

PROJET INFORMATION RETRIEVAL

RECHERCHE SÉMANTIQUE DANS LE SAINT CORAN



BUSINESS INTELLIGENCE & ANALYTICS

Réalisé et présenté par:

- *Chamss Doha BEKKAI*
- *Yahya BELGHITI-ALAOUI*
- *Amin BENALI*
- *Oumaima GHAZOUAN*
- *Omar YOUB*

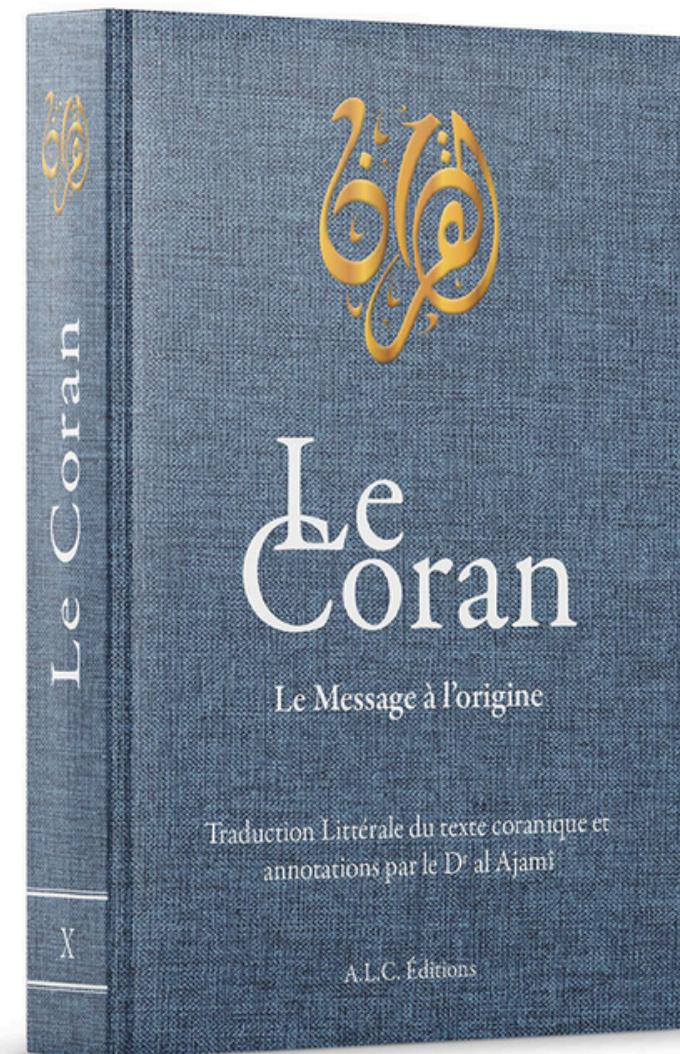
Supervisé par *Pr. Nouredine KERZAZI*

PLAN

- 1 INTRODUCTION**
- 2 CONTEXTE ET OBJECTIFS DU PROJET**
- 3 CONCEPTION DU SYSTEME**
- 4 CHOIX DES TECHNOLOGIES**
- 5 IMPLEMENTATION**
- 6 REQUÊTES ET RÉSULTATS**
- 7 CONCLUSION**



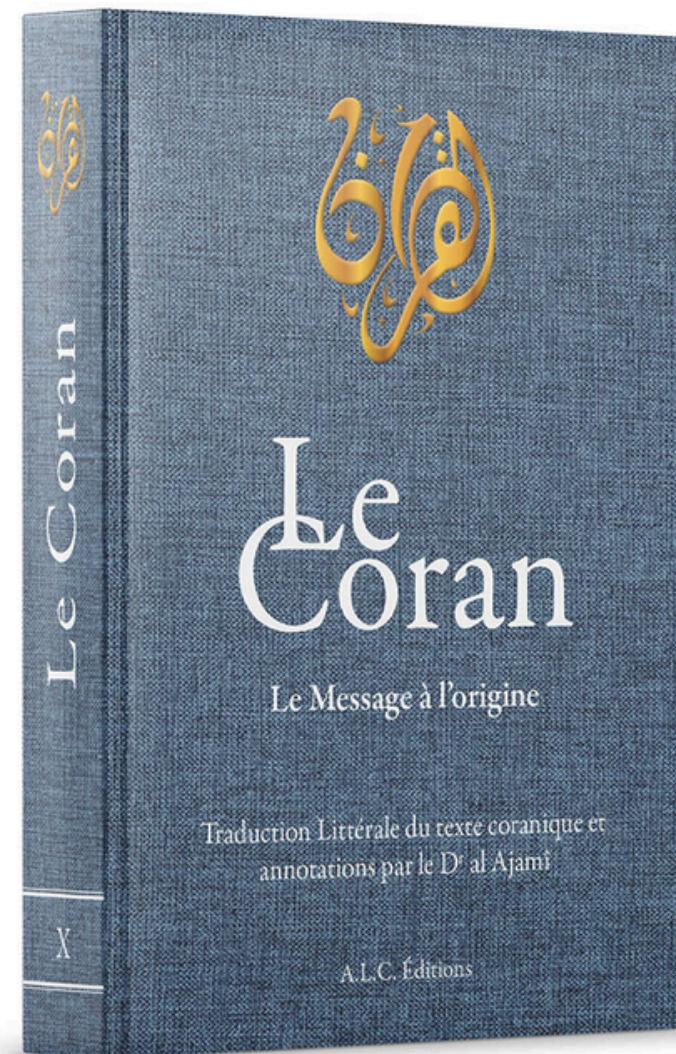
INTRODUCTION



Dans le cadre de notre projet, nous avons conçu et développé un moteur de recherche innovant, spécifiquement dédié aux textes fondamentaux de l'islam, à savoir le Coran et les Hadiths. Ce moteur ne se limite pas à une simple recherche textuelle, mais intègre également une capacité avancée de génération de réponses contextuelles, permettant aux utilisateurs d'obtenir des informations précises et adaptées à leurs requêtes.



INTRODUCTION



Notre objectif principal était de créer une solution qui offre une expérience utilisateur intuitive, rapide et efficace, répondant aux besoins d'exploration et de compréhension approfondie de ces textes riches et complexes.



CONTEXTE DU PROJET

01

La numérisation des textes religieux accentue le besoin d'outils performants pour des recherches textuelles précises et rapides.



CONTEXTE DU PROJET



01

La numérisation des textes religieux accentue le besoin d'outils performants pour des recherches textuelles précises et rapides.

02

La langue arabe pose des défis uniques : structure complexe, richesse sémantique, et nuances d'interprétation



CONTEXTE DU PROJET



- 01 La numérisation des textes religieux accentue le besoin d'outils performants pour des recherches textuelles précises et rapides.
- 02 La langue arabe pose des défis uniques : structure complexe, richesse sémantique, et nuances d'interprétation
- 03 Les systèmes actuels génèrent souvent des réponses approximatives, notamment pour les requêtes en arabe ou liées à la religion.



CONTEXTE DU PROJET



- 01 La numérisation des textes religieux accentue le besoin d'outils performants pour des recherches textuelles précises et rapides.
- 02 La langue arabe pose des défis uniques : structure complexe, richesse sémantique, et nuances d'interprétation
- 03 Les systèmes actuels génèrent souvent des réponses approximatives, notamment pour les requêtes en arabe ou liées à la religion.
- 04 Ces enjeux nécessitent des technologies avancées comme le RAG et des techniques adaptées de prétraitement linguistique.



OBJECTIFS DU PROJET



01

Concevoir une infrastructure robuste pour gérer et rechercher efficacement les données.



OBJECTIFS DU PROJET



- 01 Concevoir une infrastructure robuste pour gérer et rechercher efficacement les données.
- 02 Préparer les données avec des représentations vectorielles adaptées.



OBJECTIFS DU PROJET



- 01 Concevoir une infrastructure robuste pour gérer et rechercher efficacement les données.
- 02 Préparer les données avec des représentations vectorielles adaptées.
- 03 Intégrer un modèle de langage pour des réponses contextuelles précises.



OBJECTIFS DU PROJET



- 01 Concevoir une infrastructure robuste pour gérer et rechercher efficacement les données.
- 02 Préparer les données avec des représentations vectorielles adaptées.
- 03 Intégrer un modèle de langage pour des réponses contextuelles précises.
- 04 Créer une interface utilisateur intuitive et efficace.



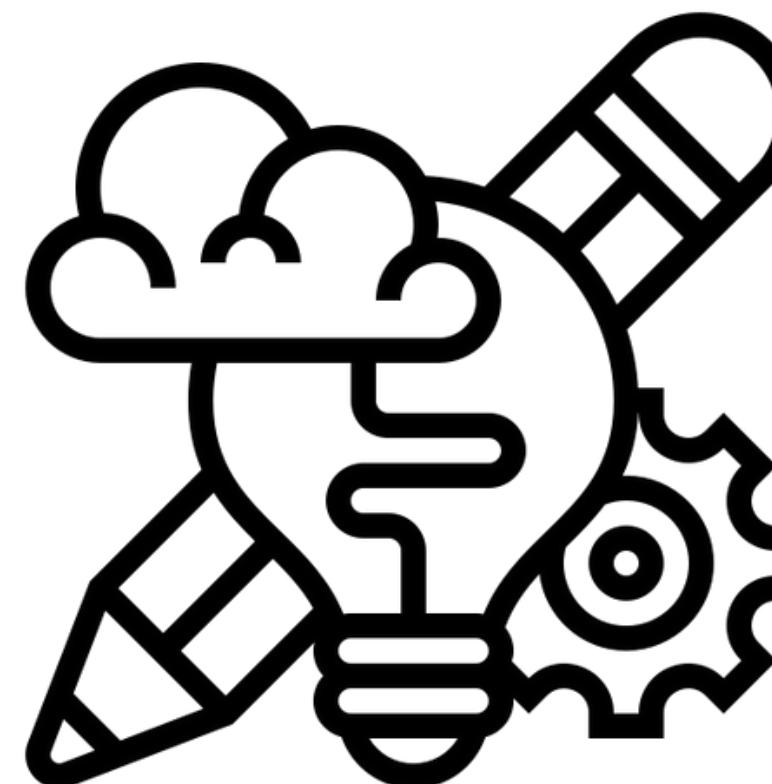
OBJECTIFS DU PROJET



- 01 Concevoir une infrastructure robuste pour gérer et rechercher efficacement les données.
- 02 Préparer les données avec des représentations vectorielles adaptées.
- 03 Intégrer un modèle de langage pour des réponses contextuelles précises.
- 04 Créer une interface utilisateur intuitive et efficace.
- 05 Assurer l'évolutivité et la facilité d'installation pour de futures améliorations.



CONCEPTION DU SYSTÈME



01

Besoins Fonctionnels

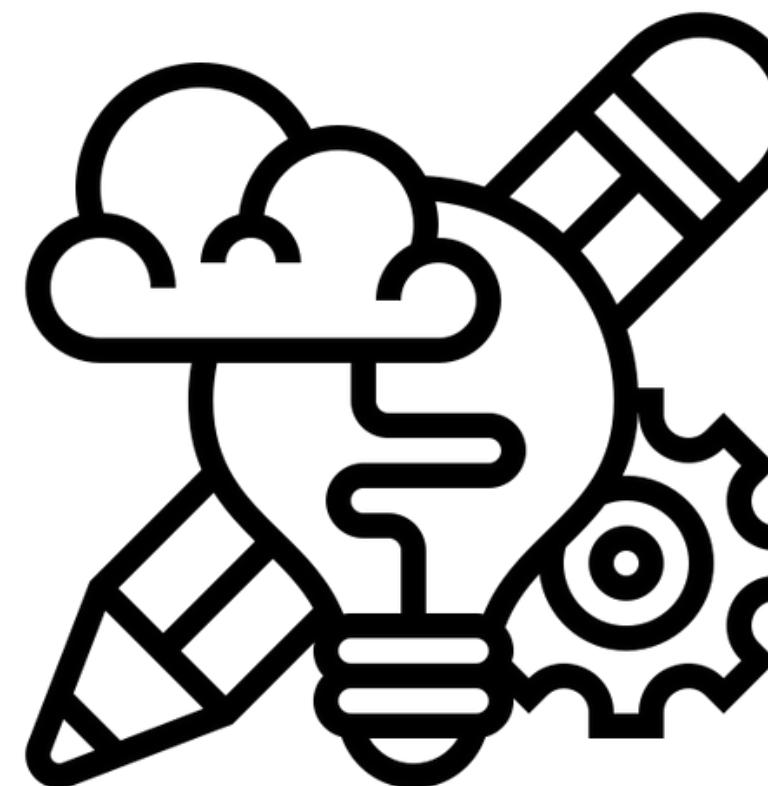
- Permettre à l'utilisateur de choisir une source de données (Coran ou Hadiths).
- Reformuler automatiquement les requêtes pour plus de clarté et de précision grâce à un LLM.
- Extraire les passages les plus pertinents en fonction des requêtes.
- Générer des réponses contextuelles basées sur les données extraites.
- Afficher les résultats de manière claire et intuitive via une interface conviviale.



02

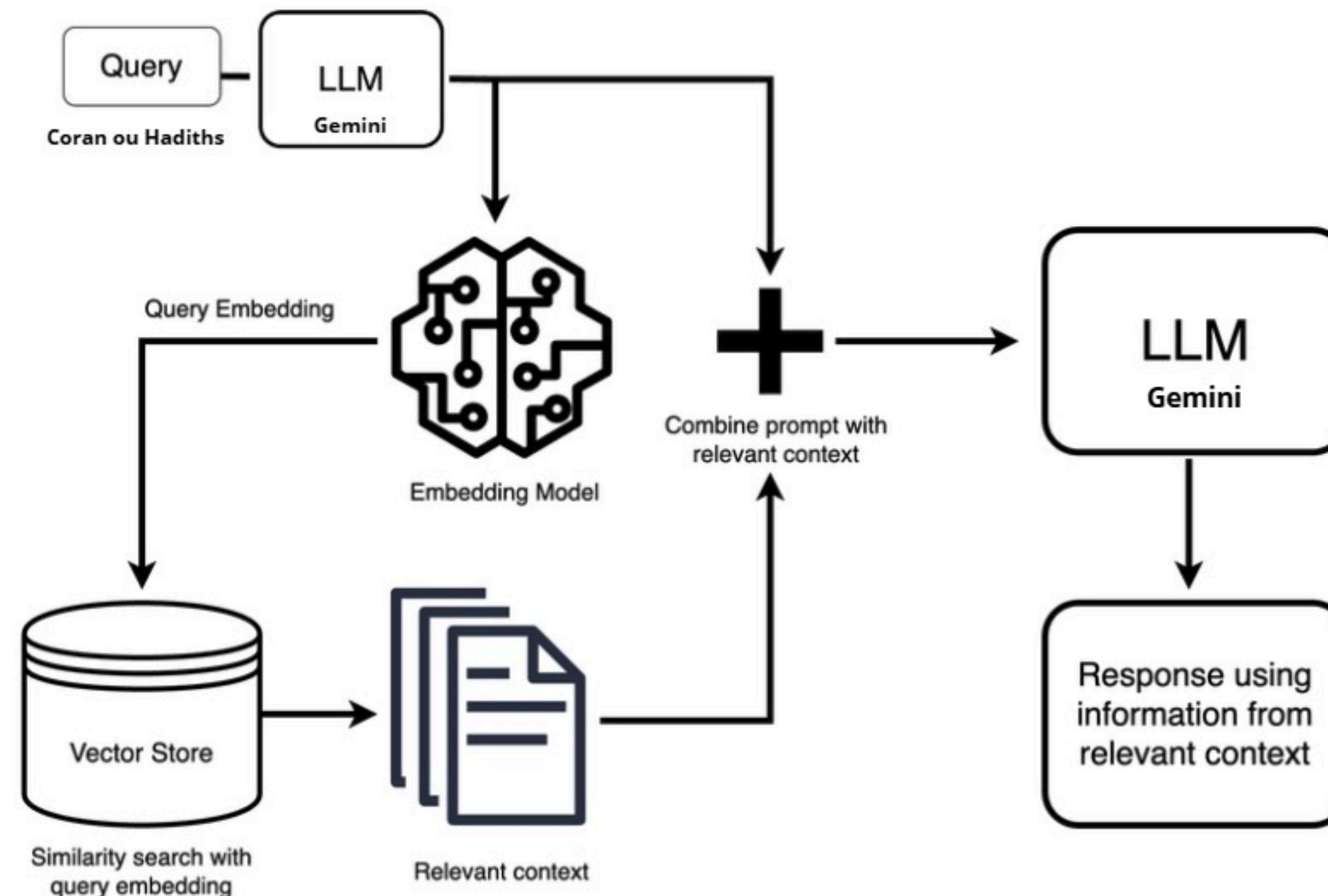
Besoins Non Fonctionnels

CONCEPTION DU SYSTÈME



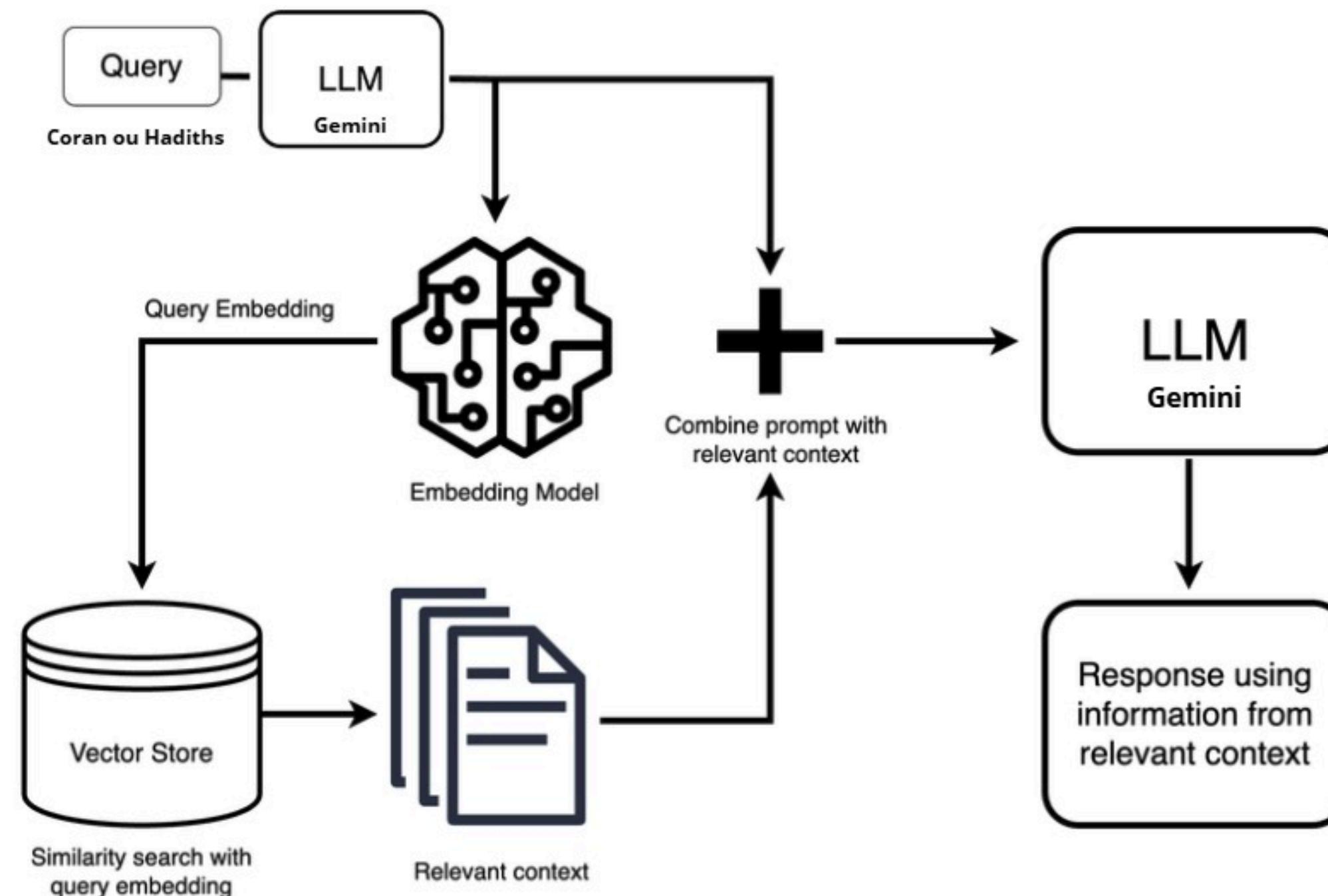
- Permettre le choix de la source (Coran ou Hadiths).
- Améliorer les requêtes avec un LLM.
- Extraire les informations pertinentes.
- Générer des réponses contextuelles.
- Afficher les résultats de façon intuitive.





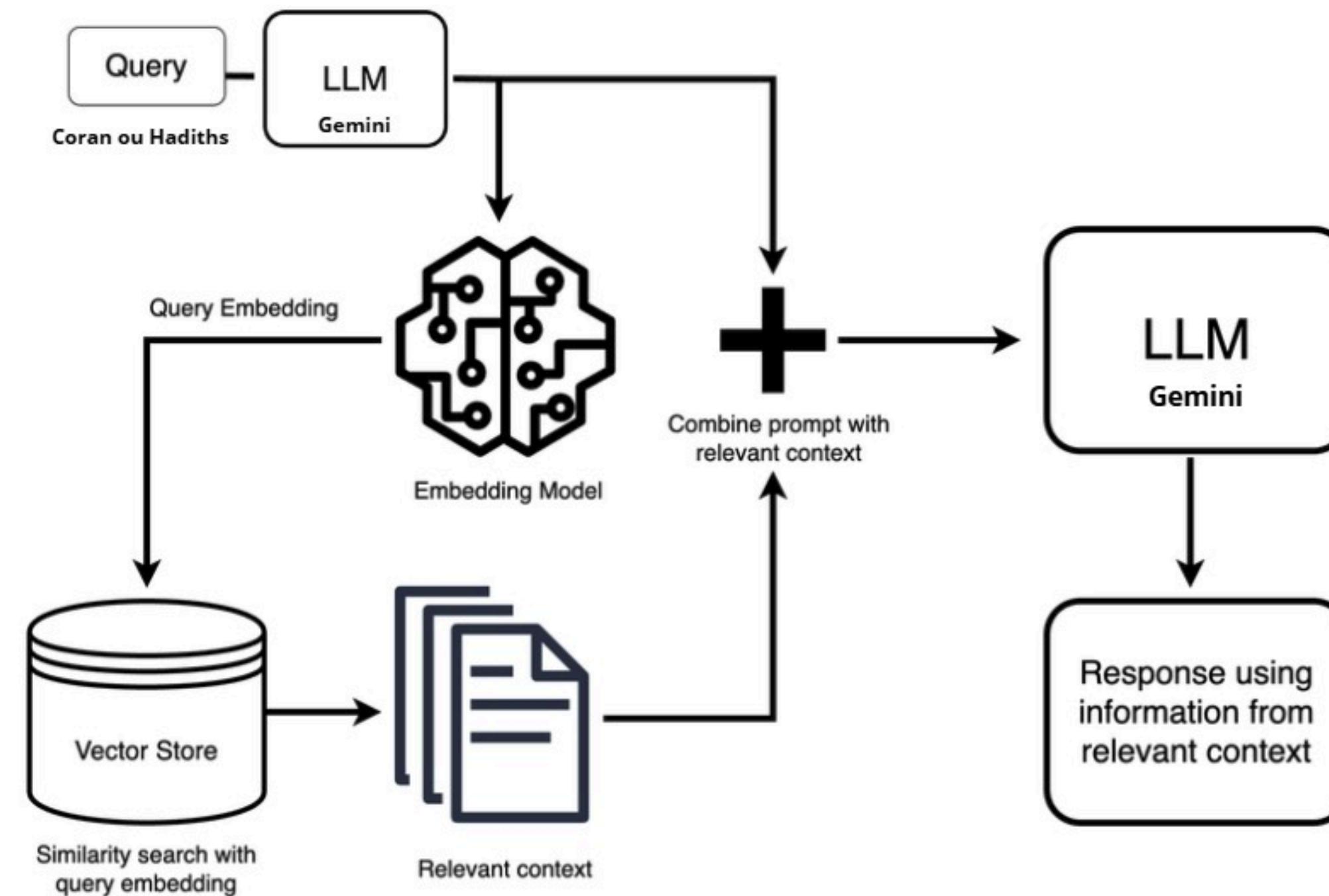
L'utilisateur saisit une requête et sélectionne une source (Coran ou Hadiths).





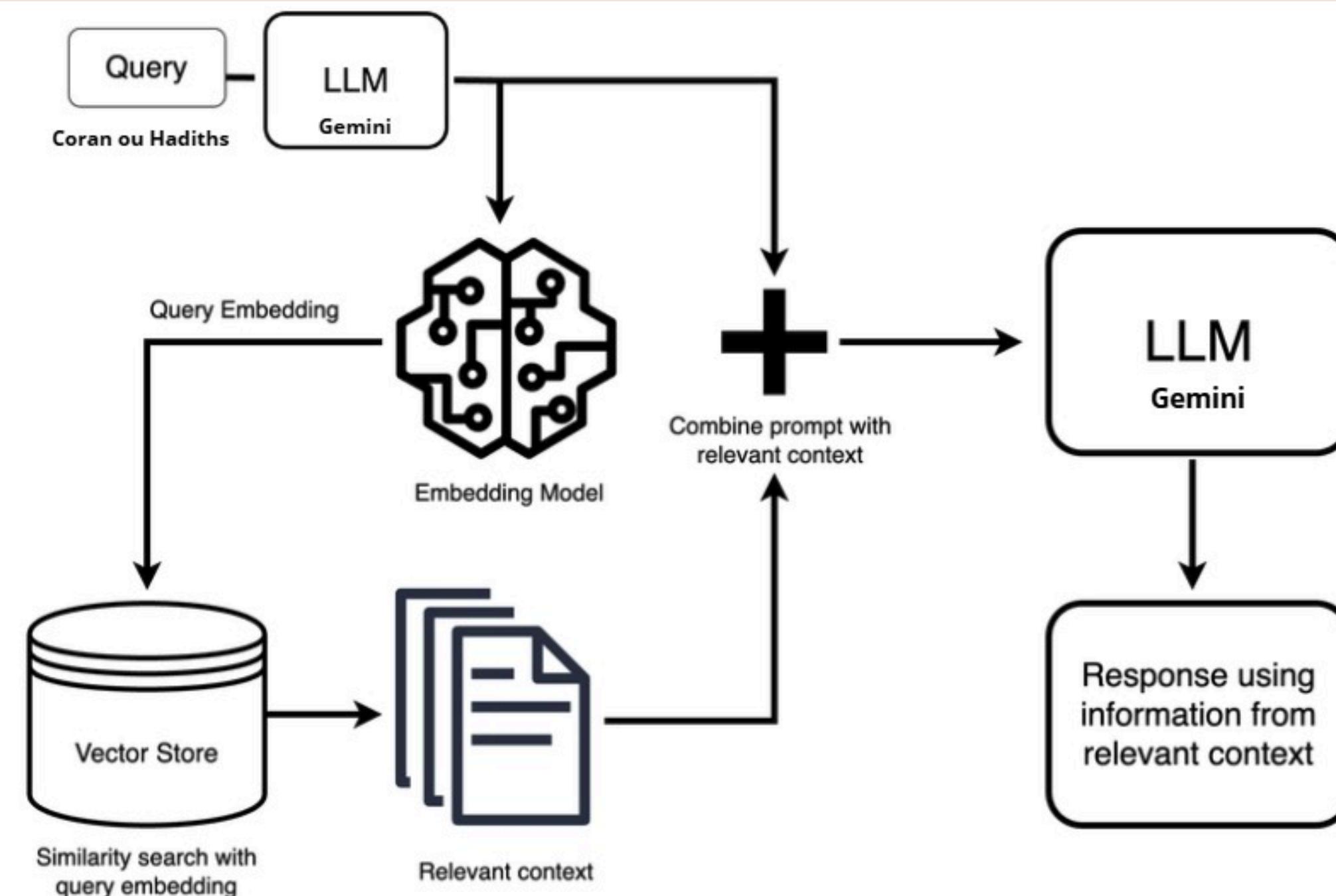
La requête est reformulée pour plus de clarté et de précision.





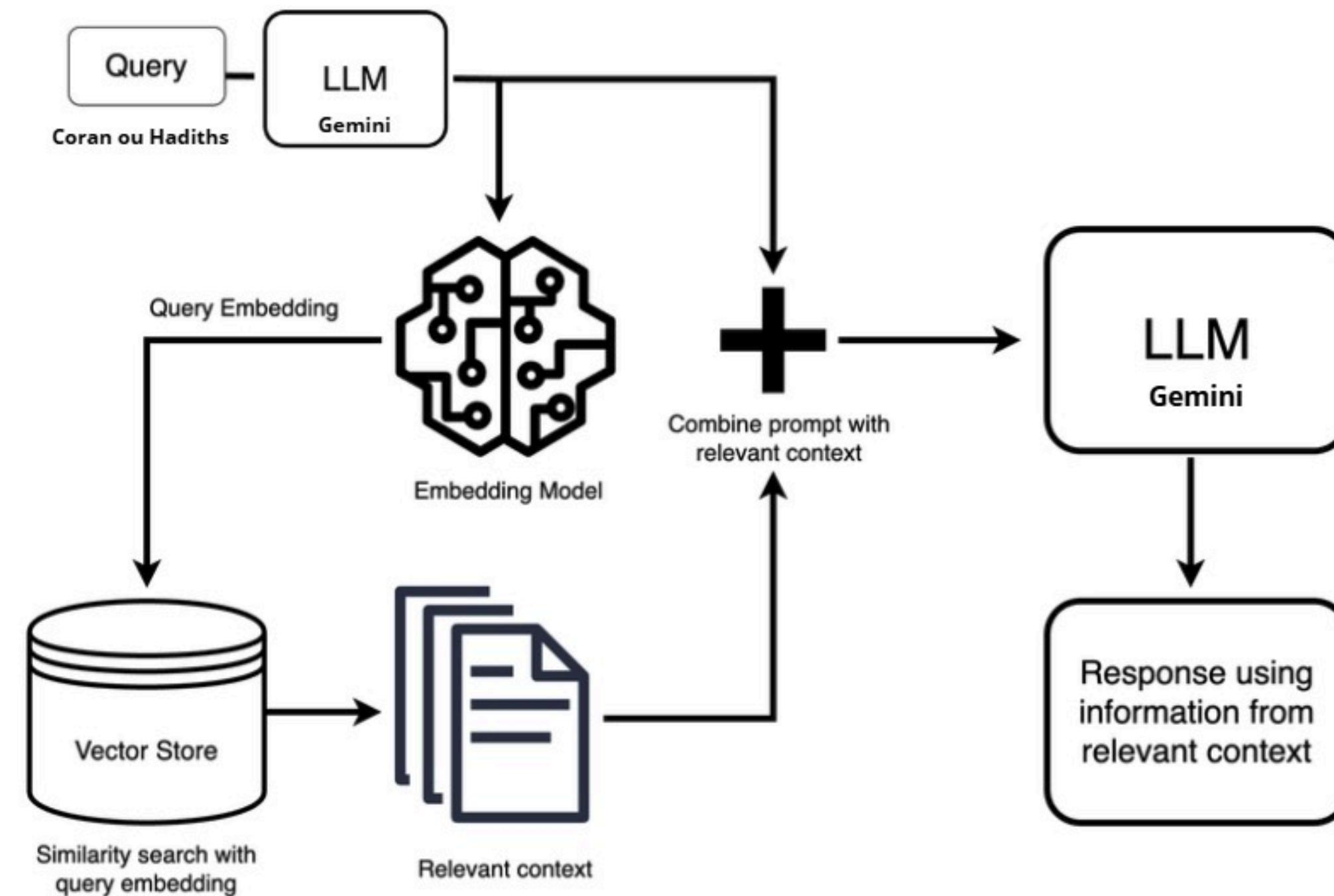
Elle est transformée en vecteur numérique pour comparaison.





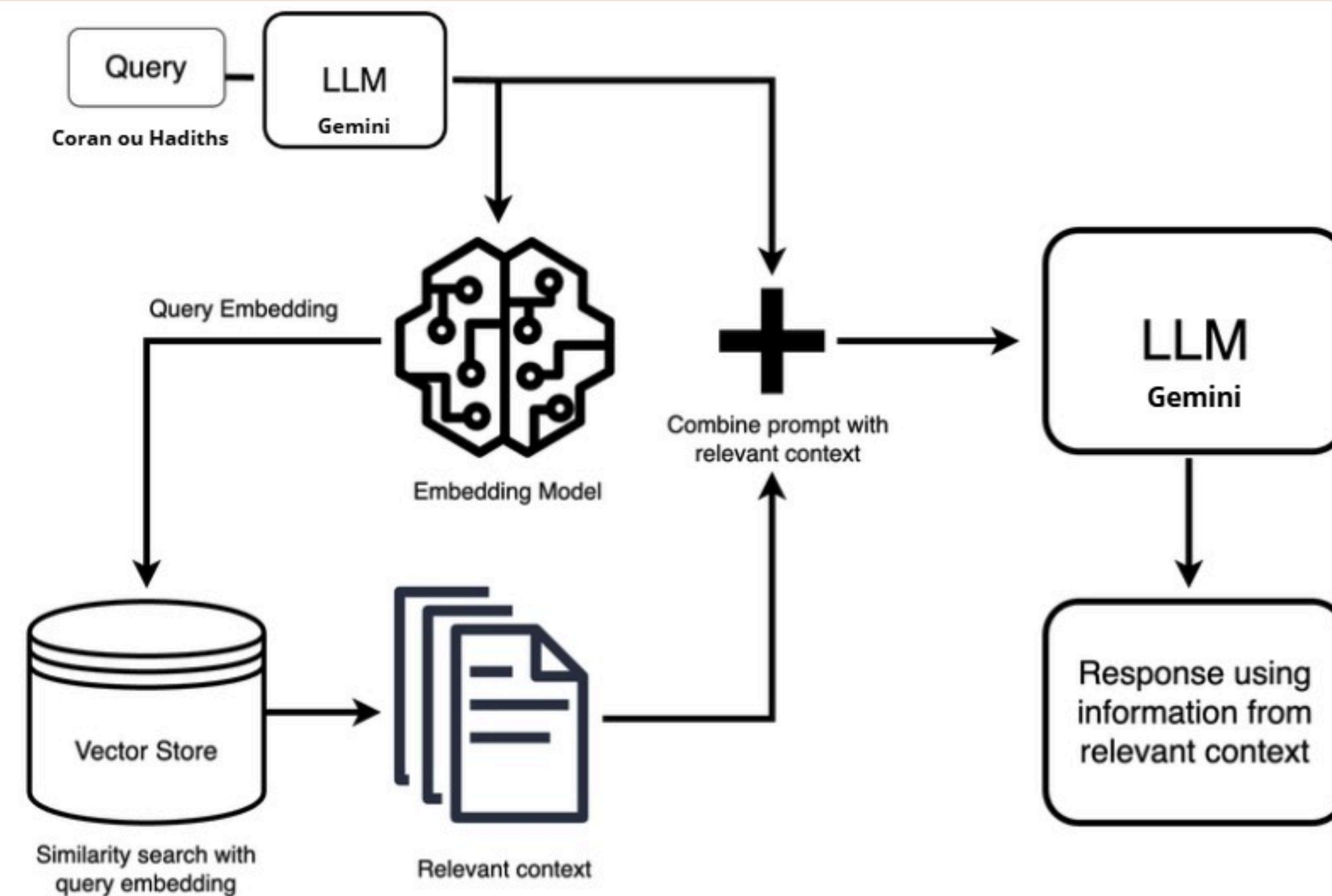
Les passages pertinents sont extraits de la base de données.





Une réponse contextuelle est générée en langage naturel.





La réponse est affichée clairement, avec la source indiquée.



MODÈLE D'EMBEDDING

- Utilisation du modèle **intfloat/multilingual-e5-large** pour générer des embeddings adaptés aux textes arabes.
- Performance démontrée dans l'article "Semantic Embeddings for Arabic Retrieval Augmented Generation (ARAG)".
- <https://thesai.org/Publications/ViewPaperVolume=14&Issue=11&Code=IJACSA&SerialNo=135>



VECTOR STORE

- Stockage des embeddings dans une base de données vectorielle **Chroma**.
- Chroma offre une gestion efficace des données vectorielles pour une recherche rapide basée sur la similarité.



Chroma



LARGE LANGUAGE MODEL (LLM)

- Utilisation de **Gemini gemini-2.0-flash-exp** via une API gratuite.
- Capacité à gérer plus d'un million de tokens et 10 requêtes par minute.
- Adapté pour générer des réponses en arabe avec précision et fluidité.



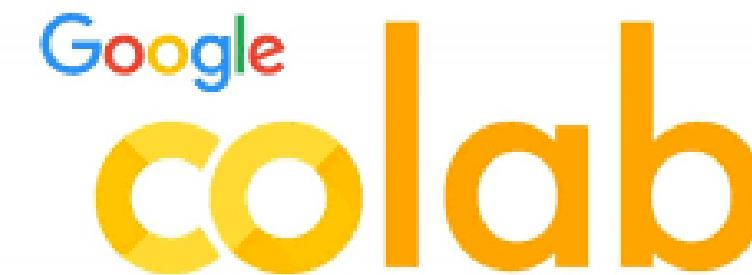
INTERFACE UTILISATEUR

- Développement d'une interface interactive avec **Streamlit**.
- Permet aux utilisateurs de saisir des requêtes, sélectionner les données (Coran ou Hadiths), et afficher les résultats (contexte et réponse).



ENVIRONNEMENT DE DÉVELOPPEMENT

- Utilisation de **Google Colab** pour héberger le serveur **FastAPI** et exécuter les étapes du pipeline.
- Intégration de **ngrok** pour générer une URL publique pour l'accès au serveur.



1

PRÉTRAITEMENT DES DONNÉES

- Les données utilisées proviennent des versets coraniques et des paroles du Prophète Mohammed (paix et bénédictions sur lui).
- Ces données ont été séparées en deux documents indépendants :
 - **quran.json** : Contenant les données du Coran.
 - **json-hadit.json** : Contenant les données des hadiths.
- Le prétraitement ayant été réalisé lors d'un projet précédent, les fichiers sont directement exploitables.



2

ALIMENTATION

- Les fichiers prétraités sont intégrés dans un **Chroma vector store**.
- Deux collections sont créées :
 - **Collection des embeddings du Coran** : Basée sur le champ **aya_chakI** dans [quran.json](#).
 - **Collection des embeddings des hadiths** : Basée sur le champ **text_ar** dans [json-hadit.json](#).
- Les embeddings sont générés en utilisant le modèle **intfloat/multilingual-e5-large**.



3

LE FLUX D'EXÉCUTION

REQUÊTE UTILISATEUR :

- La requête est saisie via l'interface et envoyée au serveur FastAPI hébergé sur Google Colab.
- L'endpoint /query traite la requête en l'améliorant grâce à un modèle LLM (Gemini)

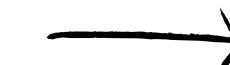


3

LE FLUX D'EXÉCUTION

REQUÊTE UTILISATEUR :

- La requête est saisie via l'interface et envoyée au serveur FastAPI hébergé sur Google Colab.
- L'endpoint /query traite la requête en l'améliorant grâce à un modèle LLM (Gemini)



TRAITEMENT DE LA REQUÊTE :

- Amélioration de la requête.
- Recherche de similarité cosinus entre l'embedding de la requête et les embeddings des collections.
- Appel à l'endpoint du LLM pour générer une réponse.

- Requête d'amélioration:

```
prompt = f"enhance this user query in arabic for better information retrieval in arabic,"
```

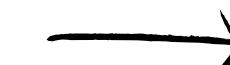
```
, and give only the enhanced query : \n{data.query}"
```

- Prompt LLM:

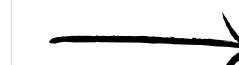
```
prompt = f":المعلومات المقدمة\nالسؤال :\nالجواب :\n:الجواب\n" :الجواب
```

3**LE FLUX D'EXÉCUTION****REQUÊTE UTILISATEUR :**

- La requête est saisie via l'interface et envoyée au serveur FastAPI hébergé sur Google Colab.
- L'endpoint /query traite la requête en l'améliorant grâce à un modèle LLM (Gemini)

**TRAITEMENT DE LA REQUÊTE :**

- Amélioration de la requête.
- Recherche de similarité cosinus entre l'embedding de la requête et les embeddings des collections.
- Appel à l'endpoint du LLM pour générer une réponse.

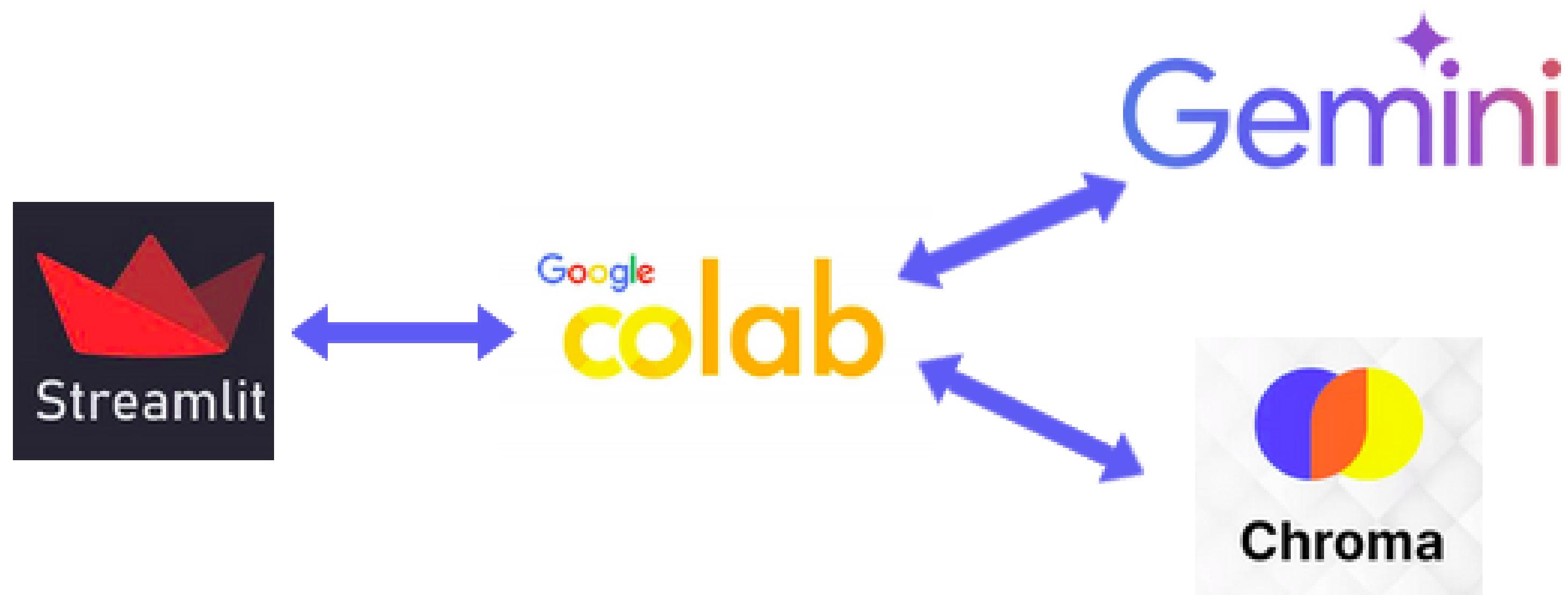
**RÉSULTAT :**

La réponse générée par le LLM, ainsi que le contexte récupéré, sont retournés et affichés dans l'interface.



3

LE FLUX D'EXÉCUTION



4

INTERFACE UTILISATEUR

- L'interface utilisateur est développée avec **Streamlit** pour être intuitive.
- Fonctionnalités principales :
 - Permettre le choix entre la collection Coran ou Hadith.
 - Affichage **organisé** en deux colonnes :
 - **Contexte** : Affiche le texte récupéré.
 - **Réponse générée** : Affiche la réponse fournie par le LLM.



SCREENSHOTS DE L'INTERFACE UTILISATEUR

Chatbot

Choose the source to search:

- Quran
- Hadith

ما قولك عن الجن

Retrieved Results

بالتأكيد، سأجيب على أسئلتك حول مفهوم الجن ومعتقدات الناس حولهم في الثقافة الإسلامية والعربية، مع الأخذ في الاعتبار المعلومات المقدمة.

Generated Response

بالتأكيد، سأجيب على أسئلتك حول مفهوم الجن ومعتقدات الناس حولهم في الثقافة الإسلامية والعربية، مع الأخذ في الاعتبار المعلومات المقدمة.

مفهوم الجن في الإسلام

الجن في الإسلام هم مخلوقات غير مرئية، خلقت من نار، ولها عقل وإرادة، مكلفة ومسئولة أمام الله مثل الإنس. وهم ليسوا بالضرورة أشراراً، فمنهم الصالح ومنهم الطالح، كما أنهم ليسوا شياطين بالضرورة. فالشيطان هو جن عاصٍ

What is your question? >

ما قولك عن الجن

Retrieved Results

[{"source": " صحيح البخاري ", "chapter_no": "59", "chapter": "Beginning of Creation - كتاب بدء الخلق", "text_ar": "حدثنا يحيى بن جعفر، حدثنا محمد بن عبد الله الأنصاري، حدثنا ابن حميد، قال أخبرني عطاء، عن جابر رضي الله عنه عن النبي صلى الله عليه وسلم قال إذا استجogn الليل أو كان جن الليل فكفوا صبيانكم، فإن الشياطين تتنشر حينئذ، فإذا ذهب ساعة من العشاء فحلوهم وأغلق يابك، واذكر اسم الله، وأطفئ مصباحك، واذكر اسم الله، وأوك سقائك، واذكر اسم الله، وخرم إبائك، واذكر اسم الله، ولو تعرض عليه شيئاً"}, {"source": " سنن النسائي ", "chapter_no": "9", "chapter": "The Book of the Qiblah - كتاب القبلة", "text_ar": "أخبرنا عمرو بن علي، قال أتيانا يزيد، قال"}]

Generated Response

بالتأكيد، سأجيب على أسئلتك حول مفهوم الجن ومعتقدات الناس حولهم في الثقافة الإسلامية والعربية، مع الأخذ في الاعتبار المعلومات المقدمة.

مفهوم الجن في الإسلام

الجن في الإسلام هم مخلوقات غير مرئية، خلقت من نار، ولها عقل وإرادة، مكلفة ومسئولة أمام الله مثل الإنس. وهم ليسوا بالضرورة أشراراً، فمنهم الصالح ومنهم الطالح، كما أنهم ليسوا شياطين بالضرورة. فالشيطان هو جن عاصٍ

What is your question? >



النار في الإسلام**Informations Récupérées (3 exemples)**

1. **Source :** سنن ابن ماجه

Chapitre : كتاب اللقطة

، حدثنا محمد بن المثنى، حدثنا يحيى بن سعيد، عن حميد الطويل

عن الحسن، عن مطرف بن عبد الله بن الشخير، عن أبيه، قال قال رسول الله صلى الله عليه وسلم ضالة المسلم حرق النار

2. **Source :** سنن ابن ماجه

Chapitre : كتاب الطهارة وسنتها

، قال القطان حدثنا أبو حاتم، حدثنا عبد المؤمن بن علي، حدثنا عبد السلام بن حرب

عن هشام بن عمرو، عن أبيه، عن عائشة، قالت قال رسول الله صلى الله عليه وسلم ويل للأععقاب من النار

3. **Source :** سنن الترمذ

Chapitre : كتاب تحريم الدم

، أخبرني محمد بن إسماعيل بن إبراهيم، عن يزيد، عن سليمان التيمي، عن الحسن

عن أبي موسى، عن النبي صلى الله عليه وسلم قال إذا تواجه المسلمين بسيفيهم فقتل أحدهما صاحبه فهما في النار

قبل يا رسول الله هذا القاتل بما بال المقتول قال أراد قتل صاحبه

Texte Généré par le LLM

النار في الإسلام هي مكان للعذاب الأليم الذي أعده الله تعالى للكافرين والعصاة من المسلمين.

الأحاديث تذكرها كعقاب شديد و"حرق" أو "ويل" لمن يستحقها. فالنار هي المصير النهائي للكافرين الذين أشركوا بالله ولم يؤمنوا به.

أما المسلمين العصاة الذين يرتكبون الذنوب الكبيرة، فقد يعذبون في النار لفترة معينة، ثم يدخلون الجنة بعد التطهير، أو قد يغفر الله لهم بعفوه ورحمته.

REQUÊTE ENFER HADITH

الجنة في الإسلام : Requête Hadith

Informations Récupérées (2 exemples)

1. **Source :** صحيح مسلم

كتاب الجنة وصفة نعيمها وأهلها

، حدثني زهير بن حرب، حدثنا عبد الرحمن بن مهدي

، حدثنا حماد بن سلمة، عن ثابت، عن أبي رافع، عن أبي هريرة

عن النبي صلى الله عليه وسلم قال من يدخل الجنة ينعم لا يبأس لا تبلى ثيابه ولا يفني شبابه

2. **Source :** صحيح البخاري

كتاب التوحيد

، حدثنا أبو اليهان، أخبرنا شعيب، حدثنا أبو الزناد، عن الأعرج، عن أبي هريرة

أن رسول الله صلى الله عليه وسلم قال إن لله تسعة وتسعين اسمًا مائة إلا واحداً، من أحصاها دخل الجنة

Texte Généré par le LLM (seulement le premier paragraphe)

الجنة في الإسلام هي دار الخلود والنعيم الأبدي التي أعدها الله تعالى لعباده المؤمنين

حيث ينعمون بلا بؤس ولا يشيخون أبداً. من يدخل الجنة ينعم لا يبأس ولا تبلى ثيابه ولا يفني شبابه

وقد ذكر النبي صلى الله عليه وسلم أن من أحصى أسماء الله الحسنى دخل الجنة

REQUETE PARADIS HADITH



الجنة في الإسلام
Informations Récupérées (exemples)

2. Retrieved Results

Sura: (Surah 22, Verse 2630) الحج

Aya: الذين إذا ذكر الله وجلت قلوبهم والصابرين على ما أصابهم والمقيمه الصلاة ومما رزقناهم ينفقون

Sura: (Surah 29, Verse 3398) العنكبوت

Aya: والذين آمنوا وعملوا الصالحات لنبؤتهم من الجنة غرفا تجري من تحتها الأنهرار خالدين فيها: نعم أجر العاملين

Sura: (Surah 57, Verse 5093) الحديد

Aya: إن المصدقون والمصدقات وأقرضوا الله فرضا حسنا يضاعف لهم ولهم أجر كريم:

Sura: (Surah 2, Verse 284) البقرة

Aya: إن الذين آمنوا وعملوا الصالحات وأقاموا الصلاة وآتوا الزكاة لهم أجرهم عند ربهم ولا خوف عليهم ولا هم يحزنون

REQUETE PARADIS QURAN



3. Generated Response

بالتأكيد، سأقدم لك شرحاً مفصلاً حول الجنة في الإسلام، مستنداً إلى المعلومات الواردة في النصوص القرآنية التي قدمتها، وإلى المعرفة العامة حول هذا الموضوع:

صفات الجنة

يمكن تلخيص أبرز صفات الجنة التي وردت في القرآن والسنة كما يلى:

دار الخلد: الجنة هي دار البقاء الأبدى، لا موت فيها ولا فناء. قال تعالى: "خَالِدُونَ فِيهَا أَبَدًا" (النساء: 57).

النعم المقيم: تتميز الجنة بالنعيم الكامل والمتجدد، فلا يوجد فيها أي نقص أو ألم أو حزن. قال تعالى: "لَهُمْ فِيهَا قَائِمُونَ خَالِدُونَ" (الفرقان: 16).

الغرف العالية: ذكر القرآن وجود غرف عالية في الجنة. كما ورد في سورة العنكبوت: "لَبَيْوَنَّهُم مِّنَ الْجَنَّةِ عُرْقًا تَجْرِي مِنْ تَحْنِهَا الْأَنْهَازُ" وهذا يدل على علو مكانتها ورفعة أهلها.

الأنهار الجارية: تجري في الجنة أنهار من ماء غير آسن، ولبن لم يتغير طعمه وخر لذة للشاربين، وعسل مصفي. قال تعالى: "فِيهَا أَنْهَازٌ مِّنْ قَاءِ عَيْنٍ آسِنٍ وَأَنْهَازٌ مِّنْ لَبِنٍ لَّمْ يَتَغَيَّرْ طَعْفَةٌ وَأَنْهَازٌ مِّنْ حَفْرٍ لَّذَّةٌ لِلشَّارِبِينَ وَأَنْهَازٌ مِّنْ عَسْلٍ فُضَّفِّيٌّ" (محمد: 15).

الأشجار والظلال: تتميز الجنة بأشجارها الكثيفة وظلالها الوارفة، وهي أشجار لا تشبه أشجار الدنيا، كما ورد في كثير من النصوص الشرعية.

المساكن الطيبة: تتميز بمساكن طيبة ومريحة، وأهل الجنة يتنقلون بين هذه المساكن كيما شاؤوا.

الجمال والبهاء: هي دار جمال وبهاء، ليس فيها شيء مكروه أو قبيح، بل كلها حسنة ومبهجة للناظرين.

السلام والأمن: لا يوجد في الجنة خوف أو حزن، بل السلام والأمن والطمأنينة. قال تعالى: "لَا يَقْسُطُونَ فِيهَا تَضَبُّ وَقَا هُمْ مِّنْهَا يُفْخَرُونَ" (الحجر: 48).

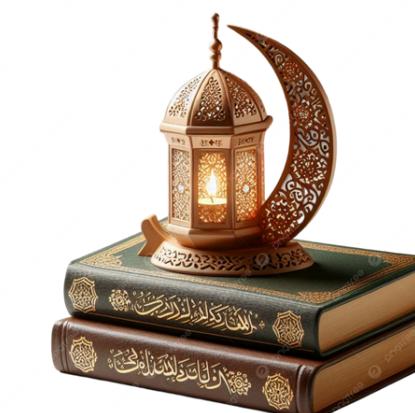
الرؤبة الكريمة: من أعظم النعيم في الجنة هو رؤبة الله تعالى، وهو فضل عظيم يختص به أهل الجنة.

REQUETE ENFER QURAN



CONCLUSION

- Ce projet a abouti à la création d'un système RAG (Retrieval-Augmented Generation) performant, combinant des bases vectorielles et l'IA générative pour offrir des réponses contextuelles et adaptées aux requêtes basées sur les corpus Hadith et Coranique.
- En exploitant les capacités des modèles LLM, ce système permet non seulement une meilleure compréhension des textes sacrés, mais aussi l'exploration de nouvelles perspectives grâce à la génération de réponses inédites.
- Cette approche ouvre des opportunités prometteuses pour l'étude approfondie des textes religieux et le développement d'applications innovantes.





THANK YOU