million d'échantillons du dataset **akhooli/arabic-triplets-1m-curated-sims-len**.
- Entraîné sur 3 époques avec une perte finale de 0,718 (**MatryoshkaLoss**).





## Embeddings models : Arabic-STS-Matryoshka-V2

- **Model Type:** Sentence Transformer
- **Base model:** [aubmindlab/bert-base-arabertv02](#)
- **Maximum Sequence Length:** 512 tokens
- **Output Dimensionality:** 768 tokens
- **Similarity Function:** Cosine Similarity



# Embeddings models : Comparaison Semantic search

```

Marquage | Exécuter tout | Redémarrer | Effacer toutes les sorties | Variables Jupyter | Structure ... Python 3.11.3

results for : ايات القسم

===== Document 1 =====
Document Content:
الآية: قاتلوا ذكرا
إِنْ مَعْبُودُكُمْ -أَيْهَا النَّاسُ- لَوَاحِدٌ لَا شَرِيكَ لَهُ، فَأَخْلَصُوا لَهُ الْعِبَادَةَ وَالطَّاعَةَ. وَيَقْسِمُ اللَّهُ بِمَا شَاءَ مِنْ خَلْقِهِ، أَمَا الْمُخْلُوقُ فَلَا يَجُوزُ لَهُ الْقُسْمَ إِلَّا بِأَيْدِيهِ، فَالْحَلْفُ بِغَيْرِ اللَّهِ شَرِكٌ
Metadata:
{السورة': 'سورة العنكبوت', 'رقم الآية': '3', 'الآية': 'إِنْ مَعْبُودُكُمْ -أَيْهَا النَّاسُ- لَوَاحِدٌ لَا شَرِيكَ لَهُ، فَأَخْلَصُوا لَهُ الْعِبَادَةَ وَالطَّاعَةَ. وَيَقْسِمُ اللَّهُ بِمَا شَاءَ مِنْ خَلْقِهِ، أَمَا الْمُخْلُوقُ فَلَا يَجُوزُ لَهُ الْقُسْمَ إِلَّا بِأَيْدِيهِ، فَالْحَلْفُ بِغَيْرِ اللَّهِ شَرِكٌ'}
Similarity Score: 0.4423079490661621

===== Document 2 =====
Document Content:
الآية: فالمقدمة أمراء
سُرُورٌ وَسُهُولٌ، فَالْمَلَائِكَةُ الَّتِي تُقْرِئُهُمْ أَمْرَهُ فِي خَلْقِهِ. إِنَّ الَّذِي تَوَعَّدُونَ بِهِ -أَيْهَا النَّاسُ- مِنَ الْبَيْعَةِ وَالْحِسَابِ لِكُلِّ أَنْشَاءٍ حَقِيقَيْنِ، وَإِنَّ الْحِسَابَ وَالْتَّوَابَ عَلَى الْأَعْمَالِ لِكُلِّ أَنْشَاءٍ لَا محَالَةٍ
Metadata:
{السورة': 'سورة النازاريات', 'رقم الآية': '4', 'الآية': 'سُرُورٌ وَسُهُولٌ، فَالْمَلَائِكَةُ الَّتِي تُقْرِئُهُمْ أَمْرَهُ فِي خَلْقِهِ. إِنَّ الَّذِي تَوَعَّدُونَ بِهِ -أَيْهَا النَّاسُ- مِنَ الْبَيْعَةِ وَالْحِسَابِ لِكُلِّ أَنْشَاءٍ حَقِيقَيْنِ، وَإِنَّ الْحِسَابَ وَالْتَّوَابَ عَلَى الْأَعْمَالِ لِكُلِّ أَنْشَاءٍ لَا محَالَةٍ'}
Similarity Score: 0.4615975618362427

===== Document 3 =====
Document Content:
الآية: لِعَمْرٍكَ إِنْهُمْ لِفِي سُكْرٍ تُهُمْ يَعْمَلُونَ
الآنَ قَدْ أَقْسِمَ اللَّهُ تَعَالَى بِحَيَاةِ مُحَمَّدٍ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ تَشْرِيفًا لَهُ. إِنَّ قَوْمَ لَوْطٍ فِي عَنْلَةٍ شَدِيدَةٍ يَتَرَدَّدُونَ وَيَتَمَادُونَ، حَتَّى حَلَّ بِهِمْ صَاعِدَةُ الْعَذَابِ وَقَتْ شَرُوقِ الشَّمْسِ
Metadata:
...
Metadata:
{السورة': 'سورة ق', 'رقم الآية': '1', 'الآية': 'لِعَمْرٍكَ إِنْهُمْ لِفِي سُكْرٍ تُهُمْ يَعْمَلُونَ'}
Similarity Score: 0.4918227195739746

```

## Arabic-STS-Matryoshka-V2

Batches: 100% [1/1] 00:00<00:00, 48.16it/s]

Input question: ايات القسم

Results (after 0.028 seconds):

Score: 0.574

Document 1:

السورة: سورة الفلم

الآية: [1] بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ نَ وَالْقَلْمَ وَمَا يَسْطِرُونَ

التفسير : (ن) سبق الكلام على الحروف المقطعة في أول سورة البقرة. أقسم الله بالقلم الذي يكتب به الملائكة والناس، وبما يكتبون من الخير والتفع والعلوم. ما أنت -أيتها الرسول- بسبب نعمة الله عليك بالرسالة بضعف العقل، ولا سفيه الرأي، وإن لك على ما تلقاه من شدائد على تبليل الرسالة لثوابها عظيماً غير منقوص ولا مقطوع، وإنك -أيتها الرسول- لعلى حلق عظيم، وهو ما اشتغل عليه القرآن من مكار م الأخلاق؛ فقد كان امتنال القرآن سجية له يأمر بأمره، وينتهي بما ينهى عنه

Score: 0.552

Document 2:

السورة: سورة الفصل

الآية: [1] بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ طَسِّ

التفسير : (طسم) سبق الكلام على الحروف المقطعة في أول سورة البقرة

Score: 0.536

Document 3:

السورة: سورة ق

الآية: [1] بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ قَ وَالْقَلْمَ وَالْمَجِيد

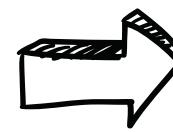
التفسير : (ة) سبق الكلام على الحروف المقطعة في أول سورة الفصل، أقسم الله تعالى بالقلم الك بما ذكره، المجد والشرف

## Arabic-Triplet-Matryoshka-V2

# Response generation with LLM



## Partie LLM: Génération de la réponse: Qwen2.5



### Origine :

- Développé par l'équipe Qwen d'Alibaba Group, incluant des modèles de langage et multimodaux(texte, video, audio, json, etc).



### Caractéristiques :

- Modèles dense et faciles à utiliser disponibles en tailles variant de 0.5B à 72B (base et **instruct**) (0.5B, 1.5B, 3B, **7B**, 14B, 32B, and 72B).
- Entraîné sur un ensemble de données de 18T tokens.
- Longueur de contexte jusqu'à 128 000 tokens et génération jusqu'à 8 000 tokens.
- Support multilingue pour plus de 29 langues, incluant le chinois, l'anglais, le français, **l'arabe**, etc.



## Qwen/Qwen2.5-7B-Instruct

LLM Leaderboard | About | Submit here!

Search for your model (separate multiple queries with `;` and press ENTER...)

Select columns to show

Model types

- pretrained
- continuously pretrained
- fine-tuned on domain-specific datasets

Precision

- chat models (RLHF, DPO, IFT, ...)
- base merges and moerges
- 

Model sizes (in billions of parameters)

?
~1.5
~3
<input checked="" type="checkbox"/> ~7
~13
~35
~60
70+

Show gated/private/deleted models

Model

Model	Average	ACVA	AlGhafa	MMLU	EXAMS	ARC Challenge	ARC Easy	BOOLQ	COPA
yellowtown/7B-v0.2	60.87	40.06	56.33	51.77	50.47	63.45	64.3	77.15	64.44
huihui-ai/Qwen2.5-7B-Instruct-abliterated-v2	55.05	53.73	49.31	51.23	50.28	51.29	48.69	76.87	57.78
Joseph717171/Llama-3.1-SuperNova-8B-Lite_TIES_with_Base	54.87	46.64	58.32	45.02	53.63	49.05	52.07	78.93	60
Orion-zhen/Qwen2.5-7B-Instruct-Uncensored	54.49	39.51	50.18	52.37	49.35	50.43	47.8	79.91	56.67
Owen/Owen2.5-7B-Instruct	54.3	39.66	49.17	51.93	50.09	49.83	46.4	76.6	56.67

```
task="text-generation",
temperature=0.2,
repetition_penalty=1.1,
return_full_text=False,
max_new_tokens=500,
```

- **temperature=0.2** : Contrôle la créativité des réponses.
- **repetition\_penalty=1.1** : Réduit la probabilité que le modèle répète les mêmes phrases ou mots en appliquant une pénalité lors des répétitions.
- **return\_full\_text=False** : Indique que le texte généré seul doit être retourné, sans inclure le prompt d'entrée.
- **max\_new\_tokens=500** : Spécifie le nombre maximal de nouveaux tokens à générer.



## Comment la réponse se génère? 🤔

### ➡ Orientation de la réponse générée :

- Répondre uniquement aux questions posées.
- Ne pas inclure de versets du Coran directement, mais refléter les principes islamiques.
- Répondre uniquement en arabe.

### ➡ Contexte:

- Est extrait à partir de la requête user et les documents pertinents extraits



Prompt se  
donne au llm



## Structure de la réponse 😐

```

print("=====Answer=====")
print(f'{answer}')
print("=====Source docs=====")
for i, (doc, score) in enumerate(relevant_docs):
    print(f'Document {i}')
    print(f'*SIM={score:.3f}')
    print(f'*Content :{doc.page_content}')
    print(f'*Metadata: {doc.metadata}')

print("\n*****")

```

=====Answer=====

هم أصحاب الجنة هم السابقون إلى الخيرات في الدنيا والذين يدخلهم ربهم في جنة هناء هم الوارثون للجنة وهم الذين أثأوا من أهلي كتاب أعماله بسم الله الرحمن الرحيم في حينه هناءة مرضية في جنة مرتفعة المكان والدرجات.

=====Source docs=====

Document 0

\*SIM=0.421404

\*Content: الآية: أولئك المقربون:

.التسير: والسابقون إلى الخيرات في الدنيا هم السابقون إلى الدرجات في الآخرة، أولئك هم المقربون عند الله، يدخلهم ربهم في جنة التسir

\*Metadata: {'السورة': 'سورة الواقعة', 'رقم الآية': 11}



## Exemples de réponses générées

Question : من هو الله؟

=====Answer=====

الله هو الخالق لكل شيء، وهو رب السموات والأرض. إنه يعلم سر العباد وجوهرهم، ويملك على كل شيء. يجب عبادته وحده دون شركاء، ويلغى أن يتذكر الناس به في صلاتهم.

=====Source docs=====

Document 0

\*SIM=0.408351

الآية: [وقالوا لخدا الله ولدا سيحانه بل له ما في السموات والأرض كل له فائلون] :Content

التفسير: وقالت اليهود والنصارى والمتركون: اتخذ الله لنفسه ولدا، تنتزه الله -سيحانه-. عن هذا القول الباطل، بل كل من في السموات والأرض ملكه وعبيده، وهم جميعاً خاضعون له، مستحرون تحت تصرفه.

\*Metadata: {'السورة': 'سورة البقرة', 'رقم الآية': 116}

\*\*\*\*\*

Document 1

\*SIM=0.409279

الآية: [نَّلَمَّكُمُ اللَّهُ رِبُّكُمْ لَا إِلَهَ إِلَّا هُوَ خَالِقُ كُلِّ شَيْءٍ فَاعْبُدُوهُ وَهُوَ عَلَى كُلِّ شَيْءٍ وَكُلِّ[Content]

التفسير: نلكم -أيها المشركون- هو ربكم جل وعلا لا معبود بحق سواء، خالق كل شيء فانقادوا واجبوا له بالطاعة والعبادة. وهو سيعانه على كل شيء وكل وحيط يدير أمور خلقه.

\*Metadata: {'السورة': 'سورة الأنعام', 'رقم الآية': 102}

\*\*\*\*\*

Document 2

\*SIM=0.412197

الآية: [وَهُوَ اللَّهُ فِي السَّمَاوَاتِ وَفِي الْأَرْضِ يَعْلَمُ مَا تَكْشِفُونَ] :Content

التفسير: والله سيعانه هو الإله المعبود في السموات والأرض. ومن دلائل ألوهيته أنه يعلم جميع ما تخونه -أيها الناس-. وما تعلونه، ويعلم جميع أعمالكم من خير أو شر، ولهم فائه -جل وعلا-. وحده هو الإله المستحق للعبادة.

\*Metadata: {'السورة': 'سورة الأنعام', 'رقم الآية': 3}

\*\*\*\*\*

Document 3

\*SIM=0.422213

الآية: [إِنِّي أَنَا اللَّهُ لَا إِلَهَ إِلَّا أَنَا فَاعْبُدُنِي وَاقْرَأْ الصَّلَاةَ لِذِكْرِي] :Content

التفسير: إنتي أنا الله لا معبود بحق إلا أنا، لا شريك لي، فاعبديني وحدي، واقرأ الصلاة لذكرني قبها

\*Metadata: {'السورة': 'سورة طه', 'رقم الآية': 14}

من هو الله؟



## Exemples de réponses générées

Question : من هم المنافقون :

=====Answer=====

من المنافقون هم المنافقون ذُكروا في الآية التي رقها 87 من سورة التوبة. وفقاً للتفسير، فإنهم رضوا بأنفسهم بالعار، وهو أن يقدروا في البوت مع النساء والصبيان وأصحاب الأعذار. كما أن الله ختم على قلوبهم بسبب نفاقهم وتخلصهم عن الجهاد والخروج مع النبي محمد صلى الله عليه وسلم في سبيل الله.

=====Source docs=====

Document 0

\*SIM=0.482481

\*Content: [قتل الخرامن]:

الآية: أُولئِنَّ الظَّانُونَ غَيْرُ الْحَقِّ، الَّذِينَ هُمْ فِي لُجُّهٍ مِّنَ الْكُفَّارِ وَالصَّابِلَةِ عَاقِلُونَ مُتَمَاهِرُونَ

\*Metadata: {'السورة': 'سورة الناريات', 'رقم الآية': '10'}

\*\*\*\*\*

Document 1

\*SIM=0.508187

\*Content: [الذين هم يراغبون]:

الآية: الظَّانُونَ الَّذِينَ هُمْ يَتَظَاهِرُونَ بِأَعْمَالِ الْخَيْرِ مُرَاوِدُ الْأَنْسَارِ

\*Metadata: {'السورة': 'سورة الماعون', 'رقم الآية': '6'}

\*\*\*\*\*

Document 2

\*SIM=0.516965

\*Content: [رضوا بأن يكونوا مع الخوارف وطبع على قلوبهم لهم لا يقهرون]:

الآية: [رضوا بأن يكونوا مع الخوارف وطبع على قلوبهم، بسبب نفاقهم وتخلصهم عن الجهاد والخروج مع رسول الله صلى الله عليه وسلم في سبيل الله، فهو لا يقهرون ما فيه صلاحهم ورشادهم]

\*Metadata: {'السورة': 'سورة التوبة', 'رقم الآية': '87'}

\*\*\*\*\*

Document 3

\*SIM=0.528900

\*Content: [يلقون السمع وأكثرهم كاذبين]:

الآية: هَلْ أَخِرْكُمْ - أَيُّهَا النَّاسُ - عَلَىٰ مَنْ تَنْزَلُ الشَّيْطَانُ؟ تَنْزَلُ عَلَىٰ كُلِّ كَذَّابٍ كَثِيرٍ الْأَكْلَامِ مِنَ الْكَهْنَةِ، يَشَرُّفُ الْمُبَاطِلِينَ السَّمْعَ، يَتَخَلَّفُونَ مِنَ الْمَلَأِ الْأَعْلَىِ، يَلْقَوْنَهُ إِلَى الْكَهْنَةِ، وَمَنْ جَرِيَ مُجَاهِمَهُ مِنَ الْفَسَقَةِ، وَأَكْثَرُ هُؤُلَاءِ كَذَّابِيُّونَ، يَسْتَأْذِنُ أَحَدَهُمْ فِي كَلْمَةٍ، فَيَزِيدُ فِيهَا أَكْثَرُ مِنْ مَائَةَ كَلْمَةٍ

\*Metadata: {'السورة': 'سورة الشراء', 'رقم الآية': '223'}

# من هم المنافقون؟

## Exemples de réponses générées



=====Answer=====

هم أصحاب الجنة هم المتقون والأبرار الذين قدموا الأعمال الصالحة في حياتهم، ويساركون في الاستمتاع بتمار الجنة. هؤلاء المقربون والوارتون للجنة هم السابقون إلى الخيرات في الدنيا والدرجات في الآخرة، وهو م الذين سيسعدون بالعيش الهنيء فيها.

=====Source docs=====

Document 0

\*SIM=0.421404

\*Content : الآية: [أولئك المقربون]

التفسير: والسابقون إلى الخيرات في الدنيا هم السابقون إلى الدرجات في الآخرة، أولئك هم المقربون عند الله، يدخلهم ربهم في جنات النعيم

\*Metadata: {'السورة': 'سورة الواقعة', 'رقم الآية': 11}

\*\*\*\*\*

Document 1

\*SIM=0.422851

\*Content : الآية: [أولئك هم الوارتون]

التفسير: هؤلاء المؤمنون هم الوارتون الجنة

\*Metadata: {'السورة': 'سورة المؤمنون', 'رقم الآية': 10}

\*\*\*\*\*

Document 2

\*SIM=0.428290

\*Content : الآية: [في جنة عالية]

التفسير: فأئم من أعطى كتاب أعماله بيدهيه، فيقول ابنهاجا ورسورا: خذوا أقرؤوا كتابي، إنني أينت في الدنيا بأني سألقى جزائي يوم القيمة، فأعذبت له العدة من الإيمان والعمل الصالح، فهو في عيشة هنية من ضبية، في جنة مرتفعة المكان والدرجات، تمارها قربة يتناولها القائم والقاعد والمضطجع. يقال لهم: كلوا أكلًا واتسربوا شرپا بعيدًا عن كل أذى، سالمين من كل مكره، بسبب ما قدّمت من الأعمال الصالحة في أيام الدنيا الماضية.

\*Metadata: {'السورة': 'سورة الحاقة', 'رقم الآية': 22}

\*\*\*\*\*

Document 3

\*SIM=0.444577

\*Content : الآية: [يشهد المقربون]

التفسير: حقا إن كتاب الأبرار -وهم المتقون- لفي المراتب العالية في الجنة. وما أدراك -أيها الرسول- ما هذه المراتب العالية؟ كتاب الأبرار مكتوب مفروغ منه، لا يزداد فيه ولا يتقص، يطلع عليه المقربون من

# من هم أصحاب الجنة؟



## Exemples de réponses générées

Question : من هم أصحاب النار؟

=====Answer=====

من الصعب تحديد أصحاب النار بدقة من خلال هذه النصوص، لكن يمكن الاستنتاج أن أهل النار هم الذين يحسون الله ورسوله، وكتبو بالذات الإلهي كما هو واضح من قوله تعالى "يُبَرِّدُنَّ إِلَى نَارِ جَهَنَّمَ دُعَا" في سورة الطور

=====Source docs=====

Document 0

\*SIM=0.493031

\*Content: الآية: [وما أدرك ما الحلمة]

التفسير: وما أدرك -أيها الرسول- ما حقيقة النار؟

\*Metadata: {السورة: 'سورة الهمزة', 'رقم الآية': 5}

\*\*\*\*\*

Document 1

\*SIM=0.499982

\*Content: الآية: [النار ذات الوقد]:

التفسير: أقسم الله تعالى بالسماء ذات المتأذل التي وعد الله الخلق أن يرحمهم فيه، وتشاهد بيته، ومتهم به، ومتهم به، وما يشاء من مخلوقاته، أما المخلوق فلا يجوز له أن يقسم بيته الله، فإن القسم بيته الله تبارك الله، إنهم على ما يفعلون بالمؤمنين من تحكيم وتحكيم حسر، وما لخزوم يمثل هذا العقاب التحكيم إلا أن كانوا مؤمنين بالله العزيز الذي لا يغافل، العميد هي أقواله وأفعاله ولو معاذه، الأرض شفاعة عظيمة، لتجذب المؤمنين، وأوكرنا النار التسديدة ذات الوقد، إذ هم مفود على الآخرين ملذتهم لهم، وهو على ما يفعلون بالمؤمنين من تحكيم وتحكيم حسر، وما لخزوم يمثل هذا العقاب التحكيم إلا أن كانوا مؤمنين بالله العزيز الذي لا يغافل، العميد هي أقواله وأفعاله ولو معاذه، الذي له ملك السموات والأرض، وهو سبحانه على كل شيء شهيد، لا يخفى عليه شيء

\*Metadata: {السورة: 'سورة البروج', 'رقم الآية': 5}

\*\*\*\*\*

Document 2

\*SIM=0.507336

\*Content: الآية: [كسلى نارا حامية]:

التفسير: وجوه الكفار يومئذ ذليلة بالذنب، مجده بالعمل متغيرة، تصيبها نار تسبّب الحرارة، ليس لأصحاب النار طعام إلا من ثبت ذمي ثوك لاصق بالأرض، وهو من شر الطعام وأخبثه، لا يُشْمَن بين صاحبه من الهُنْأ، ولا يُسْتَجْعَل جوعه ورمقه

\*Metadata: {السورة: 'سورة الشاثنة', 'رقم الآية': 4}

من هم أصحاب النار؟

# Interface Streamlit

# RAG-based Quran Chatbot



من هم المنافقون



من هم المنافقون حسب السياق هو أشخاص يظهر منهم الصدق خلاف ما في قلوبهم، حيث يكذبون ويغشون. كما أنهم راضيون بأنفسهم بالعار، وهو أن يقعدوا في البيوت مع النساء والصبيان وأصحاب الأعذار، وختم الله على قلوبهم بسبب نفاقهم وتخلفهم عن jihad والخروج مع النبي صلى الله عليه وسلم. كما أن المنافقين يحلفون با الله كذباً وباطلاً ليقنعون الناس بأنهم من المسلمين بينما ليسوا كذلك.

Retrieved Documents



Enter your query - Arabic :



من هو المنافقون؟

# Interface Streamlit

من هم اصحاب الكهف

اصحاب الكهف هم شباب مؤمنون فروا إلى الكهف لتجنب فتن قومهم وعبادة الأصنام، وطلب الرحمة والرشاد من الله تعالى.

Retrieved Documents

Document 1 (Score: 0.433)

Content:

الآية: [وتحسبهم أيقاظاً وهم رقود ونقلبهم ذات اليمين ذات الشمال وكلبهم باسط ذراعيه بالوصيد لو اطلعت عليهم لوليت منهم فرارا ولملئت منهم رعبا] التفسير: وتظن -أيها الناظر- أهل الكهف أيقاظاً، وهم في الواقع نائمون، ونتعهد لهم بالرعاية، فقلّبهم حال نومهم مرة للجانب الأيمن ومرة للجانب الأيسر؛ لئلا تأكلهم الأرض، وكلبهم الذي صاحبهم ماد ذراعيه بفناء الكهف، لو عاينتهم لأدبرت عنهم هارباً، ولم يلتفت نفسك

Enter your query - Arabic : >

من هم أصحاب الكهف؟

## Interface Streamlit

# RAG-based Quran Chatbot

لمن تعطى الصدقات

تعطى الصدقات للقراء والمساكين والعاملين عليها والمؤلفة قلوبهم وفي الرقاب والغارمين وفي سبيل الله وابن السبيل

Retrieved Documents

Document 1 (Score: 0.377)

Similarity Score: 0.377

Content:

الآية: إنما الصدقات للقراء والمساكين والعاملين عليها والمؤلفة قلوبهم وفي الرقاب والغارمين وفي سبيل الله وابن السبيل فريضة من الله والله عالم حكيم التفسير: إنما تعطى الزكوات الواجبة للمحتاجين الذين لا يملكون شيئاً، وللمساكين الذين لا يملكون كفايتهم، وللسعاة الذين

لمن تعطى الصدقات؟

# Interface Streamlit



ما أهمية الصبر في مواجهة المصاعب في القرآن؟



:أهمية الصبر في مواجهة المصاعب في القرآن تتجلى في عدة نقاط

1. الصبر هو سلاح المؤمن في وجه الأذى والإساءة.
2. الصبر يؤدي إلى الفوز برضوان الله ونجاح في الدنيا والآخرة.
3. الصبر يساعد على الانتصار على الأعداء وعدم الاستسلام لهم.
4. الصبر يجعل الإنسان يثق في الله ويستعيض عنه في الشدة.
5. الصبر يؤدي إلى العفو عن الأذى وتجاوزه، وهو فعل حميد.

فالصبر في القرآن الكريم هو طريق المؤمنين نحو النصر والسعادة.

Retrieved Documents

Retrieved Documents

Document 1 (Score: 0.406)

Content:

الآية: [فاصبر صبرا جميلا] التفسير: فاصبر -أيها الرسول- على استهزائهم واستعجالهم العذاب، صبرا لا جزع فيه، ولا شكوى منه لغير الله.

Metadata:

{السورة: 'اسورة المعارج' ، رقم الآية: 5}

Document 2 (Score: 0.431)

Content:

الآية: [ولمن صبر وغفر إن ذلك لمن عزم الأمور] التفسير: ولمن صبر على الأذى، وقابل الإساءة بالعفو والصفح والستر، إن ذلك من

Enter your query - Arabic :

# Interface Streamlit



ما هي الحكمة من فريضة الصيام وفقاً للقرآن؟



وفقاً للقرآن، فرض الصيام لعل الناس يتقوّن الله. وهذا يعني أن الصيام يجعل الناس يتجنّبون المعاصي ويحافظون على طاعة الله وعبادة الله وحده.

## Retrieved Documents



### Document 1 (Score: 0.420)

#### Content:

الآية: [يا أيها الذين آمنوا كتب عليكم الصيام كما كتب على الذين من قبلكم لكم تتقون] التفسير: يا أيها الذين صدّقوا الله ورسوله وعملوا بشرعه، فرض الله عليكم الصيام كما فرضه على الأمم قبلكم؛ لكم تتقون ربكم، فتجلعون بينكم وبين المعاصي وقایة بطاعته وعبادته وحده.

Enter your query - Arabic :



# **Conclusion:**



**Recherche Syntaxique avec Elasticsearch**



**Recherche Sémantique avec le Système RAG**

## **Perspectives:**



**Amélioration des performances grâce aux ressources GPU**



**Combinaison des approches syntaxique et sémantique**



**Evaluation approfondie**

MERCI POUR  
VOTRE ATTENTION !