

UNIVERSITE D'ÉVRY PARIS SACLAY
M1 COMPUTER & NETWORK SYSTEMS
SYSTEMES AUTONOMES

TRAVAIL D'ÉTUDE ET DE RECHERCHE

AMÉLIORER LES ENVIRONNEMENTS ÉDUCATIFS : APPROCHE BASÉE SUR LES LLMs

Présenté par

Mohammed Yacine BRAHMIA
Nabil DJELLOUDI

brahmia.my@gmail.com
nabildjelloudi772@gmail.com

Encadré par

Mr. Massinissa HAMIDI



Contents

1	INTRODUCTION	4
2	ETAT DE L'ART	5
2.1	SYSTÈMES ÉDUCATIFS CLASSIQUES	5
2.2	IA DANS LES ENVIRONNEMENTS ÉDUCATIFS	5
2.3	LLM DANS LES ENVIRONNEMENTS ÉDUCATIFS	6
2.4	IoT DANS LES ENVIRONNEMENTS ÉDUCATIF	6
2.5	IoT & LLMs	6
2.6	GÉNÉRATION AUGMENTÉE DE RÉCUPÉRATION (RAG)	7
3	ECOSYSTÈME ÉDUCATIF BASÉ SUR LES LLM ET IoT	9
3.1	SCÉNARIO PRINCIPAL	9
3.1.1	Rôle du Grand Modèle de Langage	9
3.1.2	Rôle du Système d'Internet des Objets	10
3.2	SCÉNARIOS SECONDAIRES	10
3.2.1	Automatisation de la liste de présence	10
3.2.2	En utilisant des lunettes intelligentes pour étudiants malvoyants	10
3.2.3	En utilisant des oreillettes intelligentes pour la traduction instantanée	11
3.2.4	Immersion totale en utilisant des casques de réalité virtuelle	11
4	FLUX PROPOSÉ POUR LA GÉNÉRATION DES CONTENUS INTERACTIFS	13
4.1	COMPOSANTES DE NOTRE FLUX	13
4.1.1	L'enseignant fournit son support pédagogique	13
4.1.2	LangChain & UnstructuredPDFLoader	13
4.1.3	RecursiveCharacterTextSplitter	13
4.1.4	Ollama & OllamaEmbeddings	14
4.1.5	Chroma DB	14
4.1.6	User Query & MultiQueryRetriever	14
4.1.7	Prompt Engineering	14
4.1.8	Groq & la génération de contenus	14
4.1.9	Interaction avec l'étudiant	14
5	IMPLÉMENTATION	15
5.1	INITIALISATION ET CONFIGURATION	15
5.1.1	Traitement des Fichiers PDF	15
5.2	INTERFACE	15
5.2.1	Interface Utilisateur avec ChainLit	15
5.2.2	Réponses Dynamiques et Interactivité	15

5.3	OPTIONS ET FONCTIONNALITÉS	16
5.3.1	Uploader le PDF	16
5.3.2	Poser des Questions	16
5.3.3	Générer des Quizzes	17
5.3.4	Générer des Questions	18
5.3.5	Générer des Exemples	18
6	DISCUSSION & PERSPECTIVES	20
6.1	TEMPÉRATURE	20
6.1.1	À basse température	20
6.1.2	À haute température	20
6.2	INGÉNIEURIE DES PROMPTS	21
6.3	PERSPECTIVES POUR LE FUTUR	22
7	CONCLUSION	23
	Glossaire	24
	References	25

ABSTRACT

Dans le cadre complexe et exigeant de l'environnement éducatif, la diversité des aptitudes individuelles crée des défis uniques en termes de concentration et de compréhension. Afin d'améliorer ces environnements éducatifs, notre recherche envisage l'intégration de deux technologies innovantes, les grands modèles de langage (**Large Language Model (LLM)**) et l'Internet des objets (**Internet of Things (IoT)**). Nous visitons les travaux antérieurs et l'apport des deux techniques dans le sujet, nous proposons des scénarios exploratoires afin de conquérir ces défis en se concentrant principalement sur l'utilisation des grands modèles de langage (**LLM**) pour optimiser l'expérience éducative. Nous développons un modèle de langage simulant le rôle d'un assistant virtuel, qui utilise les **LLM** pour analyser des textes éducatifs numérisés et qui génère des contenus pédagogiques personnalisés tels que des quiz, des questions de réflexion et des exemples illustratifs. Ces contenus sont conçus pour maintenir et renforcer la concentration et l'engagement des étudiants. Notre approche se focalise sur le potentiel des **LLM** à créer une certaine dynamique dans les interactions entre l'enseignant et l'étudiant répondant aux besoins spécifiques des apprenants, et l'intégration de l'**IoT** à notre approche offre des potentialités pour mesurer l'engagement de l'étudiant mais aussi pour influencer l'environnement d'apprentissage par le biais de la température, de la luminosité, du bruit et d'autres facteurs. Cette initiative vise non seulement à améliorer les pratiques pédagogiques actuelles mais aussi à anticiper et à répondre aux exigences des futurs apprenants, contribuant ainsi à enrichir l'expérience éducative à long terme. Nous discutons de nos résultats et des paramètres importants à notre approche, notamment la température et le prompting, et nous explorons les perspectives qui nous intéressent pour le futur de cette étude.

1 INTRODUCTION

L'évolution rapide de l'environnement éducatif, caractérisée par une diversité croissante des aptitudes individuelles, présente des défis majeurs pour les éducateurs en matière de concentration et de compréhension des élèves. Ce contexte met en lumière les nombreux axes d'améliorations pour le domaine éducatif, des axes que notre projet ambitionne d'explorer en créant un environnement éducatif plus dynamique et adaptatif aux besoins individuels des étudiants.

L'intention derrière notre recherche repose sur la nécessité d'aborder les défis majeurs liés aux environnements éducatifs de manière innovante. Notre approche intègre l'Internet des objets (**IoT**) et les grands modèles de langage (**LLM**), dans la perspective d'améliorer l'expérience éducative. À travers cette exploration, nous aspirons à apporter des avantages concrets tant aux enseignants qu'aux étudiants, en fournissant une réponse proactive aux défis de la concentration et de l'engagement.

Nous développons un modèle de langage simulant le rôle d'un assistant virtuel, qui utilise les **LLM** pour analyser des textes éducatifs numérisés et qui génère des contenus pédagogiques personnalisés tels que des quiz, des questions de réflexion et des exemples illustratifs. Ces contenus sont conçus pour maintenir et renforcer la concentration et l'engagement des étudiants. Notre approche se focalise sur le potentiel des **LLM** à créer une certaine dynamique dans les interactions entre l'enseignant et l'étudiant répondant aux besoins spécifiques des apprenants, et l'intégration de l'**IoT** à notre approche offre des potentialités pour mesurer l'engagement de l'étudiant mais aussi pour influencer l'environnement d'apprentissage par le biais de la température, de la luminosité, du bruit et d'autres facteurs.

Cette initiative vise non seulement à améliorer les pratiques pédagogiques actuelles mais aussi à anticiper et à répondre aux exigences des futurs apprenants, contribuant ainsi à enrichir l'expérience éducative à long terme.

ORGANISATION DU RAPPORT

Nous commençons par présenter une revue de la littérature pertinente, où nous examinerons les travaux antérieurs et les approches similaires dans le domaine. Nous présenterons les différents scénarios imaginés pour notre approche, nous détaillerons également le flux de notre méthodologie, puis l'implémentation de cette dernière en mettant en avant la conception, les outils et technologies déployés, ainsi que l'aspect technique. Par la suite, nous discutons de nos résultats et des paramètres importants à notre approche, notamment la température et le prompting. Enfin, nous discutons des résultats obtenus et les perspectives futures de ce travail de recherche.

OBJECTIF DU TRAVAIL D'ÉTUDE ET DE RECHERCHE

L'objectif principal de notre étude est de développer et améliorer les environnements éducatifs, après s'être familiarisé avec l'état de l'art concernant l'utilisation des techniques de l'intelligence artificielle (**Intelligence Artificielle (IA)**) pour ce but, nous proposons cette approche innovante utilisant les grands modèles de langage (**LLM**) et les technologies de l'Internet des Objets (**IoT**) afin de bonifier l'expérience éducative. Nous visons à créer un système simulant un assistant intelligent capable de dynamiser les interactions entre enseignants et étudiants par la génération de contenu pédagogique interactif et personnalisé, qui serait aussi capable de répondre aux besoins spécifiques des apprenants, augmentant ainsi leur engagement, leur concentration et leur réussite académique. Notre but est de démontrer comment l'intégration de l'**IA** peut transformer l'éducation en rendant l'apprentissage plus dynamique, interactif et efficace.

2 ETAT DE L'ART

2.1 SYSTÈMES ÉDUCATIFS CLASSIQUES

L'éducation est un processus conscient visant à faciliter et à guider le développement individuel (1). Elle implique la transmission de connaissances, le développement de compétences et la formation des attitudes et des valeurs, dans le but de promouvoir la croissance holistique et la réalisation de soi chez les apprenants. Le but de l'éducation est de cultiver l'adaptabilité intellectuelle, émotionnelle, morale, créative et sociale des individus, leur permettant ainsi de contribuer positivement à la société.

Ce système classique se montre peu évolutif, et met les besoins de l'individu en retrait face aux besoins du groupe. C'est pour cela que nous devons trouver une solution pour "dynamiser" ce système éducatif.

Les modèles éducatifs traditionnels sont confrontés à des défis tels que les différences individuelles entre les étudiants, l'allocation insuffisante des ressources pédagogiques, et l'évaluation de l'efficacité de l'enseignement (2). Ces enjeux représentent des obstacles significatifs à la réalisation d'une expérience éducative efficace et équitable.

2.2 IA DANS LES ENVIRONNEMENTS ÉDUCATIFS

L'application de l'IA dans l'éducation a fait l'objet d'un grand nombre de recherches au cours des 30 dernières années.

Le marché mondial de l'IA dans l'éducation est estimé à 1,1 milliard de dollars US en 2020, et les experts prévoient que celui-ci atteindra une taille révisée de 12,6 milliards de dollars US d'ici 2027 voir 41,4 % de taux de croissance annuel composé (CAGR) sur la période 2020-2027 (3).

A. Sadiku et al. (4) définissent le concept d'intelligence artificielle comme la capacité d'un système informatique à réaliser des tâches humaines (telles que penser et apprendre) qui ne peuvent habituellement être accomplies que grâce à l'intelligence humaine. La technologie de l'IA dans l'éducation offre un degré de flexibilité et d'adaptation jamais atteint auparavant. L'IA est prête à révolutionner l'éducation, et même les salles de classe, facilitant grandement le travail des enseignants.

Grâce à l'IA, les instructeurs ont pu faciliter une variété de tâches, telles que la révision et la notation des devoirs des étudiants de manière plus efficace, atteignant ainsi une qualité supérieure dans leurs activités d'enseignement. D'autre part, puisque les systèmes prennent en charge l'apprentissage automatique et l'adaptabilité, le curriculum et le contenu sont adaptés et personnalisés en fonction des besoins des étudiants, améliorant ainsi l'expérience des étudiants et la qualité globale de l'apprentissage.

Nguyen et al. (5) ont catégorisé les approches de l'IA dans l'éducation en 3 catégories. D'abord, les approches orientées "Guidance" qui assistent les étudiants et enseignants dans la prise de décision, fournissant des conseils basés sur des données pour améliorer l'accès aux opportunités éducatives, ce qui est crucial pour réduire les inégalités éducatives. Ensuite, les approches orientées "Étudiant" qui se concentrent sur l'amélioration de la qualité de l'éducation en rendant l'apprentissage plus engageant et personnalisé. Cette catégorie inclut des systèmes tels que le tutorat intelligent et l'apprentissage basé sur le jeu, axés sur la compréhension et la rétention des connaissances par les étudiants. Enfin, les approches dites "Enseignant" qui servent d'outil complémentaire pour les enseignants, en améliorant l'efficacité et l'efficience de leurs méthodes pédagogiques, par exemple en permettant la notation automatique ou en introduisant de nouvelles méthodes d'enseignement grâce à l'IA. Chaque approche reflète une facette unique de l'intégration de l'IA dans l'éducation, contribuant à un système éducatif plus adaptable, personnalisé et efficace.

L'approche que nous proposons tombe dans les deux catégories 'Étudiant' et 'Enseignant'.

2.3 LLM DANS LES ENVIRONNEMENTS ÉDUCATIFS

Large Language Model (llm) sont des modèles d'apprentissage profond qui apprennent les motifs et les règles sous-jacents du langage en s'entraînant sur de vastes corpus. Ils possèdent des capacités puissantes pour générer et comprendre le langage naturel et sont largement utilisés dans le traitement du langage naturel (NLP) (6).

Ils représentent un type de grands modèles d'apprentissage, impliquant souvent des milliards de paramètres. L'essence des grands modèles réside dans leur capacité à gérer des tâches complexes et des données à grande échelle, leur permettant d'apprendre des motifs de langage plus riches et des représentations de connaissances plus élaborées (7). Cela rend les LLMs hautement applicables dans le domaine de l'éducation.

Dans le cadre d'une éducation intelligente, les LLMs éducatifs (Edullm) font référence à des modèles d'application éducative basés sur les LLMs. En apprenant à partir de nombreuses données et corpus pédagogiques, les Edullm peuvent fournir un soutien d'apprentissage personnalisé (8) et des capacités bien meilleures d'évaluation pédagogique des étudiants. L'état de recherche des Edullm démontre un potentiel et des opportunités importants.

Premièrement, les Edullm peuvent identifier les modèles et les caractéristiques d'apprentissage des étudiants en apprenant à partir de quantités massives de données éducatives, permettant ainsi de fournir un soutien d'apprentissage personnalisé et des recommandations de ressources pédagogiques.

Deuxièmement, les Edullm ont le potentiel d'évaluation pédagogique, évaluant automatiquement la maîtrise des connaissances, les résultats d'apprentissage et les capacités d'expression des étudiants, fournissant ainsi une évaluation plus complète des étudiants et un retour d'information pédagogique aux éducateurs.

De plus, les Edullm peuvent être appliqués au tutorat intelligent, fournissant une résolution de problèmes en temps réel, des conseils d'apprentissage et une orientation académique par le biais du dialogue et de l'interaction avec les étudiants (9), et ceci est le but même de notre approche.

2.4 IOT DANS LES ENVIRONNEMENTS ÉDUCATIF

Le terme "Internet des objets" (iot) est récemment devenu populaire pour souligner la vision d'une infrastructure mondiale connectant des objets physiques, en utilisant le même protocole Internet, leur permettant de communiquer et de partager des informations (10). Selon le cabinet d'analystes Gartner, 8,4 milliards d'"objets" étaient connectés à l'Internet en 2017, à l'exclusion des ordinateurs portables, ordinateurs, tablettes et téléphones mobiles, atteignant 17,08 milliards en 2024.

L'étude de Khaula Zeeshan et al. (11) met en évidence l'importance de l'IoT dans la création d'une éducation intelligente et durable. Ils soulignent que les technologies disruptives comme l'IoT ont un grand potentiel pour révolutionner l'enseignement actuel en fournissant des expériences d'apprentissage enrichies, en augmentant l'efficacité opérationnelle et en collectant des informations en temps réel et exploitables sur la performance des étudiants.

De plus, l'article de I. A. Ghashim et al. (12) explore l'enseignement et l'apprentissage basés sur l'IoT, mettant en évidence les tendances modernes et les défis ouverts. Ils discutent de l'importance croissante de l'IoT dans les systèmes éducatifs et de son impact sur la performance des apprenants, le travail quotidien des enseignants, la gestion des bâtiments scolaires, les systèmes de transport scolaire et la fourniture d'apprentissage à distance.

2.5 IOT & LLMs

Inspiré par la théorie de la spécialisation fonctionnelle du cerveau, (13) adopte une approche similaire pour la conception et le développement de systèmes intelligents, Hongwei Cui et al. (13) ont proposé un framework qu'ils appellent LLMind, qui transforme le contrôle conventionnel des dispositifs IoT en intégrant un LLM comme orchestrateur pour accomplir des tâches complexes. Cette intégration conduit à la construction d'un écosystème IoT renforcé par LLM.

2.6 GÉNÉRATION AUGMENTÉE DE RÉCUPÉRATION (RAG)

La génération augmentée de récupération (RAG) est le processus consistant à optimiser le résultat d'un grand modèle de langage. Elle fait donc appel à une base de connaissances fiable externe aux sources de données utilisées pour l'entraîner avant de générer une réponse. Les grands modèles de langage (**LLM**) sont entraînés avec d'importants volumes de données et utilisent des milliards de paramètres pour générer des résultats originaux pour des tâches telles que répondre à des questions, traduire des langues et compléter des phrases. La RAG étend les capacités déjà très puissantes des **LLMs** à des domaines spécifiques ou à la base de connaissances interne d'une organisation, le tout sans qu'il soit nécessaire de réentraîner le modèle. Il s'agit d'une technologie qui comble aussi les défauts des **LLMs** tel que l'hallucination pour qu'ils restent cohérents, précis et utiles dans de nombreux contextes.

La génération augmentée de récupération (RAG) fonctionne en intégrant un composant de récupération d'informations externes à un grand modèle de langage (**LLM**). Lorsqu'une requête utilisateur est reçue, en plus de la base de données du **LLM**, le système recherche aussi dans une base de données dite externe (par exemple le PDF fourni de l'enseignant dans notre cas) pour extraire des informations pertinentes. Ces informations, avec la requête elle-même, sont ensuite fournies au **LLM**. Le modèle utilise ces détails supplémentaires pour générer des réponses plus précises et contextuelles. Cette méthode améliore la qualité des réponses en incorporant des connaissances à jour et spécifiques sans nécessiter de réentraîner le **LLM**. On peut décrire son fonctionnement comme suit :

Embedding

Les documents ou les données, typiquement des mots ou des phrases, doivent être transformés en une forme qui peut être efficacement recherchée et interrogée, généralement en vecteurs mathématiques de dimension réduite. Ces vecteurs captent des aspects sémantiques des données et permettent au modèle de traiter le texte de manière plus efficace. Les embeddings aident à révéler les relations entre les mots en projetant les données textuelles dans un espace où la distance entre les vecteurs reflète la proximité sémantique.

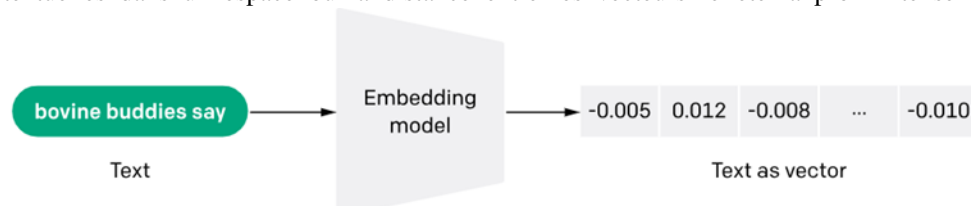


Figure 1: Transformation de Texte en Vecteur

Recherche

Lorsqu'une requête est reçue, le système recherche dans son index pour trouver des segments de texte ou des documents qui sont pertinents et qui ressemblent à la requête. Cette étape est cruciale car la qualité de la récupération affecte directement la qualité de la réponse finale.

Génération

Une fois les informations pertinentes trouvées, elles sont utilisées pour augmenter le modèle, en plus de sa propre connaissance interne, pour générer une réponse ou un contenu qui non seulement répond à la requête initiale mais le fait de manière plus informée, précise et contextuellement plus enrichie (14).

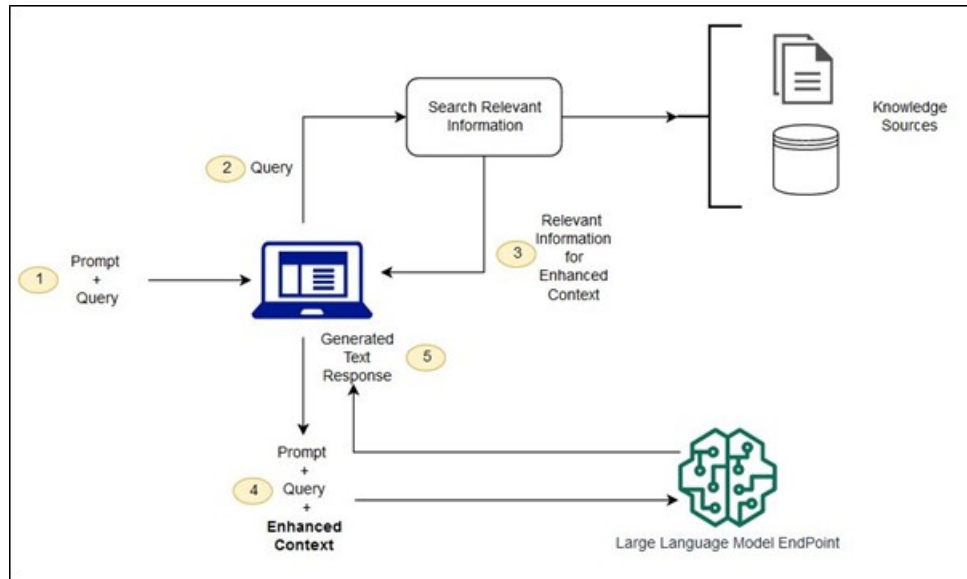


Figure 2: Processus de Recherche et Génération de Réponse Améliorée

3 ECOSYSTÈME ÉDUCATIF BASÉ SUR LES LLM ET IoT

Cette section montre la manière dont nous imaginons l'intégration des grands modèles de langages (**LLM**) dans les environnements éducatifs, combinée à l'utilisation d'objets connectés (**IoT**), afin de dynamiser l'expérience éducative pour l'étudiant. Nous verrons dans cette section les différents scénarios imaginés pour notre approche.

3.1 SCÉNARIO PRINCIPAL

Nous commencerons par le scénario principal imaginé pour notre approche : (voir figure 3)

L'enseignant lance son cours en fournissant son support pédagogique au **LLM**, pendant que les étudiants se concentrent sur l'ensemble des explications de l'enseignant, le modèle d'apprentissage automatique présente à celui-ci un tableau de bord détaillé des informations des étudiants présents et de leurs niveaux de concentration. Simultanément, des objets **IoT** surveillent en temps réel les changements de ces niveaux.

Lorsqu'une baisse de concentration est détectée, cette information est d'abord mise à jour sur le tableau de bord, puis elle est communiquée au **LLM** qui, après avoir compris le PDF de l'enseignant, commence à générer du contenu dynamique afin de regagner l'attention des étudiants. Ce contenu est ensuite directement accessible par l'étudiant via l'intermédiaire d'un autre objet **IoT**, qui l'affiche directement pour les étudiants. Ce flux d'interactions permet aux étudiants de se réengager dans le cours, favorisant une expérience d'apprentissage plus interactive et personnalisée.

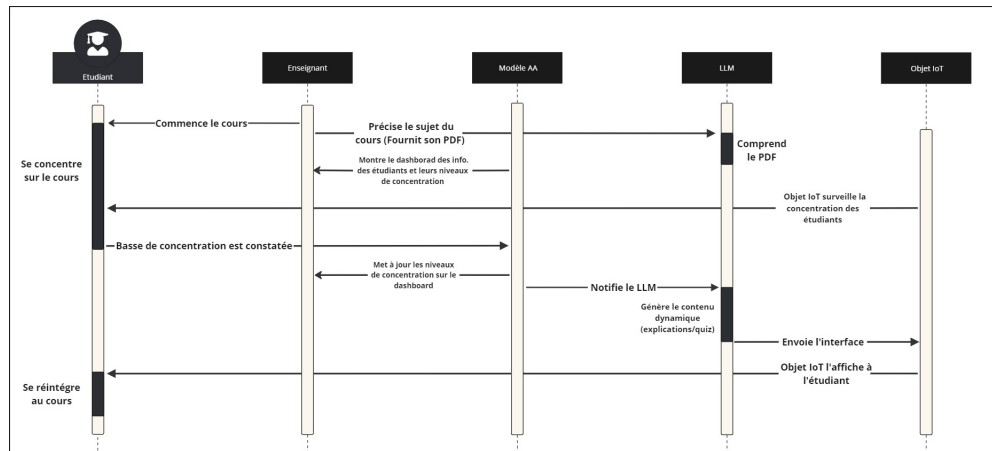


Figure 3: Diagramme de séquençage de notre scénario principal

3.1.1 Rôle du Grand Modèle de Langage

Interprétation des ressources pédagogiques

Le **LLM** commence son intervention en comprenant le contenu du PDF fourni par l'enseignant. Cette étape implique l'analyse des données textuelles du document pour extraire les informations pertinentes qui seront utilisées pour générer des réponses adaptées aux besoins des étudiants.

Génération du contenu dynamique

La génération de contenu interactif et dynamique constitue la pierre angulaire de notre approche. Ces contenus peuvent être des explications plus approfondies du point spécifique choisi par l'étudiant, des exemples supplémentaires, des questions engageantes à poser à l'enseignant afin de mieux comprendre, des résumés de cours ou même des quiz personnalisés lui permettant de mieux comprendre et assimiler les informations précédentes.

3.1.2 Rôle du Système d'Internet des Objets

Surveillance de la concentration des étudiants

Les objets **IoT** sont utilisés pour surveiller en continu le niveau de concentration des étudiants pendant le cours. Cette fonctionnalité est essentielle pour détecter les moments où les étudiants peuvent perdre l'intérêt ou avoir du mal à suivre le contenu présenté, permettant une intervention rapide et ciblée. Pour ceci, divers objets **IoT** peuvent être utilisés, par exemple, les bracelets connectés qui peuvent utiliser différents paramètres afin de détecter la concentration tels que les variations du rythme cardiaque (15), la température (16), et même la réponse galvanique de la peau (RGP) (17).

Affichage du Contenu Dynamique Généré

Cela peut inclure l'utilisation d'écrans intelligents, de tablettes ou d'autres formes d'interfaces utilisateur connectées qui présentent le contenu interactif aux étudiants.

3.2 SCÉNARIOS SECONDAIRES

Voici différents scénarios secondaires imaginés pour notre approche, qui représentent des dérivations du scénario principal, en modifiant légèrement les rôles du **LLM** ou de l'objet **IoT** par exemple.

3.2.1 Automatisation de la liste de présence

Dans ce scénario (voir figure 4), le seul changement qu'on effectue c'est d'ajouter un autre objet **IoT** qui est une carte **Radio Frequency Identification (RFID)**, que l'étudiant scanne avant d'entrer en classe. Ceci permettra de remplir automatiquement une liste de présence, ce qui modifiera le tableau de bord de l'enseignant afin d'afficher uniquement les étudiants présents au lieu de l'intégralité de classe.

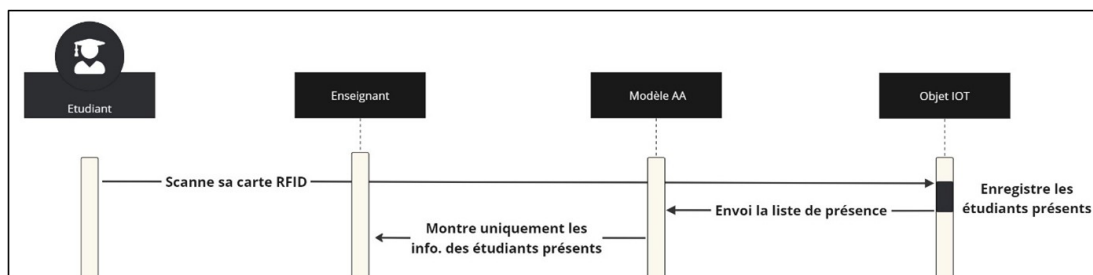


Figure 4: Scénario d'automatisation de la présence en classe via RFID et **IoT**

3.2.2 En utilisant des lunettes intelligentes pour étudiants malvoyants

Les lunettes intelligentes (voir figure 5), portées par l'étudiant, capturent des images du tableau. Ces images capturées sont envoyées au **LLM** qui applique la reconnaissance optique de caractères (OCR) pour transformer le texte extrait en messages audio descriptifs de ce qu'on voit sur le tableau. Le modèle envoie ces messages audio à l'objet **IoT** qui les diffuse à l'étudiant. Ce dernier pourra ensuite lui poser des questions verbalement, auxquelles le **LLM** répondra comme dans le scénario principal.

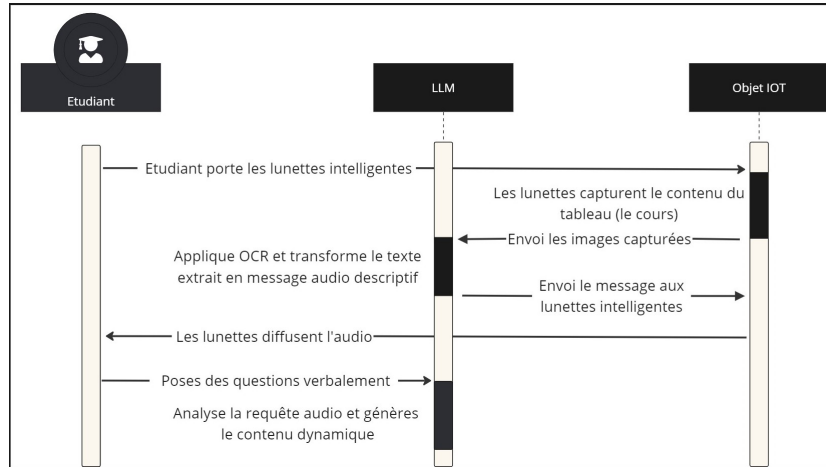


Figure 5: Scénario en utilisant des lunettes intelligentes

3.2.3 En utilisant des oreillettes intelligentes pour la traduction instantanée

L'oreillette intelligente, portée par l'étudiant, capture le son dans la salle de classe, spécialement celui du professeur. Ce son capturé est ensuite envoyé au **LLM**. Le **LLM** effectue une extraction du texte et le traduit vers la langue préférée de l'étudiant, puis génère le contenu qui est l'audio traduit. Une fois la traduction terminée, le son traduit est renvoyé à la pièce d'oreille intelligente afin qu'elle le diffuse pour l'étudiant (voir figure 6).

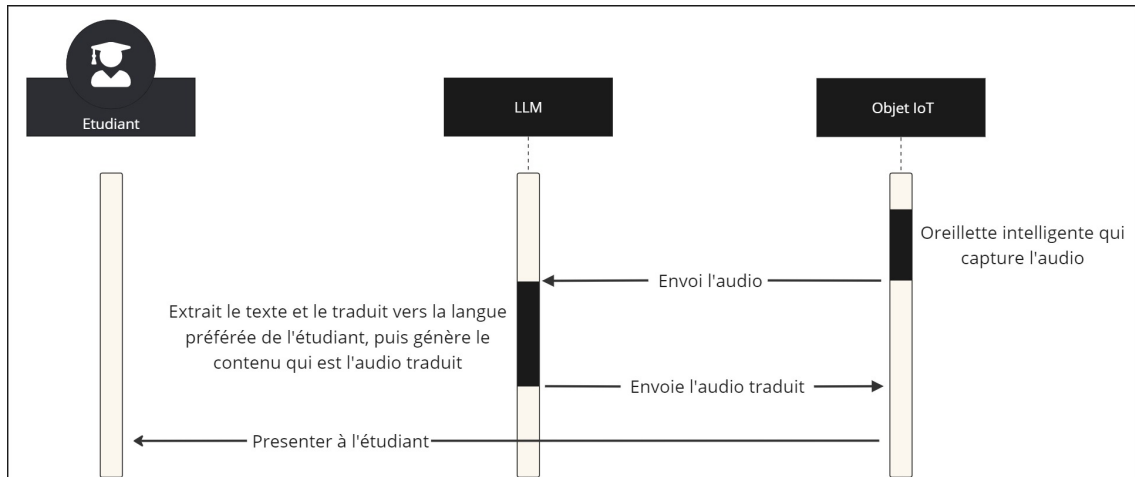


Figure 6: Scénario en utilisant des oreillettes intelligentes

3.2.4 Immersion totale en utilisant des casques de réalité virtuelle

L'étudiant commence par porter l'objet **IoT** qui est le casque de réalité virtuelle équipé de capteurs et de connectivité. Une fois le casque en place, le **LLM** commence à générer des narrations pour renforcer l'immersion dans l'expérience. Cela pourrait impliquer de fournir un contexte historique, une trame narrative ou des éléments interactifs rendant l'expérience virtuelle plus captivante. L'étudiant, immergé dans l'environnement virtuel, pose une question à haute voix lorsqu'il rencontre quelque chose d'intéressant ou s'il a une question. Le casque VR, en tant qu'objet **IoT**, capte la question verbale de l'étudiant. Cette entrée audio est ensuite envoyée du casque VR au **LLM**. Le **LLM** traite cette entrée audio, en utilisant la compréhension du langage naturel pour interpréter la question et formuler une réponse appropriée. Il peut également convertir la question de la parole en texte dans le cadre de ce processus. Le **LLM** renvoie ensuite la réponse au casque VR. Cette réponse pourrait être sous forme d'une réponse parlée ou d'un texte affiché dans l'environnement VR. Finalement, le

casque VR affiche ou fait entendre la réponse à l'étudiant, assurant ainsi l'immersion de l'utilisateur (voir figure 7).

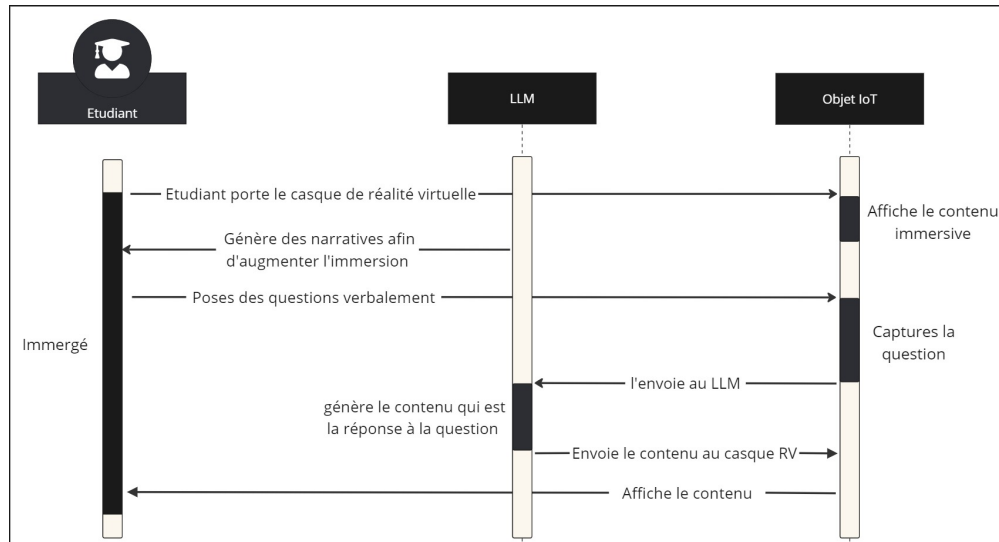


Figure 7: Scénario d'une immersion en utilisant un casque de réalité virtuelle

4 FLUX PROPOSÉ POUR LA GÉNÉRATION DES CONTENUS INTERACTIFS

Lors de notre approche, nous nous sommes focalisés sur le scénario principal, plus précisément sur la partie génération de contenu interactif en utilisant les **LLMs**. Dans cette section, nous présentons le flux proposé pour la génération.

L'architecture de notre flux (voir figure 8) est basée sur la technologie Retrieval-Augmented Generation (RAG) (2.6). Cette technologie avancée nous permet de générer des contenus éducatifs interactifs qui ne sont pas seulement précis mais aussi profondément enrichis de contexte, facilitant ainsi une expérience d'apprentissage personnalisée et intuitive pour chaque étudiant.

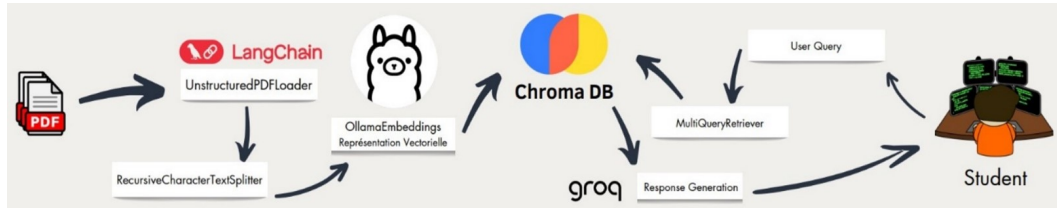


Figure 8: Composantes de notre flux

4.1 COMPOSANTES DE NOTRE FLUX

Le flux commence avec l'enseignant qui fournit son support pédagogique en PDF (4.1.1). LangChain, avec UnstructuredPDFLoader, charge et traite ce contenu (4.1.2), puis RecursiveCharacterTextSplitter segmente le texte en fragments gérables (4.1.3). OllamaEmbeddings transforme ces fragments en vecteurs numériques (4.1.4), stockés dans Chroma DB pour faciliter la récupération d'informations pertinentes (4.1.5).

Le MultiQueryRetriever gère les requêtes des utilisateurs, récupère des informations de Chroma DB et assure des réponses précises selon ce que demande l'utilisateur (4.1.6). La structure et la manière de répondre sont définies par des prompts spécifiques à chaque fonctionnalité (4.1.7). Enfin, Groq orchestre la génération des réponses dynamiques et du matériel pédagogique personnalisé (4.1.8).

4.1.1 L'enseignant fournit son support pédagogique

Pour commencer notre processus, l'enseignant doit fournir son support pédagogique en format PDF. Ceci représente le premier déclencheur de toute génération de contenu interactif. Ce contenu servira alors de base pour générer des réponses et des matériaux pédagogiques adaptés, en fonction de la demande (query) de l'étudiant, qui représente le deuxième déclencheur de notre architecture.

4.1.2 LangChain & UnstructuredPDFLoader

Le processus procède ensuite avec le **LangChain** ; Un outil qui intègre divers services de traitement du langage naturel pour faciliter le développement de solutions nécessitant une compréhension et une génération avancées du langage, intégrant l'outil **UnstructuredPDFLoader**, qui est utilisé pour charger et traiter le contenu des documents PDF non structurés. Cette étape est essentielle pour préparer le texte à une analyse plus approfondie.

4.1.3 RecursiveCharacterTextSplitter

Après le chargement, le texte est passé à travers le **RecursiveCharacterTextSplitter**, un outil conçu pour segmenter et organiser le texte en fragments gérables. Cette segmentation est cruciale pour la prochaine étape de traitement sémantique.

4.1.4 Ollama & OllamaEmbeddings

Ollama est une plateforme conçue pour faciliter l'utilisation des grands modèles de langage. Elle permet aux utilisateurs d'exécuter, de personnaliser et de créer leurs propres modèles, y compris des modèles comme Llama 3, Phi 3, Mistral et Gemma. Les segments de texte sont ensuite traités par **OllamaEmbeddings**, un sous-service de **Ollama** qui transforme les données textuelles complexes en formats vectoriels numériques (similairement à la technique RAG). Cela permet de convertir le texte en un format qui peut être facilement utilisé pour la récupération d'informations et la génération de réponses.

4.1.5 Chroma DB

Chroma DB est une base de données open-source conçue pour gérer et optimiser l'utilisation des embeddings. Elle facilite le stockage, la recherche et le filtrage d'embeddings vectoriels, facilitant ainsi la récupération rapide d'informations pertinentes lors de la génération de contenu.

4.1.6 User Query & MultiQueryRetriever

Le **MultiQueryRetriever** prend en charge la formulation des requêtes utilisateur (**User Query**) par exemple une question, et récupère de celle-ci des informations pertinentes à stocker et à comparer avec ceux présents sur **Chroma DB**. Cela assure que les réponses générées sont bien précises et bien ciblées.

4.1.7 Prompt Engineering

Un **prompt** est une séquence textuelle spécifique utilisée pour diriger la génération de réponses par un modèle de traitement du langage naturel. Il sert d'instruction claire, indiquant le thème, le ton et la direction du contenu à produire, ce qui est crucial pour obtenir des réponses adaptées, pertinentes et bien structurées. Pour notre modèle, un prompt unique est élaboré pour chaque fonctionnalité, clarifiant ainsi les attentes précises et assurant la pertinence du contenu généré par rapport aux requêtes des utilisateurs (voir section options et fonctionnalité (5.3) pour les prompts spécifiques utilisés).

4.1.8 Groq & la génération de contenus

Avec les informations récupérées, **Groq** entre en jeu pour orchestrer la génération de réponses dynamiques. Cela comprend la création de matériel pédagogique personnalisé tel que des explications détaillées, des quiz et d'autres ressources interactives.

4.1.9 Interaction avec l'étudiant

Enfin, le contenu généré est présenté à l'étudiant via une interface interactive, permettant une expérience d'apprentissage immersive et personnalisée. L'étudiant peut interagir avec ce contenu, ce qui stimule l'engagement et favorise une meilleure compréhension du cours proposé par l'enseignant.

5 IMPLÉMENTATION

5.1 INITIALISATION ET CONFIGURATION

L'assistant intelligent est implémenté en utilisant Python et tire parti de plusieurs bibliothèques pour traiter et interagir avec des documents PDF, ainsi que pour gérer la conversation. Le processus débute par le chargement des variables d'environnement depuis un fichier `.env`, qui contient la clé d'**Application Programming Interface (API)** nécessaire pour accéder à l'**API** Groq. Le modèle de l'assistant intelligent est initialisé avec des paramètres spécifiques, tels que le nom du modèle et la température, qui influencent le style et la diversité des réponses générées.

5.1.1 Traitement des Fichiers PDF

À l'aide de la bibliothèque PyPDF2, le système lit et extrait le texte des fichiers PDF téléchargés par l'utilisateur. Ce texte est ensuite segmenté en morceaux plus petits grâce à l'outil RecursiveCharacterTextSplitter, afin de maintenir une cohérence dans les fragments de texte traités.

5.2 INTERFACE

5.2.1 Interface Utilisateur avec ChainLit

L'interface est conçue avec ChainLit, qui permet une interaction asynchrone avec l'utilisateur. Ce dernier est invité à télécharger un document PDF via une interface graphique épurée. Une fois le document chargé, l'assistant propose des boutons d'action vers différentes fonctionnalités, permettant à l'utilisateur de poser des questions, de demander des exemples, ou de générer des quiz basés sur le texte du PDF.

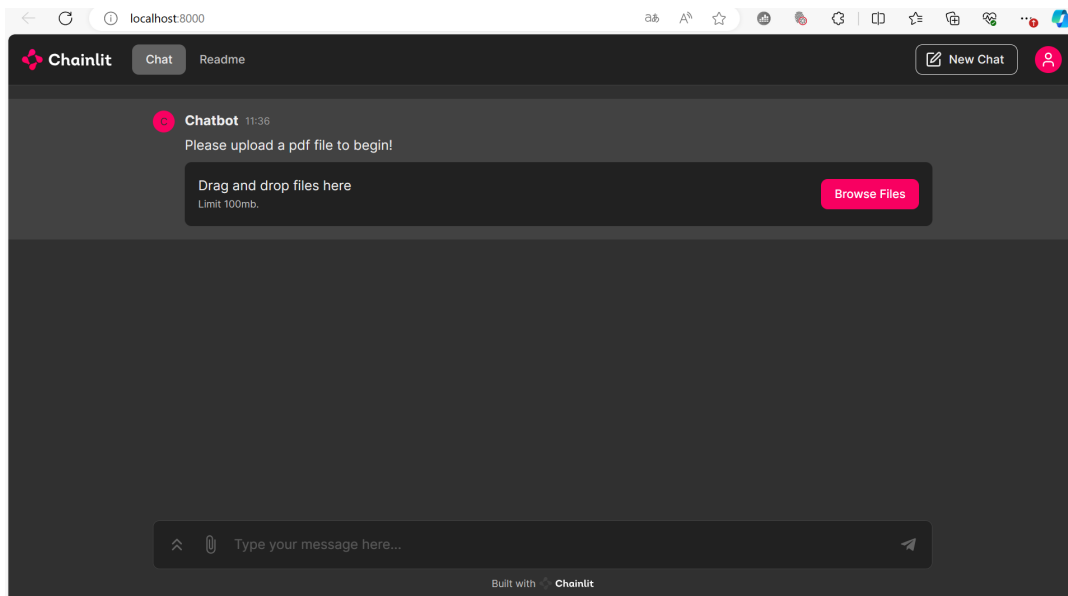


Figure 9: Notre interface

5.2.2 Réponses Dynamiques et Interactivité

Chaque interaction avec l'utilisateur, qu'il s'agisse d'une question posée ou d'une action sélectionnée via un bouton, est traitée par des callbacks qui génèrent des réponses appropriées. Le système exploite des capacités de génération de langage pour créer des réponses pertinentes et informatives basées sur le contenu du document.

5.3 OPTIONS ET FONCTIONNALITÉS

5.3.1 Uploader le PDF

Les utilisateurs commencent leur interaction avec l'assistant en téléchargeant un fichier PDF. Cette fonctionnalité est essentielle pour permettre au système d'analyser le contenu spécifique soumis par l'utilisateur.

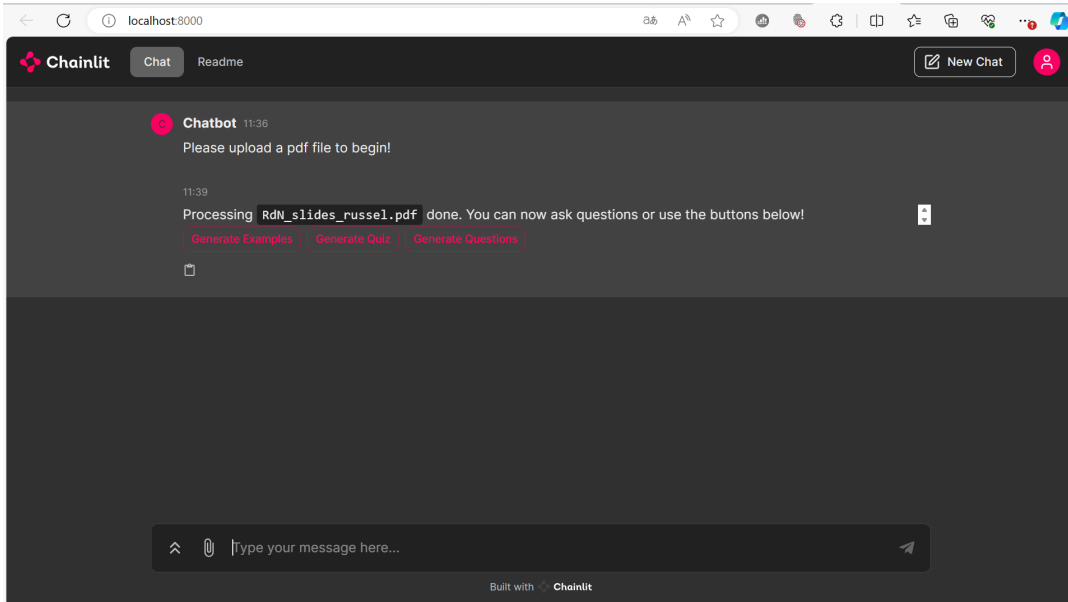


Figure 10: Uploader PDF

Il est possible de cliquer sur l'un des trois boutons ci-dessous pour générer du contenu général à partir du document, plutôt que de poser une question ciblée sur un aspect précis, une fonctionnalité que nous démontrerons prochainement.

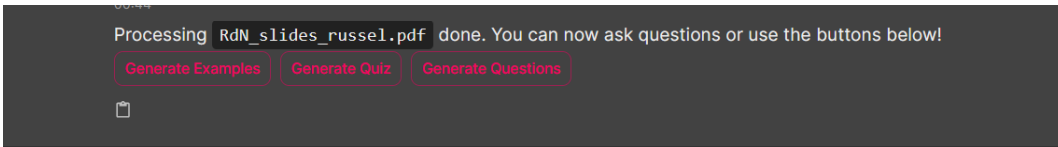


Figure 11: Illustration de nos fonctionnalités sous forme de boutons

5.3.2 Poser des Questions

Les utilisateurs peuvent poser des questions directement liées au contenu du PDF. L'assistant intelligent, configuré avec le ConversationalRetrievalChain, répond en extrayant des informations pertinentes du texte ou en générant des réponses contextualisées.

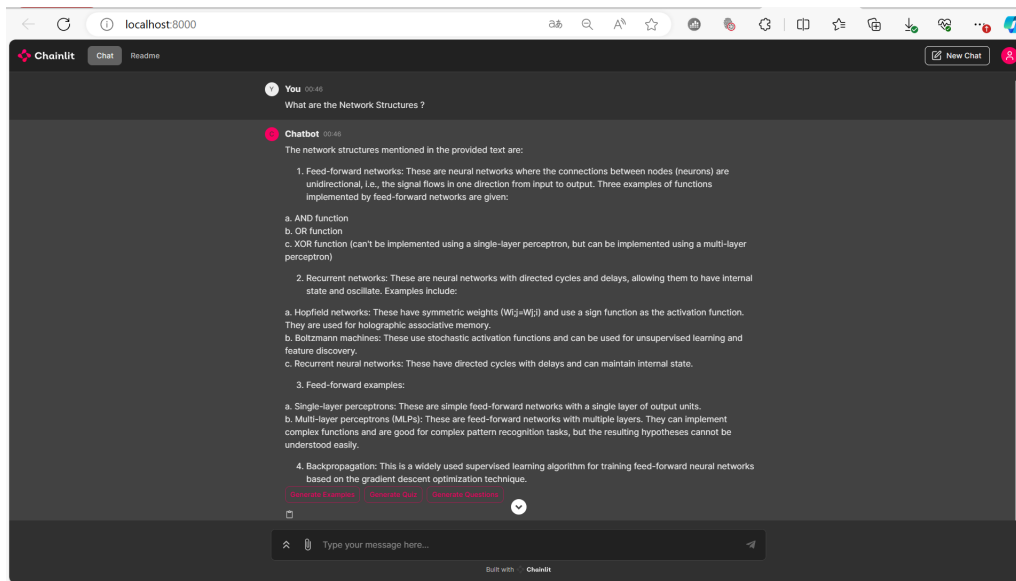


Figure 12: Exemple d'une réponse à une question

En plus de pouvoir poser des questions, l'utilisateur possède d'autres fonctionnalités avancées, ces fonctionnalités sont facilitées par l'emploi des **prompts** (voir 4.1.7).

5.3.3 Générer des Quizzes

Cette option permet de créer des quizzes visant à tester la compréhension du contenu à travers des questions à choix multiples. Chaque question est accompagnée d'une explication détaillée pour renforcer la compréhension des réponses correctes et des concepts sous-jacents.

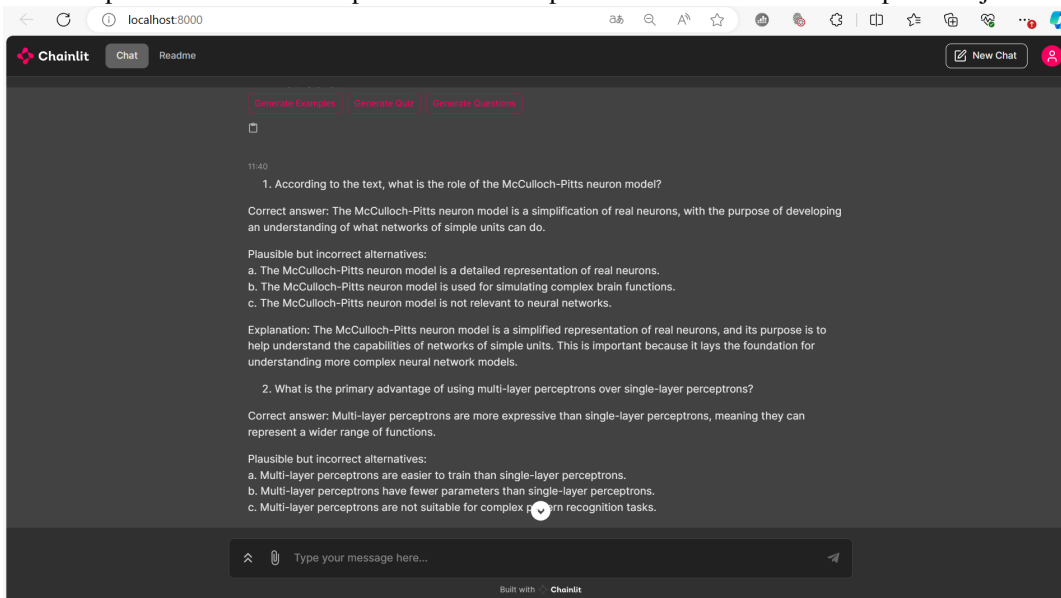


Figure 13: Exemple de génération de Quiz

Prompt utilisé : [Using seed {random_seed}, generate exactly five quiz questions based on the key points of the given document. Each question should include one correct answer, three plausible but incorrect alternatives, provide a detailed explanation for why the correct answer is right, emphasizing the relevance

and implications of the answer. Ensure the questions cover a range of topics such as factual details, analytical insights, application-based scenarios, and theoretical understanding].

5.3.4 Générer des Questions

Cette fonctionnalité permet de générer une variété de questions pertinentes que l'étudiant pourrait poser à son enseignant basées sur le texte du document PDF. Les questions visent à explorer des détails factuels, des interprétations, et des évaluations critiques du contenu, en utilisant des graines aléatoires pour diversifier les interactions.

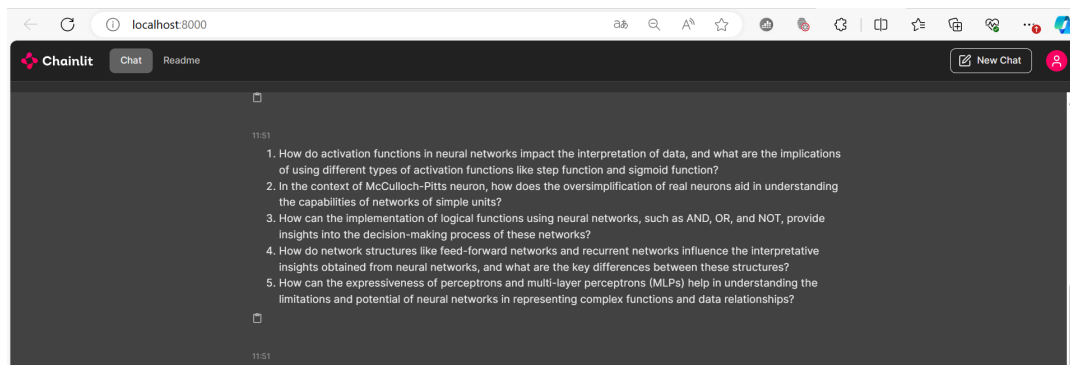


Figure 14: Exemple de génération de questions

Prompt utilisé : [Using seed {random_seed}, please generate exactly 5 unique questions focusing on key themes featured in the given document. Explore new aspects not previously covered and ensure that questions span different facets of the themes].

5.3.5 Générer des Exemples

Il est possible aussi de générer des exemples illustratifs pour clarifier les concepts clés du texte. Ces exemples incluent des analogies, des applications pratiques et des scénarios hypothétiques, enrichissant ainsi la compréhension de l'utilisateur.

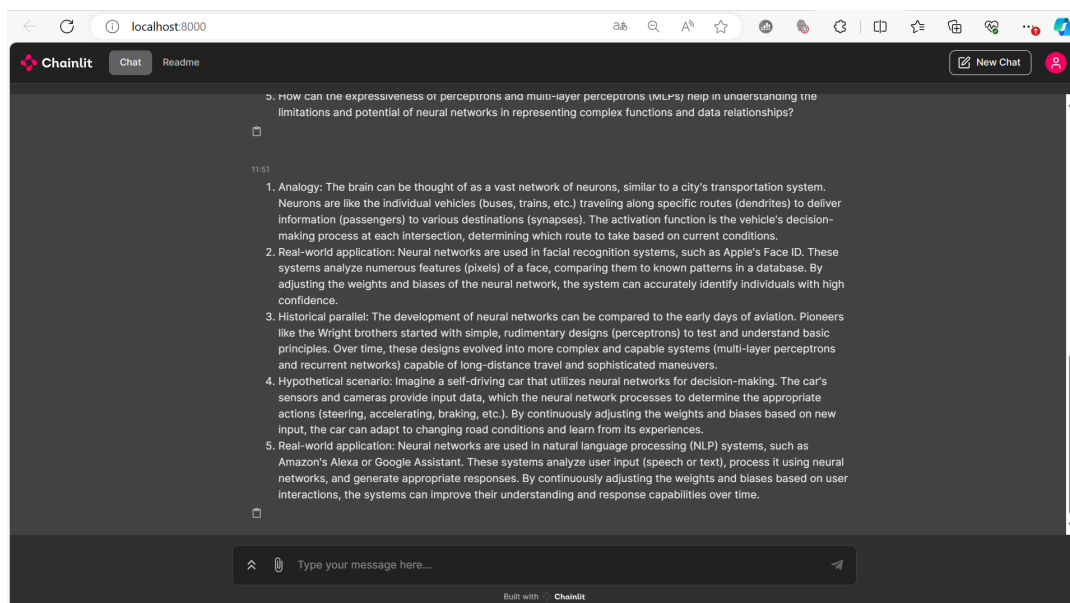


Figure 15: Exemple de génération d'exemples

Prompt utilisé : [Using seed {random_seed}, generate exactly 5 creative examples that clearly illustrate different key concepts from the following text. Include examples that are analogies, real-world applications, historical parallels, and hypothetical scenarios to enrich understanding].

Cette structure modulaire et interactive rend notre assistant non seulement facile à utiliser mais permet également une analyse détaillée et personnalisée des documents PDF. Chaque fonctionnalité est conçue pour maximiser l'engagement et l'utilité pour l'utilisateur, faisant de cet assistant intelligent un outil puissant pour l'exploration et la manipulation des documents textuels.

6 DISCUSSION & PERSPECTIVES

6.1 TEMPÉRATURE

Dans le contexte des grands modèles de langage (LLM), la "température" est un paramètre utilisé lors de la génération de texte pour contrôler le niveau de créativité et de diversité des réponses produites par le modèle. La modélisation de la température est donc essentielle pour affiner les performances du modèle.

6.1.1 À basse température

Pour des valeurs basses (on l'a initialisé à 0.2 pour notre modèle), cela produit des réponses plus précises et ciblées, strictement alignées sur le contenu du PDF fourni, favorisant une concentration accrue des étudiants sur le cours spécifique. Ce choix de température minimise les divergences potentielles. Par exemple, si l'étudiant demande le contenu d'un autre cours ou d'un sujet hors contexte, notre modèle répondra simplement que la question de l'étudiant est loin du contexte étudié.

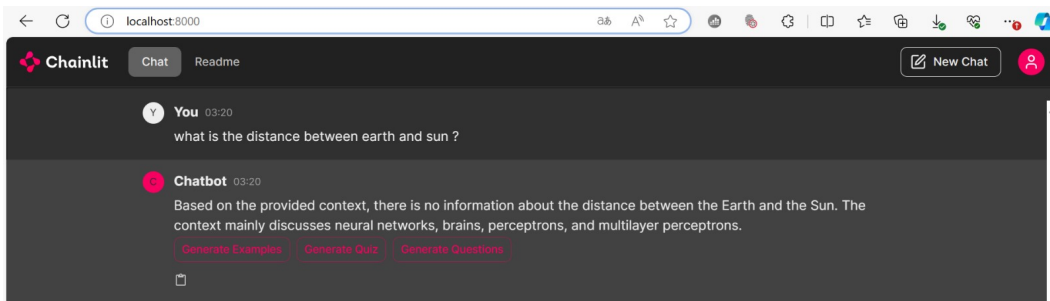


Figure 16: Exemple de réponse à basse température

6.1.2 À haute température

En revanche, une température élevée mène à générer des réponses plus créatives et étendues, parfois éloignées du contenu du PDF, ce qui peut être utile dans des sujets interdisciplinaires ou une réponse plus standardisée ne suffirait pas pour couvrir la profondeur du sujet.

Comme on peut le constater à travers des figures 16 et 17, pour la même question, à basse température, le modèle reste fidèle au document et évite de répondre hors contexte, mais à haute température, il mentionne toujours que cette question semble peu reliée au contexte du document, mais sa 'créativité' le pousse quand même à donner une réponse.

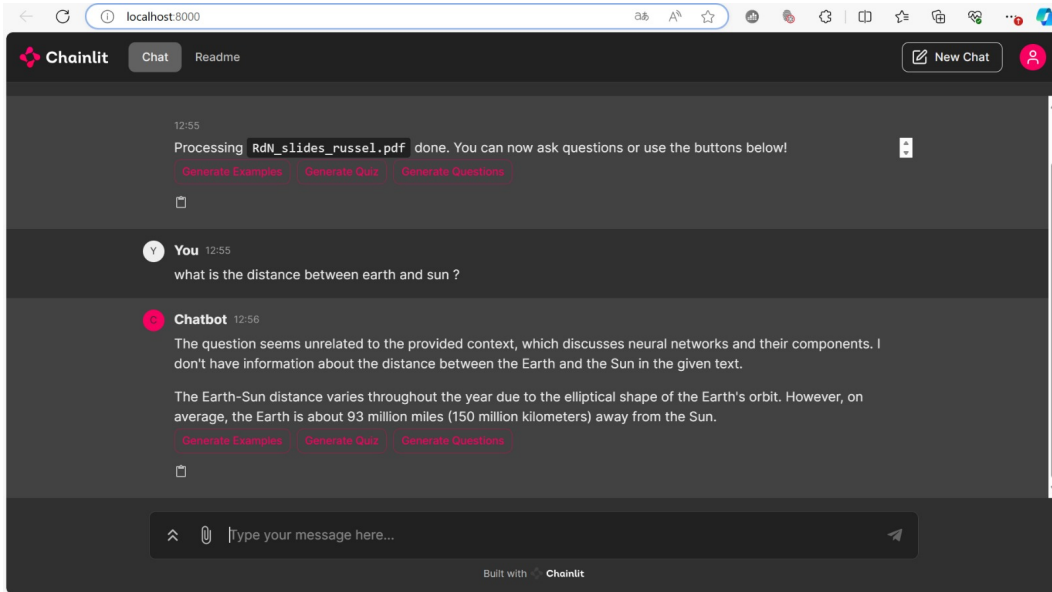


Figure 17: Exemple de réponse à haute température

6.2 INGÉNIEURIE DES PROMPTS

Lors de la conception de notre assistant intelligent, nous avons expérimenté avec une variété de prompts afin d’orienter notre modèle vers le meilleur ton à adapter, et la meilleure façon de structurer et de présenter ces réponses afin de générer le meilleur contenu possible pour l’utilisateur. Prenons par exemple le prompt utilisé pour générer les quiz. Si on utilise tout simplement le prompt générique [Based on the provided document, generate a quiz], la réponse du modèle aura plusieurs défauts et produira des résultats incohérents et peu structurés. Elle aura tendance à prendre un format différent lors de chaque exécution (par exemple différence dans le nombre de questions, de propositions, etc.), l’importance du contenu de chaque question variera aussi. Pour remédier à cela, nous utilisons des directives claires et précises du type :

[Generate exactly five quiz questions with four propositions each] assurant une uniformité dans la présentation des questions et des propositions, [Based on the key points of the given document] pour que les quiz restent centrés sur les notions pertinentes du document, [Have one correct answer and three plausible but incorrect propositions] ceci introduit un certain niveau de difficulté aux quiz, en assurant que les propositions qui ne sont pas correctes ne sont pas trop faciles à repérer, demandant une certaine réflexion de l’étudiant afin de trouver celle qui est correcte tout en leur fournissant [Provide an explanation for why the correct answer is right] pour que l’étudiant reçoive une explication détaillée de pourquoi c’est cette réponse qui est correcte.

On peut aussi choisir de cacher la réponse juste afin de laisser le temps à l’étudiant de réfléchir ou d’effectuer ses calculs. Cet exemple avec ces ajustements illustre la puissance des prompts et l’importance de les choisir minutieusement afin de générer des contenus éducatifs de haute qualité qui répondent précisément aux besoins de chaque étudiant.

6.3 PERSPECTIVES POUR LE FUTUR

ÉTENDRE L'ASPECT INTERNET DES OBJETS

Dans nos scénarios imaginés, nous avons envisagé l'intégration de plusieurs objets **IoT** pour enrichir notre solution. Jusqu'à présent, nous nous sommes principalement concentrés sur le développement des aspects liés aux **LLMs**. Cependant, pour l'avenir, nous prévoyons d'explorer et d'intégrer activement les technologies **IoT** afin d'étendre les capacités de notre système. L'objectif est de créer un environnement éducatif encore plus interactif et responsive, où les objets connectés pourraient, par exemple, surveiller l'engagement des étudiants en temps réel afin d'automatiser le déclenchement du processus de génération de contenu dynamique au moment de la perte de concentration sur le cours. Ils seraient aussi utilisés comme moyens d'affichage intelligent pour les étudiants.

ÉTENDRE L'ASPECT APPRENTISSAGE PERSONNALISÉ

Nous prévoyons d'expérimenter avec une variété de prompts pour adapter précisément les réponses en fonction de chaque étudiant dans le cadre d'un apprentissage personnalisé.

L'apprentissage personnalisé est une approche pédagogique centrée sur les besoins, les intérêts, les objectifs et les capacités de chaque apprenant. Pour réaliser cela, nous pouvons utiliser différents prompts personnalisés qui prennent en compte le niveau académique de chaque étudiant et de son progrès personnel dans une matière donnée, améliorant ainsi sa compréhension en fonction de son niveau et comblant ses lacunes de façon optimisée.

ÉTENDRE L'ASPECT APPRENTISSAGE ADAPTÉ VERS UNE ÉDUCATION SPÉCIALISÉE

Nous envisageons d'étendre les capacités de notre système pour inclure des solutions d'apprentissage adapté, spécialement conçues pour les étudiants ayant des besoins spécifiques.

L'apprentissage adapté ou spécialisé fait référence à des méthodes éducatives personnalisées qui prennent en compte les particularités individuelles des apprenants tels que les étudiants menés de troubles visuels ou auditifs, d'handicap physique, sensoriel ou cognitif, ou de troubles de l'apprentissage comme la dyslexie ou le TDAH, ou même les étudiants surdoués.

Par exemple, pour un étudiant avec des troubles visuels, notre système pourrait intégrer des fonctionnalités de synthèse vocale pour lire le texte à haute voix, tandis qu'un élève surdoué pourrait recevoir des défis supplémentaires pour le stimuler au-delà du curriculum standard.

INTÉGRATION COMME UN PLUG-IN

Enfin, nous envisageons d'intégrer notre système comme un plugin pour des plateformes d'e-learning telles qu'eCampus, où nous pourrions tirer pleinement parti des données déjà disponibles sur ces plateformes, telles que les informations détaillées sur les étudiants, leur niveau d'études, et leurs performances scolaires. Cette intégration facilitera non seulement l'implémentation et l'accessibilité à notre solution, mais optimisera également l'apprentissage personnalisé et adapté, car en exploitant directement ces données, notre système pourra identifier précisément les lacunes et les besoins spécifiques de chaque étudiant, permettant ainsi une personnalisation plus fine des prompts personnalisés à chaque apprenant. De plus, la présence conjointe des deux partenaires essentiels de notre solution, l'enseignant et l'étudiant sur ces plateformes, rendra l'adoption de notre technologie plus pratique et immédiatement bénéfique.

7 CONCLUSION

En conclusion, dans le but d'améliorer l'expérience éducative, cette étude a été élaborée en ciblant la dynamisation des contenus pédagogiques et des interactions dans les environnements éducatifs par l'intégration des grands modèles de langage (**LLMs**) et de l'Internet des Objets (**IoT**). Pour cela, nous avons mis en œuvre plusieurs scénarios différents pour notre approche et, lors de notre implémentation, nous avons réalisé les aspects liés aux **LLMs** de ces scénarios. Nous avons discuté notre approche en abordant les paramètres clés de son implémentation, tels que l'impact de la température dans la génération de contenu et l'utilisation stratégique des prompts pour optimiser les réponses du système. Notre mise en œuvre a prouvé que l'approche n'est pas seulement techniquement viable et implémentable, mais qu'elle est également prometteuse pour enrichir les interactions et le contenu pédagogique utilisé en classe, afin d'atteindre un apprentissage personnalisé à la faveur de chaque apprenant, avec tous les défis que cela peut représenter. Ces possibilités prometteuses nous motivent à poursuivre la mise en œuvre de ce projet en optant le concrétiser les environnements éducatifs actuels.

Acronyms

API Application Programming Interface

IA Intelligence Artificielle

IoT Internet of Things

LLM Large Language Model

RAG Retrieval-Augmented Generation

RFID Radio Frequency Identification

References

- [1] Shute, V. J., & Zapata-Rivera, D., 2012. Adaptive educational systems. In: *Adaptive Technologies for Training and Education*, 7(27), 1-35.
- [2] F. M. Aldhafeeri and A. A. Alotaibi. "Effectiveness of digital education shifting model on high school students' engagement," *Education and Information Technologies*, vol. 27, no. 5, pp. 6869–6891, 2022.
- [3] TeachTarget. "Exploring AI Use Cases Across Education and Government." Retrived 26/07/2022 from: <https://www.techtarget.com/searchcio/Flexibl>
- [4] Sadiku, M. N., Ashaolu, T. J., Ajayi-Majebi, A., & Musa, S. M. (2021). Artificial Intelligence in Education. *International Journal of Scientific Advances (IJSCIA)*, 2(1), 5-11.
- [5] Nguyen, Nathan. (2023). *Exploring the role of AI in education*. London Journal of Social Sciences. pp. 84-95. DOI: 10.31039/ljss.2023.6.108.
- [6] Wu, J., Gan, W., Chen, Z., Wan, S., & Lin, H. (2023). AI-generated content (AIGC): A survey. *arXiv preprint*, arXiv:2304.06632.
- [7] Bhargava, P., & Ng, V. (2022). Commonsense knowledge reasoning and generation with pre-trained language models: A survey. In *The AAAI Conference on Artificial Intelligence*, pp. 12317–12325.
- [8] Raj, N. S., & Renumol, V. (2022). Une revue systématique de la littérature sur les recommandations de contenu adaptatif dans les environnements d'apprentissage personnalisés de 2015 à 2020. *Journal of Computers in Education*, vol. 9, no. 1, pp. 113-148.
- [9] Wang, Z., Yan, W., Zeng, C., Tian, Y., Dong, S. et al. (2023). Un cadre de diagnostic d'apprentissage intelligent interprétable unifié pour la prédiction des performances d'apprentissage dans les systèmes de tutorat intelligents. *International Journal of Intelligent Systems*, vol. 2023.
- [10] Sula, A., Spaho, E., Matsuo, K., Barolli, L., Miho, R., & Xhafa, F. (2013). An IoT-based system for supporting children with autism spectrum disorder. In *2013 Eighth International Conference on Broadband and Wireless Computing, Communication and Applications*, pp. 282–289.
- [11] Zeeshan, K., Hämmäläinen, T., & Neittaanmäki, P. (2021). Internet of Things for Sustainable Smart Education: An Overview. *Sustainability*, 14(7), 4293. <https://doi.org/10.3390/su14074293>
- [12] Ghashim, I. A., & Arshad, M. (2022). Internet of Things (IoT)-Based Teaching and Learning: Modern Trends and Open Challenges. *Sustainability*, 15(21), 15656. <https://doi.org/10.3390/su152115656>
- [13] Cui, H., Du, Y., Yang, Q., Shao, Y., Liew, S. C., 2024. LLMind: Orchestrating AI and IoT with LLM for Complex Task Execution. *arXiv preprint arXiv:2312.09007*.
- [14] Amazon Web Services (AWS). "Retrieval-Augmented Generation." Accessed in 2023. <https://aws.amazon.com/fr/what-is/retrieval-augmented-generation/>
- [15] Luque-Casado A, Zabala M, Morales E, Mateo-March M, Sanabria D. Cognitive performance and heart rate variability: the influence of fitness level. *PLoS One*. 2013;8(2):e56935. doi: 10.1371/journal.pone.0056935. Epub 2013 Feb 20. PMID: 23437276; PMCID: PMC3577676.
- [16] Wright KP Jr, Hull JT, Czeisler CA. Relationship between alertness, performance, and body temperature in humans. *Am J Physiol Regul Integr Comp Physiol*. 2002 Dec;283(6):R1370-7. doi: 10.1152/ajpregu.00205.2002. Epub 2002 Aug 15. PMID: 12388468.
- [17] Shi, Y., et al., 2007. Galvanic skin response (gsr) as an index of cognitive load. In: *CHI'07 extended abstracts on Human factors in computing systems*, 2651–2656.