

Generative AI

Quantum Insights ⚡

Dans les domaines en constante évolution, il est essentiel de comprendre quels sont les principaux enjeux scientifiques du moment. Ce projet vise à explorer les larges langages models accessibles en open-source pour ***synthétiser, résumer, traduire et vulgariser des travaux de recherches***, via la conception d'un modèle RAG spécialisé dans les domaines de la recherche sur la physique quantique.

- Marius Ayrault -

Sorbonne Data Analytics - 2024/2025



Fonctionnalités Clés



Recherche

Extraction d'informations sur les recherches scientifiques depuis ArXiv.org



Synthèse

Exploration de sujets, réalisation de résumés explicatifs, synthèse, comparaison, etc..



Extraction de Code

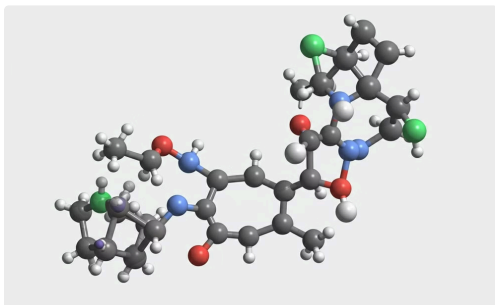
Identification et explication des snippets.



Vulgarisation

Explications adaptées à tous niveaux (de débutant à expert)

Rendre la Science Accessible



Niveau Expert

Détails techniques pour les plus aguerris.



Niveau Avancé

Explications approfondies pour les professionnels du domaine.



Niveau Intermédiaire

Concepts clés pour les non initiés.



Niveau Débutant

Vulgarisation simple pour le grand public.

Use cases

Appropriation

Introduction aux **concepts de base** et aux **principes fondamentaux**.

Explications des Découvertes Historiques et Récentes

Présentation des **percées scientifiques majeures**. Aperçu des **dernières avancées** dans le domaine.

Recherche d'Applications Industrielles

Exploration des **innovations technologiques** découlant de cette science. Identification des **secteurs industriels impactés**.

Apprentissage des Conventions Mathématiques

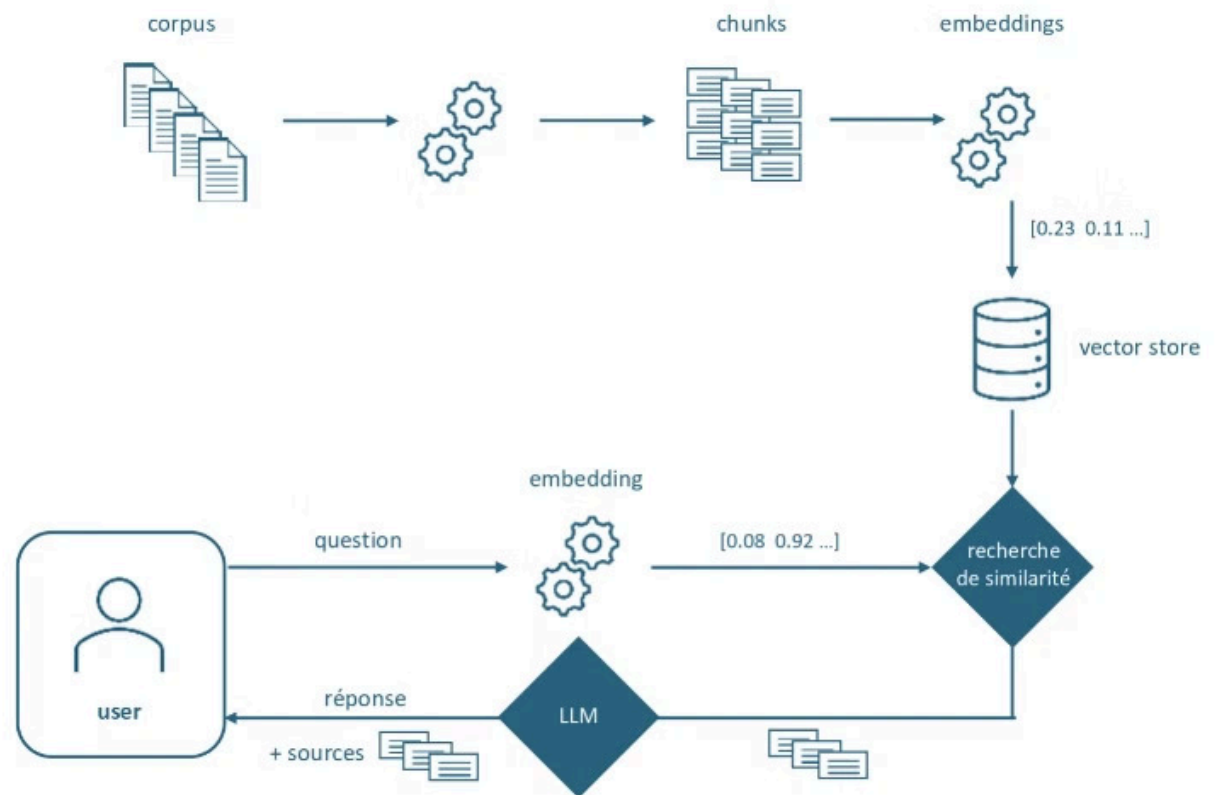
Explication des **notations** et **formules inhérentes** au domaine.

Suggestions de Fragments de Code

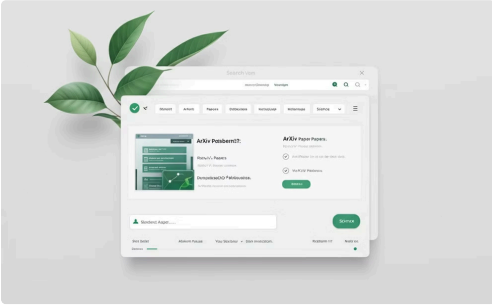
Exemples pratique pour **illustrer les concepts**.

Informations sur **librairies open-source** et service **cloud proposés**.

Méthodologie RAG



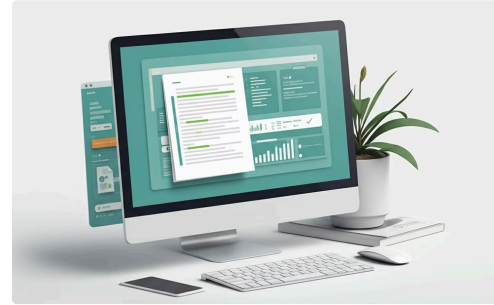
Stack Technologique



Intégration ArXiv

Recherche et récupération des derniers articles scientifiques.

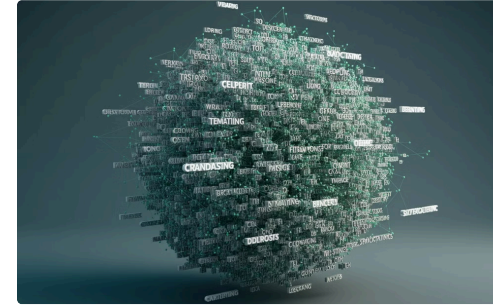
ArXiv API



Traitement de Documents

Détection de structure via NLP

SpaCy `en_core_web_sm` et analyse de sentiment via DistilBERT



Embeddings

Génération d'embeddings de documents.

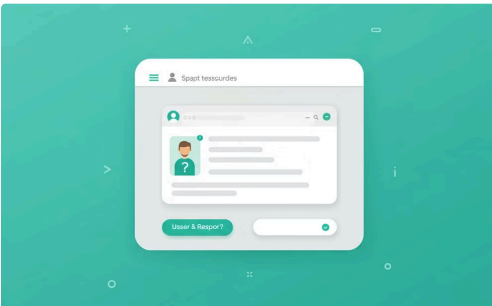
sentence-transformers/
all-MiniLM-L6-v2



Stockage Vectoriel et système RAG

Stockage des embeddings et recherche de similarité.

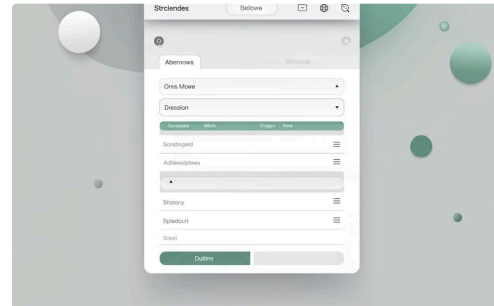
ChromaDB



Text-to-text

Génération de texte via LLM.

Meta/Llama-2-7b-chat-hf



Interface Utilisateur

UI/UX avec pour une interaction fluide. Streamlit



CI/CD

Intégration et déploiement continu. GitHub Actions Docker



Déploiement Cloud

Hébergement et déploiement de l'application. AWS

Remaining work

1

Recherche arXiv

Amélioration du système de recherche d'articles.

(aborder tous les domaines de recherches)

2

Conversation

Amélioration de l'UX.

(user session, context window size, re-prompt ...)

3

Traduction

Evolution du chat en multilingues.

(accessibilité)

4

Appel à contribution

Organisation du référentiel en open-source.

(permettre la contribution)

5

Hébergement

Création d'un nom de domaine dédié.



Lessons Learned

Petits modèles performants

Un modèle de 7 milliards de paramètres peut être très performant avec une bonne intégration.

Fenêtre de contexte

Une méthode fiable pour ajuster la fenêtre de contexte est indispensable.

Communauté open-source

De nombreuses ressources utiles pour les projets GenAI sont accessibles gratuitement.

1

2

3

4

5

Prompt robuste

Un prompt d'instructions bien défini est essentiel pour des résultats de qualité.


Test et intégration continue

Permet de gagner du temps tout au long du projet.

Sources

Repository

mriusero/gen-ai-quantum-insight





1Contributor

0Issues

0Stars

0Forks



 GitHub


GitHub - mriusero/gen-ai-quantum-insight

Contribute to mriusero/gen-ai-quantum-insight development by creating an account on GitHub.

Open-source models


distilbert

/distilbert-base-uncased

 huggingface.co distilbert/distilbert-base-...


sentence-transformers

/all-MiniLM-L6-v2

 huggingface.co sentence-transformers/all-...

meta-llama

/Llama-2-7b-chat-hf

 huggingface.co meta-llama/Llama-2-7b-chat-...