

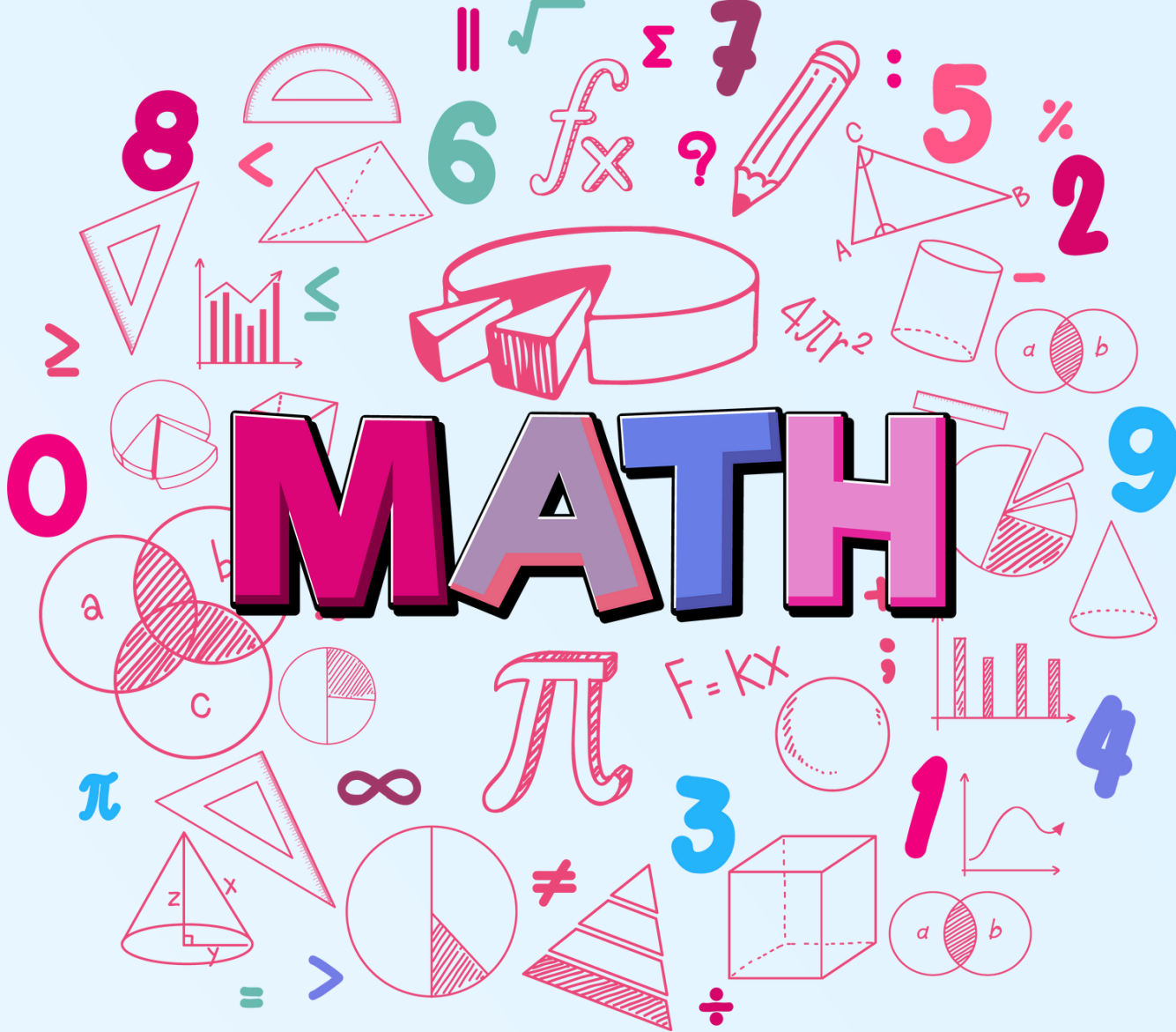
Mettre en place un jeu de données contenant des paires Question-Réponse en mathématiques et des textes explicatifs.

Réaliser par : **BAHRI ILHAME** - **Aldiebes Ghanem Israa** - **Bounadar Douaa**

Demander par : **Pr.EL Habib BENLAHMAR**

Introduction

Dans le cadre de notre projet de fin de module en Deep Learning, nous avons entrepris de répondre à un besoin crucial dans l'éducation : la création de ressources pédagogiques interactives et personnalisées pour les mathématiques au collège. Notre objectif est de développer un dataset riche et structuré de questions-réponses (Q&R), chacune accompagnée d'explications détaillées et de raisonnements pas à pas.



Face à la demande croissante pour des outils de soutien scolaire efficaces, notre approche est résolument innovante, car elle s'appuie sur les dernières avancées en Intelligence Artificielle. Nous visons à construire une base de données fiable et de haute qualité qui servira de pilier à des systèmes intelligents. Ces systèmes permettront non seulement aux élèves de vérifier leurs réponses, mais surtout de comprendre les concepts mathématiques en profondeur et de développer leur capacité de raisonnement de manière autonome. Ce dataset est la pierre angulaire pour faciliter un apprentissage plus interactif et personnalisé des mathématiques.

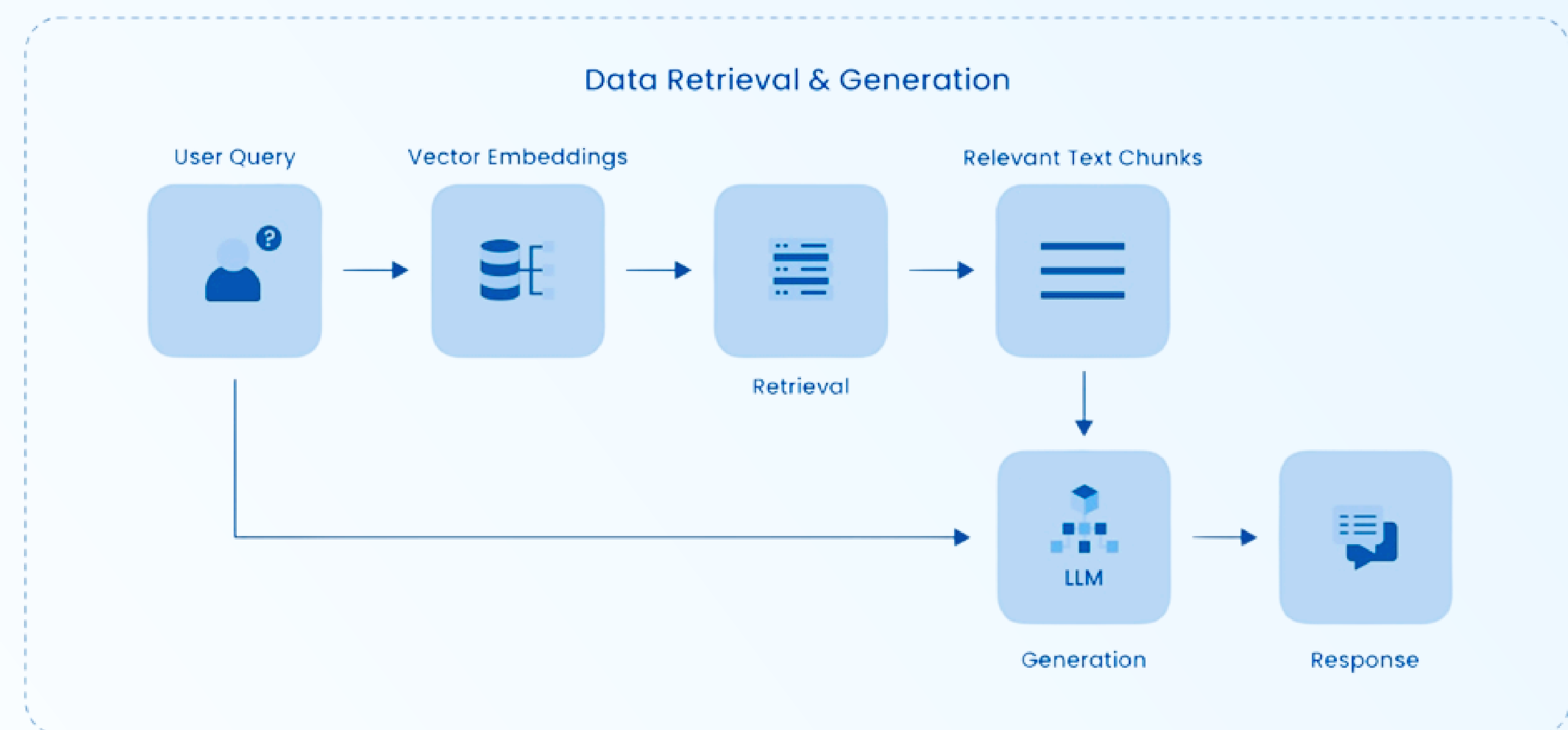
Méthodologie

L'architecture RAG (Retrieval-Augmented Generation)

Pour générer automatiquement des paires Question-Réponse en mathématiques à partir de documents pédagogiques, nous avons mis en œuvre l'architecture RAG, qui combine récupération d'information (IR) et génération de texte (NLG).

Les cours PDF sont d'abord découpés et vectorisés via des embeddings (GoogleGenerativeAIEmbeddings), puis stockés dans une base FAISS. Lorsqu'une requête est soumise (ex. : "Génère 20 questions avec réponses et explications"), les passages les plus pertinents sont extraits et fournis comme contexte à un LLM (ChatGoogleGenerativeAI), permettant une génération contextualisée, précise et traçable.

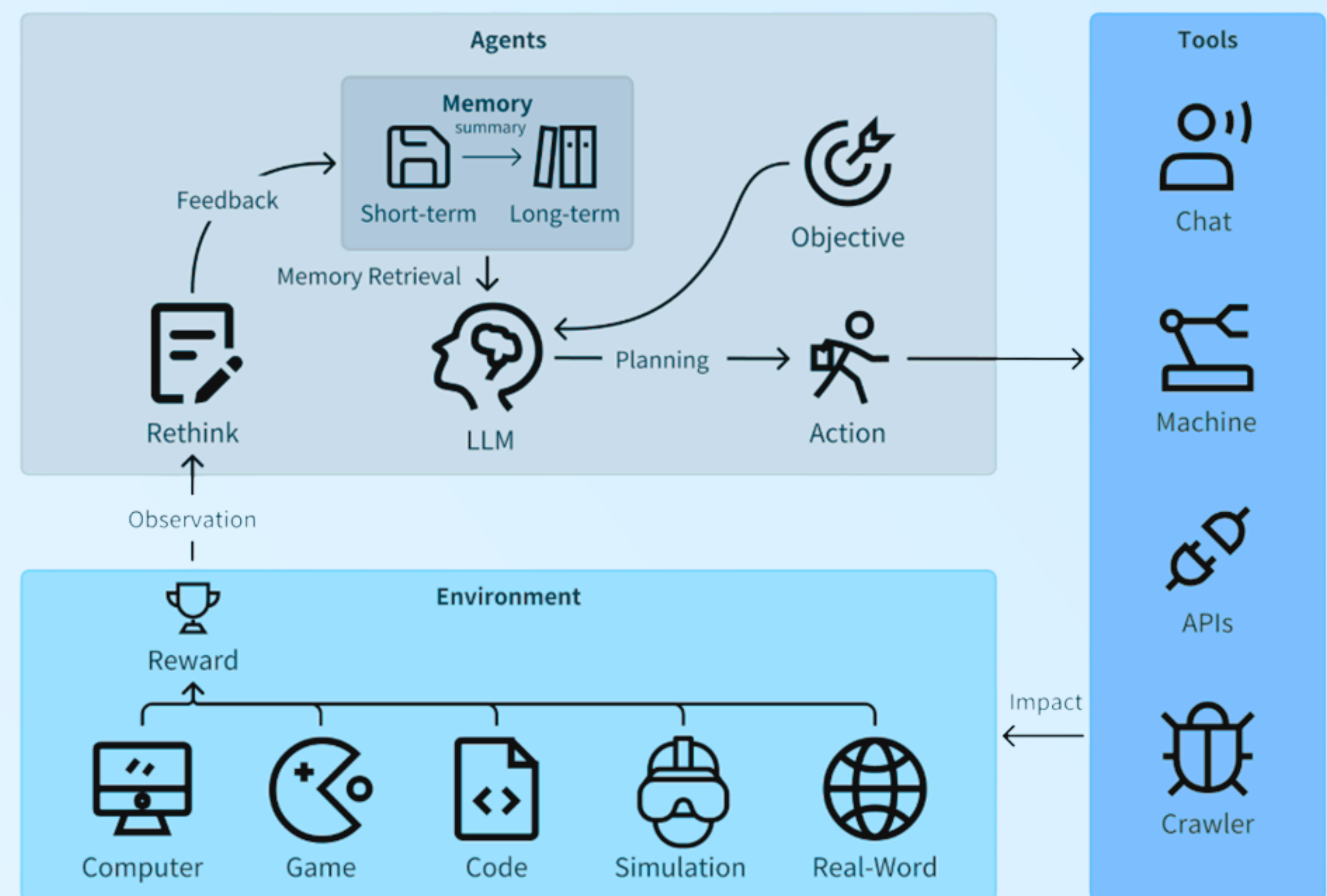
Cette approche réduit les erreurs d'hallucination, améliore la pertinence pédagogique et permet une évolution continue du système sans réentraînement du modèle.



Le LLM (Large Language Model)

- Dans le cadre de notre projet, nous avons spécifiquement utilisé le modèle gemini-2.0-flash via l'API Gemini de Google Generative AI. Ce choix est dû à sa capacité à générer rapidement des contenus textuels courts et structurés, parfaitement adaptés à nos besoins pédagogiques.
- Le LLM est au cœur de la création de notre dataset en étant responsable de :
- La génération automatique des questions de mathématiques.
- La formulation des réponses exactes.
- La production des explications détaillées et progressives qui accompagnent chaque Q&R, essentielles pour la compréhension de l'élève.
- Ce processus est piloté par des prompts dynamiques et précis, qui orientent le LLM en fonction du chapitre ou du sujet mathématique désiré, assurant ainsi une génération cohérente et diversifiée.

Méthodologie



Résultats

Notre projet a généré un dataset JSON de 5000 entrées, chacune représentant une paire Question-Réponse détaillée en mathématiques (niveau collège).

Chaque entrée est structurée comme un objet contenant :

- "cours": Nom du chapitre mathématique.
- "question": L'énoncé de la question.
- "reponse": La solution correcte.
- "explication": Le raisonnement détaillé.

```
"cours": "Équations du premier degré",  
"question": "Résoudre l'équation : -3x + 7 = 1",  
"reponse": "x = 2",  
"explication": "Soustraire 7 des deux côtés : -3x = -6. Diviser par -3 : x = 2."
```

Impact et Perspectives Futures

Le projet ouvre des voies prometteuses pour l'intégration de l'IA dans l'éducation. Voici comment notre travail peut faire la différence et les directions que nous envisageons.

Applications Pédagogiques Concrètes

- Plateformes d'Apprentissage Interactives** : Notre dataset peut alimenter des applications web et mobiles, offrant des exercices dynamiques et des retours instantanés aux élèves.
- Chatbots Pédagogiques Intelligents** : Imaginez un assistant virtuel capable de répondre aux questions mathématiques des collégiens en temps réel, avec des explications personnalisées.
- Outils d'Analyse pour Enseignants** : Le dataset permettrait aux éducateurs d'identifier les notions les plus problématiques, d'adapter leurs cours et d'analyser les performances des modèles d'IA.

Avantages Clés pour l'Éducation

- Précision Accrue** : Grâce à l'ancrage RAG, les réponses sont fidèles aux contenus des cours, réduisant les erreurs.
- Scalabilité** : Génération rapide de milliers de questions, un gain de temps considérable pour les créateurs de contenu pédagogique.
- Personnalisation** : Possibilité d'adapter les questions et explications au niveau et aux besoins spécifiques de chaque élève.

Contact

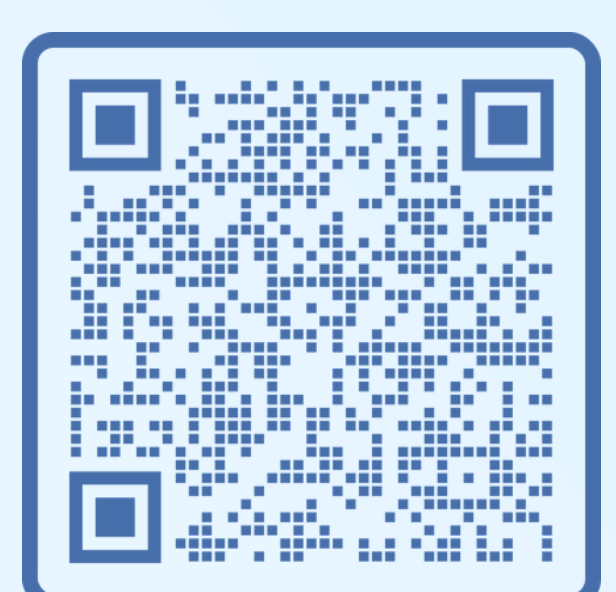
Scannez pour consulter nos profils LinkedIn



Ilhame



Israa



Douaa