



UNIVERSITÉ TÉLUQ
UNIVERSITÉ DU QUÉBEC À MONTRÉAL

AGENTS CONVERSATIONNELS PÉDAGOGIQUES BASÉS SUR L'INTELLIGENCE ARTIFICIELLE GÉNÉRATIVE
POUR AMÉLIORER L'APPRENTISSAGE EN LIGNE

THÈSE
PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN INFORMATIQUE COGNITIVE

PAR
FATMA MILADI



<https://r-libre.teluq.ca/3816>

JUILLET 2025

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude à mon directeur de thèse, Daniel Lemire, et à ma codirectrice, Valéry Psyché, professeurs à l'Université TÉLUQ. Merci de m'avoir offert l'opportunité de réaliser cette thèse, pour votre confiance, votre encadrement bienveillant et vos conseils éclairés tout au long de ce parcours. Votre disponibilité, vos interventions pertinentes et votre écoute attentive ont grandement contribué à l'aboutissement de ce travail. Ce fut un réel plaisir d'évoluer à vos côtés.

À mes parents, Amira et Mounir, je dois tout. Merci pour votre patience, vos sacrifices et votre amour inconditionnel, qui ont toujours été une source de force, de stabilité et de courage. À ma sœur Emna, je suis profondément reconnaissante pour ton soutien constant. Tu as su m'épauler avec douceur et générosité, remonter mon moral dans les moments les plus difficiles et être présente quand j'en avais le plus besoin. À mon mari Abderrahim, je te remercie du fond du cœur. Ton soutien constant, ta foi en nous et ton optimisme inépuisable m'ont portée dans les moments les plus exigeants de cette aventure. Je suis comblée de t'avoir comme partenaire de vie. Je remercie chaleureusement ma belle-famille, Wafa, Abdeljalil, Wissal, Abderrahim, Fatma et Mohamed Ali, pour leur bienveillance et leur soutien tout au long de ce parcours.

Je souhaite exprimer toute ma reconnaissance à mon amie Henda, pour son soutien continu tout au long de ces années de thèse, ses encouragements sans limites et sa présence bienveillante à chaque étape de ce parcours. Je remercie également Salwa et Riadh pour leur soutien amical, leurs conseils et leur écoute attentive, qui m'ont été d'un grand réconfort.

Je remercie sincèrement les membres du jury pour le temps accordé à l'évaluation de ce travail et pour leurs retours constructifs. Merci à toute l'équipe du Dot-Lab, en particulier les professeurs Nicolas Bélanger et Wassim Bouachir, ainsi que mes collègues Pierre, Faten, Ahmed, Safwen, Hela et Alaa, pour les échanges stimulants, l'entraide et l'ambiance collaborative.

Je remercie également Nour El Mawas et Awa Diattara pour leur précieuse collaboration et leur aide généreuse. Leur disponibilité, leur rigueur et leur esprit d'équipe ont été d'une grande valeur tout au long du projet, et j'ai beaucoup apprécié nos échanges constructifs et leur engagement sincère. Enfin, j'adresse

mes remerciements à Richard Hotte, Karim Ait-Seddik, Djamila Abbas, Samia Sekher et Mélanie Samson pour leur accompagnement, leur soutien et leur implication à différents moments de mon parcours.

TABLE DES MATIÈRES

TABLE DES FIGURES	viii
LISTE DES TABLEAUX	ix
ACRONYMES	x
RÉSUMÉ	xi
INTRODUCTION	1
CHAPITRE 1 PROBLÉMATIQUE	6
1.1 Questions de recherche	11
1.2 But général	11
1.2.1 Objectifs de recherche	12
1.3 Conclusion	13
CHAPITRE 2 CADRE CONCEPTUEL ET REVUE DE LA LITTÉRATURE	14
2.1 Fondements conceptuels et théoriques.....	14
2.1.1 Les grands modèles de langage	14
2.1.2 Génération augmentée par récupération	30
2.1.3 Les modèles théoriques et leurs implications	33
2.2 Revue de la littérature	36
2.2.1 Historique et évolution des agents conversationnels	37
2.2.2 Agents conversationnels pédagogiques en éducation	41
2.2.3 Agents conversationnels basés sur les grands modèles de langage en éducation	47
2.2.4 Agents conversationnels basés sur la génération augmentée par récupération en éducation	48
2.3 Conclusion	50
CHAPITRE 3 CONCEPTION ET ÉVALUATION DE L'AGENT CONVERSATIONNEL BASÉ SUR GPT-4 ET RAG	52

3.1	Conception de l'Agent Conversationnel Basé sur GPT-4 et RAG.....	53
3.1.1	GPT-4.....	53
3.1.2	RAG	53
3.2	Objectifs.....	58
3.3	Hypothèses de recherche	59
3.4	Méthodologie d'évaluation.....	60
3.4.1	Ressources pédagogiques et données	60
3.4.2	Protocole expérimental.....	61
3.5	Résultats	64
3.5.1	Analyse des performances des modèles GPT avec l'intégration de la RAG dans un MOOC	64
3.5.2	Comparaison des performances du GPT-4 enrichi par la RAG avec celles des apprenants dans un MOOC.....	66
3.6	Discussion	68
3.6.1	(RQ1) : Dans quelle mesure l'intégration de la RAG dans les modèles GPT permet-elle de réduire les erreurs d'hallucination et d'améliorer la précision des réponses générées dans un contexte éducatif, notamment lors de l'utilisation d'exercices issus d'un MOOC sur l'intelligence artificielle ?.....	68
3.6.2	Implications pédagogiques de l'intégration du RAG dans les modèles GPT	70
3.6.3	Limites.....	71
3.7	Conclusion	72
CHAPITRE 4 MESURE DE L'EFFET D'UN AGENT CONVERSATIONNEL AMÉLIORÉ PAR LA RAG SUR L'ACQUISITION DES CONNAISSANCES ET L'EXPÉRIENCE DES APPRENANTS.....		74
4.1	Objectifs.....	75
4.2	Hypothèses de recherche	76
4.3	Méthodologie de recherche	77

43.1	Contexte de l'étude	77
43.2	Configuration du prompt pour l'agent conversationnel	78
43.3	Méthodes de collecte de données	79
43.4	Participants à la recherche	79
43.5	Protocole expérimental.....	80
43.6	Instruments de recherche	82
44	Résultats	84
44.1	Approche quantitative de la recherche	85
44.2	Approche qualitative de la recherche	88
45	Discussion	91
45.1	Effet du robot conversationnel sur l'apprentissage des apprenants (H1)	91
45.2	Perception de l'utilisabilité du robot conversationnel (H2)	92
45.3	Expérience d'interaction avec le robot conversationnel (H3).....	93
46	Limites de l'étude expérimentale.....	93
47	Conclusion	94
CHAPITRE 5 ANALYSE DES EFFETS D'UN AGENT CONVERSATIONNEL MOTIVANT SUR LES CONNAIS-		
SANCES ET LES ÉMOTIONS DANS UN MOOC.....		
51	Hypothèses de recherche	97
52	Méthodologie de recherche	98
52.1	Participants	98
52.2	Protocole expérimental.....	99
52.3	Instruments.....	103
53	Résultats	104

53.1	Résultats de l'acquisition de connaissances	104
53.2	Résultats des émotions	105
53.3	Analyse des perceptions de l'intelligence des participants	107
54	Discussion	107
54.1	Efficacité sur l'acquisition des connaissances (H1)	108
54.2	Effet sur les émotions des apprenants (H2a, H2b, H2c)	108
55	Conclusion	110
CHAPITRE 6 CONCLUSION ET PERSPECTIVES		111
6.1	Contributions scientifiques	112
6.2	Discussion	112
6.3	Perspectives de recherche	115
ANNEXE A CERTIFICAT D'ÉTHIQUE		118
ANNEXE B IMPLÉMENTATION.....		120
BIBLIOGRAPHIE		121

TABLE DES FIGURES

Figure 2.1	Architecture simple d'un réseau de neurones artificiels avec une seule couche cachée.	16
Figure 2.2	Illustration d'une analogie vectorielle dans l'espace des représentations.	20
Figure 2.3	Projection en 3D des représentations vectorielles. Les mots masculins et féminins apparaissent respectivement sur les moitiés gauche et droite de l'axe x , tandis que les mots relatifs aux adultes et aux jeunes se situent dans les moitiés supérieure et inférieure de l'axe y (Bandyopadhyay <i>et al.</i> , 2022).....	22
Figure 2.4	L'architecture du transformeur (Vaswani <i>et al.</i> , 2017).	24
Figure 3.1	Architecture complète de l'agent conversationnel GPT-RAG.....	54
Figure 3.2	Prompt utilisé dans notre agent conversationnel	59
Figure 3.3	Scores en pourcentage de la moyenne des apprenants (bleu) et du modèle GPT-4 augmenté par RAG (vert) pour différents types d'exercices	67
Figure 4.1	Prompt utilisé dans notre agent conversationnel	77
Figure 4.2	Procédure expérimentale	82
Figure 4.3	Scores moyens des pré-tests et post-tests pour les groupes expérimental et contrôle pour le module 1 du MOOC	86
Figure 4.4	Boxplot des scores SUS pour le robot conversationnel (Médiane 82.5, Moyenne 80.4).....	88
Figure 4.5	Échelle SUS de Bangor (Bangor <i>et al.</i> , 2009) et Score SUS pour l'Agent Conversationnel (Valeur Moyenne).	92
Figure 5.1	Procédure expérimentale	100
Figure B.1	Exemple de Code Python pour créer le contexte.....	120
Figure B.2	Exemple de Code Python pour générer la réponse.....	120

LISTE DES TABLEAUX

Table 2.1	Comparaison des méthodes de prompt.....	29
Table 3.1	Types d'exercices utilisés pour évaluer les performances des modèles GPT	61
Table 3.2	Résultats de l'évaluation des exercices Vrai/Faux.....	65
Table 3.3	Résultats de l'évaluation des QCMs.....	65
Table 3.4	Résultats de l'évaluation des exercices d'appariement	66
Table 3.5	Résultats de l'évaluation des exercices à trous	66
Table 3.6	Médiane et mode pour chaque type d'exercice.....	68
Table 4.1	Concepts de connaissances abordés dans les questions de pré-test et post-test.....	83
Table 4.2	Questions de pré-test.....	84
Table 4.3	Questions de post-test	85
Table 4.4	Analyse des acquisitions des connaissances au pré-test	87
Table 4.5	Analyse des acquisitions des connaissances au post-test.....	87
Table 4.6	Questionnaire SUS et statistiques pour chaque affirmation.....	87
Table 5.1	Résultats statistiques pour le pré-test de connaissances.....	104
Table 5.2	Résultats statistiques du post-test de connaissances	105
Table 5.3	Résultats statistiques pour les émotions	106

ACRONYMES

AI Artificial Intelligence.

IAGén L'intelligence artificielle générative.

NLP Natural Language Processing.

LLM Large Language Model (Grand Modèle de Langage).

GPT Generative Pre-trained Transformer.

RAG Retrieval-Augmented Generation (Génération augmentée par récupération).

MOOC Massive Open Online Course.

SUS System Usability Scale.

RÉSUMÉ

L'intelligence artificielle (IA) a considérablement évolué ces dernières années, notamment avec l'apparition des modèles de langage de grande taille (Large Language Models, LLMs) tels que GPT. Ces modèles, capables de générer des réponses textuelles cohérentes et fluides, offrent un potentiel prometteur pour transformer divers secteurs, y compris l'éducation. Cependant, leur application dans des contextes éducatifs soulève des difficultés importantes, notamment les « hallucinations » — ou génération de contenus incorrects — qui peuvent compromettre la qualité de l'apprentissage des apprenants. En outre, ces modèles ne proposent souvent pas un accompagnement pédagogique adapté, pourtant essentiel pour stimuler la motivation des apprenants et répondre efficacement à leurs besoins spécifiques.

Pour surmonter ces limitations, cette thèse présente le développement d'un agent conversationnel pédagogique, conçu comme un véritable compagnon d'apprentissage. Cet agent intègre la Génération Augmentée par Récupération (RAG), une méthode qui combine les LLMs avec un accès en temps réel à une base de connaissances externe validée, issue d'un corpus pédagogique associé au cours en ligne. Grâce à cette intégration, l'agent améliore la précision, la pertinence et la fiabilité des réponses fournies. De plus, des fonctionnalités pédagogiques ont été intégrées pour accompagner et soutenir les apprenants dans leur parcours d'apprentissage. Cependant, l'évaluation empirique de cet agent dans des contextes éducatifs réels reste peu explorée, notamment en ce qui concerne son influence sur l'acquisition des connaissances et l'expérience globale des apprenants.

Cette étude s'articule autour de la conception et de l'évaluation d'un agent conversationnel dans le cadre d'un cours en ligne ouvert et massif (MOOC) sur l'IA. Ces cours, caractérisés par un apprentissage à distance, asynchrone et autonome, sont souvent associés à un taux d'abandon élevé. Dans ce contexte, les agents conversationnels jouent un rôle crucial en compensant l'absence de l'enseignant et en offrant un accompagnement aux apprenants. Notre étude explore trois axes principaux : améliorer la qualité des réponses fournies par l'agent, examiner son influence sur l'acquisition des connaissances, et comprendre son influence potentielle sur les émotions des apprenants, ces deux derniers éléments étant essentiels pour soutenir leur réussite. Contrairement aux recherches existantes, cette étude évalue l'agent conversationnel pédagogique en interaction directe avec des apprenants réels dans un contexte d'apprentissage réel.

Notre thèse s'appuie sur plusieurs expériences réalisées avec des étudiants dans le cadre d'un MOOC sur l'intelligence artificielle. Nos résultats montrent que l'intégration de la RAG dans l'agent conversationnel (1) réduit les erreurs d'hallucination et améliore la précision des réponses fournies dans le contexte d'un MOOC, (2) a un effet positif sur l'acquisition des connaissances des apprenants, et (3) augmente leur satisfaction globale. En revanche, l'ajout d'éléments motivationnels dans l'agent conversationnel n'a pas eu d'effet significatif sur les émotions des apprenants, telles que la confiance, l'enthousiasme et l'intérêt. Fondée sur des données empiriques, notre thèse est une avancée dans le domaine de l'application des LLMs en éducation.

Mots clés: Intelligence Artificielle (IA), Modèles de Langue de Grande Taille (LLMs), Génération Augmentée par Récupération (RAG), Emotion, Agents conversationnels pédagogiques, MOOCs, Acquisition des connaissances.

INTRODUCTION

L'émergence des modèles de langage de grande taille (LLMs), qui sont des réseaux de neurones entraînés sur d'immenses volumes de données textuelles, a révolutionné le domaine de l'intelligence artificielle (IA) (Brown *et al.*, 2020; Vaswani *et al.*, 2017). Ces modèles, comme GPT, sont conçus pour comprendre et générer le langage humain de manière naturelle, facilitant des interactions fluides et contextuelles. Ils ouvrent la voie à une large gamme d'applications, telles que la traduction automatique, le résumé de texte ou l'analyse de sentiments. Leur capacité à fournir des réponses cohérentes a transformé les interactions homme-machine et ouvert de nouvelles perspectives dans divers secteurs.

Dans le domaine de la finance, par exemple, les modèles de langage de grande taille facilitent l'analyse prédictive et stratégique en exploitant de vastes ensembles de données financières pour améliorer les décisions stratégiques (Wu *et al.*, 2023). Dans le secteur de la santé, les LLMs permettent des prédictions personnalisées en exploitant des données physiologiques, comportementales et contextuelles pour améliorer la gestion des soins et prévenir les maladies (Kim *et al.*, 2024).

Dans le domaine de l'éducation, les LLMs se sont révélés particulièrement prometteurs, facilitant la création de contenus pédagogiques, tels que des quiz (Dijkstra *et al.*, 2022) et l'accompagnement des apprenants par des agents conversationnels (Abdelghani *et al.*, 2022; Xie *et al.*, 2024). Ces outils offrent un soutien individualisé, clarifient des concepts complexes (Griol et Callejas, 2013), et réduisent l'isolement des apprenants grâce à des interactions engageantes et adaptées (Roos, 2018). En intégrant ces agents conversationnels dans les environnements numériques, ils participent à la création d'un cadre d'apprentissage interactif et motivant, renforçant l'engagement des apprenants tout en contribuant à réduire les taux d'abandon (Winkler et Söllner, 2018; Roos, 2018).

Cependant, malgré leurs nombreux avantages, les LLMs présentent des limites importantes, notamment leur tendance aux « hallucinations »—la génération de contenus incorrects ou hors contexte (Alkaissi et McFarlane, 2023; Ji *et al.*, 2023). Dans un cadre éducatif, où la fiabilité et la précision des informations sont essentielles, ces erreurs peuvent induire des incompréhensions chez les apprenants, fausser leur apprentissage et compromettre leur progression. De plus, elles risquent d'affaiblir la confiance des apprenants dans

les outils pédagogiques, réduisant ainsi leur engagement et leur capacité à atteindre les objectifs éducatifs visés. Ces effets négatifs sur l'expérience et les résultats académiques rendent la gestion des hallucinations essentielle pour garantir l'efficacité de l'apprentissage dans un contexte soutenu par l'IA.

Pour surmonter ces limites, la Génération Augmentée par Récupération (RAG) a été introduite (Lewis *et al.*, 2020). Contrairement aux LLMs traditionnels, qui s'appuient uniquement sur les données accumulées durant leur entraînement, la RAG permet d'accéder en temps réel à des sources externes validées, augmentant ainsi la précision et la pertinence des réponses (Lewis *et al.*, 2020). En intégrant des documents pertinents dans les réponses générées, les systèmes RAG visent à réduire les erreurs d'hallucination, tout en garantissant un contenu adapté aux besoins des apprenants. Ces systèmes se révèlent particulièrement utiles dans les contextes éducatifs, où l'accès à des informations fiables et contextualisées est essentiel.

L'application de la RAG aux LLMs ouvre la voie au développement d'agents conversationnels pédagogiques performants et fiables. Ces agents conversationnels, conçus pour interagir avec les utilisateurs en langage naturel, présentent un potentiel unique pour l'éducation. Ils permettent de mener des dialogues personnalisés avec les apprenants et de fournir un soutien individualisé tout au long du processus d'apprentissage (Kestin *et al.*, 2024). Ce soutien individualisé est crucial pour répondre aux besoins variés des apprenants et les accompagner de manière efficace dans leurs parcours éducatifs. Les recherches récentes ont démontré l'efficacité des agents conversationnels enrichis par la RAG pour améliorer la qualité des réponses, réduire les erreurs d'hallucination et offrir des explications précises et transparentes (Taneja *et al.*, 2024; Yan *et al.*, 2024).

Il devient indispensable d'évaluer empiriquement les agents conversationnels enrichis par la RAG non seulement en termes de précision des réponses, mais aussi en examinant leur influence sur l'acquisition des connaissances, l'expérience des apprenants et les émotions associées à l'apprentissage. Pourtant, malgré les avancées technologiques, les études empiriques dans des environnements éducatifs restent rares. La majorité des recherches existantes se concentrent principalement sur des métriques techniques, telles que la précision ou la pertinence des réponses, sans explorer en profondeur leur contribution réelle à l'amélioration des processus d'apprentissage et à l'enrichissement de l'expérience des apprenants (Wang *et al.*, 2023; Taneja *et al.*, 2024; Liu *et al.*, 2024). Ces dimensions, encore peu étudiées, sont pourtant essentielles

pour garantir des environnements éducatifs efficaces et engageants.

Les cours en ligne ouverts et massifs (MOOCs), avec leur diversité d'apprenants et leur apprentissage autonome, offrent un cadre idéal pour évaluer l'efficacité des agents conversationnels enrichis par la RAG. Ces environnements, souvent caractérisés par un apprentissage à distance asynchrone, exigent une grande autonomie de la part des apprenants. Dans ce cadre, les agents conversationnels jouent un rôle clé en apportant un soutien personnalisé et interactif, capable de compenser l'absence de l'enseignant en temps réel. La précision des réponses, l'amélioration des connaissances, et la capacité à encourager des émotions positives chez les apprenants sont autant d'éléments cruciaux pour garantir un apprentissage de qualité.

Ainsi, des agents conversationnels capables de fournir des réponses pédagogiquement solides, adaptées et contextualisées peuvent clarifier des concepts complexes, réduire les erreurs d'hallucination, et améliorer la compréhension des contenus pédagogiques. Ces outils peuvent également influencer sur la motivation des apprenants en stimulant des émotions positives comme la confiance, l'intérêt et l'enthousiasme, des facteurs essentiels pour la persévérance et la réussite académique (Pekrun *et al.*, 2007). Inspirés par des travaux sur les états d'esprit des apprenants (Dweck, 2002), ces agents aident les étudiants à transformer les difficultés en opportunités, renforçant ainsi leur engagement et leur progression.

Cette recherche s'appuie sur des agents conversationnels basés sur l'IA générative, conçus comme des compagnons d'apprentissage (Chan et Baskin, 1988). Grâce à des consignes pédagogiques spécifiques, ces agents visent à fournir des réponses contextualisées et stimuler des émotions positives. Ils proposent un cadre d'apprentissage interactif, adapté aux défis rencontrés dans les environnements numériques.

Cette thèse se concentre sur l'évaluation des effets des agents conversationnels enrichis par la RAG dans un MOOC dédié à l'intelligence artificielle, à travers plusieurs expériences réalisées avec des étudiants. Elle examine leurs effets sur l'acquisition des connaissances et sur l'expérience des apprenants, en tenant compte de leur perception des agents et de leur utilité. De plus, elle explore l'effet des stratégies motivationnelles intégrées, basées sur la RAG, sur des émotions clés pour l'apprentissage, telles que la confiance, l'intérêt et l'enthousiasme. Ces émotions, essentielles à la persévérance et à la réussite académique (Pekrun *et al.*, 2007), font l'objet d'une analyse approfondie. Cette recherche vise à concevoir une expérience éducative

en ligne engageante, fiable et adaptée aux besoins des apprenants.

Cette thèse est organisée en six chapitres pour suivre une progression logique et cohérente, allant des fondements théoriques à l'évaluation pratique des agents conversationnels enrichis par RAG. Chaque chapitre a été conçu pour répondre aux différentes questions de recherche et approfondir progressivement les aspects clés étudiés.

- *Chapitre 1* : Problématique. Ce chapitre présente le contexte général de la recherche, introduit la problématique et définit les objectifs de la thèse.
- *Chapitre 2* : Cadre conceptuel et revue de la littérature. Ce chapitre explore les bases théoriques des LLMs, de la RAG et de leurs applications en éducation, en analysant leur utilisation dans ce domaine.
- *Chapitre 3* : Conception et évaluation de l'agent conversationnel basé sur GPT-4 et RAG. Ce chapitre décrit le processus de conception de l'agent conversationnel et évalue sa capacité à fournir des réponses exactes et contextuellement pertinentes, en utilisant des exercices d'évaluation formative comme étape préliminaire avant de tester l'agent avec de vrais apprenants. Les résultats présentés dans ce chapitre ont donné lieu à deux contributions scientifiques publiées. La première communication, intitulée « Leveraging GPT-4 for Accuracy in Education : A Comparative Study on Retrieval-Augmented Generation in MOOCs », a été publiée. La deuxième communication, intitulée « Comparative Performance of GPT-4, RAG-Augmented GPT-4, and Students in MOOCs », a également été publiée.
- *Chapitre 4* : Mesure de l'effet d'un agent conversationnel enrichi par la RAG sur l'acquisition des connaissances et l'expérience des apprenants. Ce chapitre explore les résultats expérimentaux sur l'acquisition des connaissances, l'utilisabilité, et l'expérience d'interaction avec l'agent conversationnel, basés sur une approche mixte combinant des méthodes quantitatives (pré-tests, post-tests, questionnaires) et qualitatives (entretiens semi-structurés). Les résultats présentés dans ce chapitre ont donné lieu à plusieurs contributions scientifiques. Ils font l'objet d'un article soumis pour publication, intitulé « Conception et évaluation d'un agent conversationnel enrichi par la génération augmentée par récupération sur l'acquisition des connaissances et l'expérience des apprenants », ainsi qu'une communication soumise pour publication, intitulé « Evaluating a GPT-4 and Retrieval-Augmented Generation-Based Conversational Agent to Enhance Learning Experience in a MOOC ».

De plus, une communication orale a été réalisée lors de la Journée annuelle sur l'intelligence artificielle multidisciplinaire.

- *Chapitre 5* : Analyse des Effets d'un Robot Conversationnel Motivant sur les Connaissances et les Émotions dans un MOOC. Ce chapitre examine l'effet de l'utilisation d'un robot conversationnel motivant sur l'acquisition des connaissances et les émotions des apprenants dans un MOOC dédié à l'intelligence artificielle. L'objectif est d'évaluer si l'intégration des éléments motivationnels peut favoriser l'acquisition des connaissances et renforcer des émotions positives chez les apprenants telles que la confiance, l'enthousiasme et l'intérêt.
- *Chapitre 6* : Conclusion générale et perspectives de recherche. Ce chapitre résume les contributions de la thèse et propose des pistes pour des recherches futures afin d'améliorer l'intégration des LLMs dans les systèmes éducatifs.

Cette structure a été choisie pour garantir une analyse progressive et approfondie des différentes facettes de l'utilisation de la RAG dans les environnements éducatifs, tout en permettant une évaluation globale des effets sur la précision des réponses, l'expérience des apprenants et leur aspects émotionnels.

CHAPITRE 1

PROBLÉMATIQUE

L'intelligence artificielle générative (IAGén) désigne des systèmes d'intelligence artificielle capables de générer de nouvelles données ou productions, telles que des images, de la musique ou du texte, en utilisant l'apprentissage profond (Gimpel *et al.*, 2023). Les modèles de langage de grande taille (LLMs) appartiennent à cette catégorie, car ils peuvent générer du texte en langage naturel en s'appuyant sur des modèles entraînés sur de vastes corpus de données (Gimpel *et al.*, 2023). Ces capacités reposent sur le traitement du langage naturel (NLP), qui permet aux ordinateurs de comprendre, analyser et produire le langage humain.

Des avancées majeures ont été réalisées dans ce domaine grâce à l'introduction des modèles transformateurs (Vaswani *et al.*, 2017). Ces modèles, tels que BERT et GPT, s'appuient sur un mécanisme d'attention capable d'analyser chaque mot en relation avec tous les autres dans un texte, offrant ainsi une compréhension contextuelle plus précise et une vitesse de traitement nettement améliorée.

Les agents conversationnels, appelés en anglais « chatbots », sont des systèmes d'intelligence artificielle conçus pour interagir avec les humains en langage naturel. Dans un contexte éducatif, les agents conversationnels pédagogiques intelligents se distinguent par leur capacité à favoriser l'apprentissage en ligne en jouant un rôle essentiel dans l'accompagnement des apprenants. Ces outils offrent un soutien personnalisé en répondant aux besoins spécifiques de chaque apprenant, tout en améliorant la qualité de l'apprentissage grâce à des retours immédiats et adaptés. Ils facilitent la compréhension de concepts complexes, personnalisent les processus d'apprentissage et enrichissent l'environnement éducatif (Griol et Callejas, 2013). Particulièrement adaptés aux environnements d'apprentissage à distance ou asynchrones, ils ont la capacité de compenser l'absence d'un enseignant en temps réel, favorisant ainsi un apprentissage autonome

tout en réduisant l'isolement des apprenants (Winkler et Söllner, 2018). Grâce à leur capacité à structurer les interactions pédagogiques, ils augmentent l'engagement et la sociabilité des apprenants en rendant les échanges plus interactifs et dynamiques (Roos, 2018).

En s'appuyant sur les avancées des LLMs, ces agents conversationnels peuvent désormais offrir des interactions plus fluides et contextuellement riches, élargissant ainsi leurs applications dans l'éducation (Gimpel *et al.*, 2023). ChatGPT, un exemple d'agent conversationnel reposant sur des LLMs, s'appuie sur les modèles GPT-3.5 et GPT-4, développés par OpenAI ¹. Depuis son lancement en novembre 2022, les LLMs ont montré des perspectives prometteuses dans divers domaines, y compris en éducation (Kasneci *et al.*, 2023).

Dans le domaine éducatif, les agents conversationnels basés sur les LLMs soulèvent un intérêt croissant pour leur capacité potentielle à faciliter des interactions naturelles et engageantes avec les apprenants. Bien que leur efficacité soit encore peu explorée, des travaux initiaux suggèrent qu'ils pourraient enrichir les processus d'apprentissage. Par exemple, Abdelghani *et al.* (2022) ont intégré GPT-3 dans une plateforme éducative pour encourager les enfants à poser des questions divergentes, stimulant leur réflexion critique et créative. Leur étude a montré que les instructions générées par GPT-3 favorisaient une plus grande variété de questions ouvertes, soulignant son potentiel à enrichir l'apprentissage interactif. De plus, Xie *et al.* (2024) ont analysé l'effet des agents conversationnels sur l'autonomie d'apprentissage. Leur étude montre que les compagnons virtuels, grâce à leur capacité à créer une présence sociale accrue, favorisent l'autonomie des apprenants. En revanche, les outils principalement conçus pour fournir des réponses directes aux questions des apprenants peuvent entraîner une dépendance, limitant ainsi la capacité des apprenants à apprendre de manière autonome. De plus, les LLMs peuvent renforcer les interactions sociales dans les forums des MOOCs. Li et Xing (2021) ont ainsi montré que GPT-2, en générant des réponses automatisées émotionnelles, informatives et communautaires, augmente la satisfaction et l'engagement des apprenants.

1. <https://openai.com/>

Malgré leur potentiel, les LLMs présentent des limites importantes, notamment leur tendance aux « hallucinations »—la génération de contenus incorrects ou non pertinents (Alkaissi et McFarlane, 2023; Ji *et al.*, 2023). Par exemple, ChatGPT a été observé produisant des références académiques et des articles fictifs, dotés d’une apparence crédible mais totalement inventés (Alkaissi et McFarlane, 2023). En outre, comme le souligne Qadir (2023), les informations produites par ChatGPT et autres modèles génératifs dépendent fortement de la qualité des données d’entraînement, ce qui peut entraîner la propagation de désinformation ou de biais. Ces erreurs sont particulièrement problématiques dans des contextes tels que l’éducation, où l’exactitude et la fiabilité des informations sont fondamentales. Elles peuvent non seulement perturber la compréhension des apprenants, mais également compromettre la qualité de leur apprentissage en introduisant des données erronées ou non vérifiables.

Pour remédier à ces limitations, plusieurs approches ont été développées, notamment l’ajustement fin (fine-tuning) et la génération augmentée par récupération (retrieval-augmented generation). L’ajustement fin (fine-tuning) consiste à adapter un LLM pré-entraîné avec un jeu de données spécifique (Wei *et al.*, 2021). Bien que cette méthode améliore la précision dans des domaines particuliers, elle est coûteuse et longue, nécessitant des ressources importantes en termes de données et de calcul Touvron *et al.* (2023).

La Génération Augmentée par Récupération (RAG) est une méthode hybride qui combine les capacités génératives des grands modèles de langage avec des techniques avancées de récupération d’informations (Gao *et al.*, 2023). Cette approche permet de réduire les hallucinations en fournissant au modèle des documents pertinents extraits de bases de connaissances externes validées. Elle repose sur la recherche et la récupération de documents correspondant sémantiquement à une requête utilisateur, ces documents étant ensuite intégrés dans le prompt du modèle pour enrichir les réponses. Ce processus permet d’améliorer la fiabilité, la pertinence, et la précision des réponses tout en renforçant leur contextualisation (Lewis *et al.*, 2020). La RAG se révèle particulièrement efficace dans des contextes nécessitant des informations

actualisées et validées, notamment dans le domaine éducatif, où l'exactitude des réponses est cruciale.

La RAG joue un rôle crucial dans l'amélioration de la précision et de la pertinence des modèles de langage dans des contextes éducatifs. Par exemple, Wang *et al.* (2023) ont développé ChatEd, un agent conversationnel combinant la récupération contextuelle avec ChatGPT, permettant de fournir des réponses adaptées aux besoins spécifiques des cours tout en maintenant une vérifiabilité grâce à des références explicites aux documents sources. L'évaluation de ChatEd, basée sur des critères tels que la pertinence, l'exactitude et la vérifiabilité des réponses, a montré que l'intégration de la récupération contextuelle renforce l'efficacité pédagogique. De même, Liu *et al.* (2024) ont développé le CS50 Duck, un agent conversationnel basé sur GPT-4 et enrichi par RAG, intégré dans le cadre du cours afin d'accompagner les étudiants. Cet outil s'est révélé plus performant que ChatGPT pour fournir des réponses précises et alignées sur le contenu spécifique du cours. Taneja *et al.* (2024) ont présenté Jill Watson, une assistante virtuelle utilisant la RAG pour fournir des réponses plus précises et fiables, tout en réduisant les hallucinations. Cet outil a démontré une qualité de réponse nettement supérieure à celle des assistants dépourvus de cette technique RAG.

Toutefois, malgré ces avancées, les études empiriques sur l'impact des agents conversationnels enrichis par la RAG dans les environnements éducatifs restent limitées. Ces recherches se concentrent principalement sur des métriques techniques, telles que la précision, la pertinence ou la vérifiabilité des réponses, sans évaluer leur influence sur l'acquisition des connaissances, la motivation des apprenants ou les émotions suscitées par ces interactions. Ce manque de recul sur leur efficacité pédagogique est d'autant plus préoccupant dans les environnements d'apprentissage à distance tels que les MOOCs, où l'interaction humaine est souvent réduite. Dans ce contexte, où l'autonomie est essentielle, l'intégration d'agents conversationnels pourrait jouer un rôle clé dans l'accompagnement et le soutien des apprenants.

Les MOOC offrent un apprentissage flexible et autonome à un large public (Yuan et Powell, 2013). Toutefois,

cette autonomie s'accompagne de défis majeurs, notamment en matière de suivi et de soutien aux apprenants. L'absence de rétroaction immédiate et de soutien personnalisé peut entraîner une démotivation et une baisse des taux de rétention (Hone et El Said, 2016). De plus, le sentiment d'isolement est particulièrement marqué dans ce contexte (Bulger *et al.*, 2015). L'éloignement des pairs et des enseignants et enseignantes réduit l'engagement et augmente le risque d'abandon (Papi *et al.*, 2022). Bien que les forums de discussion visent à pallier ce manque d'interaction, leur taux de participation demeure faible, variant de 5 % à 12 % (Chiu et Hew, 2018). Face à ces limites, le développement d'outils pédagogiques interactifs apparaît comme une solution nécessaire pour mieux accompagner les apprenants dans les MOOC.

Parmi ces outils, les agents conversationnels enrichis par la RAG se distinguent par leur capacité à fournir des réponses contextualisées, adaptées aux besoins des apprenants et susceptibles d'améliorer leur engagement dans l'apprentissage. Pourtant, leur influence sur l'acquisition des connaissances des apprenants et l'expérience d'interaction reste peu étudiée. Des recherches approfondies sont nécessaires pour mieux comprendre leur contribution à l'apprentissage en ligne et optimiser leur conception pour un accompagnement plus efficace des apprenants.

Par ailleurs, des expériences émotionnelles positives telles que l'intérêt et l'enthousiasme jouent un rôle déterminant dans l'apprentissage, notamment dans des disciplines comme les mathématiques et les sciences, influençant les résultats à long terme bien au-delà de l'acquisition ponctuelle de connaissances (Ceci *et al.*, 2009). De plus, des études montrent que des émotions comme la confiance, l'ennui ou la confusion influencent directement la réussite des étudiants et permettent de prédire leur performance académique (Pekrun *et al.*, 2007).

Pourtant, ces dimensions émotionnelles demeurent largement sous-explorées dans les approches actuelles. À ce jour, aucune étude n'a exploité de manière empirique la synergie entre les capacités des LLM et de la

RAG pour développer des solutions visant à favoriser des émotions positives, telles que l'intérêt, la confiance ou l'enthousiasme, chez les apprenants. Une telle approche pourrait non seulement enrichir l'expérience émotionnelle des apprenants, mais également optimiser la qualité des interactions éducatives en offrant des réponses plus précises, contextualisées et adaptées aux besoins spécifiques des apprenants.

1.1 Questions de recherche

Cette thèse explore les questions de recherche suivantes :

1. *RQ1* : Dans quelle mesure l'intégration de la RAG dans les modèles GPT permet-elle de réduire les erreurs d'hallucination et d'améliorer la précision des réponses générées dans un contexte éducatif, notamment lors de l'utilisation d'exercices issus d'un MOOC sur l'intelligence artificielle ?
2. *RQ2* : Quels sont les effets d'un agent conversationnel enrichi par la RAG sur l'acquisition des connaissances, l'utilisabilité perçue et l'expérience des apprenants dans un contexte de MOOC sur l'intelligence artificielle ?
3. *RQ3* : Quels sont les effets d'un agent conversationnel motivationnel enrichi par GPT et la RAG sur l'acquisition des connaissances et les émotions des apprenants (confiance, intérêt, enthousiasme), comparativement à un agent neutre, dans un contexte de MOOC ?

1.2 But général

Cette thèse vise à explorer et à contribuer à l'intégration des agents conversationnels enrichis par la RAG dans des environnements éducatifs numériques, en prenant comme cas d'application un MOOC sur l'intelligence artificielle. Elle a pour but de répondre à des enjeux critiques : améliorer la précision des réponses fournies par les LLMs, soutenir l'acquisition des connaissances des apprenants, et encourager des émotions positives essentielles à leur engagement et à leur réussite.

En s'appuyant sur les avancées récentes en LLMs et en RAG, cette recherche cherche à combler les lacunes identifiées dans la littérature, notamment l'absence d'études empiriques approfondies sur leurs effets éducatifs. Elle ambitionne de concevoir et d'évaluer des agents conversationnels capables de réduire les erreurs d'hallucination, de fournir des réponses pédagogiquement solides et contextualisées, et de favoriser des émotions positives telles que la confiance, l'intérêt et l'enthousiasme.

1.2.1 Objectifs de recherche

Cette recherche poursuit trois objectifs complémentaires, en lien direct avec les trois questions de recherche formulées précédemment. Le premier objectif de cette thèse est d'évaluer l'effet d'un agent conversationnel reposant sur un LLM enrichi par la RAG sur la précision des réponses fournies aux apprenants. Cet objectif répond à la nécessité d'améliorer la fiabilité des réponses générées par les LLMs, en réduisant les erreurs d'hallucination qui peuvent induire des confusions ou compromettre l'apprentissage des concepts complexes. La capacité de l'agent à fournir des réponses précises et pédagogiquement pertinentes est essentielle pour garantir un apprentissage efficace et éviter des malentendus susceptibles de freiner la progression des apprenants. Cet objectif constitue une étape clé avant d'intégrer l'agent dans un MOOC.

Le deuxième objectif est de mesurer l'effet de l'agent conversationnel sur l'acquisition des connaissances des apprenants et d'analyser son influence sur leur expérience globale d'apprentissage dans un MOOC. En combinant des méthodes quantitatives, telles que les pré-tests et post-tests pour évaluer les gains de connaissances, et des approches qualitatives, comme des entretiens semi-structurés pour recueillir les perceptions des apprenants, cette thèse offre une analyse approfondie des effets de l'agent. Cette méthodologie mixte permet de capturer non seulement les résultats académiques, mais aussi les interactions et perceptions des apprenants, fournissant ainsi une compréhension complète de l'efficacité de l'agent.

Enfin, un troisième objectif central de cette thèse est d'intégrer des stratégies motivationnelles dans l'agent

conversationnel afin de favoriser à la fois l'acquisition des connaissances et l'émergence d'émotions positives chez les apprenants, telles que la confiance, l'intérêt et l'enthousiasme. Inspirée par la théorie de Dweck sur les états d'esprit (Dweck, 2002), cette approche vise à encourager les apprenants à percevoir les difficultés comme des opportunités d'apprentissage, tout en valorisant l'effort et la persévérance. Ainsi, l'agent conversationnel ne se limite donc pas à fournir des réponses ou à clarifier des concepts, il cherche également à stimuler des émotions positives tout au long du parcours d'apprentissage, contribuant ainsi à un engagement durable et à une expérience éducative enrichissante.

13 Conclusion

Dans ce chapitre, nous avons examiné les enjeux associés à l'intégration des agents conversationnels enrichis par la RAG dans des environnements éducatifs, en soulignant leur potentiel pour enrichir l'apprentissage tout en prenant en considération leurs limites. Ces analyses constituent le fondement de notre recherche, qui a pour objectif de concevoir et d'évaluer des agents conversationnels pédagogiques adaptés aux besoins des apprenants dans un MOOC sur l'intelligence artificielle.

CHAPITRE 2

CADRE CONCEPTUEL ET REVUE DE LA LITTÉRATURE

Dans ce chapitre, nous établissons les bases conceptuelles nécessaires pour guider notre recherche sur les agents conversationnels pédagogiques. Les concepts fondamentaux relatifs aux grands modèles de langage (LLMs) et à la génération augmentée par récupération (RAG) sont explorés pour mettre en lumière leur pertinence dans le contexte éducatif. Ensuite, nous présentons les modèles théoriques et leurs implications, en soulignant le lien entre les concepts de motivation et d'émotions et leur rôle dans l'atteinte des objectifs pédagogiques.

Par ailleurs, nous présentons la revue de la littérature en abordant l'évolution des agents conversationnels, leur contribution à l'apprentissage en ligne, ainsi que l'apport des agents conversationnels basés sur les grands modèles de langage et ceux enrichis par la génération augmentée par récupération dans les contextes éducatifs.

2.1 Fondements conceptuels et théoriques

Cette section examine les concepts techniques et théoriques fondamentaux de cette recherche. Elle aborde les LLMs, la RAG ainsi que les modèles théoriques liés à la motivation et aux émotions dans l'apprentissage.

2.1.1 Les grands modèles de langage

Les LLMs représentent une avancée révolutionnaire dans le domaine de l'intelligence artificielle, offrant des capacités sans précédent pour traiter et générer du texte de manière cohérente et contextuelle. Dans cette section, nous examinerons leur évolution, depuis les approches traditionnelles de l'intelligence artificielle jusqu'à l'avènement des architectures modernes comme les transformeurs. Nous aborderons également

les concepts clés, tels que les réseaux de neurones artificiels, les représentations vectorielles, l'architecture des transformeurs et l'ingénierie des consignes, qui sous-tendent leurs performances exceptionnelles.

2.1.1.1 Les réseaux de neurones artificiels

Les réseaux de neurones artificiels (Artificial Neural Networks, ANNs) représentent une avancée majeure en apprentissage automatique, en s'inspirant du fonctionnement des neurones biologiques présents dans le cerveau humain. Cette idée trouve ses racines dans les travaux pionniers de McCulloch et Pitts qui, dès 1943, ont proposé une première modélisation mathématique du fonctionnement des neurones, mettant en lumière leur capacité à effectuer des calculs logiques simples (McCulloch et Pitts, 1943).

En 1949, Hebb a introduit un principe fondamental d'apprentissage, connu sous le nom de « règle de Hebb », qui stipule que les connexions entre deux neurones se renforcent lorsqu'elles sont activées simultanément (Hebb, 1949). En 1958, Rosenblatt a proposé le perceptron, considéré comme l'une des premières applications concrètes des réseaux de neurones artificiels, capable d'apprendre et de reconnaître des motifs dans des données complexes (Rosenblatt, 1958).

Ces premières recherches s'inscrivent dans le courant dit « connexionniste » de l'intelligence artificielle, qui vise à simuler le comportement du cerveau humain par l'intermédiaire de réseaux de neurones artificiels. Cette approche s'est développée en opposition aux systèmes dits « symboliques », dominants dans les premières décennies de l'IA, qui reposaient sur des règles explicites et des raisonnements logiques formels (Nilsson, 1991).

Cependant, les limites de cette approche ont été mises en évidence par Minsky et Papert en 1969, qui ont montré que les perceptrons étaient incapables de résoudre des problèmes non linéaires tels que la fonction XOR (Minsky et Papert, 1969). Cette critique a conduit à une période de désintérêt pour les réseaux de

neurones, souvent appelée « premier hiver de l'IA ».

Ce n'est qu'au début des années 1980 que l'approche connexionniste suscite un regain d'intérêt, grâce aux travaux de Rumelhart, Hinton et Williams, qui proposent l'algorithme de rétropropagation de l'erreur (back-propagation) (Rumelhart *et al.*, 1986). Cet algorithme a permis l'entraînement efficace de réseaux multi-couches, surmontant ainsi les limites qui avaient freiné leur développement dans les années précédentes. Ce progrès a marqué le début d'une deuxième vague du connexionnisme, où les réseaux de neurones se sont imposés dans de nouvelles applications.

Les réseaux de neurones artificiels sont constitués de plusieurs couches, chacune jouant un rôle spécifique dans le traitement des informations (Le Cun *et al.*, 2015). La couche d'entrée reçoit des données brutes (images, textes, etc.) et les transforme en représentations numériques exploitables par le modèle. Les couches intermédiaires, souvent appelées couches cachées, appliquent des transformations non linéaires, notamment via des fonctions d'activation comme la fonction Sigmoid ou ReLU (Rectified Linear Unit) (Nair et Hinton, 2010), permettant de modéliser des relations complexes. Enfin, la couche de sortie génère des prédictions adaptées à la tâche visée, qu'il s'agisse de classification, de régression ou d'autres objectifs spécifiques. La Figure 2.1 illustre une architecture simple d'un réseau de neurones artificiels, comprenant les couches d'entrée, une couche cachée et une couche de sortie.

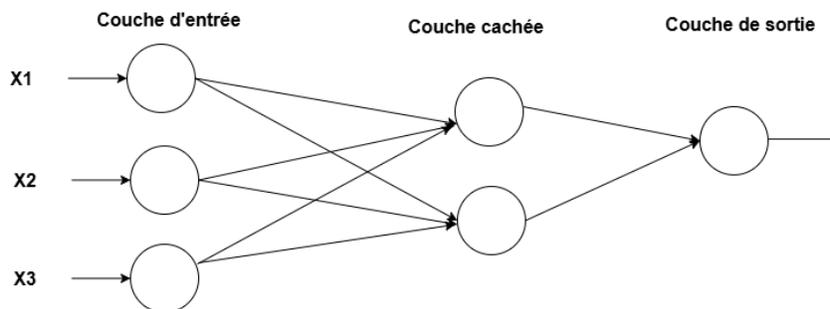


Figure 2.1 Architecture simple d'un réseau de neurones artificiels avec une seule couche cachée.

Malgré des succès notables dans des domaines variés tels que la vision par ordinateur et le traitement auto-

matique du langage naturel, les réseaux de neurones classiques présentent plusieurs limites. Leur efficacité dépend largement de la disponibilité de grandes quantités de données et leur entraînement requiert des ressources computationnelles considérables. De plus, ils se révèlent peu efficaces pour le traitement des dépendances temporelles et des séquences longues, ce qui a motivé le développement de nouvelles architectures, telles que les réseaux de mémoire à long terme (LSTM) et, plus récemment, les transformeurs. Ces innovations seront présentées dans les sections suivantes.

2.1.12 Évolution de l'intelligence artificielle vers les grands modèles de langage

L'intelligence artificielle (IA) est un domaine vaste, qui vise à créer des machines capables de réaliser des tâches demandant une intelligence humaine, comme résoudre des problèmes ou prendre des décisions. Les premières approches de l'IA utilisaient des systèmes symboliques, c'est-à-dire des programmes informatiques fonctionnant avec des règles précises pour résoudre des problèmes. Le General Problem Solver (GPS) développé par Newell et Simon (1959) ainsi que les systèmes experts comme MYCIN dans le domaine médical (Shortliffe, 1975) illustrent ces premières tentatives.

Bien que ces systèmes aient permis de franchir une étape importante dans le développement de l'IA, ils présentaient de fortes limitations : leur rigidité face à des environnements complexes, leur dépendance à une programmation humaine fastidieuse, et surtout leur incapacité à apprendre et à s'adapter automatiquement (Nilsson, 1991).

L'émergence des réseaux de neurones artificiels dans les années 1980 a marqué une avancée importante, permettant de traiter des données complexes via des couches interconnectées de neurones artificiels (Rumelhart *et al.*, 1986). Ces modèles, inspirés du fonctionnement du cerveau humain, se sont révélés efficaces pour des tâches telles que la classification et la reconnaissance d'images. Cependant, leur performance restait limitée pour le traitement de séquences longues ou de données temporelles (Elman, 1990).

Pour pallier ces problèmes, les réseaux de mémoire à long terme, ou Long Short-Term Memory (LSTM), introduits par Hochreiter et Schmidhuber (1997), ont offert une solution innovante. Grâce à une structure de mémoire interne et des mécanismes de régulation, ces modèles pouvaient conserver des informations sur de longues séquences. Toutefois, ces modèles restaient limités dans leur capacité à exécuter plusieurs calculs en parallèle (parallélisation), car leur architecture séquentielle nécessitait de traiter chaque étape de la séquence dans un ordre strict. Cette contrainte rendait difficile pour ces modèles de mémoriser des informations distantes dans les données (Bengio *et al.*, 1994).

Le véritable tournant a eu lieu au début des années 2010 avec l'essor de l'apprentissage profond (deep learning), qui a bénéficié de trois catalyseurs majeurs : l'accès à de vastes quantités de données (big data), l'augmentation considérable de la puissance de calcul grâce aux unités de traitement graphique (GPU), et des avancées algorithmiques significatives (Le Cun *et al.*, 2015). Ces évolutions ont permis de concevoir des architectures neuronales plus profondes et plus efficaces, capables d'apprendre des représentations complexes.

Dans ce contexte, l'architecture du transformeur introduite par Vaswani *et al.* (2017) a marqué une rupture décisive dans le traitement des séquences. Contrairement aux modèles séquentiels comme les RNN et LSTM, les transformeurs permettent un traitement parallèle des séquences, ce qui améliore considérablement la vitesse d'entraînement.

L'émergence des grands modèles de langage (LLMs), tels que BERT (Devlin *et al.*, 2019) et GPT (Brown *et al.*, 2020), a ensuite ouvert de nouvelles perspectives. Ces modèles, préentraînés sur de vastes corpus textuels, sont capables de générer du langage naturel de manière cohérente et contextuelle, de répondre à des questions, de produire des résumés ou d'expliquer des concepts. Leur succès repose sur leur capacité à produire des représentations vectorielles contextuelles du langage, ce qui dépasse largement les approches

classiques.

Cette évolution, des systèmes symboliques aux grands modèles de langage, illustre un progrès continu visant à dépasser les contraintes des approches antérieures. Les concepts fondamentaux, tels que les représentations vectorielles de mots et l'architecture du transformeur, jouent un rôle clé dans ces avancées et seront explorés en détail dans les sections suivantes.

2.1.13 Représentations vectorielles de mots

Les représentations vectorielles de mots constituent un fondement essentiel du traitement automatique du langage naturel. Elles consistent à associer à chaque mot un vecteur multidimensionnel de nombres réels permettant de capturer et de quantifier son sens (Jurafsky, 2000).

Ces représentations peuvent être **statiques**, où chaque mot reçoit un vecteur unique et fixe, ou **contextuelles**, où la représentation du mot varie en fonction de son contexte. Les premières approches pour créer des représentations statiques incluaient des méthodes statistiques comme le *TF-IDF* (*Term Frequency-Inverse Document Frequency*) (Sparck Jones, 1972) et le modèle *GloVe* (*Global Vectors for Word Representation*) (Pennington *et al.*, 2014). Ces approches, bien que pionnières, attribuaient la même représentation vectorielle à un mot, indépendamment de son sens dans une phrase donnée. Par exemple, le mot « Java » aurait le même vecteur qu'il fasse référence à un langage de programmation ou à une île en Indonésie.

Les progrès en réseaux neuronaux ont profondément transformé cette approche, permettant d'apprendre directement ces représentations à partir des données. Le modèle *Word2Vec*, introduit par Mikolov *et al.* (2013), en est un exemple marquant. Ce modèle repose sur deux algorithmes principaux : *CBOW* (*Continuous Bag of Words*) prédit un mot cible à partir des mots environnants, tandis que *Skip-Gram* prédit les mots du contexte à partir d'un mot cible. *Word2Vec* a permis de produire des représentations plus perti-

nelles, capables de capturer des régularités sémantiques au sein des textes.

Une propriété remarquable des représentations vectorielles apprises par des modèles comme Word2Vec est leur capacité à capturer des relations analogiques entre les mots dans l'espace vectoriel. Par exemple, la relation « homme est à femme ce que roi est à reine » peut être modélisée par une simple opération mathématique sur les vecteurs correspondants (Mikolov *et al.*, 2013). Cette capacité au raisonnement analogique repose sur la structure géométrique de l'espace des représentations, où les relations sémantiques entre les mots se traduisent par des directions et des distances cohérentes. La Figure 2.2 illustre visuellement cette propriété analogique des représentations vectorielles, où la relation entre « roi » et « homme » est comparable à celle entre « reine » et « femme ».

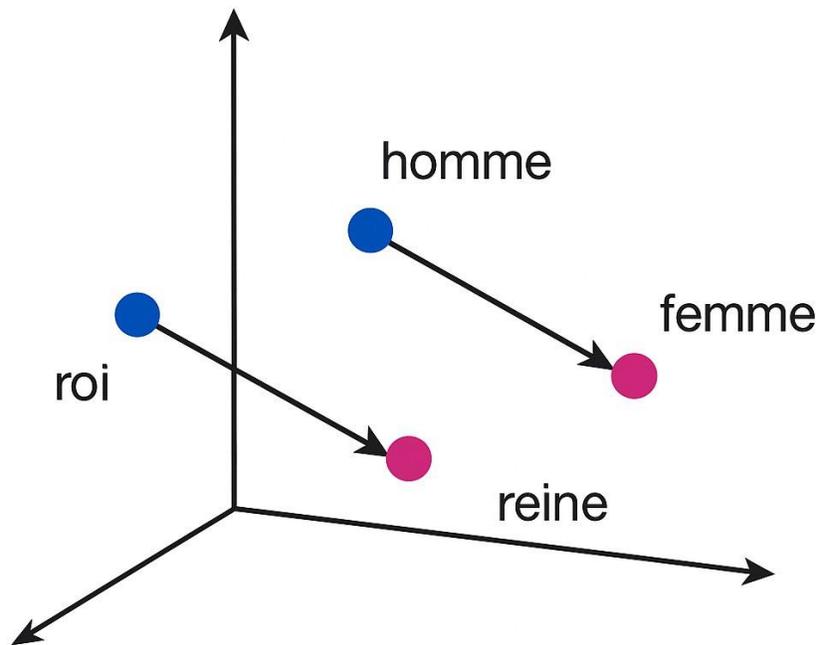


Figure 2.2 Illustration d'une analogie vectorielle dans l'espace des représentations.

Cependant, ces représentations vectorielles apprises par Word2Vec présentent une limite importante : elles restent statiques, c'est-à-dire que le même vecteur est attribué à un mot, quel que soit son contexte d'utilisation. Par exemple, dans les phrases « Java est un langage de programmation populaire » et « Java est une île située en Indonésie », le modèle associera au mot « Java » le même vecteur, sans prendre en compte les différences de sens.

Pour gérer plus finement les cas de mots rares, inconnus ou morphologiquement complexes, d'autres approches ont été développées, basées cette fois sur les caractères. Les modèles tels que les réseaux de neurones convolutifs au niveau des caractères (Zhang et Le Cun, 2015) ou les réseaux de neurones récurrents sensibles aux caractères (Kim *et al.*, 2016) permettent de construire des représentations vectorielles à partir des lettres qui composent un mot. Cela permet aux modèles de reconnaître des mots qu'ils n'ont jamais vus, en se basant sur leur structure interne. Ce type d'analyse est particulièrement utile dans les langues à morphologie riche, où un mot peut apparaître sous de nombreuses formes, ou encore pour traiter des néologismes ou des noms propres.

Des modèles plus avancés, tels que *BERT (Bidirectional Encoder Representations from Transformers)* (Devlin *et al.*, 2019) et *GPT (Generative Pre-trained Transformer)* (Brown *et al.*, 2020), ont permis de dépasser ces limitations. Ces modèles produisent des représentations **contextuelles**, où la signification d'un mot est ajustée en fonction des mots qui l'entourent. Par exemple, dans les phrases précédentes, « Java » serait représenté différemment selon qu'il désigne un langage de programmation ou une île. Cette flexibilité améliore considérablement la performance des modèles dans des tâches telles que la compréhension linguistique et la génération de texte.

Les représentations vectorielles, qui capturent des relations sémantiques et contextuelles entre les mots, peuvent être explorées visuellement en réduisant leur espace multidimensionnel à deux ou trois dimen-

sions. Par exemple, Bandyopadhyay *et al.* (2022) ont projeté un ensemble de représentations vectorielles dans un espace 3D pour illustrer des dimensions sémantiques significatives, comme le genre et l'âge. La Figure 2.3 montre cette visualisation, où les mots masculins et féminins se situent respectivement sur les moitiés gauche et droite de l'axe x , tandis que les mots relatifs aux adultes et aux jeunes se trouvent en haut et en bas de l'axe y . Cette approche rend les relations entre les mots plus intuitives et accessibles.

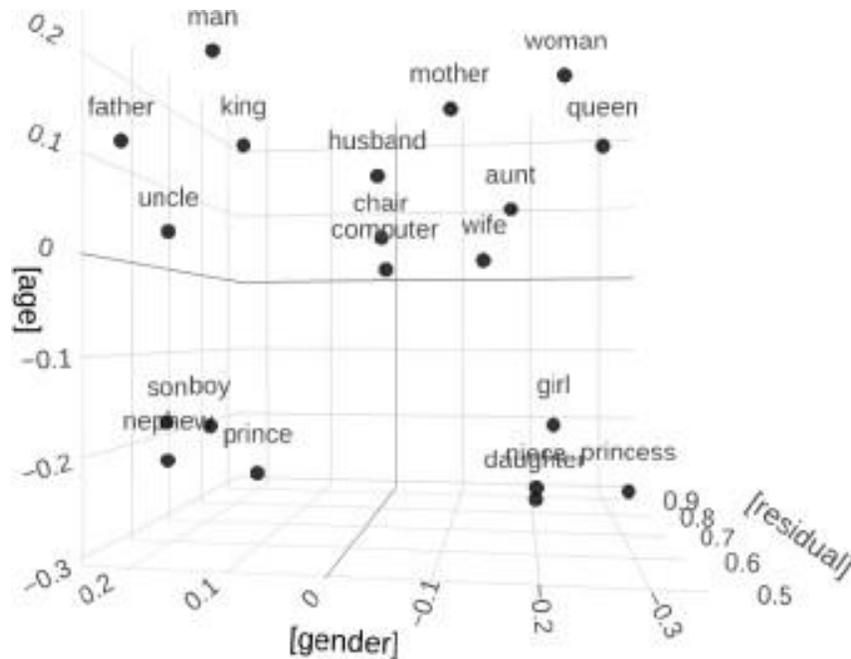


Figure 2.3 Projection en 3D des représentations vectorielles. Les mots masculins et féminins apparaissent respectivement sur les moitiés gauche et droite de l'axe x , tandis que les mots relatifs aux adultes et aux jeunes se situent dans les moitiés supérieure et inférieure de l'axe y (Bandyopadhyay *et al.*, 2022).

Ces progrès s'appuient sur l'architecture *Transformer*, qui a transformé le traitement du langage naturel en capturant efficacement les relations complexes entre les mots. Cette architecture sera examinée plus en détail dans la section suivante.

2.1.14 L'architecture du transformeur

L'architecture du transformeur est une architecture avancée d'apprentissage profond conçue pour générer des représentations vectorielles contextuelles des mots. L'apprentissage profond est un sous-domaine de

l'intelligence artificielle qui repose sur l'utilisation de réseaux de neurones artificiels comportant de nombreuses couches (d'où le terme profond). Ces réseaux permettent d'extraire automatiquement des caractéristiques complexes à partir de grandes quantités de données (Le Cun *et al.*, 2015). Introduit par Vaswani et ses collègues en 2017 (Vaswani *et al.*, 2017), le modèle de transformeur repose sur des blocs empilés, appelés encodeurs et décodeurs, qui utilisent ces couches de réseaux neuronaux pour convertir des séquences textuelles en représentations numériques. Lorsqu'une séquence passe par un bloc encodeur ou décodeur, les valeurs de sortie sont influencées par les mots environnants, grâce à un mécanisme d'« attention ». Ce mécanisme permet d'établir des relations contextuelles entre les mots en évaluant leur importance mutuelle dans la séquence (Vaswani *et al.*, 2017). La distinction majeure entre les encodeurs et les décodeurs réside dans leur manière de gérer le contexte : les encodeurs prennent en compte le contexte situé avant et après un mot donné, tandis que les décodeurs se limitent au contexte précédent.

La Figure 2.4 illustre l'architecture du transformateur, introduite par Vaswani et ses collègues en 2017 (Vaswani *et al.*, 2017). Cette architecture repose sur deux composants principaux : le bloc encodeur et le bloc décodeur, qui travaillent ensemble pour traiter et générer des séquences textuelles.

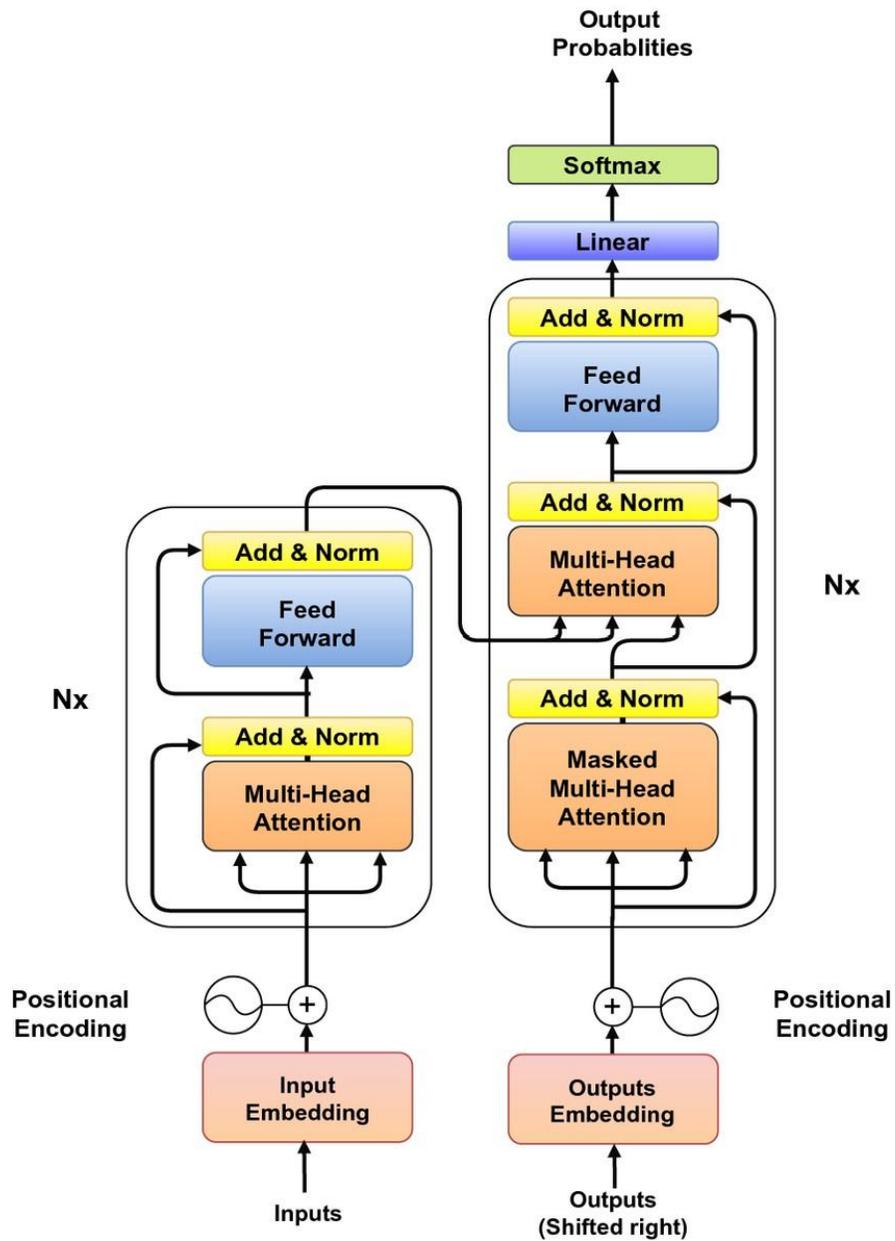


Figure 2.4 L'architecture du transformeur (Vaswani *et al.*, 2017).

Dans le bloc encodeur, une séquence d'entrée est d'abord convertie en une représentation vectorielle grâce à un encodage positionnel. Cet encodage ajoute des vecteurs positionnels aux représentations vectorielles des mots, permettant ainsi au modèle d'intégrer des informations sur l'ordre des éléments dans la séquence. En combinant ces vecteurs positionnels avec les représentations vectorielles des mots, le modèle

est capable de distinguer les mots non seulement en fonction de leur sens sémantique, mais aussi en fonction de leur position dans la séquence. Cette technique garantit que la structure logique et séquentielle des données textuelles est préservée, offrant au modèle une meilleure compréhension des relations syntaxiques et contextuelles entre les mots.

Ensuite, un mécanisme d'auto-attention multi-tête, central dans cette architecture, est appliqué. Ce mécanisme permet au modèle d'évaluer le poids relatif de chaque mot ou token dans une séquence en fonction de ses relations avec les autres mots. Contrairement aux architectures séquentielles comme les réseaux neuronaux récurrents, l'auto-attention analyse simultanément l'ensemble des mots, capturant ainsi des dépendances à longue portée et des interactions complexes. Les résultats sont ensuite transmis à un réseau de neurones à propagation directe (Feedforward Neural Network), entraîné par rétropropagation (Rumelhart *et al.*, 1986), afin d'affiner les représentations. Ces sorties sont soit utilisées comme entrée pour les couches suivantes de l'encodeur, soit transmises au décodeur pour la phase de génération.

Dans le bloc décodeur, les étapes sont similaires, mais avec une spécificité importante : un mécanisme d'attention masquée est ajouté avant le calcul de l'attention multi-tête standard. Ce masque empêche le modèle de prendre en compte les tokens futurs dans la séquence, garantissant que la prédiction s'effectue de manière progressive, token par token. Les résultats de chaque étape d'attention et réseau de neurones à propagation directe sont combinés, normalisés, puis utilisés pour produire les sorties finales.

Les transformateurs se distinguent par leur capacité à effectuer des calculs en parallèle grâce au mécanisme d'auto-attention et à leur structure en blocs (Vaswani *et al.*, 2017). Contrairement aux architectures séquentielles traditionnelles, où les mots sont traités un par un dans un ordre linéaire, les transformateurs permettent de traiter tous les mots d'une séquence simultanément. Ce parallélisme réduit considérablement le temps d'entraînement tout en maintenant des performances élevées dans les tâches de traitement

du langage naturel (Vaswani *et al.*, 2017). Cette efficacité, combinée à l'exploitation de larges ensembles de données textuelles, a conduit au développement des grands modèles de langage basés sur les transformateurs, tels que BERT (Devlin *et al.*, 2019), GPT-3 (Brown *et al.*, 2020) et GPT-4 (Zhao *et al.*, 2023; Radford *et al.*, 2019). Ces modèles ont profondément transformé des domaines variés tels que la traduction automatique, la génération de texte, la réponse aux questions et le développement des agents conversationnels.

2.1.15 Les grands modèles de langage et les transformateurs génératifs pré-entraînés

Les grands modèles de langage (Large Language Models, LLMs) sont des modèles d'apprentissage automatique spécialisés dans le traitement du langage naturel. Ils génèrent du texte en prédisant, mot par mot, la suite logique d'une séquence donnée à partir d'une entrée (Zhao *et al.*, 2023; Radford *et al.*, 2019). Ces modèles apprennent à produire des textes sémantiquement et syntaxiquement cohérents grâce à un entraînement sur de vastes corpus textuels diversifiés. Cette diversité des données leur permet de résoudre efficacement des tâches telles que répondre à des questions, résumer des documents ou générer des dialogues conversationnels (Chang *et al.*, 2024).

Parmi ces modèles, la série GPT (Generative Pretrained Transformer), développée par OpenAI, représente une avancée majeure (Brown *et al.*, 2020; Radford *et al.*, 2019; Achiam *et al.*, 2023). Ces modèles, reposant sur l'architecture Transformer, se distinguent par leur capacité à produire du texte naturel et cohérent (Raffel *et al.*, 2020; Brown *et al.*, 2020). Les modèles GPT s'appuient sur des blocs décodeurs améliorés de l'architecture Transformer, combinés à des algorithmes avancés, pour traiter efficacement des ensembles de données toujours plus vastes et variés. Depuis le lancement du modèle initial GPT, les versions suivantes, comme GPT-2, GPT-3, GPT-3.5 et GPT-4, ont accru leur précision et leur fluidité textuelle (Radford *et al.*, 2019; Mao *et al.*, 2023). GPT-4, en particulier, se distingue par sa capacité à produire des textes de qualité comparable à ceux rédigés par des humains. Il est également capable de fournir des réponses précises à

des questions complexes, atteignant des performances suffisantes pour réussir des examens académiques et professionnels (Achiam *et al.*, 2023).

Une méthode clé pour adapter les modèles de langage à des tâches spécifiques est le fine-tuning (ou ajustement fin). Cette méthode ajuste uniquement les poids des dernières couches d'un modèle pré-entraîné, tout en laissant les premières couches inchangées, afin de le spécialiser pour une tâche ou un domaine particulier (Wei *et al.*, 2021). Par exemple, le fine-tuning a permis d'adapter des modèles pré-entraînés à des corpus scientifiques, améliorant leurs performances pour des tâches spécifiques comme l'analyse et la classification des textes académiques (Beltagy *et al.*, 2019). Dans le domaine des mathématiques, le fine-tuning des modèles multi-modaux permet d'évaluer des réponses ouvertes en combinant textes et images. Ces systèmes génèrent des scores et des commentaires cohérents, aidant les apprenants à identifier leurs erreurs et à renforcer leur compréhension des concepts mathématiques (Li *et al.*, 2024).

Cependant, bien que cette méthode permette d'améliorer significativement la précision et les performances dans des domaines spécialisés, elle peut s'avérer coûteuse et exigeante en termes de ressources. Le fine-tuning demande souvent beaucoup de données spécifiques et de haute qualité, ainsi qu'une puissance de calcul importante, ce qui peut rendre son utilisation difficile à grande échelle (Touvron *et al.*, 2023).

Une autre avancée majeure des LLMs réside dans leur capacité à comprendre et à exécuter des consignes formulées en langage naturel. Ce procédé, connu sous le nom de prompting, simplifie l'interaction avec le modèle en éliminant le besoin d'un apprentissage spécifique pour chaque tâche. En formulant une consigne claire, il devient possible d'obtenir des réponses adaptées et pertinentes (Brown *et al.*, 2020; Liu *et al.*, 2023). Ce mécanisme constitue une composante essentielle du présent travail.

2.1.16 Ingénierie des consignes

L'ingénierie des consignes (prompt engineering) consiste à concevoir des instructions textuelles destinées à guider un modèle de langage avancé, afin d'optimiser la qualité des résultats sans nécessiter un nouvel entraînement sur des jeux de données supplémentaires (Reynolds et McDonell, 2021). Cette approche prometteuse s'est imposée comme une technique clé pour améliorer les performances des modèles de langage sur des tâches variées.

Un prompt se compose généralement d'une consigne précisant le comportement attendu du modèle, parfois enrichie d'exemples illustratifs (few-shot learning) (Mann *et al.*, 2020), ou d'instructions directes sans exemple (zero-shot learning) (Bommarito *et al.*, 2023). Une méthode avancée, appelée chain-of-thought prompting, invite le modèle à raisonner étape par étape. Cette approche améliore significativement sa capacité à résoudre des tâches complexes nécessitant une réflexion logique ou des explications détaillées (Wei *et al.*, 2022). Dans le tableau 2.1, nous comparons trois approches courantes de prompt : Zero-Shot, Few-Shot et Chain of Thought, en fournissant une description et un exemple pour chacune.

L'ingénierie des consignes trouve également des applications dans des contextes pratiques, tels que l'extraction de données. Par exemple, la méthode ChatExtract, basée sur des consignes structurées, permet d'automatiser l'extraction précise de données à partir d'articles de recherche. Elle se distingue par l'utilisation de questions de suivi, qui vérifient l'exactitude des informations extraites et réduisent les erreurs potentielles (Polak et Morgan, 2024). Par ailleurs, des catalogues structurés de modèles d'ingénierie des consignes ont été développés pour optimiser l'utilisation des modèles conversationnels comme ChatGPT. Ces modèles permettent de personnaliser les réponses, d'adapter le ton ou le style, et de contrôler le contexte des interactions, améliorant ainsi la qualité et la pertinence des résultats (White *et al.*, 2023). Ils sont appliqués à des tâches variées, telles que la mise en forme des réponses ou la génération de résumés précis, soulignant l'im-

Méthode	Description	Exemple
Zero-Shot (Bommarito <i>et al.</i> , 2023)	Fournir une instruction directement sans donner d'exemples spécifiques.	Q : Additionne les nombres suivants : $a = 7, b = 5$ R :
Few-Shot (Mann <i>et al.</i> , 2020)	Fournir un ou plusieurs exemples pour guider le modèle dans la tâche demandée.	Traduire l'anglais vers le français : Exemple : Anglais : Hello, how are you? Français : Bonjour, comment ça va? Traduction : Anglais : What time is it? Français :
Chain of Thought (Wei <i>et al.</i> , 2022)	Expliciter le raisonnement étape par étape avant de fournir une solution finale.	Q1 : Quel est le x dans cette équation : $2x + 4 = 12$? R1 : Pour résoudre $2x + 4 = 12$, suivez les étapes suivantes : 1. Soustraire 4 des deux côtés de l'équation : $2x + 4 - 4 = 12 - 4$ $\Rightarrow 2x = 8$ 2. Diviser les deux côtés par 2 pour isoler x : $\frac{2x}{2} = \frac{8}{2}$ $\Rightarrow x = 4$ Réponse : $x = 4$ Q2 : Quel est le x dans cette équation : $5x - 2 = 13$? R2 :

Table 2.1 Comparaison des méthodes de prompt.

portance croissante de l'ingénierie des consignes dans l'exploitation des modèles conversationnels avancés.

Bien que l'ingénierie des consignes améliore les interactions avec les modèles de langage, ces derniers rencontrent des limites telles que le phénomène des hallucinations (Alkaiissi et McFarlane, 2023; Ji *et al.*, 2023). De plus, ces modèles s'appuient uniquement sur les informations apprises au cours de leur entraînement, ce qui limite leur capacité à fournir des réponses actualisées ou adaptées à des besoins spécifiques.

La génération augmentée par récupération (RAG) s'impose comme une solution complémentaire (Gao *et al.*, 2023). En intégrant des bases de connaissances externes validées, cette approche réduit les erreurs tout en améliorant la pertinence des réponses générées. En combinant les capacités avancées des LLMs avec des informations fiables, RAG constitue une méthode efficace pour renforcer la précision et la pertinence des résultats.

2.1.2 Génération augmentée par récupération

La RAG est une méthode combinant les capacités des grands modèles de langage LLMs avec des données externes pour fournir des réponses pertinentes et ancrées dans des sources fiables (Gao *et al.*, 2023; Lewis *et al.*, 2020). Trois paradigmes principaux structurent la RAG : la RAG naïve, la RAG avancée et la RAG modulaire. Chaque paradigme apporte des solutions spécifiques aux défis rencontrés dans les systèmes d'intelligence artificielle. Cette section introduit ces paradigmes et justifie le choix de la RAG naïve pour cette étude.

2.1.2.1 Génération augmentée par récupération naïve

La RAG naïve est le paradigme le plus élémentaire et constitue une approche fondatrice dans ce domaine. Ce modèle suit un processus traditionnel structuré autour des étapes suivantes : l'indexation, la récupération et la génération (Gao *et al.*, 2023; Lewis *et al.*, 2020).

La première étape, l'indexation, consiste à organiser un corpus documentaire pour permettre une recherche efficace et précise. Ce processus débute par la collecte et le nettoyage de documents provenant de sources variées, tels que des fichiers PDF, HTML ou Word, qui sont convertis en un format texte uniforme et standardisé. Ensuite, ces documents sont segmentés en unités plus petites, comme des phrases ou des paragraphes, pour améliorer la précision des recherches. Ces segments sont ensuite transformés en représentations numériques appelées vecteurs, à l'aide de modèles avancés tels que SBERT (Reimers et Gurevych, 2019) ou text-embedding-ada-002 d'OpenAI (Neelakantan *et al.*, 2022). Ces vecteurs permettent de représenter le sens et le contexte des segments, plaçant ceux qui sont similaires proches les uns des autres dans un espace mathématique multidimensionnel. Enfin, les vecteurs générés sont stockés dans des bases de données vectorielles spécialisées, comme Pinecone, Weaviate ou FAISS (Johnson *et al.*, 2019). Ces bases permettent d'effectuer des recherches rapides et précises grâce à des algorithmes qui comparent les vec-

teurs pour évaluer leur similarité.

La seconde étape, la récupération, intervient lorsqu'une requête est soumise par un utilisateur. La requête est d'abord transformée en vecteur à l'aide du même modèle utilisé lors de l'indexation. Le système compare ensuite ce vecteur de requête avec les vecteurs des segments indexés et stockés dans une base de données vectorielle. À cette étape, des scores de similarité sont calculés pour chaque segment afin de mesurer leur pertinence par rapport à la requête. Les K segments ayant les scores de similarité les plus élevés sont sélectionnés. Ces segments pertinents, souvent appelés contexte enrichi, servent ensuite de base pour compléter le prompt destiné au modèle de génération. Ce contexte enrichi améliore la capacité du modèle à produire une réponse plus précise et contextualisée.

Enfin, la dernière étape est celle de la génération. Les segments sélectionnés sont combinés avec la question initiale de l'utilisateur pour construire un prompt enrichi, qui est ensuite soumis à un modèle de langage de grande taille, tel que GPT-4. Le modèle formule une réponse qui peut se limiter aux informations extraites ou s'enrichir de ses propres connaissances internes. Cette méthode permet de réduire les risques d'hallucinations en s'appuyant sur des données externes validées, offrant ainsi des réponses plus fiables et adaptées au contexte.

Malgré sa simplicité, la RAG naïve présente plusieurs inconvénients qui limitent son efficacité dans certains contextes. La phase de récupération peut manquer de précision, conduisant à la sélection de segments non pertinents. De plus, lors de la génération, le modèle peut produire des réponses biaisées ou incohérentes, en raison des biais présents dans les données initiales ou d'un manque de pertinence des segments récupérés.

2.1.2.2 Génération augmentée par récupération avancée

Pour répondre aux limites de la RAG naïve, la RAG avancée propose des améliorations dans les étapes d'indexation, de récupération et de post-récupération. Lors de l'indexation, les documents sont divisés en segments plus petits et détaillés, appelés segmentation fine. Ces segments sont enrichis avec des informations supplémentaires, comme des mots-clés ou des résumés automatiques, qui facilitent leur identification lors de la recherche. Dans la phase de récupération, des recherches hybrides sont utilisées. Cela signifie que la RAG avancée combine des techniques traditionnelles, comme la recherche de mots-clés, avec des méthodes plus modernes basées sur des modèles de langage, pour identifier les segments les plus proches de la requête de l'utilisateur. Enfin, dans la post-récupération, les segments sélectionnés sont triés et réorganisés pour mettre en avant les informations les plus pertinentes. Le contenu est également condensé pour éviter une surcharge d'informations dans le prompt final, ce qui améliore la pertinence des réponses générées. Bien que ces améliorations permettent une meilleure qualité des réponses, elles nécessitent des ressources computationnelles importantes et des configurations complexes, qui dépassent les objectifs et les contraintes de cette étude.

2.1.2.3 Génération augmentée par récupération modulaire

La RAG modulaire se distingue des paradigmes précédents par son architecture flexible et sa capacité à intégrer des modules spécialisés répondant à des besoins variés (Gao *et al.*, 2023). Par exemple, un module de recherche peut être configuré pour interroger des bases de données spécifiques ou des graphiques de connaissances, tandis qu'un module de mémoire peut conserver des informations contextuelles pour enrichir les interactions futures. Cette approche modulaire permet une personnalisation avancée, chaque module pouvant être adapté ou remplacé selon les exigences d'une tâche particulière.

Cependant, cette complexité accrue s'accompagne de défis, tels que des besoins élevés en ressources com-

putationnelles et une configuration sophistiquée. Bien que ces caractéristiques soient avantageuses dans des contextes nécessitant une gestion complexe des données ou des tâches multimodales, elles la rendent moins adaptée aux environnements éducatifs où la simplicité, l'efficacité, et la rapidité de déploiement sont prioritaires

Dans le cadre de cette thèse, la RAG naïve a été retenue pour son adéquation aux besoins spécifiques d'un MOOC, sa simplicité d'implémentation et sa compatibilité avec les ressources limitées disponibles. Contrairement aux paradigmes avancés et modulaires, elle ne nécessite ni optimisation poussée ni gestion de volumes massifs de données. En intégrant des supports de cours spécifiques dans les étapes d'indexation et de récupération, la RAG naïve permet d'assurer une précision et une contextualisation suffisantes pour les besoins identifiés dans cette étude. Cette approche répond efficacement aux objectifs de cette étude, à savoir fournir des réponses pédagogiquement pertinentes et adaptées au contexte éducatif. Dans le cadre de notre thèse, l'intégration de la RAG dans notre agent conversationnel vise à renforcer l'exactitude et la pertinence des réponses fournies aux apprenants, en exploitant les informations issues des supports de cours du MOOC. Une présentation détaillée de l'architecture et des étapes de fonctionnement de cet agent conversationnel utilisant la RAG sera proposée dans le chapitre 3.

2.1.3 Les modèles théoriques et leurs implications

Cette section explore deux modèles théoriques fondamentaux liés à l'apprentissage. Elle met en lumière l'influence des émotions sur l'apprentissage ainsi que l'influence des croyances des apprenants sur la nature de leur intelligence dans leur motivation et leurs performances.

2.1.3.1 Les émotions et la théorie du contrôle-valeur

Les émotions jouent un rôle déterminant dans le processus d'apprentissage, influençant directement la motivation, l'engagement, et les performances académiques (Pekrun *et al.*, 2007). Selon la théorie du contrôle-valeur (*Control-Value Theory*), les émotions liées à l'apprentissage peuvent être classées en fonction de leur nature (positives ou négatives) et de leur niveau d'activation (élevé ou faible) (Pekrun *et al.*, 2007). Cette théorie met en évidence que les émotions positives, telles que l'intérêt et la joie, soutiennent les processus cognitifs en favorisant l'attention et la mémorisation. À l'inverse, les émotions négatives, comme l'anxiété ou l'ennui, peuvent perturber l'apprentissage en diminuant la concentration et en augmentant le risque d'abandon.

Des recherches, telles que celles de Arroyo *et al.* (2009), ont structuré les émotions liées à l'apprentissage en axes cognitifs et affectifs, organisés sous forme de dimensions bipolaires. Par exemple, ces axes incluent des oppositions comme « anxiété. . . confiance ». Ces dimensions illustrent l'interaction entre les émotions et les processus cognitifs dans un contexte éducatif.

Ces travaux montrent comment les émotions influencent des aspects clés de l'apprentissage, comme l'attention et l'engagement actif. Par exemple, des émotions positives, comme la confiance, sont associées à une meilleure concentration, tandis que des émotions négatives, telles que l'anxiété, peuvent perturber la résolution de problèmes.

De manière similaire, Pekrun *et al.* (2002) soulignent que des émotions positives telles que la joie, l'intérêt et la fierté favorisent l'effort cognitif, la motivation et conduisent à de meilleures performances académiques. À l'inverse, des émotions négatives comme l'anxiété, le découragement ou l'ennui sont susceptibles de réduire la motivation, de perturber les processus d'apprentissage et de favoriser un désengagement progressif, pouvant aller jusqu'au retrait des activités académiques. Ces résultats soulignent l'importance cru-

ciale de prendre en compte les émotions dans la conception d'environnements d'apprentissage favorisant la réussite.

2.1.3.2 Théorie de l'état d'esprit

La théorie de l'état d'esprit (mindset) développée par Dweck (2002) offre un cadre conceptuel solide pour concevoir des interventions motivationnelles. Cette théorie met en lumière l'importance de la perception de l'intelligence chez les apprenants. Ceux ayant un état d'esprit fixe considèrent leurs capacités comme immuables, ce qui peut les amener à éviter les défis et à abandonner face aux échecs. À l'inverse, un état d'esprit de développement les encourage à persévérer et à voir les erreurs comme des opportunités d'apprentissage (Dweck, 2002).

Les environnements numériques d'apprentissage, comme les systèmes tutoriels intelligents, ont intégré les principes de cette théorie pour renforcer la résilience et la motivation des apprenants. Par exemple, le système Wayang Outpost a démontré l'efficacité des messages valorisant l'effort pour réduire l'anxiété et promouvoir une attitude positive envers l'apprentissage, en particulier chez les étudiants à faibles performances (Woolf *et al.*, 2010).

Cependant, cette théorie a également fait l'objet de critiques. Des chercheurs, tels que Macnamara et Burgoyne (2023), soulignent que les effets des interventions basées sur le mindset restent souvent modestes et difficiles à généraliser, notamment dans des contextes éducatifs de courte durée ou auprès d'apprenants déjà motivés. En outre, des failles méthodologiques se manifestent dans la conception des groupes témoins ou contrôles. L'absence d'interventions alternatives comparables dans ces groupes limite les possibilités d'évaluation rigoureuse et peut conduire à attribuer de manière erronée les résultats positifs à l'intervention basée sur le mindset. Ces lacunes soulignent la nécessité de concevoir des études plus rigoureuses pour isoler précisément l'effet de ces approches.

Dans le cadre de cette thèse, la théorie de Dweck est intégrée dans la conception d'un agent conversationnel pédagogique visant à cultiver des émotions positives, telles que la confiance et l'enthousiasme. Cet agent valorise les efforts des apprenants en fournissant des retours motivationnels adaptés, tout en adoptant une méthodologie exploratoire pour évaluer son effet et contribuer à surmonter les limites identifiées dans la littérature.

22 Revue de la littérature

Dans le domaine de l'éducation, les interventions éducatives désignent des actions planifiées, structurées et mises en œuvre dans un environnement d'apprentissage, dans le but d'améliorer l'expérience éducative des apprenants. Ces actions sont stratégiques, car elles sont spécifiquement conçues pour atteindre des objectifs pédagogiques précis, tels que l'acquisition des connaissances des apprenants. Elles s'appuient sur des approches pédagogiques fondées sur la recherche et des outils technologiques adaptés, permettant ainsi de personnaliser l'apprentissage en fonction des besoins spécifiques des apprenants. De plus, ces interventions tiennent compte des contextes spécifiques dans lesquels elles sont déployées, qu'il s'agisse d'apprentissage en présentiel, en ligne ou dans un format hybride

Dans ce cadre, Vygotsky (1978) souligne l'importance primordiale des interactions sociales et du contexte culturel dans le développement cognitif. Selon cet auteur, l'apprentissage ne se réalise pas uniquement par des expériences individuelles, mais surtout dans la Zone Proximale de Développement (ZPD), qui désigne l'écart entre ce qu'un apprenant peut accomplir seul et ce qu'il peut accomplir avec l'aide d'une personne plus expérimentée. Cette théorie met en évidence le rôle crucial de l'accompagnement et des interactions sociales dans le processus d'apprentissage, en soulignant que l'apprentissage optimal survient lorsque l'apprenant reçoit un soutien approprié lui permettant de franchir cette zone de développement.

Dans ce contexte, les technologies éducatives jouent un rôle central en enrichissant l'expérience d'appren-

tissage, notamment en offrant des interactions plus flexibles et mieux adaptées aux besoins des apprenants. Parmi ces technologies, les agents conversationnels basés sur l'intelligence artificielle se révèlent particulièrement efficaces pour favoriser l'interaction et l'accompagnement. Grâce à l'utilisation du langage naturel, les agents adaptent leurs réponses en temps réel, créant ainsi une expérience d'apprentissage fluide et interactive. En s'ajustant aux besoins spécifiques des apprenants, ces agents fournissent un soutien ciblé, ce qui renforce leur efficacité dans les environnements éducatifs.

2.2.1 Historique et évolution des agents conversationnels

Les *chatbots*, ou agents conversationnels, sont des systèmes d'interaction humain-machine capables d'interpréter et de générer du contenu en langage naturel pour communiquer avec les utilisateurs, à travers différents modes incluant le texte, la voix et les gestes (Allouch *et al.*, 2021). Depuis leur création, les *chatbots* ont connu plusieurs étapes d'évolution, débutant par des systèmes simples basés sur des règles, avant de se tourner vers des modèles d'IA générative plus avancés.

2.2.1.1 Systèmes basés sur des règles

Les agents conversationnels basés sur des règles constituaient la technologie principale dans les premières étapes du développement des agents conversationnels, à l'exemple d'ELIZA, développé en 1966 (Weizenbaum, 1966). Conçu pour simuler le rôle d'un psychothérapeute, ELIZA a marqué le début des *chatbots* en illustrant le potentiel de systèmes de dialogue simples fondés sur des règles. Ce système reposait sur une architecture simple, où des mots-clés spécifiques étaient identifiés dans les entrées de l'utilisateur pour déclencher des réponses prédéfinies, suivant des règles et des modèles établis manuellement par des humains. L'avantage des systèmes basés sur des règles est qu'il peuvent fournir des réponses précises. Les systèmes basés sur des règles présentent plusieurs limites. Leur capacité à prendre en compte le contexte est limitée, ce qui conduit souvent à des réponses génériques et peu adaptées. Leur efficacité dépend for-

tement de la clarté des messages d'entrée, les rendant peu robustes face à des formulations ambiguës. De plus, leur conception nécessite une programmation manuelle intensive, rendant leur mise à jour et leur adaptation complexes.

Développé dans les années 1990 par Richard Wallace, ALICE (Artificial Linguistic Internet Computer Entity) représente une avancée par rapport à ELIZA grâce à l'utilisation du langage AIML (Artificial Intelligence Markup Language) (Wallace, 2003). Ce langage permettait de structurer les règles sous forme de modèles de correspondance (pattern matching) pour associer des entrées utilisateur à des réponses prédéfinies. Contrairement à ELIZA, ALICE offrait une flexibilité accrue grâce au langage AIML, qui permettait de prendre en compte des variations linguistiques et des structures hiérarchiques. Cela facilitait l'ajout ou la modification de réponses, rendant le système accessible, adaptable et capable de traiter un plus large éventail de questions. Ces caractéristiques enrichissaient les interactions et offraient des conversations plus fluides et engageantes pour l'utilisateur. Cependant, comme ELIZA, ALICE était limité par son incapacité à interpréter le contexte des interactions ou à conserver une mémoire des échanges. Chaque réponse était générée de manière indépendante, sans lien avec les messages précédents, ce qui entraînait des dialogues répétitifs et souvent déconnectés. Bien que des avancées comme l'utilisation d'AIML aient permis une meilleure structuration des règles et une certaine flexibilité dans les réponses, ces systèmes restaient peu adaptés aux interactions complexes nécessitant une compréhension contextuelle approfondie.

2.2.1.2 Systèmes basés sur la récupération

Les *chatbots* basés sur la récupération (retrieval-based chatbots) représentent une avancée par rapport aux systèmes à base de règles. Ils s'appuient sur une base de connaissances contenant des réponses prédéfinies, sélectionnées à l'aide de techniques comme la correspondance par mots-clés et des métriques avancées, telles que la similarité cosinus, pour améliorer leur précision (Banchs et Li, 2012; Wu *et al.*, 2016).

Leur principal avantage réside dans leur capacité à fournir des réponses précises et adaptées à une grande variété de requêtes. Ces systèmes sont couramment utilisés dans des applications pratiques, telles que les assistants virtuels (Siri, Alexa, Google Assistant), grâce à leur efficacité pour récupérer des informations provenant de diverses sources (Grudin et Jacques, 2019).

Cependant, ces systèmes présentent des limites importantes qui restreignent leur efficacité. Leur principal inconvénient réside dans leur dépendance à la qualité et à l'exhaustivité de leur base de connaissances. Si une réponse appropriée n'est pas disponible dans cette base, le système est incapable de répondre ou fournit une réponse inexacte. En outre, ces *chatbots* manquent de flexibilité et ne peuvent pas adapter leurs réponses à des questions imprévues ou mal formulées, limitant ainsi leur robustesse. De plus, ces systèmes traitent chaque requête isolément, sans considérer l'historique ou le contexte global, les rendant peu adaptés aux échanges prolongés ou complexes. Enfin, leur capacité à personnaliser les réponses est limitée, car ils s'appuient uniquement sur des correspondances fixes, ce qui réduit leur pertinence dans des contextes nécessitant une adaptation aux besoins spécifiques des utilisateurs (Suhaili *et al.*, 2021).

2.2.1.3 Chatbots basés sur des modèles génératifs

Les *chatbots* basés sur des modèles génératifs marquent une avancée importante dans le domaine des agents conversationnels. Contrairement aux systèmes basés sur des règles ou la récupération d'information, ces *chatbots* produisent des réponses en temps réel en s'appuyant sur des modèles d'apprentissage automatique entraînés sur de vastes corpus de données textuelles. Ces modèles génèrent des réponses adaptées au contexte de la conversation, permettant de simuler des interactions plus naturelles et proches de celles d'un humain (Vinyals et Le, 2015). Les premiers modèles génératifs s'appuyaient sur des réseaux neuronaux, notamment les LSTM (Long Short-Term Memory), conçus pour gérer les informations sur de longues séquences comme les phrases ou les dialogues. Ces modèles ont permis de mieux comprendre

le contexte des échanges et d'intégrer l'historique pour produire des réponses plus pertinentes (Su *et al.*, 2017; Xu *et al.*, 2017). Cependant, ils ont montré leurs limites dans la gestion de séquences très longues ou de relations complexes, ce qui a conduit au développement des Transformers, une nouvelle architecture conçue pour surmonter ces défis. Ces derniers, introduits par Vaswani et ses collègues (Vaswani *et al.*, 2017), utilisent un mécanisme appelé « attention », conçu pour identifier et établir les relations clés entre les éléments d'une séquence, même lorsque ces éléments sont éloignés dans le texte. Ce mécanisme permet au modèle de se concentrer sur les informations les plus pertinentes, facilitant ainsi la compréhension et le traitement des données, y compris dans des phrases longues ou des contextes complexes (Vaswani *et al.*, 2017). Les Transformers, grâce à ce mécanisme d'attention, ont révolutionné le domaine des modèles de langage. Des modèles comme GPT (Generative Pre-trained Transformer), souvent qualifiés de Modèles de Langue de Grande Taille (LLM), ont significativement amélioré les performances des *chatbots*. En s'appuyant sur un pré-entraînement massif et des mécanismes d'attention avancés, ces modèles génèrent des réponses flexibles et cohérentes dans divers contextes (Radford, 2018). Cependant, les *chatbots* génératifs présentent des défis importants. Leur efficacité dépend de la disponibilité de données d'entraînement, et leurs performances peuvent être limitées dans des domaines spécialisés où ces données sont insuffisantes. Par ailleurs, le risque de réponses incohérentes ou inexactes demeure.

Dans cette thèse, nous explorons une approche combinant des techniques de récupération d'information et des modèles génératifs, appliquée à la conception d'un agent conversationnel capable de produire des réponses précises et contextuelles. Notre contribution réside dans l'adaptation de cet agent à un cadre éducatif, en particulier dans un MOOC, et dans l'évaluation empirique de ses effets sur la qualité des réponses, l'acquisition des connaissances et l'expérience globale des apprenants. Cette étude se distingue par une évaluation empirique réalisée directement auprès d'apprenants dans un contexte réel.

2.2.2 Agents conversationnels pédagogiques en éducation

Les agents conversationnels pédagogiques, également appelés *chatbots* éducatifs, sont des outils numériques interactifs conçus pour interagir avec les apprenants dans un cadre éducatif (Shawar et Atwell, 2007; Kerlyl *et al.*, 2006). Ils utilisent des technologies avancées, telles que le traitement du langage naturel et l'intelligence artificielle, pour répondre aux questions, fournir des rétroactions personnalisés et adapter les contenus en fonction des besoins spécifiques des apprenants (Winkler et Söllner, 2018; Woolf, 2010). Ces outils jouent un rôle clé dans le soutien de l'apprentissage collaboratif et individuel tout en stimulant l'engagement des apprenants (Winkler et Söllner, 2018; Woolf, 2010).

Ces technologies ne visent pas à remplacer les enseignants mais plutôt à compléter leur rôle en enrichissant les expériences d'apprentissage. Comme l'ont souligné Shawar et Atwell (2007), les agents conversationnels pédagogiques peuvent alléger la charge des enseignants en fournissant des réponses rapides et contextuelles, tout en facilitant un apprentissage individualisé à grande échelle. De plus, leur capacité à s'adapter aux besoins spécifiques des apprenants permet d'automatiser des tâches répétitives, comme répondre aux questions fréquentes ou identifier les lacunes éducatives (Kerlyl *et al.*, 2006).

Plusieurs initiatives majeures des années 2000 et 2010 ont également marqué l'évolution des agents conversationnels en contexte éducatif. Les travaux de Lewis Johnson et ses collègues (Johnson *et al.*, 2000) ont ainsi introduit des agents pédagogiques animés pionniers tels qu'ADELE et STEVE, capables d'accompagner les apprenants dans des environnements d'apprentissage interactifs en combinant des interactions verbales et non verbales. Ces agents se distinguaient par leur capacité à démontrer des procédures complexes, à utiliser des gestes et des mouvements pour guider l'attention des étudiants, et à exprimer des réactions émotionnelles, renforçant ainsi la motivation et l'engagement des apprenants. Par exemple, STEVE a été utilisé dans des environnements virtuels immersifs, tels qu'un navire ou une salle des machines virtuelle,

pour enseigner l'utilisation d'équipements techniques. De son côté, ADELE proposait une assistance pédagogique dans un contexte d'apprentissage en ligne via un navigateur web standard, en aidant les étudiants à apprendre des notions médicales telles que les maladies infectieuses.

De leur côté, Arthur Graesser et ses collègues (Graesser *et al.*, 2001) ont développé AutoTutor, un système tutoriel intelligent en langage naturel conçu pour simuler les stratégies pédagogiques d'un tuteur humain. En posant des questions complexes, AutoTutor encourage les étudiants à formuler des réponses élaborées et à approfondir leur raisonnement. Le système intègre un agent animé affichant des expressions faciales, des gestes, une intonation vocale et un discours synthétisé, afin de rendre l'interaction plus naturelle. Des évaluations ont montré qu'AutoTutor contribue à améliorer l'apprentissage et la mémorisation des concepts.

Par ailleurs, les recherches de Kurt VanLehn et de ses collègues (Vanlehn *et al.*, 2005) ont conduit au développement de ANDES, un système tutoriel intelligent (STI) conçu pour accompagner les étudiants dans la résolution de problèmes de physique en leur fournissant un feedback immédiat à chaque étape. Dans le cadre de ces recherches, un module appelé ATLAS (Freedman *et al.*, 2000) a été proposé pour permettre la gestion de dialogues en langage naturel dans les environnements tutoriels intelligents. Conçu comme une architecture modulaire et indépendante du domaine, ATLAS visait à rendre les interactions tutorales plus naturelles tout en étant intégrable dans différents STI. Ces travaux ont ainsi contribué à faire émerger des approches plus interactives et adaptatives dans les environnements d'apprentissage.

Avec l'évolution des technologies, les agents conversationnels ont progressivement intégré des fonctionnalités avancées qui permettent des interactions plus souples et mieux adaptées aux besoins variés des apprenants. Cette capacité d'adaptation leur permet également de relever de nombreux défis éducatifs, en analysant les besoins en temps réel et en proposant des réponses personnalisées (Ruan *et al.*, 2019). Grâce

à cette souplesse, ces agents peuvent être déployés dans des contextes éducatifs très diversifiés, allant des systèmes traditionnelles aux environnements numériques massifs.

En outre, les agents conversationnels pédagogiques ont démontré leur capacité à transformer les méthodes éducatives traditionnelles grâce à des interactions engageantes et personnalisées. Par exemple, Woolf (2010) ont exploré des stratégies de personnalisation permettant de cibler les zones de difficulté des apprenants et de maintenir leur engagement. Ces approches montrent comment les agents conversationnels pédagogiques peuvent compléter les enseignants et enrichir l'expérience éducative des apprenants.

2.2.2.1 Rétroaction et personnalisation avec les agents conversationnels pédagogiques

La rétroaction (ou feedback) est considérée comme l'un des leviers les plus puissants pour favoriser l'apprentissage et améliorer les performances des apprenants (Hattie et Timperley, 2007). Son objectif principal est de réduire l'écart entre la compréhension actuelle et les objectifs d'apprentissage visés. Un feedback efficace fournit des informations sur la performance, aide à clarifier les progrès accomplis et oriente les actions nécessaires pour progresser davantage. Il peut ainsi accroître l'effort, la motivation ou l'engagement des apprenants.

Dans les environnements éducatifs numériques, l'intégration d'une rétroaction adaptative et personnalisée constitue un levier essentiel pour améliorer l'efficacité des agents conversationnels pédagogiques. En facilitant des interactions dynamiques et adaptées aux besoins individuels, ces agents soutiennent la progression des apprenants tout en enrichissant leur expérience d'apprentissage (Winkler et Söllner, 2018).

Par exemple, l'outil MentorChat décrit par Tegos *et al.* (2011) a démontré comment un agent conversationnel peut adapter ses interventions en fonction des contributions des apprenants dans des activités de collaboration. Ce type d'interaction guidée stimule les échanges, oriente les apprenants vers des solutions

approfondies et contribue à l'amélioration de leur réflexion critique.

De plus, les agents conversationnels pédagogiques tels que ceux décrits par Shawar et Atwell (2007) offrent aux apprenants un moyen autonome de pratiquer des interactions adaptées à leurs besoins linguistiques spécifiques, favorisant ainsi une approche individualisée de l'apprentissage.

Par ailleurs, Fryer *et al.* (2017) ont mis en évidence le rôle des agents conversationnels pédagogiques dans le maintien de l'engagement des apprenants à travers des interactions dynamiques et stimulantes. Ces stratégies interactives illustrent le potentiel des agents conversationnels pédagogiques à transformer les expériences éducatives en les rendant plus immersives et engageantes.

Dans des environnements éducatifs massifs tels que les MOOCs, les agents conversationnels pédagogiques jouent un rôle crucial en fournissant des rétroactions en temps réel, adaptés aux besoins individuels des apprenants. Ces interactions ciblées renforcent leur motivation et leur autonomie tout en palliant les limites des interactions humaines traditionnelles. Winkler et Söllner (2018) et González-Castro *et al.* (2021) mettent en lumière la capacité des agents conversationnels pédagogiques à offrir un support personnalisé, permettant ainsi d'optimiser les expériences d'apprentissage tout en répondant aux défis spécifiques de ces environnements à grande échelle.

Selon Wollny *et al.* (2021), les *chatbots* éducatifs démontrent un potentiel significatif en fournissant des retours adaptés aux besoins spécifiques des apprenants. Ces retours personnalisés contribuent non seulement à renforcer leurs compétences, mais jouent également un rôle essentiel dans l'amélioration de leur motivation et de leur autonomie.

De plus, Winkler et Söllner (2018) identifient des éléments essentiels pour maximiser l'efficacité des *chatbots* éducatifs, notamment leur adaptabilité aux besoins individuels, la qualité des interactions qu'ils faci-

litent et leur capacité à enrichir le processus d'apprentissage. Ces caractéristiques renforcent l'impact des agents conversationnels pédagogiques en proposant des expériences interactives adaptées à chaque apprenant, tout en améliorant leur engagement.

En plus de leurs fonctions cognitives, les agents conversationnels pédagogiques intègrent une dimension affective par l'utilisation de signaux émotionnels. Par exemple, Guo *et al.* (2015) ont étudié les agents pédagogiques affectifs dans un tutoriel en ligne, démontrant que des expressions émotionnelles comme le sourire ou les encouragements verbaux augmentent l'attention et la satisfaction des apprenants, bien que leur impact sur la rétention des connaissances soit limité.

De manière similaire, les travaux de Kim *et al.* (2007) montrent que les réponses empathiques des agents pédagogiques augmentent significativement l'intérêt et la confiance des apprenants. Ces interactions empathiques contribuent à renforcer l'engagement des apprenants, tout en offrant un soutien pédagogique ciblé adapté à leurs besoins émotionnels.

2.2.2.2 Les rôles éducatifs diversifiés des agents conversationnels

Les agents conversationnels pédagogiques jouent un rôle central dans l'éducation en répondant aux besoins spécifiques des apprenants et aux exigences des différents contextes d'apprentissage. Ainsi, Roos (2018) distinguent clairement les rôles des agents conversationnels pédagogiques, tels que tuteurs, évaluateurs et facilitateurs, de leurs caractéristiques fonctionnelles, comme la capacité à fournir des rétroactions ou à simuler des interactions sociales engageantes. Ces agents peuvent agir en tant que tuteurs virtuels ou facilitateurs de collaboration, tout en simulant des interactions sociales engageantes.

Par ailleurs, le concept de « Learning Companion System », inspiré des travaux de Chan et Baskin (1988), propose un cadre où les apprenants interagissent avec un compagnon d'apprentissage. Ce compagnon encou-

rage une participation active et favorise l'autonomie grâce à des approches adaptatives. Ces agents conversationnels se distinguent par leur capacité à simuler des interactions sociales collaboratives, soutenant ainsi l'apprentissage collectif et individuel. Ces interactions se traduisent également par des opportunités pour personnaliser les apprentissages, en adaptant les contenus et les approches aux besoins spécifiques des apprenants (Woolf, 2010).

En outre, les agents conversationnels pédagogiques ne se limitent pas aux interactions sociales, mais peuvent intégrer des stratégies pédagogiques avancées qui encouragent un apprentissage profond et interactif. Par exemple, le concept de « learning by teaching » initialement proposé dans des environnements tutoriels intelligents est particulièrement pertinent. Cette approche engage les apprenants dans une réflexion active en les incitant à enseigner un agent virtuel. Woolf (2010) explique que ces agents « enseignables » permettent aux étudiants de construire des cartes conceptuelles ou des diagrammes causaux, puis d'utiliser l'agent pour prendre des décisions, explorer les interactions qualitatives entre concepts ou répondre à des questions. En structurant leurs connaissances de cette manière, les apprenants renforcent leur rétention cognitive et leur capacité à articuler des concepts complexes.

De manière similaire, l'approche du « learning by disturbing » (Aimeur *et al.*, 1997), où le tuteur agit parfois comme un « perturbateur » en émettant des recommandations erronées ou contradictoires, vise à inciter les apprenants à justifier leurs réponses et à évaluer leurs propres opinions. Ce processus repose sur des conflits intentionnels qui stimulent une confrontation constructive et encouragent la réflexion critique. Woolf (2010) souligne que ces conflits peuvent enrichir l'apprentissage coopératif en exposant des divergences d'idées et en les résolvant par la discussion. En adaptant ces stratégies aux interactions avec les agents conversationnels pédagogiques, il est possible d'enrichir l'expérience éducative en stimulant l'autonomie, l'exploration critique, et la confiance des apprenants dans leur raisonnement (Woolf, 2010).

Enfin, l'émergence des grands modèles de langage constitue une avancée significative. Ces modèles permettent aux agents conversationnels pédagogiques d'offrir des réponses plus précises, contextuelles et adaptées aux besoins individuels des apprenants. Lorsqu'elles sont combinées aux stratégies pédagogiques avancées, leurs capacités peuvent révolutionner les interactions éducatives en proposant des expériences d'apprentissage personnalisées et engageantes.

2.2.3 Agents conversationnels basés sur les grands modèles de langage en éducation

L'émergence des LLMs, comme ChatGPT, a significativement amélioré les solutions éducatives, offrant des interactions plus riches et adaptées aux besoins variés des apprenants.

Par exemple, Abdelghani *et al.* (2022) ont montré que GPT-3 peut être intégré dans une plateforme éducative pour entraîner les enfants à poser des questions divergentes, un aspect essentiel pour stimuler la réflexion critique et l'apprentissage actif. Ces chercheurs ont conçu un agent conversationnel capable de générer automatiquement des indices pédagogiques adaptés, et ont montré que ces indices encouragent la curiosité et améliorent les compétences des enfants. Cette étude met en lumière l'importance de combiner les capacités des LLMs avec des cadres pédagogiques structurés afin d'optimiser leur effet sur l'apprentissage.

De même, Xie *et al.* (2024) ont étudié l'effet des agents conversationnels basés sur les LLMs sur l'autonomie d'apprentissage. Leur étude montre que, pour les apprenants recherchant une compagnie virtuelle, l'autonomie est renforcée via la présence sociale, tandis que pour ceux axés sur l'acquisition de connaissances, des interactions fréquentes réduisent à la fois la présence sociale et l'autonomie. Ces résultats soulignent la nécessité d'équilibrer soutien émotionnel et aide cognitive pour maximiser les bénéfices pédagogiques.

Les robots conversationnels basés sur les LLMs ont également été explorés pour leur rôle dans le renfor-

gement des interactions sociales dans les MOOC, avec un potentiel pour réduire les taux d'abandon. Par exemple, Li et Xing (2021) ont démontré que GPT-2 peut soutenir les échanges dans les forums de discussion en générant des réponses à la fois émotionnelles, informatives et communautaires. Leur étude a évalué la qualité de ces réponses sur des critères tels que la lisibilité, la cohérence et la pertinence, et a constaté qu'elles étaient proches de celles générées par des pairs humains. De plus, les apprenants ont évalué l'agent conversationnel à travers des questionnaires, mettant en avant leur satisfaction quant à l'utilité et à la pertinence des réponses.

Malgré leur potentiel, les LLMs présentent des limites importantes, notamment leur tendance aux « hallucinations »—la génération de contenus incorrects ou non pertinents (Alkaissi et McFarlane, 2023; Ji *et al.*, 2023). Ces erreurs, particulièrement problématiques dans un cadre éducatif où la fiabilité des informations est essentielle, peuvent perturber la compréhension des apprenants et compromettre la qualité de leur apprentissage. Pour remédier à ces limitations, la Génération Augmentée par Récupération (RAG) a été introduite. Cette méthode permet aux LLMs de s'appuyer sur des bases de connaissances externes validées, améliorant ainsi la précision, la pertinence et la fiabilité des réponses (Lewis *et al.*, 2020).

2.2.4 Agents conversationnels basés sur la génération augmentée par récupération en éducation

La RAG contribue de manière significative à améliorer la précision et la pertinence des modèles de langage dans les environnements éducatifs. Par exemple, Wang *et al.* (2023) ont développé ChatEd, un agent conversationnel pour l'enseignement supérieur, combinant un système de récupération d'information contextuelle avec ChatGPT. L'évaluation de ChatEd repose sur trois critères principaux : la pertinence, l'exactitude et l'utilité des réponses fournies. Comparé à ChatGPT seul, ChatEd a obtenu de meilleurs scores sur ces dimensions grâce à l'utilisation d'une base de données contextuelle, permettant d'aligner les réponses sur le contenu spécifique des cours. Ces résultats mettent en avant les avantages de coupler des modèles de

langage avancés à des bases de données contextuelles pour améliorer l'expérience d'apprentissage. En complément, Miladi *et al.* (2024b) ont comparé les modèles standard GPT-3.5 et GPT-4 à leurs versions enrichies par la RAG dans le contexte d'un MOOC consacré à l'intelligence artificielle. Les résultats ont confirmé que l'intégration de la RAG entraîne des améliorations significatives de la précision des réponses.

De même, Liu *et al.* (2024) ont développé le CS50 Duck, un agent conversationnel basé sur GPT-4 enrichi par la RAG, intégré au cours pour soutenir les étudiants. Cet agent a démontré une meilleure capacité à fournir des réponses précises et adaptées au contenu du cours, surpassant ChatGPT seul. Les étudiants ont perçu cet outil comme un tuteur personnel efficace et engageant, exprimant une forte satisfaction dans les questionnaires d'évaluation. Ils ont particulièrement apprécié la clarté des explications et la pertinence des réponses.

En parallèle, Taneja *et al.* (2024) ont présenté Jill Watson, une assistante virtuelle utilisant la RAG pour formuler des réponses contextualisées à partir de ressources pédagogiques. Ce système réduit les risques d'hallucinations tout en améliorant la qualité des réponses. L'étude a révélé une nette supériorité de Jill Watson par rapport à des assistants virtuels non enrichis par la RAG, tant en termes de qualité des réponses que de réduction des erreurs. De manière similaire, Yan *et al.* (2024) ont proposé VizChat, un agent conversationnel qui applique la RAG pour interpréter des tableaux de bord d'analytique de l'apprentissage. Ce système a permis de produire des explications plus précises et plus compréhensibles des données visuelles, favorisant ainsi une meilleure compréhension de la part des apprenants.

De plus, la précision des réponses produites par les assistants virtuels basés sur la RAG s'améliore de manière significative lorsque les segments de documents appropriés sont récupérés et intégrés dans le prompt, comme l'ont démontré Kuratomi *et al.* (2024). De manière similaire, le système CyberRAG développé par Zhao *et al.* (2024) pour l'enseignement de la cybersécurité renforce la pertinence et la fiabilité des réponses

généérées en récupérant des documents validés et en utilisant un cadre de validation basé sur une ontologie.

Bien que ces travaux mettent en évidence des résultats prometteurs sur des dimensions techniques telles que la précision et la pertinence, ils n'évaluent pas l'impact des agents conversationnels enrichis par la RAG sur l'acquisition des connaissances ni sur l'expérience d'apprentissage globale des apprenants dans des environnements éducatifs authentiques. Cette limite souligne la nécessité d'études centrées sur les effets pédagogiques concrets, ce à quoi répond la présente thèse. Celle-ci vise à examiner l'effet d'un agent conversationnel enrichi par la RAG dans un MOOC, en évaluant l'acquisition des connaissances, l'utilisabilité perçue et l'expérience d'apprentissage globale, à travers des dimensions telles que l'utilité perçue et la satisfaction.

Enfin, les émotions jouent un rôle central dans le processus d'apprentissage, influençant directement la réussite académique des étudiants en orientant leur engagement et leurs stratégies. De plus, ces états affectifs permettent de prédire la performance future, notamment grâce à l'effet des émotions positives, telles que l'intérêt et l'enthousiasme, sur la motivation et l'autonomie (Pekrun *et al.*, 2007). Pourtant, leur prise en compte dans les approches actuelles reste limitée. Aucun travail à ce jour n'a combiné LLM et RAG pour élaborer des stratégies visant à encourager ces émotions positives. Cette lacune met en évidence un potentiel de recherche pour concevoir des agents conversationnels intégrant des dimensions motivationnelles tout en assurant la précision et la pertinence contextuelle des réponses grâce à la RAG.

23 Conclusion

Ce chapitre a établi les fondations conceptuelles nécessaires pour comprendre les agents conversationnels enrichis par la RAG dans un contexte éducatif. L'exploration des LLMs et de la RAG a permis de mettre en lumière leur complémentarité, notamment en réduisant les hallucinations et en améliorant la précision ainsi que la pertinence des réponses.

Par ailleurs, la revue de la littérature a mis en évidence l'évolution des agents conversationnels, soulignant le rôle croissant des LLMs et de la RAG dans l'éducation. Cependant, bien que les recherches actuelles apportent des avancées notables, elles se concentrent principalement sur des indicateurs techniques, négligeant souvent des aspects essentiels comme l'effet pédagogique direct et les émotions des apprenants. Ce constat souligne la nécessité d'études approfondies pour mieux exploiter le potentiel éducatif de ces technologies, une lacune que ce travail de recherche vise à combler.

CHAPITRE 3

CONCEPTION ET ÉVALUATION DE L'AGENT CONVERSATIONNEL BASÉ SUR GPT-4 ET RAG

Ce chapitre répond au premier objectif de la thèse, qui consiste à évaluer la précision des réponses générées par un agent conversationnel basé sur GPT-4 enrichi par la RAG, dans le cadre d'un MOOC sur l'intelligence artificielle (RQ1). Il s'appuie sur les fondements théoriques présentés dans le chapitre 2, et décrit la mise en œuvre concrète de l'agent, avant son utilisation par les apprenants.

Le chapitre est structuré comme suit : nous présentons d'abord la conception technique de l'agent conversationnel, incluant l'intégration de GPT-4 et de la méthode RAG. Nous définissons ensuite les objectifs de l'évaluation, les hypothèses formulées et la méthodologie employée. Puis, nous exposons les résultats obtenus et en proposons une analyse critique, avant de conclure.

Ce chapitre fait l'objet de deux communications scientifiques valorisant les résultats de cette étude. La première, intitulée *Leveraging GPT-4 for Accuracy in Education : A Comparative Study on Retrieval-Augmented Generation in MOOCs* (Miladi et al., 2024b), a été acceptée à la conférence internationale *Artificial Intelligence in Education (AIED 2024)*. Elle met en lumière les bénéfices de l'intégration de la RAG dans les agents conversationnels, en démontrant que cette approche permet d'améliorer la précision des réponses générées dans un contexte éducatif.

La seconde communication, intitulée *Comparative Performance of GPT-4, RAG-Augmented GPT-4, and Students in MOOCs* (Miladi et al., 2024a), a été présentée à la conférence internationale *Intelligent Tutoring Systems (ITS 2024)*. Elle propose une étude comparative rigoureuse des performances de GPT-4, de GPT-4 enrichi par la RAG, et d'apprenants humains dans un MOOC sur l'intelligence artificielle.

3.1 Conception de l'Agent Conversationnel Basé sur GPT-4 et RAG

Notre agent conversationnel repose sur l'intégration d'un LLM et de la RAG. Il vise à offrir des interactions précises, contextualisées et adaptées aux besoins des apprenants, notamment dans des environnements éducatifs comme les MOOCs. L'intégration de la RAG permet de récupérer des informations pertinentes en temps réel, améliorant ainsi la qualité des réponses.

3.1.1 GPT-4

Le modèle de langage utilisé dans notre agent conversationnel repose sur GPT-4, disponible via l'API d'OpenAI¹. GPT-4 est un modèle de langage multimodal développé par OpenAI, qui représente une évolution significative par rapport à ses prédécesseurs, GPT-1, GPT-2 et GPT-3, tant en termes d'échelle que de capacités.

GPT-4 dispose d'une capacité de traitement du contexte pouvant atteindre 32 768 tokens, contre 4 096 tokens pour GPT-3. Les tokens représentent des unités de texte, telles que des mots, des ensembles de caractères ou des combinaisons de mots et de ponctuation, que les LLMs utilisent pour décomposer le texte et le traiter. Cette extension permet à GPT-4 de gérer des textes plus longs tout en maintenant une cohérence accrue dans ses réponses, facilitant ainsi des interactions plus complexes avec les utilisateurs.

3.1.2 RAG

La RAG exploite la puissance des LLMs en les enrichissant avec une base de connaissances externe. Son objectif principal est de surmonter certaines des limitations inhérentes aux LLMs, telles que le phénomène d'hallucination, où ces modèles peuvent produire des informations inexacts ou non pertinentes (Gao *et al.*, 2023). En combinant les capacités de récupération d'informations provenant de sources externes avec celles

1. <https://openai.com/index/gpt-4/>

de génération de texte, la RAG permet de générer des réponses plus précises, pertinentes et adaptées au contexte des apprenants.

L'architecture de notre agent conversationnel combine les capacités de l'IA générative, telle que GPT-4, pour la compréhension des langues naturelles, avec les avancées récentes de la RAG. Le processus repose sur plusieurs étapes séquentielles (voir Figure 3.1), illustrant les différentes phases de l'intégration de la RAG avec GPT-4 :

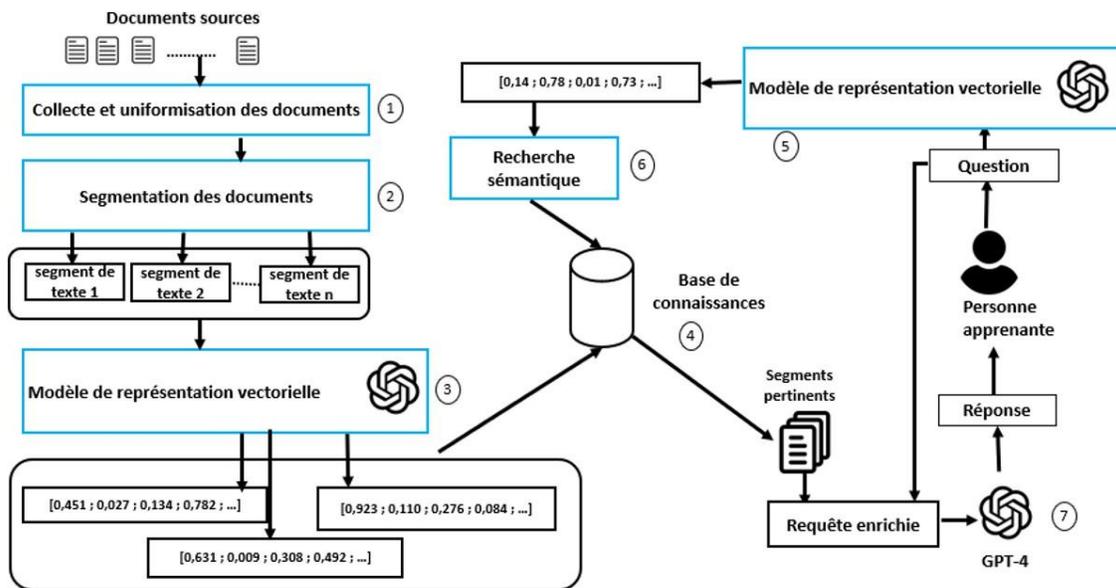


Figure 3.1 Architecture complète de l'agent conversationnel GPT-RAG.

1. *Collecte des Données* : Nous avons extrait les documents du MOOC sur l'intelligence artificielle (Psyché, 2020), comme source principale d'informations. Ces documents incluent des textes explicatifs, des transcriptions de vidéos ainsi que des tableaux.

Le choix de cette source s'explique par la richesse et la diversité du contenu textuel, qui offre une base complète pour fournir des réponses adaptées aux requêtes des apprenants. En particulier, les transcriptions de vidéos permettent de capturer des informations présentées oralement, offrant

ainsi un complément essentiel aux documents purement textuels. Cela enrichit la base de connaissances du modèle en intégrant à la fois des descriptions formelles et des informations supplémentaires provenant des transcriptions des vidéos.

Les données textuelles, qu'il s'agisse de textes ou de transcriptions, ont été extraites et converties en texte uniforme (voir Figure 3.1, étape 1). Cette uniformisation est cruciale pour garantir une cohérence dans les traitements ultérieurs, notamment dans la génération des représentations vectorielles et la recherche d'informations.

2. *Segmentation des documents* : Comme illustré dans la Figure 3.1, étape 2, une fois les documents préparés, l'étape suivante consiste à segmenter les données en unités plus petites. Nous avons choisi d'utiliser le découpage à longueur fixe à l'aide du diviseur de texte récursif par caractère de Langchain, avec une taille de segment de 2000 caractères et un chevauchement de 200 caractères entre chaque segment, conformément aux paramètres d'Aymeric Roucher². Les séparateurs utilisés pour déterminer les points de division sont les sauts de ligne doubles, simples, les points, les espaces, et enfin, aucune séparation si les autres séparateurs ne conviennent pas.

Ce choix se justifie par plusieurs raisons. Tout d'abord, le découpage à longueur fixe garantit que les segments ne dépassent pas une certaine taille, ce qui est crucial pour maintenir l'efficacité du processus d'encodage avec des modèles de type transformeur. Les modèles d'encodage, comme GPT-4, ont des limites en termes de longueur de séquences qu'ils peuvent traiter efficacement. Segmenter le texte en unités fixes permet de respecter ces contraintes et d'éviter des coûts en ressources trop élevés lors de l'encodage de longues séquences.

Ensuite, cette méthode est particulièrement adaptée aux cas où le texte doit être divisé rapidement

2. https://huggingface.co/learn/cookbook/en/rag_evaluation

et de manière uniforme, ce qui facilite le traitement de grandes quantités de données. Bien que cette méthode ne prenne pas en compte la structure sémantique du texte, elle assure une cohérence entre les segments en utilisant des séparateurs tels que des sauts de ligne, des points et des espaces pour découper le texte de manière logique et éviter de tronquer les phrases de manière aléatoire.

Enfin, le chevauchement de 200 caractères entre chaque segment permet de préserver le contexte entre les segments, minimisant ainsi les risques de perte d'information lors du découpage. En résumé, le découpage à longueur fixe offre un compromis efficace entre la gestion des ressources, la vitesse de traitement, et la conservation du contexte pertinent pour le processus de récupération d'informations.

3. *Conversion en vecteurs* : Nous avons utilisé le modèle de représentation vectorielle text-embedding-ada-002 (Neelakantan *et al.*, 2022) d'OpenAI pour convertir les documents segmentés en vecteurs (voir Figure 3.1, étape 3). Ces représentations vectorielles jouent un rôle central en traitement du langage naturel (NLP), permettant de représenter des éléments textuels tels que des mots, des phrases ou des documents sous forme de vecteurs denses de nombres réels. Ces vecteurs capturent les significations sémantiques, les nuances contextuelles et les relations entre les entités dans un espace vectoriel compact (Lebret, 2016). En transformant des données éparses et de haute dimension en un espace vectoriel dense et plus compact, ces représentations optimisent la capacité du système à fournir des réponses pertinentes aux requêtes des utilisateurs.
4. *Stockage dans une base de connaissances* : Après leur conversion en vecteurs, ces derniers sont stockés dans une base de connaissances spécialement conçue pour le domaine du MOOC, permettant une recherche rapide des informations pertinentes en fonction des requêtes des apprenants (voir Figure 3.1, étape 4). La création de cette base de connaissances contextuellement adaptée est essentielle pour garantir que le modèle d'IA fournisse des réponses précises, tout en réduisant les risques

de réponses erronées ou hors sujet.

Le choix de ne pas utiliser de techniques d'indexation des vecteurs s'explique par le volume modéré des données traitées. Cela nous permet de maintenir une recherche performante et rapide sans avoir besoin de méthodes d'indexation complexes, qui sont généralement requises pour de plus grandes bases de données ou lorsqu'un grand nombre d'utilisateurs doit accéder simultanément aux informations. En optant pour cette approche plus simple, nous assurons une efficacité optimale du système tout en préservant la pertinence des résultats obtenus.

5. *Conversion de la requête* : La requête soumise par l'apprenant est également convertie en vecteur à l'aide du modèle text-embedding-ada-002 (voir Figure 3.1, étape 5). Ce processus facilite une recherche sémantique en capturant l'intention et le contexte de la question, rendant ainsi possible l'identification de réponses pertinentes même si les termes exacts ne sont pas présents.

En utilisant cette approche, il devient possible de trouver des réponses pertinentes même si les mots-clés utilisés dans la requête ne correspondent pas exactement à ceux présents dans la base de connaissances. Ce procédé est particulièrement utile dans des environnements d'apprentissage comme un MOOC, où les apprenants peuvent formuler leurs questions de manière variée. La capacité à capturer le contexte et les nuances sémantiques permet de répondre aux requêtes avec une plus grande précision, garantissant ainsi une expérience apprenant optimisée.

6. *Recherche sémantique* : Le système effectue une recherche au sein de la base de connaissances en calculant la similarité cosinus entre le vecteur de la requête et ceux présents dans la base de connaissances (voir Figure 3.1, étape 6). Les résultats de cette recherche sont ensuite triés par ordre de pertinence, ce qui permet de sélectionner les réponses les plus adaptées aux besoins des apprenants.

Dans notre cas, nous avons opté pour une sélection des trois vecteurs les plus pertinents ($k=3$), afin de maximiser les chances de trouver des informations exactes tout en évitant une surcharge d'informations. Les segments de texte associés à ces trois vecteurs sont identifiés et ensuite combinés pour fournir un contexte enrichi à la requête initiale. Cette combinaison assure que les réponses générées soient à la fois précises et contextuellement adaptées, répondant aux questions des apprenants.

7. *Génération de réponse par GPT-4* : La requête enrichie est transmise à GPT-4, qui génère une réponse en s'appuyant sur le contexte des informations récupérées, comme illustré dans la Figure 3.2. Ce choix de transmission restreinte garantit que la réponse est construite à partir de données spécifiques et pertinentes, ce qui permet de réduire les risques d'erreurs ou de réponses inappropriées.

En utilisant des prompts enrichis par les segments de texte les plus pertinents, GPT-4 est capable de fournir une réponse à la fois plus précise et adaptée aux besoins de l'apprenant, même lorsque la formulation des questions varie. Ce processus est illustré dans la Figure 3.1, étape 7), qui montre comment GPT-4 utilise un prompt enrichi pour générer une réponse pertinente, répondant ainsi aux exigences spécifiques de l'apprenant.

Des extraits du code, portant spécifiquement sur les procédures de recherche des segments de texte pertinents et de génération de réponses à l'aide du modèle, sont présentés en Annexe B. Ces exemples montrent comment les requêtes des utilisateurs sont traitées à l'aide de la similarité cosinus pour identifier les segments les plus pertinents et comment ces informations sont ensuite utilisées pour générer une réponse adaptée à la question posée.

3.2 Objectifs

L'objectif principal de cette étude est d'évaluer l'efficacité du modèle GPT-4 enrichi par la technologie de génération augmentée par récupération (RAG) dans le cadre d'un MOOC sur l'intelligence artificielle. Cette

```
Prompt_template = """"  
Question : """" + question + """"  
Contexte : {context}  
Seul le contexte ci-dessus est utilisé et reformulé pour répondre à la question. Si le contexte ne  
permet pas de répondre, indiquez que vous ne connaissez pas la réponse.""""
```

Figure 3.2 Prompt utilisé dans notre agent conversationnel

recherche vise à comprendre comment l'intégration de la RAG influence la précision des réponses fournies par les modèles GPT et à comparer leurs performances à celles des apprenants humains.

Plus spécifiquement, l'étude évalue comment la RAG impacte la précision des réponses générées par les modèles GPT-4 et GPT-3.5. En comparant les performances des versions augmentées par RAG à leurs versions standard, nous cherchons à déterminer si cette technologie permet de produire des réponses plus précises et contextualisées dans le cadre d'un MOOC.

En outre, nous comparons la performance du modèle GPT-4 enrichi par RAG à celle des apprenants humains dans la résolution des exercices du MOOC. Cette analyse vise à évaluer dans quelle mesure le modèle peut égaler ou surpasser les apprenants en termes de précision des réponses.

33 Hypothèses de recherche

Cette étude examine l'effet de l'intégration de la génération augmentée par récupération (RAG) avec les modèles GPT sur la précision des réponses dans un contexte de MOOC sur l'intelligence artificielle. Les hypothèses sont formulées en lien avec la question de recherche suivante *RQ1* : Dans quelle mesure l'intégration de la RAG dans les modèles GPT permet-elle de réduire les erreurs d'hallucination et d'améliorer la précision des réponses générées dans un contexte éducatif, notamment lors de l'utilisation d'exercices issus d'un MOOC sur l'intelligence artificielle ?

Les hypothèses sont les suivantes :

- *H1 (précision des réponses)* : Les modèles GPT-3.5 et GPT-4, lorsqu'ils sont améliorés par la génération augmentée par récupération (RAG), surpasseront leurs versions standard respectives en termes de précision des réponses générées.
- *H2 (performance comparative)* : Le modèle GPT-4 enrichi par RAG surpassera à la fois le modèle GPT-3.5, sa version augmentée par RAG, ainsi que le modèle GPT-4 standard en termes de précision des réponses.
- *H3 (performance par rapport aux apprenants)* : Le GPT-4 enrichi par RAG dépassera les scores moyens des apprenants dans les exercices du MOOC.
- *H4 (meilleures performances)* : Le GPT-4 enrichi par RAG dépassera les scores les plus élevés obtenus par les apprenants dans les exercices du MOOC.

Ces hypothèses permettent d'évaluer les différences de précision entre les différents modèles GPT et de comparer leurs résultats avec ceux des apprenants dans un cadre éducatif.

34 Méthodologie d'évaluation

Dans cette section, nous décrivons d'abord l'ensemble des données utilisées pour nos expériences, en détaillant les exercices issus du MOOC. Nous présentons ensuite la méthodologie de la recherche, qui inclut l'évaluation de l'agent conversationnel à travers deux expériences distinctes visant à analyser les performances des modèles GPT enrichis par la génération augmentée par récupération (RAG).

34.1 Ressources pédagogiques et données

Dans cette étude, nous avons exploité les ressources pédagogiques du MOOC sur l'intelligence artificielle (Psyché, 2020). Ce cours est structuré en quatre modules principaux : introduction générale à l'IA, IA symbolique, IA connexionniste, et applications de l'IA en éducation.

Ces modules sont accompagnés d’une variété de ressources pédagogiques, notamment des vidéos explicatives, des textes approfondis, des définitions de concepts, et des exercices. Au total, 115 exercices d’évaluation formative ont été conçus pour tester et renforcer les apprentissages, répartis en quatre formats principaux : 24 exercices de type vrai/faux, 24 questions à choix multiples (QCM), 13 exercices d’appariement, et 54 exercices à trous. Le Tableau 3.1 présente des exemples de ces exercices utilisés pour évaluer les performances des modèles GPT.

Type d'exercice	Exemple d'énoncé
Vrai/Faux	Indiquez si la déclaration suivante est vraie ou fausse : Un agent intelligent ne peut ni adapter ses actions par rapport à son environnement ni agir sur celui-ci. 1. Vrai 2. Faux
Question à choix multiples	Sélectionnez la bonne réponse : Selon Yann LeCun, rendre une machine intelligente, c’est lui permettre : A. de rêver. B. de mémoriser. C. d’apprendre. D. de percevoir.
Exercice d'appariement	Associez chaque définition à son terme correspondant parmi les suivants : Définitions : 1. Différentes composantes technologies numériques, mathématiques et autres permettant la conception d’une voiture autonome. 2. Capacité d’un réseau de neurones à s’ajuster par lui-même, modifiant son comportement en fonction d’un environnement, cette capacité pouvant servir durant la phase d’apprentissage. 3. Main robotisée ayant appris par manipulation par essais et erreurs à manipuler un cube Rubik. Termes : A. Intelligence artificielle B. Adaptabilité C. Agent intelligent
Exercice à trous	Complétez la phrase : Pour réussir le test de ..., l’ordinateur doit être équipé d’un dispositif de vision artificielle pour percevoir des objets et d’une capacité robotique pour manipuler des objets et se déplacer.

Table 3.1 Types d’exercices utilisés pour évaluer les performances des modèles GPT

3.4.2 Protocole expérimental

Dans le cadre de cette étude, nous avons mené deux expériences distinctes. La première expérience vise à comparer les performances des modèles GPT dans leurs versions standard et enrichies par RAG, en se focalisant sur la précision des réponses générées. La seconde expérience, quant à elle, se concentre sur la comparaison des performances du modèle GPT-4 enrichi par RAG avec celles des apprenants humains.

3.4.2.1 Expérience 1 : Évaluation comparative des performances des modèles GPT avec et sans RAG

Dans cette section, nous décrivons l'expérimentation menée pour comparer les performances des modèles GPT avec et sans l'intégration de la génération augmentée par récupération. Nous présentons d'abord la procédure utilisée pour tester les modèles, puis le processus d'évaluation des réponses générées.

3.4.2.1.1 Procédure

Nous avons utilisé les 115 exercices d'évaluation formative du MOOC sur l'intelligence artificielle comme requêtes pour tester les quatre versions des modèles GPT (GPT-3.5 standard, GPT-3.5 enrichi par RAG, GPT-4 standard et GPT-4 enrichi par RAG). Le prompt du modèle GPT enrichi par RAG a été configuré pour inclure des segments contextuels pertinents, sélectionnés à partir d'une base de connaissances d'embeddings textuels spécifiques au MOOC (voir section 3.1.2).

3.4.2.1.2 Processus d'évaluation

Les exercices ont été traités de manière séquentielle, selon leur ordre d'apparition dans le MOOC, sans rétroaction sur les réponses incorrectes. Un évaluateur indépendant, doctorante en informatique cognitive, a évalué les réponses à l'aide d'une échelle binaire (0-1), où « 1 » correspondait à une réponse correcte et « 0 » à une réponse incorrecte ou comportant des fabrications (hallucinations). Ce système de notation binaire a simplifié le processus d'évaluation, facilitant ainsi la comparaison directe des performances des quatre modèles GPT sur l'ensemble des exercices.

3.4.2.2 Expérience 2 : Évaluation comparative des performances du GPT-4 enrichi par RAG et des apprenants

Cette section présente une étude expérimentale visant à comparer les performances du modèle GPT-4 enrichi par la RAG avec celles d'apprenants humains dans le cadre du premier module d'un MOOC sur

l'intelligence artificielle. Nous décrivons successivement les caractéristiques des participants, la procédure expérimentale, ainsi que le processus d'évaluation des performances.

3.4.2.2.1 Participants

L'étude a impliqué 20 apprenants, dont 7 hommes et 13 femmes, âgés de 22 à 35 ans. Les participants étaient issus de diverses disciplines : 60% en informatique, 25% en éducation et 15% en ingénierie. Ils ont été recrutés par un appel à volontaires diffusé auprès d'étudiants de niveau licence, master et doctorat. Tous ont participé de manière volontaire, après avoir signé un formulaire de consentement éclairé. L'étude a été préalablement approuvée par le Comité d'éthique de la recherche de l'Université TÉLUQ (certificat n° 10/2023).

3.4.2.2.2 Procédure

L'expérimentation s'est déroulée à distance. Les participants ont été invités à s'inscrire sur la plateforme du MOOC, à suivre le contenu du module 1, puis à compléter un total de 33 exercices. Afin de garantir l'équité et l'authenticité des réponses, il leur a été explicitement demandé de ne pas recourir à une aide extérieure ni à des recherches sur Internet. Les exercices ont été réalisés dans des conditions comparables à celles imposées au modèle GPT-4 enrichi par RAG, assurant ainsi une base équitable de comparaison entre les performances humaines et celles du modèle.

3.4.2.2.3 Processus d'évaluation

Les exercices ont été complétés par les apprenants et le modèle GPT-4 augmenté par RAG dans des conditions identiques. Un évaluateur indépendant, doctorante en informatique cognitive, a évalué les réponses sur une échelle binaire (0-1), où « 1 » correspond à une réponse correcte et « 0 » à une réponse incorrecte ou à une hallucination, comme indiqué dans la Section 3.4.2.1.2. Ce système d'évaluation a permis d'assurer

une analyse équitable des performances du modèle et des apprenants.

3.5 Résultats

Dans cette section, nous présentons les résultats de l'intégration de la génération augmentée par récupération sur les performances des modèles GPT en termes de précision des réponses. Nous analysons également la manière dont la performance du modèle GPT-4 enrichi par RAG se compare à celles des apprenants dans les exercices du MOOC.

3.5.1 Analyse des performances des modèles GPT avec l'intégration de la RAG dans un MOOC

Nous avons comparé les performances des modèles GPT standard et des modèles GPT enrichis par RAG en utilisant un ensemble de données de 115 exercices issus d'un MOOC sur l'IA. Les résultats, résumés dans les tableaux 3.2, 3.3, 3.4, et 3.5 montrent une amélioration progressive de la précision à travers les différentes versions des modèles GPT. Le modèle GPT-3.5 a atteint un taux de réussite de base de 60%. Cette performance a été améliorée par le GPT-3.5 enrichi par RAG, qui a atteint un taux de réussite de 74%. Le modèle GPT-4 a encore augmenté cette précision avec un taux de réussite de 77%, et le modèle GPT-4 enrichi par RAG a atteint le taux de réussite le plus élevé, à savoir 80%.

3.5.1.1 Exercices vrai/faux

Dans les exercices Vrai/Faux, comme indiqué dans le Tableau 3.2, le modèle GPT-3.5 a montré un taux de réussite initial de 65%. Ce taux a été amélioré par le modèle GPT-3.5 enrichi par RAG, qui a atteint 85%. Les modèles GPT-4 et sa version enrichie ont encore amélioré ces performances, atteignant un taux de réussite de 87%, représentant ainsi le plus haut niveau de précision parmi les modèles testés.

Thème du module du MOOC	Exercices Vrai/Faux			
	GPT-3.5	GPT-3.5 enrichi par RAG	GPT-4	GPT-4 enrichi par RAG
Concepts généraux de l'IA	7/8 (87%)	8/8 (100%)	8/8 (100%)	8/8 (100%)
Concepts de l'IA symbolique	3/4 (75%)	3/4 (75%)	4/4 (100%)	4/4 (100%)
Concepts de l'IA connexionniste	3/6 (50%)	5/6 (83%)	4/6 (67%)	5/6 (83%)
Concepts de l'IA en éducation	3/6 (50%)	5/6 (83%)	5/6 (83%)	4/6 (67%)
Total	65%	85%	87%	87%

Table 3.2 Résultats de l'évaluation des exercices Vrai/Faux

3.5.12 Questions à Choix Multiples (QCM)

Comme indiqué dans le Tableau 3.3, le modèle GPT-3.5 standard a atteint un taux de réussite de 60%, qui a été amélioré à 73% avec le modèle GPT-3.5 enrichi par RAG. Par la suite, les modèles GPT-4 et sa version enrichie par RAG ont encore amélioré ce taux de réussite, atteignant 76%.

Thème du module du MOOC	Exercices QCM			
	GPT-3.5	GPT-3.5 enrichi par RAG	GPT-4	GPT-4 enrichi par RAG
Concepts généraux de l'IA	4/7 (57%)	5/7 (71%)	5/7 (71%)	5/7 (71%)
Concepts de l'IA symbolique	5/7 (71%)	5/7 (71%)	5/7 (71%)	5/7 (71%)
Concepts de l'IA connexionniste	5/8 (62%)	4/8 (50%)	5/8 (62%)	5/8 (62%)
Concepts de l'IA en éducation	1/2 (50%)	2/2 (100%)	2/2 (100%)	2/2 (100%)
Total	60%	73%	76%	76%

Table 3.3 Résultats de l'évaluation des QCMs

3.5.13 Exercices d'appariement

Le modèle GPT-3.5 a atteint un taux de réussite de 67%, qui a été amélioré à 75% avec le modèle GPT-3.5 enrichi par RAG, comme indiqué dans le Tableau 3.4. Le modèle GPT-4 a poursuivi cette tendance d'amélioration, atteignant un taux de réussite de 81%, tandis que le modèle GPT-4 enrichi par RAG a enregistré la meilleure précision avec un taux de réussite de 87%.

Thème du module du MOOC	Exercices d'appariement			
	GPT-3.5	GPT-3.5 enrichi par RAG	GPT-4	GPT-4 enrichi par RAG
Concepts généraux de l'IA	2/4 (50%)	2/4 (50%)	3/4 (75%)	4/4 (100%)
Concepts de l'IA symbolique	2/2 (100%)	2/2 (100%)	2/2 (100%)	2/2 (100%)
Concepts de l'IA connexionniste	2/4 (50%)	2/4 (50%)	2/4 (50%)	2/4 (50%)
Concepts de l'IA en éducation	2/3 (67%)	3/3 (100%)	3/3 (100%)	3/3 (100%)
Total	67%	75%	81%	87%

Table 3.4 Résultats de l'évaluation des exercices d'appariement

35.14 Exercices à trous

Dans les exercices à trous, comme indiqué dans le Tableau 3.5, les performances des modèles ont montré une progression significative. Le modèle GPT-3.5 a débuté avec un taux de réussite de 48%, qui a été nettement amélioré à 63% avec le modèle GPT-3.5 enrichi par RAG. Par la suite, le modèle GPT-4 a atteint un taux de réussite de 65%, tandis que le modèle GPT-4 enrichi par RAG a encore amélioré ces performances, atteignant 72%.

Thème du module du MOOC	Exercices à trous			
	GPT-3.5	GPT-3.5 enrichi par RAG	GPT-4	GPT-4 enrichi par RAG
Concepts généraux de l'IA	9/14 (64%)	8/14 (71%)	11/14 (79%)	10/14 (71%)
Concepts de l'IA symbolique	8/13 (61%)	11/13 (85%)	9/13 (69%)	13/13 (100%)
Concepts de l'IA connexionniste	5/13 (38%)	7/13 (54%)	7/13 (54%)	8/13 (62%)
Concepts de l'IA en éducation	4/14 (29%)	6/14 (43%)	8/14 (57%)	8/14 (57%)
Total	48%	63%	65%	72%

Table 3.5 Résultats de l'évaluation des exercices à trous

35.2 Comparaison des performances du GPT-4 enrichi par la RAG avec celles des apprenants dans un MOOC

Nous avons évalué les performances du modèle GPT-4 augmenté par RAG sur 33 types d'exercices différents dans le MOOC et comparé ses résultats à ceux des apprenants. Le modèle a démontré des performances supérieures, surpassant la moyenne des apprenants dans tous les types d'exercices. Il a atteint un taux de

réussite globale de 85% contre une moyenne de 60% pour les apprenants, comme illustré à la Figure 3.3. Notamment, le modèle a obtenu un score parfait de 100% dans les exercices Vrai/Faux et d'appariement, tandis que sa performance dans les exercices à trous a atteint 71%, surpassant de manière considérable la moyenne des apprenants, fixée à seulement 35%.

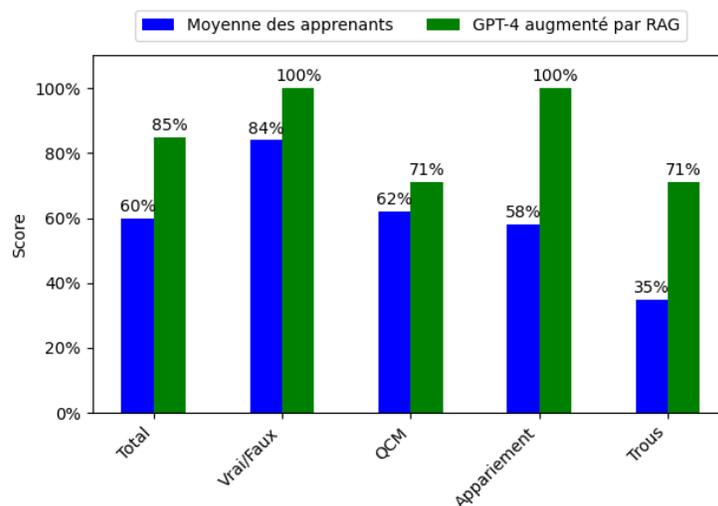


Figure 3.3 Scores en pourcentage de la moyenne des apprenants (bleu) et du modèle GPT-4 augmenté par RAG (vert) pour différents types d'exercices

Comme détaillé dans le Tableau 3.6, l'analyse des performances des apprenants montre des résultats variés selon les types d'exercices. La médiane, représentant la valeur centrale d'un ensemble de données ordonnées, et le mode, indiquant la valeur la plus fréquente, offrent un aperçu précis des performances des apprenants. Pour les exercices Vrai/Faux, la médiane et le mode atteignent un score parfait de 100%. Pour les exercices à choix multiples, le taux de réussite médian diminue à 57%, tandis que le mode est encore plus faible à 43%. Les exercices d'appariement montrent un taux de réussite médian de 75% et un mode de 100%. Enfin, les exercices à trous, plus complexes, montrent des résultats médians et modes de 43%.

Une analyse plus approfondie des performances du modèle GPT-4 augmenté par RAG révèle ses capacités supérieures par rapport aux scores des apprenants humains. Par exemple, dans les exercices Vrai/Faux et

Type d'exercice	Taux de réussite médian (%)	Taux de réussite mode (%)
Vrai/Faux	100	100
QCM	57	43
Appariement	75	100
Trous	43	43

Table 3.6 Médiane et mode pour chaque type d'exercice

à choix multiples, le modèle se classe au 52e percentile, ce qui signifie qu'il surpasse les scores de 52% des participants humains. En ce qui concerne les exercices d'appariement, le GPT-4 augmenté par RAG atteint le 67e percentile, démontrant sa capacité à surpasser la majorité des apprenants. Enfin, dans les exercices à trous, le modèle se distingue encore davantage en se positionnant au 95e percentile, ce qui le place parmi les 5% des meilleures performances. Cela souligne clairement sa supériorité sur presque la totalité des apprenants dans cette catégorie d'exercices.

3.6 Discussion

Cette section analyse les résultats de l'étude en réponse aux questions de recherche. Nous examinons l'impact de l'intégration de la RAG dans les modèles GPT sur la précision des réponses, comparons les performances du GPT-4 enrichi par RAG à celles des apprenants, et discutons des implications pédagogiques de ces résultats dans un contexte éducatif.

3.6.1 (RQ1) : Dans quelle mesure l'intégration de la RAG dans les modèles GPT permet-elle de réduire les erreurs d'hallucination et d'améliorer la précision des réponses générées dans un contexte éducatif, notamment lors de l'utilisation d'exercices issus d'un MOOC sur l'intelligence artificielle ?

Notre étude a évalué les modèles GPT enrichis par RAG à travers 115 exercices d'évaluation. Nos résultats montrent que le modèle GPT-3.5, lorsqu'il est enrichi par RAG, dépasse le modèle GPT-3.5 standard en termes de précision des réponses. De même, le modèle GPT-4 enrichi par RAG présente des performances supérieures à celles du modèle GPT-4 standard. Ces résultats peuvent s'expliquer par l'extraction d'informations pertinentes à partir d'une base de données externe spécifique au MOOC, ce qui permet de fournir

des réponses précises et contextualisées. Ainsi, notre hypothèse (H1), selon laquelle le GPT enrichi par RAG surpasse les versions standard en termes de précision des réponses, est validée.

Toutefois, bien que les performances globales soient en hausse avec l'intégration de la RAG, certaines exceptions méritent d'être discutées. D'une part, l'ajout de la RAG au modèle GPT-4 n'a pas systématiquement entraîné une augmentation du taux de réussite. Par exemple, les taux de précision pour les exercices de type QCM et Vrai/Faux sont identiques entre GPT-4 et GPT-4 enrichi par RAG (respectivement 76% et 87%). Cela suggère que, dans le cas de tâches relativement simples, le modèle GPT-4, déjà très performant grâce à son entraînement sur un large corpus, atteint un certain plafond de performance, rendant l'apport de la RAG moins significatif.

De plus, nos résultats montrent que le modèle GPT-4 enrichi par RAG a dépassé non seulement le modèle GPT-3.5 standard et enrichi, mais également le modèle GPT-4 standard. Cette performance valide notre hypothèse (H2), suggérant que le GPT-4 enrichi par RAG est particulièrement efficace pour produire des réponses précises dans un contexte éducatif.

L'analyse des résultats montre que le modèle GPT-4 enrichi par RAG a largement dépassé les scores moyens des apprenants, validant ainsi l'hypothèse (H3). Bien que l'hypothèse (H4) ne soit que partiellement confirmée, les performances du modèle restent impressionnantes, notamment dans les exercices à trous, où il se classe parmi les 5% des meilleurs scores.

Le modèle GPT-4 enrichi par RAG excelle dans ces exercices, qui exigent la mémorisation d'informations factuelles et leur application contextuelle. Ces compétences, souvent difficiles à maîtriser pour les apprenants, sont reflétées par leurs scores médians et modes relativement bas. En revanche, le modèle GPT-4 enrichi par RAG démontre une compétence notable dans ces tâches complexes. Grâce à ses capacités avancées

de récupération et de traitement de données contextuelles, il surmonte des défis qui posent souvent des difficultés aux apprenants humains.

Notre étude a principalement exploré les capacités des modèles GPT enrichis par RAG, en se concentrant sur l'amélioration de la précision et de la contextualisation des réponses dans un cadre éducatif. Contrairement à une évaluation plus large des technologies d'IA générative telles que Gemini ou Copilot, notre approche s'est focalisée sur l'effet spécifique de l'intégration de RAG au sein du modèle GPT-4.

3.6.2 Implications pédagogiques de l'intégration du RAG dans les modèles GPT

Les résultats de notre étude montrent que l'intégration du modèle GPT-4 enrichi par RAG peut considérablement améliorer les pratiques pédagogiques dans un contexte d'apprentissage en ligne. Ce modèle, avec ses capacités accrues en matière de précision et de contextualisation des réponses, surpasse non seulement les versions standard des modèles GPT, mais également les performances des apprenants humains dans plusieurs types d'exercices. Ces résultats soulèvent des implications pédagogiques importantes tant pour les enseignants que pour les étudiants.

Pour les enseignants, le modèle GPT-4 enrichi par RAG pourrait constituer un outil stratégique pour concevoir des contenus pédagogiques adaptés aux besoins cognitifs des apprenants. En modulant la complexité des supports selon leur niveau de compréhension ou leur rythme d'apprentissage, il devient possible de favoriser un apprentissage plus efficace tout en prévenant la surcharge cognitive (Sweller, 2011).

Il est recommandé que les enseignants adoptent des approches d'évaluation axées sur le développement des compétences cognitives de haut niveau, telles que la pensée critique, la résolution de problèmes et la créativité, qui ne sont pas facilement reproductibles par les modèles GPT. Plutôt que de se limiter aux évaluations traditionnelles, l'intégration d'activités complexes, comme la résolution de problèmes et l'ap-

prentissage collaboratif, permet aux étudiants de démontrer activement leur compréhension.

Pour les étudiants, le modèle GPT-4 enrichi par RAG pourrait transformer l'expérience éducative en offrant un dialogue immersif et interactif. En agissant comme un "compagnon d'apprentissage", tel qu'envisagé par Chan et Baskin (Chan et Baskin, 1988), ce modèle peut fournir des clarifications instantanées et des explications adaptées au niveau de compréhension de chaque étudiant. Cette interaction personnalisée favorise l'apprentissage actif, la pensée critique, et permet aux étudiants d'apprendre à leur propre rythme tout en renforçant la rétention des connaissances et la motivation.

Cependant, il est essentiel de noter que, malgré ses capacités avancées, le modèle GPT-4 enrichi par RAG ne remplace pas les enseignants humains. En complément des enseignants, il peut néanmoins constituer un outil puissant pour améliorer la réactivité de l'apprentissage et faciliter l'acquisition des connaissances pour les apprenants dans les environnements éducatifs numériques.

3.6.3 Limites

Malgré les résultats prometteurs de notre étude, il est essentiel de reconnaître ses limites pour obtenir une compréhension complète. Cette recherche a été menée dans le cadre d'un seul MOOC, sur une plateforme spécifique, et portait sur des exercices d'évaluation en langue française, incluant des questions à choix multiples, des exercices vrai/faux, d'appariement et des exercices à trous.

Par ailleurs, la taille de l'échantillon de participants ($N = 20$), ainsi que le corpus de 115 exercices utilisés pour mesurer les performances des modèles GPT, représentent un volume restreint. Ce volume de données demeure limité d'un point de vue statistique, ce qui restreint la généralisation des conclusions à d'autres domaines disciplinaires. Des recherches complémentaires, menées sur des échantillons plus larges, diversifiés et multilingues, permettraient de renforcer la robustesse des résultats et d'en évaluer la transférabilité

dans des contextes variés.

3.7 Conclusion

Dans ce chapitre, nous avons présenté la conception de l'agent conversationnel basé sur la RAG avec GPT-4, destiné à accompagner les apprenants dans le cadre du MOOC sur l'intelligence artificielle. Cet agent a été conçu pour répondre aux besoins des apprenants en leur offrant des réponses précises et adaptées, issues d'une base de connaissances validée, et en surmontant les limites des modèles génératifs traditionnels, comme les erreurs d'hallucination.

Nous avons ensuite évalué l'effet de l'intégration de la génération augmentée par récupération dans les modèles GPT-4 et GPT-3.5 sur la précision des réponses générées, ainsi que la performance comparative du modèle GPT-4 enrichi par RAG par rapport aux apprenants dans le cadre d'un MOOC sur l'intelligence artificielle.

Pour évaluer cet effet, nous avons mené deux expériences distinctes. La première avait pour objectif de comparer les performances des versions standard et enrichies par RAG des modèles GPT, en se concentrant sur la précision des réponses. La seconde visait à comparer les résultats obtenus par le modèle GPT-4 enrichi par RAG à ceux des apprenants humains, afin de mieux comprendre son efficacité dans un contexte éducatif.

Les résultats confirment en grande partie nos hypothèses. Les modèles GPT enrichis par la génération augmentée par récupération ont montré une précision nettement supérieure à leurs versions standard. De plus, le modèle GPT-4 enrichi par RAG a non seulement surpassé ses versions antérieures, mais il a également obtenu de meilleures performances que les apprenants humains, notamment dans les exercices les plus complexes. Ces résultats suggèrent que l'intégration du RAG dans les modèles GPT renforce leur capacité à fournir des réponses plus précises et adaptées, tout en répondant potentiellement mieux aux besoins des

apprenants dans un environnement éducatif en ligne.

CHAPITRE 4

MESURE DE L'EFFET D'UN AGENT CONVERSATIONNEL AMÉLIORÉ PAR LA RAG SUR L'ACQUISITION DES CONNAISSANCES ET L'EXPÉRIENCE DES APPRENANTS

Ce chapitre s'inscrit dans la continuité du travail présenté au chapitre 3, où nous avons montré que le modèle GPT-4 enrichi par la RAG surpasse les versions classiques de GPT ainsi que les performances moyennes des apprenants dans les exercices du MOOC.

Il répond ici au deuxième objectif de la thèse, qui vise à évaluer les effets de cet agent conversationnel sur l'acquisition des connaissances, l'utilisabilité perçue et l'expérience d'apprentissage dans un MOOC sur l'intelligence artificielle (RQ2).

Nous y présentons une étude expérimentale centrée sur l'interaction entre des étudiants et l'agent conversationnel. Le chapitre est structuré comme suit : nous commençons par exposer les objectifs de l'étude et les hypothèses formulées. Nous décrivons ensuite la méthodologie adoptée, incluant les participants, les procédures et les instruments de recherche. Les résultats des analyses quantitatives et qualitatives sont ensuite présentés, suivis d'une discussion des résultats au regard des hypothèses et d'une conclusion.

Ce chapitre est associé à deux publications scientifiques valorisant les résultats de cette recherche. La première communication, intitulée *Evaluating a GPT-4 and Retrieval-Augmented Generation-Based Conversational Agent to Enhance Learning Experience in a MOOC* (Miladi et al., 2025b), a été publiée dans les actes de la conférence internationale *CSEDU 2025* (International Conference on Computer Supported Education). Elle repose sur une analyse quantitative approfondie, portant notamment sur l'acquisition des connaissances et l'utilisabilité perçue d'un agent conversationnel enrichi par la RAG.

Le deuxième article, publié dans la *Revue internationale des technologies en pédagogie universitaire (RITPU)* (Miladi et al., 2025a), est intitulé *Conception et évaluation d'un agent conversationnel enrichi par la génération augmentée par récupération : effet sur l'acquisition des connaissances des personnes apprenantes, l'utilisabilité perçue et l'expérience d'interaction*. Il s'appuie sur une méthodologie mixte combinant des analyses quantitatives et qualitatives, incluant notamment des entretiens semi-structurés réalisés dans le cadre d'un MOOC sur l'intelligence artificielle. Ces deux publications illustrent l'ampleur de la contribution de cette thèse en matière d'évaluation empirique d'un agent conversationnel enrichi par la RAG dans des contextes éducatifs concrets.

4.1 Objectifs

Cette étude a pour objectif d'évaluer l'effet d'un agent conversationnel enrichi par la RAG sur l'apprentissage et l'expérience des apprenants dans un MOOC. Plus précisément, il s'agit de mesurer comment l'utilisation de cet agent influence l'acquisition des connaissances en intelligence artificielle. Pour ce faire, les résultats en termes d'acquisition des connaissances des apprenants utilisant le robot conversationnel seront comparés à ceux d'un groupe de contrôle n'utilisant pas cet outil, afin de déterminer si l'agent conversationnel favorise une meilleure assimilation des concepts.

Nous évaluons également l'utilisabilité de l'agent conversationnel à l'aide de l'échelle SUS (System Usability Scale) (Brooke, 1996), afin de mesurer dans quelle mesure l'agent est perçu comme facile à utiliser, intuitif et fonctionnel. Cette notion renvoie aux dimensions classiques de l'ergonomie et de la facilité d'usage, telles que définies dans la littérature sur l'utilisabilité (Nielsen, 1994). Un score élevé en termes d'utilisabilité indiquerait que l'agent conversationnel est perçu comme simple d'usage et bien conçu pour répondre aux besoins des apprenants.

En complément, cette étude explore l'expérience d'interaction globale des apprenants avec l'agent conver-

sationnel. Ce concept dépasse la simple utilisabilité et englobe des aspects subjectifs tels que la satisfaction, le ressenti émotionnel, la perception de l'utilité pédagogique et le confort dans l'interaction. Pour cela, les perceptions des apprenants sont recueillies à travers des entretiens semi-structurés, afin de fournir une compréhension approfondie de l'effet de l'agent conversationnel sur l'expérience globale d'apprentissage, au-delà de la seule mesure de son utilisabilité.

4.2 Hypothèses de recherche

Dans le cadre de cette étude, nous cherchons à comprendre l'effet d'un agent conversationnel amélioré par GPT et RAG sur l'acquisition des connaissances, son utilisabilité et l'expérience des apprenants dans un MOOC. Les hypothèses sont formulées en lien avec la question de recherche *RQ2* suivante : Quels sont les effets d'un agent conversationnel enrichi par la RAG sur l'acquisition des connaissances, l'utilisabilité perçue et l'expérience des apprenants dans un contexte de MOOC sur l'intelligence artificielle ?

Les hypothèses sont les suivantes :

- *H1 (Apprentissage)* : Les apprenants utilisant l'agent conversationnel enrichi par GPT et RAG montreront une amélioration significative de l'acquisition de leurs connaissances, comparativement à ceux qui ne bénéficient pas de cet agent.
- *H2 (utilisabilité)* : L'agent conversationnel enrichi par GPT et RAG sera perçu comme ayant une utilisabilité élevée par les apprenants.
- *H3 (expérience d'interaction)* : Les apprenants évalueront l'expérience d'interaction avec l'agent conversationnel enrichi par GPT et RAG de manière globalement positive.

Ces hypothèses serviront de base à l'analyse des données collectées et permettront de tester empiriquement l'effet de l'agent conversationnel sur les différents aspects de l'apprentissage et de l'expérience utili-

sateur dans un environnement de MOOC.

4.3 Méthodologie de recherche

Cette section présente la méthodologie de l'étude, qui évalue l'effet d'un agent conversationnel amélioré par RAG sur l'apprentissage et l'expérience des apprenants dans un MOOC sur l'intelligence artificielle. Nous présentons en détail la configuration du prompt utilisé, le contexte de l'étude, les méthodes de collecte de données quantitatives et qualitatives, les considérations éthiques, les participants, les procédures expérimentales ainsi que les instruments de recherche utilisés pour mesurer l'acquisition des connaissances et l'utilisabilité de l'agent.

```
messages=[{
  "rôle": "system",
  "conten": "tu es un assistant pour répondre à des questions sur la matière d'un cours en ligne
ouvert et massif sur l'intelligence artificielle. Agis comme un outil d'aide à l'apprentissage.
Réponds en français à la question en te basant d'abord sur le contexte ci-dessous
{historique}\n\n"
}, {
  "rôle": "user",
  "conten": "f'Contexte: {contexte}\n\n---\n\nQuestion: {question}\nRéponse:"
} ],
```

Figure 4.1 Prompt utilisé dans notre agent conversationnel

4.3.1 Contexte de l'étude

Notre étude repose sur un MOOC dédié à l'intelligence artificielle (Psyché, 2020). Le cours est structuré en quatre modules couvrant des thématiques variées : les concepts généraux de l'IA, l'IA symbolique, l'IA connexionniste, et les applications de l'IA dans l'éducation. Cette recherche se concentre principalement sur le premier module, qui aborde les fondements de l'intelligence artificielle.

4.3.2 Configuration du prompt pour l'agent conversationnel

Dans cette configuration de prompt (voir Figure 4.1), trois éléments fondamentaux structurent l'interaction. L'historique conserve l'ensemble des questions et réponses précédentes, offrant au modèle un contexte enrichi pour mieux interpréter la nouvelle question et produire une réponse plus précise et cohérente. Le contexte représente les segments d'information les plus pertinents pour répondre à la requête actuelle, fournissant ainsi au modèle les données nécessaires pour situer la question dans son cadre spécifique. Enfin, la question correspond à l'interrogation formulée par l'apprenant, à laquelle le modèle est chargé de répondre. Cette structure permet au modèle de s'appuyer simultanément sur les interactions passées via l'historique et sur des informations pertinentes via le contexte, pour formuler des réponses précises et pertinentes.

Ce prompt oriente le comportement du modèle GPT-4 pour qu'il devienne un véritable outil pédagogique, conçu pour accompagner les apprenants dans un MOOC sur l'intelligence artificielle. En répondant aux questions des apprenants à partir d'un contexte fourni et en maintenant une continuité dans les échanges grâce à un historique conversationnel, cet agent conversationnel remplit une mission éducative essentielle.

La mission pédagogique de cet agent est renforcée par l'intégration de consignes explicites, qui lui permettent de fournir des réponses adaptées aux besoins des apprenants et alignées avec les objectifs du cours. Ce cadre méthodologique s'inspire des principes établis par Chan et Baskin (1988) sur les compagnons d'apprentissage, en structurant les interactions pédagogiques et en facilitant l'acquisition de concepts complexes. Cette configuration du prompt constitue une base solide pour concevoir un agent conversationnel capable d'améliorer l'expérience d'apprentissage et de soutenir efficacement les apprenants dans divers contextes éducatifs.

4.3.3 Méthodes de collecte de données

Nous avons utilisé une conception de méthodes mixtes explicatives séquentielles Creswell *et al.* (2003), qui intègre à la fois des techniques de collecte de données quantitatives et qualitatives pour fournir une analyse complète des questions de recherche. Les données quantitatives ont été recueillies via des questionnaires, analysées ensuite à l'aide de statistiques descriptives. Cette approche quantitative a été choisie pour offrir un aperçu global des données et faciliter l'analyse des résultats expérimentaux, renforçant ainsi la reproductibilité de l'étude.

En parallèle, des données qualitatives ont été collectées à travers des entretiens semi-structurés avec les apprenants, afin de comprendre leurs expériences concernant l'utilisation des robots conversationnels. L'analyse thématique de ces entretiens a impliqué un processus de codage et de catégorisation des données, offrant des compréhensions détaillées sur le rôle du robot conversationnel dans le contexte éducatif.

4.3.4 Participants à la recherche

L'échantillon de cette recherche se compose de 42 étudiants inscrits en Master et Licence en Informatique dans une université publique au Sénégal. Cependant, 17 d'entre eux n'ont pas pu compléter l'expérience pour des raisons personnelles. Finalement, 25 participants ont pris part à l'étude, répartis de manière aléatoire en deux groupes : un groupe de contrôle (GC) (n=12 ; quatre femmes et huit hommes) et un groupe expérimental (GE) (n=13 ; cinq femmes et huit hommes), avec un âge compris entre 19 et 23 ans.

Les considérations éthiques ont été au cœur de notre démarche. Tous les participants ont été informés du protocole de l'étude et ont donné leur consentement éclairé. L'accès aux données recueillies a été strictement limité aux personnes autorisées. De plus, l'étude a été approuvée par le Comité d'éthique de l'Université TELUQ (numéro d'approbation 10/2023). Le certificat d'éthique est disponible en Annexe A.

Au début de l'étude, les participants des deux groupes, GC et GE, ont complété un pré-test visant à évaluer leur compréhension initiale des concepts d'intelligence artificielle. Les étudiants du groupe expérimental ont ensuite visionné une courte vidéo de démonstration expliquant l'utilisation de l'agent conversationnel. Ensuite, le groupe expérimental a complété le module 1 du MOOC avec l'assistance de l'agent, tandis que le groupe de contrôle a suivi le même module sans cette assistance. Chaque participant a travaillé de manière autonome et à distance, avec une période de trois jours pour terminer le module. Des rappels par courriel ont été envoyés aux participants n'ayant pas finalisé le module après deux jours.

À la fin de l'expérience, tous les participants ont complété le post-test, qui visait à mesurer l'acquisition des connaissances après l'apprentissage. Le post-test a été réalisé immédiatement après la finalisation du module 1, tant pour le groupe expérimental que pour le groupe de contrôle. De plus, les participants du groupe expérimental ont répondu à un questionnaire SUS (System Usability Scale) pour évaluer l'utilisabilité du robot conversationnel.

Par ailleurs, six étudiants (trois hommes et trois femmes) ont accepté de participer volontairement à des entretiens semi-structurés individuels d'une durée de 20 à 30 minutes, menés à distance via Zoom. Ces entretiens visaient à explorer en profondeur leurs perceptions et expériences avec le robot conversationnel utilisé dans le cadre du MOOC.

Le guide d'entretien, basé sur la littérature existante (Sun *et al.*, 2024; Duong et Suppasetserree, 2024) et adapté à notre contexte, comprenait les questions suivantes :

Toutes les sessions ont été enregistrées audio, et transcrites mot à mot. Par la suite, une analyse thématique a été réalisée en suivant les étapes décrites par Braun et Clarke Braun et Clarke (2006), ce qui nous a permis

- 1 Dans quel but avez-vous utilisé l'agent conversationnel ? Veuillez préciser.
- 2 Quel impact l'interaction avec l'agent conversationnel a-t-elle sur votre apprentissage ?
- 3 Avez-vous trouvé que l'agent conversationnel fournissait des réponses précises et contextuelles liées au MOOC ?
- 4 L'agent conversationnel a-t-il amélioré votre compréhension des concepts d'intelligence artificielle abordés dans le MOOC ? Si oui, comment ?
- 5 Qu'avez-vous particulièrement apprécié chez l'agent conversationnel ? Expliquez en quoi il vous a été utile.
- 6 Quels sont les principaux problèmes que vous avez rencontrés avec l'agent conversationnel ? Veuillez détailler et fournir des exemples concrets.
- 7 L'agent conversationnel a-t-il compris vos questions et fourni toutes les informations dont vous aviez besoin ?
- 8 Globalement, étiez-vous satisfait de l'agent conversationnel ?
- 9 Pensez-vous que vous utiliseriez cet agent conversationnel à nouveau ?
- 10 Comment décririez-vous votre expérience en tant qu'utilisateur de l'agent conversationnel ? Expliquez en détail.

d'identifier et d'interpréter les thèmes émergents.

Les étapes méthodologiques suivantes ont été suivies : (1) préparation des transcriptions des entretiens pour l'analyse, y compris la transcription des données et leur révision pour noter les idées initiales ; (2) attribution systématique de codes aux éléments clés de l'ensemble des données ; (3) regroupement des codes en thèmes potentiels, en rassemblant toutes les données pertinentes sous chaque thème ; (4) révision des thèmes pour s'assurer qu'ils représentent fidèlement les données, avec ajustements si nécessaire ; (5) affinement et nomination des thèmes pour clarifier la narration globale qui se dégage de l'analyse ; et (6) compilation des résultats dans un rapport final, incluant des exemples sélectionnés pour illustrer les résultats. Pour assurer la confidentialité, chaque participant a été anonymisé et identifié par un code (par exemple, S1, S2).

La procédure expérimentale est illustrée dans la Figure 4.2, qui présente un schéma des étapes clés de l'étude. Cette figure met en évidence la séquence des activités, incluant les pré-tests, les post-tests, l'évaluation SUS et les entretiens menés avec le groupe expérimental.

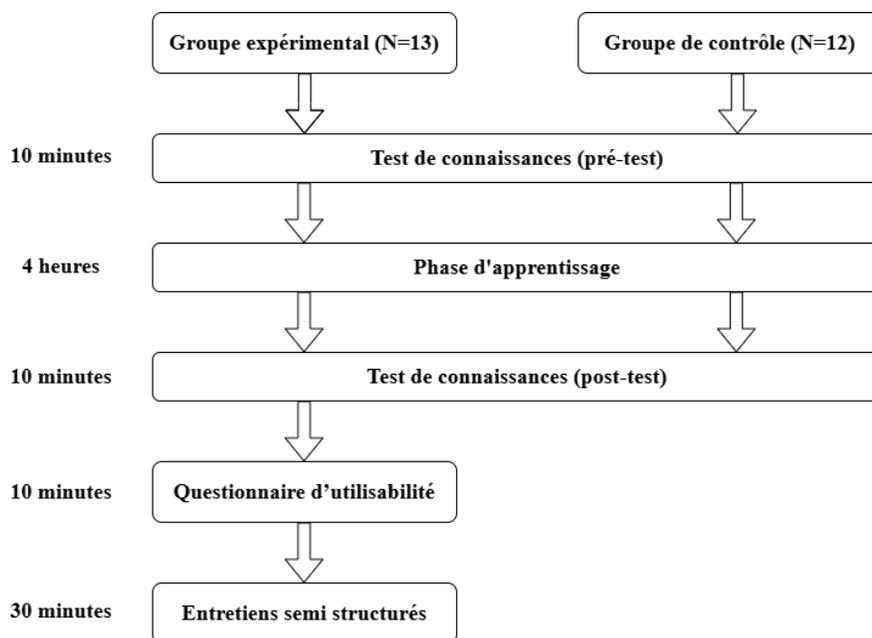


Figure 4.2 Procédure expérimentale

4.3.6 Instruments de recherche

Différents instruments ont été utilisés dans cette étude pour évaluer les connaissances des participants et l'utilisabilité du robot conversationnel. Les sections suivantes décrivent chaque instrument en détail, en commençant par le pré-test et le post-test, qui ont servi à mesurer les acquis d'apprentissage.

4.3.6.1 Pré-test et post-test

Pour évaluer la compréhension des apprenants concernant l'intelligence artificielle dans le cadre de ce MOOC dédié à l'IA, les deux groupes ont complété le même pré-test avant l'expérience. De la même manière, les deux groupes ont complété le même post-test après avoir terminé le module 1, permettant ainsi d'analyser le niveau de connaissances qu'ils avaient acquis. Les questions du pré-test et du post-test ont été soigneusement conçues pour couvrir des concepts identiques, assurant ainsi la cohérence de la mesure des acquis d'apprentissage (Tableau 4.1). Par exemple, la question 2 du pré-test et la question 8 du post-test évaluent toutes deux les connaissances liées au concept d'apprentissage supervisé, tandis que la question

4 du pré-test et la question 6 du post-test évaluent la compréhension du Test de Turing. Les tableaux 4.2 et 4.3 présentent les questions utilisées dans les pré-tests et les post-tests, qui incluait des questions à choix unique ainsi que des questions à réponses courtes.

Concept	Question du pré-test	Question du post-test
Intelligence Artificielle (IA)	Q1	Q9
Apprentissage supervisé	Q2	Q8
Réseau Neuronal Artificiel	Q3	Q7
Test de Turing	Q4	Q6
Système tutoriel intelligent	Q5	Q5
Algorithme d'apprentissage automatique	Q6	Q4
Base de connaissances	Q7	Q3
Traitement du langage naturel (NLP)	Q8	Q1
Raisonnement automatisé	Q9	Q10
Vision par ordinateur	Q10	Q2

Table 4.1 Concepts de connaissances abordés dans les questions de pré-test et post-test

4.3.6.2 Échelle d'utilisabilité du système

Plusieurs instruments existent pour mesurer l'utilisabilité, tels que le QUIS (Chin *et al.*, 1988), l'AttrakDiff (Hassenzahl, 2003) ou encore l'UMUX (Finstad, 2010). Nous avons choisi la System Usability Scale (SUS) de Brooke (1996) en raison de sa simplicité, de sa précision et de sa robustesse empirique, y compris avec de petits échantillons (Tullis et Stetson, 2004). La SUS se compose de 10 affirmations, chacune étant évaluée sur une échelle de Likert à 5 points allant de « Pas du tout d'accord » (1 point) à « Tout à fait d'accord » (5 points), produisant un score d'utilisabilité unique compris entre 0 et 100. Des scores plus élevés indiquent une meilleure utilisabilité. Les affirmations impaires représentent des attitudes positives envers le système, tandis que les affirmations paires mettent en évidence des attitudes négatives. Les réponses au questionnaire SUS ont été obtenues auprès de 13 apprenants du groupe expérimental ayant interagi avec l'agent conversationnel.

<p>Q1 : Qu'est-ce que l'intelligence artificielle (IA) ?</p> <p>a) Une technologie qui permet aux machines de reproduire les fonctions humaines et animales. b) Un type d'ordinateur très puissant. c) Un ensemble de règles strictes pour programmer des logiciels. d) Un langage de programmation spécifique. e) Je ne connais pas la réponse.</p>
<p>Q2 : L'apprentissage supervisé consiste à entraîner un modèle sans fournir des exemples étiquetés.</p> <p>a) Vrai b) Faux c) Je ne connais pas la réponse.</p>
<p>Q3 : Un ... est un ensemble de neurones artificiels interconnectés qui apprennent à partir de données.</p> <p>a) Compléter la phrase b) Je ne connais pas la réponse.</p>
<p>Q4 : Le test de Turing consiste à :</p> <p>a) Déterminer si une machine peut résoudre des problèmes mathématiques. b) Mettre en communication, à l'aveugle, un être humain et un ordinateur pour vérifier s'ils sont capables d'atteindre les mêmes niveaux de performance. c) Évaluer la vitesse de traitement d'un ordinateur. d) Tester la capacité d'un robot à effectuer des tâches physiques. e) Je ne connais pas la réponse.</p>
<p>Q5 : Un système tutoriel intelligent s'adapte à l'apprenant en fonction de ses états cognitifs, métacognitifs ou affectifs.</p> <p>a) Vrai b) Faux c) Je ne connais pas la réponse.</p>
<p>Q6 : Un ... se base sur un ensemble de mégadonnées pour modifier le comportement d'un robot, d'un agent ou d'un réseau de neurones artificiels.</p> <p>a) Compléter la phrase b) Je ne connais pas la réponse.</p>
<p>Q7 : Une ... est une collection organisée de connaissances utilisée par un système expert pour tirer des conclusions ou prendre des décisions.</p> <p>a) Compléter la phrase b) Je ne connais pas la réponse.</p>
<p>Q8 : Quel est l'objectif principal du traitement du langage naturel ?</p> <p>a) Traduire des textes d'une langue à une autre. b) Analyser et comprendre le langage humain. c) Créer des langages de programmation. d) Je ne connais pas la réponse.</p>
<p>Q9 : Le raisonnement automatique est la capacité d'un système informatique à tirer des conclusions à partir de faits et de règles.</p> <p>a) Vrai b) Faux c) Je ne connais pas la réponse.</p>
<p>Q10 : La vision par ordinateur permet aux machines de comprendre et d'interpréter des images et des vidéos.</p> <p>a) Vrai b) Faux c) Je ne connais pas la réponse.</p>

Table 4.2 Questions de pré-test

44 Résultats

Cette section vise à présenter les résultats quantitatifs et qualitatifs de notre étude. Nous explorons d'abord l'effet du robot conversationnel sur l'acquisition des connaissances et l'utilisabilité à travers l'analyse quantitative, puis nous examinons l'expérience globale des apprenants avec le robot conversationnel via l'analyse qualitative.

<p>Q1 : Le traitement du langage naturel est une technique d'IA qui permet aux machines de comprendre le langage humain.</p> <p>a) Vrai b) Faux</p>
<p>Q2 : La ... permet l'analyse, le traitement et la compréhension d'images prises à partir de caméras.</p> <p>a) Compléter la phrase</p>
<p>Q3 : Une base de connaissances est une collection de données non structurées et aléatoires utilisée par un système expert pour tirer des conclusions ou prendre des décisions.</p> <p>a) Vrai b) Faux</p>
<p>Q4 : Un algorithme d'apprentissage automatique est :</p> <p>a) Un algorithme qui suit un ensemble de règles strictes pour programmer des logiciels. b) Un algorithme qui se base sur un ensemble de mégadonnées pour modifier le comportement d'un robot, d'un agent ou d'un réseau de neurones artificiels. c) Une technique pour automatiser les tâches répétitives sur un ordinateur.</p>
<p>Q5 : Quel est l'objectif principal des systèmes tutoriels intelligents ?</p> <p>a) Réduire le besoin de tuteurs humains. b) Augmenter la vitesse d'apprentissage. c) S'adapter à l'apprenant en fonction de ses états cognitifs, métacognitifs ou affectifs.</p>
<p>Q6 : Le test de Turing consiste à évaluer la capacité d'un ordinateur à résoudre des problèmes mathématiques complexes.</p> <p>a) Vrai b) Faux</p>
<p>Q7 : Un réseau de neurones artificiels est un ensemble de neurones artificiels interconnectés qui apprennent à partir de données.</p> <p>a) Vrai b) Faux</p>
<p>Q8 : Quel type d'apprentissage nécessite des données étiquetées pour entraîner le modèle?</p> <p>a) Apprentissage supervisé b) Apprentissage non supervisé c) Apprentissage par renforcement</p>
<p>Q9 : L'... est une technologie permettant aux machines de reproduire des fonctions humaines et animales.</p> <p>a) Compléter la phrase.</p>
<p>Q10 : Le ... est un procédé par lequel un système informatique effectue un enchaînement logique, à partir de propositions de départ et d'une base de connaissances, afin d'arriver à une conclusion.</p> <p>a) Compléter la phrase.</p>

Table 4.3 Questions de post-test

4.4.1 Approche quantitative de la recherche

Cette section décrit l'analyse quantitative de l'effet du robot conversationnel sur l'acquisition des connaissances et son utilisabilité. L'acquisition des connaissances a été mesurée à travers des pré-tests et post-tests, tandis que l'utilisabilité du robot conversationnel a été évaluée à l'aide de l'échelle SUS (System Usability Scale).

4.4.1.1 Résultats du gain de connaissances

Pour évaluer l'acquisition des connaissances dans les groupes contrôle et expérimental, des pré-tests et post-tests ont été utilisés. Les résultats de l'évaluation des connaissances, illustrés dans la Figure 4.3, indiquent les pourcentages de gain de connaissances des deux groupes (expérimental et contrôle) à travers les deux types de tests. Les scores moyens initiaux de 72 % des deux groupes au pré-test montrent des niveaux de connaissances comparables avant l'activité d'apprentissage.

Après l'activité d'apprentissage, le groupe expérimental, ayant utilisé le robot conversationnel, a atteint un gain de connaissances de 17 %, tandis que le groupe contrôle, sans robot conversationnel, a obtenu un gain de 10 %. Ces résultats suggèrent que l'utilisation du robot conversationnel a significativement amélioré l'acquisition des connaissances.

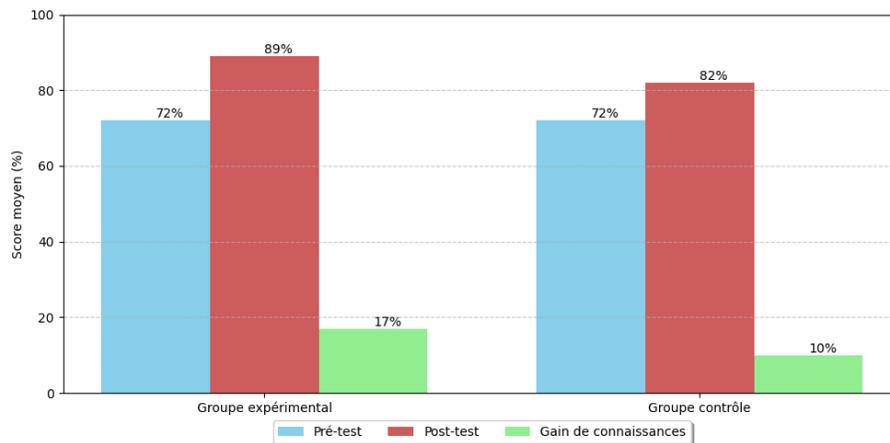


Figure 4.3 Scores moyens des pré-tests et post-tests pour les groupes expérimental et contrôle pour le module 1 du MOOC

L'analyse statistique appuie ces résultats. Bien que les deux groupes aient amélioré leurs scores au post-test, le groupe expérimental a enregistré une augmentation plus marquée. Les analyses statistiques des scores au pré-test confirment que les groupes contrôle et expérimental suivent une distribution normale (test de Shapiro-Wilk, $p > 0.05$) et présentent des variances homogènes (test de Levene, $p > 0.05$). Ces conditions

permettent l'application d'un test t de Student, adapté à la comparaison des moyennes de deux groupes indépendants lorsque les distributions sont normales et les variances équivalentes.

Le test t n'a révélé aucune différence significative entre les scores du pré-test des deux groupes ($p = 0.99 > 0.05$), indiquant que les deux groupes avaient des niveaux de connaissances similaires avant l'expérience (Tableau 4.4). Cependant, une différence significative a été observée entre les scores du post-test ($p = 0.017 < 0.05$), suggérant que l'utilisation du robot conversationnel a eu un effet positif sur l'acquisition des connaissances (Tableau 4.5).

Groupes	N	Moyenne	Écarts-types	Médiane	Minimum	Maximum	Valeur p
Groupe contrôle	12	7.2	2.19	7.5	4	10	0.99
Groupe expérimental	13	7.2	1.9	8	4	10	

Table 4.4 Analyse des acquisitions des connaissances au pré-test

Groupes	N	Moyenne	Écarts-types	Médiane	Minimum	Maximum	Valeur p
Groupe contrôle	12	8.2	1.14	8	7	10	0.017
Groupe expérimental	13	8.9	1.03	9	6	10	

Table 4.5 Analyse des acquisitions des connaissances au post-test

4.4.1.2 Résultats du SUS

Sur les 13 réponses obtenues via le questionnaire SUS, le Tableau 4.6 présente les statistiques descriptives pour chaque affirmation du questionnaire, incluant la moyenne, la médiane et les écarts-types des réponses.

Question	Affirmation	Moyenne	Médiane	Écarts-types
1	Je pense que j'aimerais utiliser cet agent conversationnel.	4.46	4	0.5
2	J'ai trouvé cet agent conversationnel inutilement complexe.	1.85	2	0.86
3	J'ai trouvé que cet agent conversationnel était facile à utiliser.	4.54	5	0.63
4	Je pense que j'aurais besoin d'une assistance technique pour utiliser cet agent conversationnel.	1.15	1	0.36
5	J'ai trouvé que les différentes fonctions de cet agent conversationnel étaient bien intégrées.	3.85	4	0.77
6	J'ai trouvé qu'il y avait trop d'incohérences dans cet agent conversationnel.	1.54	1	0.84
7	Je pense que la plupart des gens apprendraient à utiliser cet agent conversationnel très rapidement.	4.38	4	0.62
8	J'ai trouvé cet agent conversationnel très encombrant à utiliser.	2.15	2	1.1
9	Je me suis senti très confiant en utilisant cet agent conversationnel.	4.31	4	0.72
10	J'ai dû apprendre beaucoup de choses avant de pouvoir utiliser cet agent conversationnel.	2.69	3	1.43

Table 4.6 Questionnaire SUS et statistiques pour chaque affirmation

Comme indiqué par Brooke (1996), le score SUS est obtenu en ajustant les scores des questions impaires et

paires. Pour les questions impaires (1, 3, 5, 7 et 9), on soustrait 1 à chaque score, et les valeurs résultantes sont additionnées pour obtenir la variable X. De même, pour les questions paires (2, 4, 6, 8 et 10), on soustrait chaque score de 5, et ces valeurs ajustées sont additionnées pour obtenir la variable Y. Le score final du SUS est ensuite calculé en additionnant X et Y, puis en multipliant la somme par 2,5, ce qui donne un score compris entre 0 et 100.

Dans le cas de notre robot conversationnel, le score final du SUS s'élève à 80,4, ce qui reflète un haut niveau d'utilisabilité. Les scores obtenus par les participants allaient de 52,5 à 95 sur 100. La moitié des utilisateurs ont obtenu des scores compris entre 75 et 85, avec une médiane de 82,5, comme le montre le boxplot de la Figure 4.4. Le score moyen était de 80,4, avec un écart-type de 11,7.

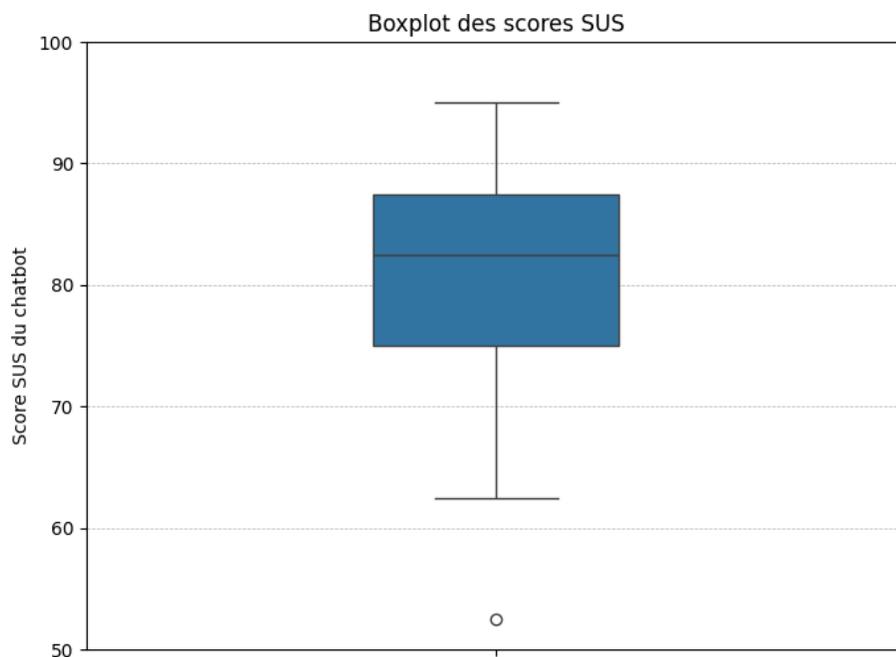


Figure 4.4 Boxplot des scores SUS pour le robot conversationnel (Médiane 82.5, Moyenne 80.4)

4.4.2 Approche qualitative de la recherche

Les données qualitatives collectées à partir des entretiens individuels semi-structurés avec six apprenants ont été catégorisées en trois thèmes principaux : l'utilité perçue, les avantages et inconvénients perçus, et

la satisfaction globale vis-à-vis du robot conversationnel. Ces thèmes ont été identifiés grâce à une analyse thématique, fournissant une compréhension approfondie des expériences et perceptions des apprenants. Cette approche qualitative a exploré les perceptions des étudiants concernant leur interaction avec le robot conversationnel.

4.4.2.1 Utilité perçue

L'agent conversationnel a été généralement perçu comme un outil utile pour favoriser l'apprentissage et la compréhension tout au long du MOOC. Plusieurs étudiants ont sollicité l'agent pour clarifier des termes et concepts complexes. Par exemple, un étudiant a déclaré : « Lorsque je rencontre des termes que je ne comprends pas, je demande au robot conversationnel de les expliquer ou de les clarifier avec des exemples concrets » (S1). Cet exemple montre l'utilité de l'agent pour apporter des explications immédiates. Un autre étudiant a mentionné : « Par exemple, le terme "langage naturel" n'était pas très clair pour moi. J'ai donc posé une question à l'agent pour voir comment il me l'expliquerait » (S4), soulignant le rôle de l'agent dans l'éclaircissement de concepts complexes.

Par ailleurs, l'agent a également permis d'optimiser le processus d'apprentissage. Comme l'a indiqué un étudiant : « Je l'ai utilisé pour mieux comprendre, mais aussi pour accélérer mon rythme d'apprentissage, car il me permet de me sentir beaucoup plus à l'aise dans la compréhension » (S5). Ces retours montrent l'utilité de l'agent, tant pour faciliter la compréhension que pour favoriser un apprentissage plus fluide et rapide.

4.4.2.2 Avantages et inconvénients perçus

Bien que l'agent conversationnel ait été reconnu pour son utilité, les étudiants ont également soulevé divers avantages et inconvénients au cours de leurs interactions. En ce qui concerne les avantages, l'agent

s'est révélé très efficace pour soutenir l'apprentissage. Un étudiant a rapporté : « Il m'a aidé dans mon apprentissage » (S1). Un autre étudiant a souligné la rapidité et la clarté des réponses, déclarant : « C'est plus rapide à comprendre que la vidéo. Je trouve que c'est plus rapide. Au lieu de regarder, par exemple, une vidéo d'une ou deux minutes, je pose juste ma question et je comprends rapidement » (S3). Cela met en évidence l'efficacité de l'agent dans la transmission de l'information.

Cependant, certains étudiants ont également rencontré des problèmes d'exactitude et de pertinence des réponses de l'agent. Par exemple, un étudiant a partagé : « Il y avait une question qui n'était pas très claire, et lorsque j'ai demandé à l'agent de m'expliquer le concept, c'était un peu confus. Par exemple, j'ai demandé de m'expliquer le concept de langage naturel, et lorsque l'agent a fourni une réponse, je lui ai redemandé de clarifier parce que ce n'était pas très clair. Mais la réponse qu'il m'a donnée était insatisfaisante et complètement hors sujet, se concentrant davantage sur des concepts symboliques et connexionnistes » (S2). Un autre étudiant a noté : « Parfois, les réponses ne sont ni exactes ni précises » (S6). Bien que ces cas soient minoritaires, ils suggèrent qu'il reste encore des améliorations à envisager pour assurer des réponses précises et adaptées au contexte. Cet objectif est d'autant plus crucial compte tenu de l'utilisation du RAG pour accroître la précision des réponses du robot conversationnel.

4.4.23 Satisfaction

Dans l'ensemble, les étudiants ont manifesté un haut niveau de satisfaction vis-à-vis de l'agent conversationnel. La majorité a apprécié sa clarté, son efficacité, ainsi que l'aide apportée lorsqu'ils recherchaient des réponses. Un étudiant a exprimé son expérience positive en déclarant : « Je suis très satisfait » (S1).

L'agent a également favorisé un environnement sans jugement, ce qui a permis aux étudiants de se sentir plus à l'aise pour poser des questions. Comme l'a souligné un étudiant : « Je me sens en confiance parce que je sais qu'il ne va pas me juger, qu'il ne va pas me dire que je ne sais rien, etc. » (S3). Cela suggère que

l'agent a joué un rôle dans la création d'un environnement d'apprentissage favorable, contribuant ainsi à la satisfaction générale.

En résumé, ces témoignages illustrent la satisfaction globale des étudiants en ce qui concerne l'agent conversationnel, notamment en termes de clarté, soutien, et efficacité dans le processus d'apprentissage. Les informations tirées de ces entretiens offrent des perspectives précieuses pour le développement et l'amélioration des outils éducatifs intégrant l'intelligence artificielle.

4.5 Discussion

D'une manière globale, les résultats issus de cette expérimentation vont dans le sens de nos hypothèses : ils suggèrent que l'utilisation du robot conversationnel enrichi par GPT-4 et RAG a eu un effet positif sur l'acquisition des connaissances des apprenants ainsi que sur leur perception de l'agent conversationnel. Nous détaillons ici les effets observés à travers les résultats de cette expérimentation.

4.5.1 Effet du robot conversationnel sur l'apprentissage des apprenants (H1)

Les résultats de cette étude suggèrent que l'utilisation du robot conversationnel basé sur GPT-4, enrichi par la RAG, a conduit à une différence statistiquement significative sur l'acquisition des connaissances des apprenants. Le groupe expérimental a montré une amélioration de 17% en termes de gain de connaissances, comparé au groupe de contrôle qui n'a réalisé qu'une amélioration de 10%. Cette différence peut être attribuée à la capacité du robot conversationnel à fournir un soutien personnalisé et contextuellement pertinent en temps réel. En récupérant des informations pertinentes provenant de sources externes, la RAG a renforcé les capacités génératives du robot conversationnel (Lewis *et al.*, 2020), permettant de délivrer des réponses à la fois précises et adaptées aux besoins individuels des apprenants. Ces résultats s'alignent avec les conclusions de Ko *et al.* (2024), qui ont démontré que les modèles LLM enrichis par RAG offrent des réponses plus

fiables et précises, ce qui améliore les résultats d'apprentissage, notamment dans des contextes complexes comme la programmation Python. De manière similaire, notre étude a montré que le robot conversationnel fournissait des informations contextuellement pertinentes et précises, facilitant une meilleure compréhension des contenus. Ainsi, le soutien apporté par le robot conversationnel a permis aux apprenants de clarifier des concepts difficiles, contribuant à une amélioration notable du gain de connaissances. Notre hypothèse selon laquelle le robot conversationnel améliorerait l'acquisition des connaissances des apprenants est donc validée.

4.5.2 Perception de l'utilisabilité du robot conversationnel (H2)

Le score SUS obtenu pour notre agent conversationnel est de 80,4. Selon Bangor *et al.* (2009), ce score correspond à une note de « B » sur l'échelle de notation du SUS. En termes de plages d'acceptabilité, notre robot conversationnel est considéré comme « Acceptable » et, dans les évaluations adjectivales, il est classé comme « Bon » (voir Figure 4.5). Cela démontre que le robot conversationnel est bien accueilli par les apprenants et qu'il possède un fort potentiel pour améliorer l'expérience utilisateur dans les environnements éducatifs. Ces résultats valident notre hypothèse concernant l'utilisabilité positive du robot conversationnel.

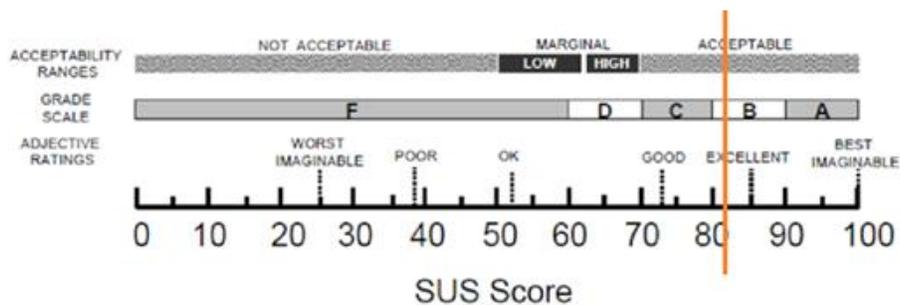


Figure 4.5 Échelle SUS de Bangor (Bangor *et al.*, 2009) et Score SUS pour l'Agent Conversationnel (Valeur Moyenne).

4.5.3 Expérience d'interaction avec le robot conversationnel (H3)

L'utilité perçue élevée du robot conversationnel et la satisfaction des étudiants peuvent s'expliquer par sa capacité à offrir un soutien immédiat et contextuellement pertinent, répondant en temps réel aux besoins des apprenants. Ce soutien a permis aux étudiants de clarifier rapidement des concepts complexes, rendant le processus d'apprentissage plus efficace. De plus, la nature non jugeante du robot conversationnel a favorisé un environnement où les apprenants se sentent à l'aise pour poser des questions sans crainte de jugement, ce qui a contribué à leur satisfaction globale. Notre hypothèse concernant l'expérience d'interaction positive avec le robot conversationnel est ainsi validée. Ces résultats s'alignent avec les conclusions de Li et Xing (2021), qui ont démontré que les apprenants perçoivent positivement les robots conversationnels basés sur les LLMs dans les forums de discussion des MOOCs. Dans leur étude, le robot conversationnel a été évalué comme utile, clair et pertinent, contribuant à la satisfaction globale des utilisateurs. De même, dans notre étude, la capacité du robot conversationnel à offrir un soutien immédiat et adapté a favorisé une expérience d'apprentissage positive, renforçant la satisfaction et l'utilité perçue chez les apprenants. En plus de son efficacité, le robot conversationnel a joué un rôle d'accompagnement éducatif, en facilitant la clarification des concepts complexes et en soutenant l'autonomie des apprenants. Ces résultats suggèrent que des interactions adaptées et contextualisées peuvent favoriser l'apprentissage (Ruan *et al.*, 2019; González-Castro *et al.*, 2021).

4.6 Limites de l'étude expérimentale

Malgré les résultats positifs obtenus, cette étude présente plusieurs limitations. Les résultats sont spécifiques à une population d'apprenants issue d'une seule institution, ce qui limite leur généralisation à des contextes éducatifs plus diversifiés. De plus, la taille de l'échantillon était relativement réduite, ce qui pourrait ne pas refléter toutes les expériences d'apprentissage. Une autre limitation réside dans l'utilisation de données auto-déclarées issues des questionnaires et des entretiens, qui peuvent être influencées par des

biais de réponse.

4.7 Conclusion

À travers cette étude, nous avons évalué l'effet de l'utilisation d'un agent conversationnel enrichi par GPT-4 et RAG sur l'acquisition des connaissances et la perception de l'agent conversationnel dans le cadre d'un MOOC. Nous avons mis en évidence l'importance d'intégrer des technologies de génération et de récupération d'informations afin d'offrir un soutien personnalisé et pertinent aux apprenants.

Pour mesurer cet effet, nous avons adopté une approche mixte (quantitative et qualitative) en mettant en place une procédure expérimentale impliquant deux groupes : un groupe utilisant l'agent conversationnel enrichi et un groupe de contrôle sans agent. Chaque groupe a suivi le premier module du MOOC de manière autonome, avec des pré-tests et post-tests pour évaluer l'acquisition des connaissances. Dans le groupe expérimental, des questionnaires SUS ont été utilisés pour mesurer l'utilisabilité de l'agent. En complément de cette approche quantitative, nous avons mené des entretiens individuels semi-structurés avec six participants du groupe expérimental pour recueillir des données qualitatives approfondies sur leur expérience avec l'agent conversationnel.

Les résultats vont globalement dans le sens de nos hypothèses concernant l'acquisition des connaissances, l'utilisabilité et l'expérience d'interaction avec l'agent conversationnel. Le groupe expérimental a montré une amélioration significative de la l'acquisition des connaissances, avec une différence statistiquement significative par rapport au groupe de contrôle. En outre, les scores élevés obtenus au SUS révèlent que l'agent conversationnel a été perçu de manière très favorable en termes d'utilisabilité. Les données qualitatives recueillies à travers les entretiens ont permis d'enrichir notre compréhension des interactions entre les apprenants et l'agent. Les participants ont particulièrement apprécié la capacité de l'agent à fournir des réponses précises et adaptées, contribuant ainsi à un soutien immédiat et pertinent, facilitant la compré-

hension des concepts complexes abordés dans le MOOC.

CHAPITRE 5

ANALYSE DES EFFETS D'UN AGENT CONVERSATIONNEL MOTIVANT SUR LES CONNAISSANCES ET LES ÉMOTIONS DANS UN MOOC

Ce chapitre prolonge les résultats présentés au chapitre 4, où nous avons montré que l'agent conversationnel enrichi par la RAG améliore l'acquisition des connaissances et contribue positivement à l'expérience d'apprentissage. Il répond ici au troisième objectif de la thèse, qui consiste à évaluer l'effet de l'ajout des éléments motivationnels inspirés de la théorie des états d'esprit de Dweck (Dweck, 2002), à l'agent conversationnel enrichi par GPT-4 et la RAG sur deux dimensions essentielles : (1) l'acquisition des connaissances et (2) les émotions positives des apprenants telles que la confiance, l'enthousiasme et l'intérêt (RQ3).

Cette théorie explique que les élèves qui voient leur intelligence comme fixe ont tendance à éviter les défis, tandis que ceux qui croient qu'elle peut se développer grâce à l'effort recherchent activement des occasions d'apprentissage. En valorisant l'effort des apprenants, un agent conversationnel motivant pourrait non seulement les aider à adopter une vision positive de leurs capacités, mais aussi à renforcer des émotions telles que la confiance, l'enthousiasme et l'intérêt, ce qui est crucial pour leur réussite dans des environnements en ligne.

Cependant, bien que cette approche ait été largement adoptée dans les contextes éducatifs, elle n'est pas exempte de critiques. Par exemple, Macnamara et Burgoyne (2023) soulignent que les interventions inspirées par cette théorie ne produisent pas toujours les effets escomptés. Ils mettent en avant que les gains observés sont souvent modestes et difficiles à généraliser, en particulier dans des programmes éducatifs de courte durée ou auprès d'apprenants déjà motivés. De plus, certains résultats manquent de robustesse en raison de failles méthodologiques, ce qui peut exagérer l'efficacité perçue de ces approches. Ces limites

soulignent la nécessité d'évaluer avec précaution l'effet des éléments motivationnels intégrés dans notre agent conversationnel.

Le chapitre est structuré comme suit : nous exposons d'abord les hypothèses formulées. Ensuite, nous détaillons la méthodologie, incluant les participants, le protocole expérimental et les instruments de mesure. Enfin, nous présentons les résultats des analyses et concluons le chapitre.

Ce chapitre fait l'objet d'une communication scientifique en préparation pour publication intitulée «Analyse des effets d'un robot conversationnel motivant sur les connaissances et émotions dans un MOOC», présentant une analyse détaillée des résultats de cette étude.

5.1 Hypothèses de recherche

Les émotions positives, telles que la confiance, l'enthousiasme et l'intérêt, jouent un rôle essentiel dans la réussite éducative, car elles influencent directement l'engagement et les performances des apprenants (Pekrun *et al.*, 2007). Dans les environnements en ligne tels que les MOOCs, où les apprenants travaillent de manière autonome, ces émotions deviennent particulièrement importantes pour maintenir leur motivation et leur persévérance face aux difficultés qu'ils peuvent rencontrer dans leur apprentissage.

Les agents conversationnels, en valorisant les efforts des apprenants et en adoptant une approche motivationnelle inspirée des travaux de Dweck, peuvent simultanément renforcer l'apprentissage des concepts tout en stimulant des émotions positives.

Dans ce contexte, cette étude explore l'effet d'un agent conversationnel motivant sur l'acquisition des connaissances et les émotions des apprenants dans un MOOC. Les hypothèses formulées s'inscrivent en réponse à la question de recherche *RQ3* suivante : Quels sont les effets d'un agent conversationnel mo-

tivationnel enrichi par GPT et la RAG sur l'acquisition des connaissances et les émotions des apprenants (confiance, intérêt, enthousiasme), comparativement à un agent neutre, dans un contexte de MOOC?

Les hypothèses sont les suivantes :

- *H1* : L'acquisition des connaissances des apprenants utilisant un robot conversationnel motivationnel sera significativement plus élevée que celle des apprenants utilisant un robot conversationnel neutre.
- *H2a* : La confiance des apprenants utilisant un robot conversationnel motivationnel sera significativement plus élevée que celle des apprenants utilisant un robot conversationnel neutre.
- *H2b* : L'enthousiasme des apprenants utilisant un robot conversationnel motivationnel sera significativement plus élevé que celui des apprenants utilisant un robot conversationnel neutre.
- *H2c* : L'intérêt des apprenants utilisant un robot conversationnel motivationnel sera significativement plus élevé que celui des apprenants utilisant un robot conversationnel neutre.

5.2 Méthodologie de recherche

Cette section décrit la méthodologie de l'étude visant à évaluer l'effet d'un agent conversationnel motivant dans le cadre d'un MOOC sur l'intelligence artificielle. L'objectif était d'examiner l'effet de cet agent sur l'acquisition des connaissances et les émotions des apprenants. Nous présentons les participants, le protocole expérimental et les instruments utilisés pour mesurer ces deux dimensions.

5.2.1 Participants

Cette expérimentation a été réalisée auprès de 35 étudiants inscrits dans des universités publiques situées au Sénégal, en France et au Québec. L'échantillon comprenait des participants issus de parcours éducatifs variés : 50 % étaient en informatique, 20 % en communication, 15 % en administration, et 15 % dans d'autres

domaines tels que la psychologie, le droit, l'enseignement et l'éducation.

Les participants ont été répartis aléatoirement en deux groupes : un groupe de contrôle (GC) (n = 13 ; sept femmes et six hommes) et un groupe expérimental (GE) (n = 22 ; 12 femmes et 10 hommes). L'âge des participants variait entre 21 et 53 ans. Tous les participants ont volontairement pris part à cette étude après avoir donné leur consentement éclairé.

L'étude a été approuvée par le Comité d'éthique de l'Université TELUQ (approbation no 10/2023), et le certificat correspondant est disponible en Annexe A.

5.2.2 Protocole expérimental

Au début de l'étude, les participants des deux groupes, GC et GE, ont complété un pré-test de connaissances pour évaluer leur compréhension initiale des concepts d'intelligence artificielle. Ils ont également répondu à un pré-questionnaire mesurant leurs émotions, ainsi qu'à une question ouverte visant à recueillir leur perception de l'intelligence.

Par la suite, le groupe expérimental a suivi le premier module du MOOC avec l'accompagnement d'un robot conversationnel motivationnel, tandis que le groupe de contrôle a utilisé un robot conversationnel neutre. Cette session d'apprentissage, d'une durée approximative de quatre heures, s'est déroulée de manière autonome et à distance. Les participants ont été amenés à lire des textes, visionner des vidéos et compléter des exercices formatifs proposés dans ce module.

À la fin de l'expérience, les participants ont complété un post-test de connaissances, répondu au même post-questionnaire sur les émotions et répondu une seconde fois à la même question ouverte sur leur perception de l'intelligence. La Figure 5.1 illustre les principales étapes de cette procédure expérimentale.

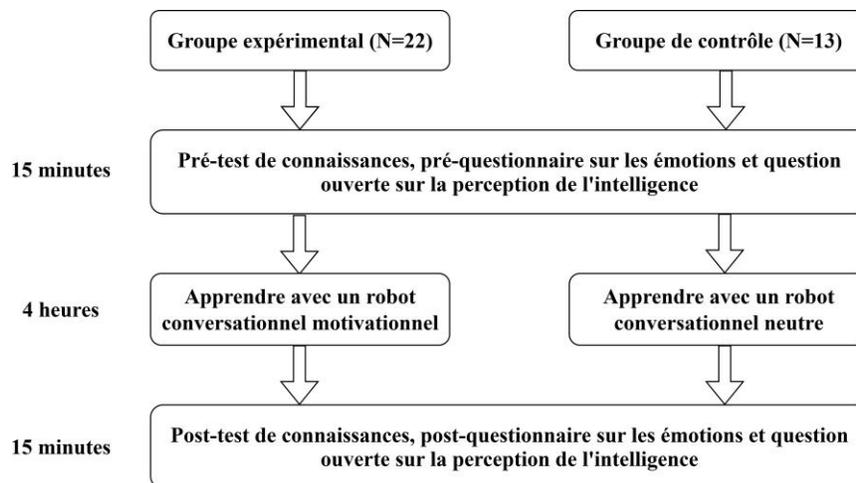


Figure 5.1 Procédure expérimentale

Pour soutenir la motivation des apprenants dans le MOOC sur l’intelligence artificielle, nous avons conçu un robot conversationnel motivant reposant sur un prompt spécifique, directement inspiré de la théorie des états d’esprit de Dweck (Dweck, 2002). Cette théorie postule que l’intelligence n’est pas une capacité fixe, mais peut se développer grâce à l’effort, à la persévérance et à l’adoption de stratégies d’apprentissage efficaces. Elle souligne également que la perception des difficultés comme une étape normale du processus d’apprentissage favorise l’engagement et la réussite.

Le prompt a été conçu pour traduire ces principes théoriques dans le comportement de l’agent conversationnel. Il met systématiquement l’accent sur la valorisation des efforts, la normalisation des défis et des erreurs, ainsi que sur le renforcement de l’idée que les compétences se développent avec la pratique et la persévérance. L’objectif est de fournir des réponses à la fois précises, bienveillantes et motivantes, afin de soutenir à la fois la compréhension des concepts et la confiance des apprenants dans leurs capacités.

Ce prompt définit les principes directeurs qui guident le comportement du robot conversationnel. Il est structuré en trois éléments principaux : le contexte de la situation, la mission de l’agent conversationnel et des instructions précises (Levy et Albertos, 2024). Ces composantes visent à garantir des interactions

motivantes et adaptées aux besoins des apprenants. Les détails de ce prompt sont présentés ci-dessous.

Contexte de la situation : L'apprenant suit un MOOC sur l'intelligence artificielle. L'objectif est de l'aider à comprendre les concepts tout en renforçant sa motivation et sa confiance. La réponse doit être claire, reconnaître les efforts de l'apprenant et l'encourager à persévérer face aux défis. L'accent doit être mis sur l'idée que l'intelligence et les compétences se développent avec l'effort et la pratique.

Mission : En tant que robot conversationnel axé sur le soutien motivationnel, ta mission est de répondre aux questions de l'apprenant en utilisant le CONTEXTE fourni. Réponds de manière claire et bienveillante, tout en valorisant l'effort de l'apprenant et en encourageant une attitude de croissance face à l'apprentissage.

Instructions :

1. Reformule la question de l'apprenant pour montrer que tu as compris à la fois la question et ses efforts.
2. Fournis une réponse claire, en reconnaissant que les défis sont une partie naturelle du processus d'apprentissage.
3. Adopte un ton motivant, même si la réponse n'est pas disponible dans le CONTEXTE.
4. À la fin de chaque réponse, félicite l'apprenant pour son effort, souligne qu'il progresse grâce à sa persévérance, et encourage-le à continuer à poser des questions et à explorer.

Exemple de réponse : Je comprends que ce concept peut sembler complexe au début, mais c'est tout à fait normal. Chaque question que tu poses montre ton engagement à apprendre et à progresser. Voyons cela ensemble : [insérer une brève explication ou un exemple]. N'oublie pas, chaque effort que tu fais te rapproche de la maîtrise de ce sujet. Continue comme ça, chaque étape te rend plus compétent!

Ce prompt est conçu pour orienter le comportement d'un agent conversationnel basé sur GPT-4, afin qu'il devienne un outil d'accompagnement pédagogique et motivationnel. Il s'appuie sur des consignes expli-

cites, intégrant des stratégies pédagogiques pour guider l'agent dans ses interactions avec les apprenants. Il favorise la clarification des concepts et la reformulation des questions pour confirmer la compréhension, tout en encourageant explicitement les apprenants à persévérer face aux difficultés. De plus, en adoptant un ton bienveillant et motivant, il stimule des émotions positives telles que la confiance, l'intérêt et l'enthousiasme, ce qui valorise les efforts accomplis. Cette approche s'aligne parfaitement avec les travaux de Chan et Baskin (1988) sur les compagnons d'apprentissage, qui mettent en avant l'importance d'un soutien continu et interactif pour l'apprenant. En combinant ces deux dimensions — pédagogique et motivationnelle —, ce prompt contribue à créer une expérience d'apprentissage interactive, engageante et adaptée, essentielle pour accompagner les apprenants dans des contextes d'apprentissage à distance et compenser l'absence d'encadrement humain en temps réel.

Cet exemple d'interaction met en lumière la manière dont le robot conversationnel accompagne l'apprenant en valorisant ses efforts et en le motivant à progresser dans son apprentissage.

Question : Qu'est-ce que le test de Turing ?

Réponse de l'agent motivationnel : Il semble que tu cherches à comprendre l'idée derrière le fameux test de Turing, une étape importante dans ta progression vers la maîtrise de l'intelligence artificielle. J'apprécie ton engagement et ta curiosité ! Le test de Turing est une procédure conçue par le mathématicien britannique Alan Turing pour déterminer si une machine peut démontrer une intelligence équivalente à celle d'un être humain... N'oublie pas que chaque question que tu poses renforce ta compréhension de l'intelligence artificielle et t'aide à développer tes compétences dans ce domaine. Continue à poser des questions et à explorer. Tu fais du très bon travail en t'efforçant de comprendre ces concepts complexes. Continue comme ça, je suis là pour t'aider dans ton parcours d'apprentissage !

Au début de l'étude, les participants des deux groupes ont complété un pré-test sur leurs connaissances en intelligence artificielle. Ce test comportait des questions à choix unique et des réponses courtes, visant à évaluer leur compréhension initiale des concepts. Les tableaux 4.2 et 4.3 présentent les questions utilisées dans le pré-test et le post-test.

En complément, les participants ont répondu à un questionnaire mesurant leurs émotions selon trois dimensions spécifiques :

- *Confiance* : « À quel point vous sentez-vous confiant(e) lorsque vous apprenez des concepts ou résolvez des exercices liés à l'intelligence artificielle ? »
- *Enthousiasme* : « À quel point vous sentez-vous enthousiaste lorsque vous apprenez des concepts ou résolvez des exercices liés à l'intelligence artificielle ? »
- *Intérêt* : « À quel point vous sentez-vous intéressé(e) lorsque vous apprenez des concepts ou résolvez des exercices liés à l'intelligence artificielle ? »

Les réponses ont été recueillies sur une échelle de 1 à 5, où « 1 » représente un niveau très faible, « 3 » correspond à une position neutre, et « 5 » indique un niveau très élevé.

Enfin, une question ouverte a invité les participants à réfléchir sur leur perception de l'intelligence : « Comment percevez-vous votre propre intelligence lorsque vous apprenez ? Pensez-vous qu'elle se développe avec l'effort et la pratique, ou qu'elle est fixe ? Pourquoi ? »

Cette section présente les résultats de l'étude, organisés en trois axes : l'acquisition de connaissances, les émotions des apprenants et leurs perceptions de l'intelligence. Les analyses statistiques comparent les groupes contrôle et expérimental à partir des pré-tests, post-tests, et questionnaires, afin d'évaluer l'effet du robot conversationnel motivant sur ces dimensions.

5.3.1 Résultats de l'acquisition de connaissances

Les résultats des analyses statistiques pour le pré-test des connaissances montrent que les scores des deux groupes, contrôle et expérimental, suivent une distribution normale (Shapiro-Wilk, $p > 0.05$) et que leurs variances sont homogènes (Levene, $p > 0.05$). Ces conditions remplissent les hypothèses nécessaires pour l'utilisation d'un test t de Student, qui est approprié pour comparer les moyennes de deux groupes indépendants lorsque les distributions sont normales et que les variances sont homogènes. Le test t réalisé indique qu'il n'y a pas de différence significative entre les scores moyens des deux groupes ($t = -1.60$, $p = 0.12$, $p > 0.05$). Les scores moyens, les écarts-types, la médiane et la statistique t sont présentés dans le Tableau 5.1. Ces résultats suggèrent que les deux groupes avaient des niveaux de connaissances similaires avant l'intervention.

Table 5.1 Résultats statistiques pour le pré-test de connaissances

<i>Groupe</i>	<i>N</i>	<i>Moyenne</i>	<i>Écart-type</i>	<i>Médiane</i>	<i>Statistique t</i>	<i>Valeur p</i>
Contrôle	13	4.69	1.97	5.00	-1.60	0.12
Expérimental	22	6.00	2.53	6.00		

Après l'activité d'apprentissage, les résultats des analyses statistiques pour le post-test de connaissances indiquent que les scores du groupe contrôle suivent une distribution normale (Shapiro-Wilk, $p > 0.05$), tandis que ceux du groupe expérimental ne suivent pas une distribution normale (Shapiro-Wilk, $p < 0.05$). De plus, les variances des groupes sont homogènes (Levene, $p > 0.05$). En raison de la non-normalité des

données pour le groupe expérimental, un test de Mann-Whitney U a été utilisé pour comparer les scores des deux groupes.

Le test de Mann-Whitney U n'a révélé aucune différence significative entre les scores des deux groupes ($U = 115.500$, $p = 0.344$, $p > 0.05$). Les scores moyens, les écarts-types, les médianes et la statistique U sont présentés dans le Tableau 5.2. Ces résultats suggèrent que l'utilisation d'un robot conversationnel motivant n'a pas significativement influencé les connaissances des apprenants par rapport à un robot conversationnel neutre.

Table 5.2 Résultats statistiques du post-test de connaissances

<i>Groupe</i>	<i>N</i>	<i>Moyenne</i>	<i>Écart-type</i>	<i>Médiane</i>	<i>Test</i>	<i>p-valeur</i>
Contrôle	13	7.92	1.50	6.00	Mann-Whitney $U = 115.500$	0.344
Expérimental	22	8.45	1.34	8.50		

5.3.2 Résultats des émotions

Les résultats des analyses statistiques pour les émotions sont présentés dans le Tableau 5.3. Ces analyses montrent qu'il n'y a pas de différences significatives entre les groupes contrôle et expérimental pour les trois émotions évaluées (*confiance*, *enthousiasme* et *intérêt*), que ce soit au pré-test ou au post-test ($p > 0.05$).

Pour la *confiance*, les résultats indiquent qu'il n'y a pas de différence significative entre les groupes au pré-test ($U = 154.500$, $p = 0.691$). Ces résultats suggèrent que les deux groupes avaient des niveaux de confiance similaires avant l'intervention. Pour le post-test, les résultats confirment également qu'il n'y a pas de différence significative entre les groupes ($U = 139.500$, $p = 0.915$). Ainsi, l'utilisation d'un robot conversationnel motivant n'a pas influencé de manière significative la confiance des apprenants comparé à un robot conversationnel neutre.

Concernant l'*enthousiasme*, les analyses révèlent qu'il n'y a pas de différence significative entre les groupes au pré-test ($U = 142.500$, $p = 1.000$), suggérant des niveaux d'enthousiasme comparables avant l'inter-

vention. Pour le post-test, les résultats indiquent également qu'il n'y a pas de différence significative entre les groupes ($U = 159.000$, $p = 0.854$). Ces résultats montrent que l'intervention n'a pas produit d'effet significatif sur l'enthousiasme des apprenants.

Enfin, pour l'intérêt, les résultats montrent qu'il n'y a pas de différence significative entre les groupes au pré-test ($U = 131.000$, $p = 0.663$). Ces résultats suggèrent que les deux groupes avaient des niveaux d'intérêt similaires avant l'intervention. Au post-test, aucune différence significative n'a également été observée ($U = 130.500$, $p = 0.656$). Ces résultats indiquent que l'utilisation d'un robot conversationnel motivant n'a pas influencé l'intérêt des apprenants comparé à un robot conversationnel neutre.

En résumé, les analyses statistiques pour les trois émotions (confiance, enthousiasme, intérêt) montrent qu'il n'y a pas de différences significatives entre les groupes contrôle et expérimental, au post-test ($p > 0.05$). Ces résultats suggèrent que l'intervention utilisant un robot conversationnel motivant n'a pas produit d'effet mesurable sur les émotions des apprenants.

Table 5.3 Résultats statistiques pour les émotions

Émotion	Phase	Groupe	N	Moyenne	Écart-type	Médiane	p-value
Confiance	Pré-test	Contrôle	13	3.54	0.88	4.00	0.691
		Expérimental	22	3.41	1.01	3.50	
	Post-test	Contrôle	13	3.62	1.04	4.00	
		Expérimental	22	3.59	1.22	4.00	
Enthousiasme	Pré-test	Contrôle	13	4.08	0.86	4.00	1.000
		Expérimental	22	4.05	0.90	4.00	
	Post-test	Contrôle	13	3.93	0.88	4.00	
		Expérimental	22	4.00	0.82	4.00	
Intérêt	Pré-test	Contrôle	13	4.23	0.44	4.00	0.663
		Expérimental	22	4.23	0.87	4.00	
	Post-test	Contrôle	13	4.08	0.76	4.00	
		Expérimental	22	4.09	1.02	4.00	

5.3.3 Analyse des perceptions de l'intelligence des participants

Les réponses recueillies avant l'expérimentation ont permis d'examiner les perceptions initiales des participants sur la nature de l'intelligence. Parmi les 35 participants, une majorité (33 sur 35) a exprimé une vision dynamique de l'intelligence, considérant qu'elle peut se développer grâce à l'effort et à la pratique. En revanche, 2 participants avaient une perception fixe de l'intelligence, la percevant comme une capacité immuable.

Après l'expérimentation, les réponses post-expérimentation ont confirmé que ces perceptions initiales sont restées stables. Les 33 participants ayant une vision dynamique ont maintenu cette conception, tandis que les 2 participants ayant une perception fixe ont également conservé leur point de vue.

Ces résultats suggèrent que, bien que l'interaction avec un robot conversationnel motivant puisse promouvoir des messages favorisant une mentalité de croissance, elle n'a pas suffi à modifier les perceptions profondément ancrées des participants ayant une vision fixe de l'intelligence. Toutefois, il est important de souligner que la majorité des participants partageait déjà une perception dynamique avant l'intervention, ce qui limite la possibilité d'observer un changement significatif dans cette population.

5.4 Discussion

Les résultats de cette étude offrent des perspectives précieuses pour comprendre l'efficacité des interventions motivationnelles dans un contexte éducatif numérique, en particulier à travers l'utilisation d'un robot conversationnel motivant dans un MOOC. Les hypothèses initiales (H1, H2a, H2b, H2c) ont été rejetées en raison de l'absence de différences significatives entre les groupes contrôle et expérimental.

5.4.1 Efficacité sur l'acquisition des connaissances (H1)

Les analyses n'ont pas montré d'amélioration significative des connaissances des apprenants suite à l'intervention avec le robot conversationnel motivant. Ces résultats rejoignent les conclusions de Macnamara et Burgoyne (2023), qui mettent en évidence que les interventions fondées sur le concept d'« état d'esprit de croissance » ne produisent pas systématiquement des gains académiques mesurables. Plusieurs raisons peuvent expliquer ces résultats. Premièrement, la durée de l'intervention, limitée dans cette étude, pourrait ne pas avoir permis de produire un effet durable, comme le souligne également Arroyo et al. dans leurs travaux sur les systèmes tutoriels intelligents (Arroyo *et al.*, 2014). Deuxièmement, la taille relativement réduite de l'échantillon pourrait avoir restreint la puissance statistique des analyses, rendant plus difficile la détection d'effets significatifs, même si ceux-ci étaient présents. Troisièmement, le pré-test et le post-test de connaissances, composés de seulement dix questions, pourraient ne pas avoir été suffisants pour évaluer avec précision les progrès des apprenants. Enfin, ces limites méthodologiques s'inscrivent dans les critiques formulées par Macnamara et Burgoyne (2023), qui mettent en évidence les lacunes fréquentes dans la rigueur méthodologique des études évaluant les interventions éducatives.

5.4.2 Effet sur les émotions des apprenants (H2a, H2b, H2c)

Les résultats indiquent que le robot conversationnel motivant n'a pas significativement influencé la confiance, l'enthousiasme ou l'intérêt des apprenants. Ces observations contrastent avec les travaux Arroyo *et al.* (2014), qui ont montré que les compagnons d'apprentissage affectifs, intégrant des principes motivationnels basés sur la théorie de Dweck, peuvent avoir un effet positif sur des émotions telles que la confiance et l'enthousiasme, notamment dans le contexte des mathématiques.

Ces observations peuvent être attribuées à plusieurs facteurs. Premièrement, les échelles utilisées pour mesurer les émotions, avec une granularité de 1 à 5, pourraient ne pas avoir capté des variations émotionnelles

plus subtiles. Il serait donc pertinent d'explorer l'utilisation d'instruments plus sensibles, tels que les capteurs d'affect ou des échelles psychométriques avancées, afin d'évaluer avec davantage de précision l'effet des interventions sur les émotions des apprenants.

Deuxièmement, la durée relativement courte de l'intervention pourrait avoir limité son effet. Arroyo *et al.* (2014) soulignent que des interactions prolongées et variées sont essentielles pour induire des changements mesurables dans les émotions des apprenants. En effet, une intervention sur une seule session de MOOC pourrait ne pas suffire à établir la confiance ou l'enthousiasme, qui nécessitent souvent un renforcement progressif.

Troisièmement, la taille relativement réduite de l'échantillon pourrait avoir limité la généralisation des résultats et ne pas refléter pleinement la diversité des expériences d'apprentissage. Une étude incluant un échantillon plus large pourrait permettre de mieux capturer les effets potentiels du robot conversationnel motivant sur les émotions des apprenants.

De manière plus globale, plusieurs contraintes méthodologiques ont pu limiter le plein potentiel de l'évaluation menée dans cette étude. En particulier, le manque de moyens d'évaluation plus fins, la durée limitée de l'intervention, la taille réduite de l'échantillon et le nombre restreint de contextes (un seul MOOC testé) ont restreint la capacité à mesurer pleinement l'impact d'un agent conversationnel pourtant doté de capacités pédagogiques augmentées, via la RAG et l'intégration d'une consigne pédagogique et comportementale avancée. Ces facteurs peuvent également expliquer en partie pourquoi l'implémentation de la théorie motivationnelle de Dweck dans cet agent n'a pas produit les effets attendus, contrairement à certains résultats observés dans des contextes plus prolongés ou plus interactifs.

Ce chapitre a examiné l'effet d'un robot conversationnel motivant sur l'acquisition des connaissances et les émotions des apprenants dans un MOOC dédié à l'intelligence artificielle. En s'appuyant sur la théorie de Dweck, l'objectif était de favoriser des émotions positives, telles que la confiance, l'enthousiasme et l'intérêt, tout en renforçant l'apprentissage des concepts.

Les résultats n'ont pas révélé de différences significatives entre le groupe expérimental et le groupe contrôle, que ce soit en termes d'acquisition des connaissances ou d'émotions. Ces observations mettent en évidence les limites des interventions motivationnelles de courte durée dans des contextes éducatifs numériques tels que les MOOCs. Elles soulignent également l'importance de développer des méthodologies plus robustes et d'explorer des approches innovantes pour mieux soutenir l'apprentissage et la motivation des apprenants.

CHAPITRE 6

CONCLUSION ET PERSPECTIVES

Cette thèse avait pour objectif d'évaluer l'intégration d'agents conversationnels enrichis par la RAG dans un environnement éducatif numérique, en s'appuyant sur trois questions de recherche.

Concernant RQ1, les résultats montrent que l'intégration de la RAG dans les modèles GPT permet de réduire significativement les erreurs d'hallucination et d'améliorer la précision des réponses générées dans un contexte éducatif. Cette avancée confirme l'intérêt d'utiliser la RAG pour concevoir des agents conversationnels capables de fournir des réponses pédagogiquement pertinentes, contextualisées et fiables.

En lien avec RQ2, l'expérimentation menée dans un MOOC sur l'intelligence artificielle démontre que l'agent conversationnel enrichi par la RAG a eu un effet positif sur l'acquisition des connaissances des apprenants, tout en améliorant leur satisfaction et leur perception de l'utilisabilité du système. Ces résultats soulignent le potentiel de ces agents pour enrichir l'expérience d'apprentissage dans des contextes où l'encadrement humain est limité.

Concernant RQ3, l'intégration de stratégies motivationnelles dans l'agent conversationnel, inspirées de la théorie des états d'esprit de Dweck, n'a pas permis d'observer d'effets significatifs ni sur l'acquisition des connaissances, ni sur les émotions positives mesurées, telles que la confiance, l'intérêt et l'enthousiasme. Cette absence d'effet observable suggère que les dynamiques affectives dans les environnements d'apprentissage en ligne sont complexes et sensibles à de nombreux facteurs. Des analyses supplémentaires seraient nécessaires, notamment en recourant à des consignes plus variées et à un échantillon plus large. Il serait également pertinent d'utiliser des instruments d'évaluation des émotions plus sensibles et plus précis, afin de mieux saisir les subtilités des réactions émotionnelles et d'évaluer plus finement l'impact des agents

conversationnels motivationnels sur l'expérience des apprenants.

6.1 Contributions scientifiques

Cette thèse se distingue par plusieurs contributions spécifiques au domaine des agents conversationnels basés sur les LLM enrichis par la RAG et de leur intégration dans des environnements éducatifs numériques :

1. *Amélioration de la fiabilité des réponses des LLMs grâce à la RAG* : L'intégration de la RAG a permis de réduire les hallucinations, en générant des réponses mieux contextualisées et plus pertinentes pédagogiquement.
2. *Évaluation empirique dans un contexte réel* : Contrairement à de nombreux travaux limités à des évaluations théoriques, cette thèse a mené une expérimentation directe dans le cadre d'un MOOC. Cette approche a permis d'évaluer l'effet de l'agent conversationnel sur l'acquisition des connaissances et l'expérience des apprenants, en fournissant des données empiriques précieuses sur son utilité pédagogique.
3. *Exploration des stratégies motivationnelles et des émotions* : Cette recherche a exploré des stratégies pour intégrer des éléments motivationnels dans les agents conversationnels. Bien que les résultats sur les émotions positives soient limités, ces travaux mettent en lumière les défis méthodologiques associés à l'étude des états affectifs dans des environnements numériques. Ils ouvrent également des perspectives pour le développement d'agents conversationnels plus empathiques et adaptatifs.

6.2 Discussion

Les résultats de cette recherche, fondée sur des données empiriques, représentent une avancée dans le domaine de l'application des LLMs en éducation. Ils fournissent des orientations concrètes pour aider les concepteurs pédagogiques à améliorer la conception des cours en ligne, tels que les MOOCs. En particulier, l'intégration d'agents conversationnels enrichis par la RAG constitue une solution prometteuse pour clarifier

les concepts complexes et offrir un soutien pédagogique fiable, même en l'absence d'encadrement humain en temps réel. Cette technologie a permis de réduire de manière significative les erreurs d'hallucination, garantissant des réponses précises et contextualisées, ce qui renforce leur pertinence pour répondre aux besoins des apprenants.

Par ailleurs, les résultats montrent que ces agents conversationnels jouent un rôle clé dans l'acquisition de connaissances et la satisfaction des apprenants, contribuant ainsi à renforcer leur engagement. Ces constats invitent les concepteurs pédagogiques à intégrer ces agents tout au long du cours ou dans des modules spécifiques, tels que les révisions ou les sections complexes, afin de fournir un soutien personnalisé et de favoriser l'autonomie des apprenants. Par exemple, un agent conversationnel pourrait accompagner les apprenants dans la révision des concepts-clés à la fin de chaque module, en proposant des explications adaptées à leur niveau.

De plus, les traces des conversations des apprenants avec ces agents enrichis par la RAG constituent une ressource précieuse pour les concepteurs pédagogiques. En analysant ces échanges, ils peuvent identifier les concepts mal compris, ajuster les explications du cours, et même enrichir le contenu pédagogique pour mieux répondre aux besoins des apprenants. En adoptant ces agents conversationnels enrichis par la RAG, les concepteurs peuvent non seulement optimiser l'expérience d'apprentissage, mais également garantir une meilleure accessibilité aux contenus pédagogiques, rendant ainsi les cours en ligne plus interactifs et efficaces.

Ce travail s'inscrit dans un changement de paradigme lié à l'IA générative, marqué par une augmentation de l'usage des agents conversationnels durant les apprentissages, particulièrement dans les contextes d'apprentissage à distance asynchrones. Ces environnements nécessitent une grande autonomie de la part des apprenants, et les agents conversationnels pourraient représenter une avancée notable en matière d'enca-

drement pédagogique. Dans ce cadre, les agents conversationnels agissent comme des compagnons d'apprentissage, offrant un soutien pédagogique adapté grâce à leurs connaissances supérieures à celles des apprenants qu'ils accompagnent, tout en demeurant susceptibles d'erreurs. Sans se substituer aux enseignants ou experts humains, ils assurent un encadrement accessible et efficace, particulièrement dans des contextes où le soutien humain est limité. Bien qu'ils puissent parfois se tromper, leur précision, estimée à 80% selon les résultats de notre étude (Miladi *et al.*, 2024b), constitue un atout majeur dans des environnements d'apprentissage où l'encadrement humain est limité ou absent. Ce rôle de compagnon est particulièrement crucial dans les cours à distance non crédités, tels que les MOOCs, où les apprenants doivent parcourir seuls à travers les contenus. En fournissant des réponses contextualisées et adaptées, les agents conversationnels enrichis par la RAG contribuent à maintenir l'engagement des apprenants, notamment dans des contextes éducatifs où le risque d'abandon est élevé.

Enfin, cette recherche s'inscrit dans la lignée des travaux sur les systèmes tutoriels intelligents (STI), qui visent à offrir un soutien pédagogique personnalisé aux apprenants grâce à l'intelligence artificielle. Les STI reposent sur l'intégration de quatre modèles complémentaires : un modèle du domaine, un modèle de l'apprenant, un modèle pédagogique et un modèle d'interface, qui interagissent pour guider l'apprentissage (Nkambou *et al.*, 2010).

Dans ce cadre, l'agent conversationnel enrichi par la RAG proposé dans cette thèse peut être analysé à la lumière de ces concepts : les données sélectionnées par la RAG constituent le modèle du domaine, tandis que le *prompt engineering* joue le rôle d'un modèle pédagogique en orientant la manière dont les réponses sont formulées. L'intégration d'un modèle de l'apprenant constituerait un élément clé pour permettre à ces agents de personnaliser davantage les réponses en fonction du profil, des connaissances et des besoins spécifiques de chaque apprenant. Cette dimension permettrait de rapprocher ces agents conversationnels des approches traditionnelles des STI, tout en conservant la flexibilité et la puissance offertes par les modèles

de langage génératifs.

6.3 Perspectives de recherche

Cette thèse met en lumière le potentiel des agents conversationnels pédagogiques dans les environnements éducatifs numériques, notamment les MOOCs, où l'autonomie des apprenants est un enjeu clé. Ces agents représentent une évolution majeure à l'ère de l'IA générative, en offrant un encadrement pédagogique adapté dans des contextes où le soutien humain est souvent limité. Grâce à des bases de connaissances validées, ils fournissent des réponses précises et participent à l'enrichissement des expériences d'apprentissage. Cependant, pour exploiter pleinement leur potentiel, plusieurs axes de recherche méritent d'être approfondis :

1. *Adaptation à divers contextes éducatifs* : Bien que cette étude se concentre sur les MOOCs, il serait pertinent d'explorer l'utilisation des agents conversationnels dans d'autres contextes pédagogiques. Par exemple, leur utilisation pourrait être étudiée dans des environnements d'apprentissage tels que les cours en ligne, les environnements hybrides, les formations professionnelles ou collaboratives. Ces contextes présentent des défis spécifiques, tels que la diversité des besoins des apprenants. L'intégration des agents conversationnels pourrait s'avérer particulièrement utile dans les situations où l'encadrement humain est limité, comme dans les environnements asynchrones. De futures recherches pourraient également valider l'efficacité des agents dans des disciplines variées, en approfondissant leur influence sur l'apprentissage et en explorant davantage leur influence émotionnelle. Cela permettrait non seulement de confirmer la robustesse des résultats obtenus, mais aussi d'élargir leur applicabilité à une gamme plus vaste de contextes éducatifs.
2. *Études longitudinales sur l'apprentissage et les émotions* : Des évaluations à long terme seraient essentielles pour approfondir la compréhension des effets durables des agents conversationnels sur

l'apprentissage des apprenants, ainsi que sur l'évolution de leurs émotions tout au long de leur parcours éducatif.

3. *Personnalisation accrue* : Le développement d'agents conversationnels basés sur l'intelligence artificielle générative capables de s'adapter au profil de l'apprenant offre des perspectives prometteuses pour l'éducation numérique. Ce profil peut inclure des dimensions telles que les compétences actuelles de l'apprenant, ainsi que son historique de réussite ou d'échec dans des activités d'apprentissage spécifiques. En exploitant ces informations, l'agent conversationnel peut personnaliser ses interactions, proposer des activités adaptées et fournir des rétroactions précises et motivantes. Cette personnalisation vise à stimuler la motivation, renforcer l'autonomie des apprenants et optimiser l'efficacité de leur apprentissage en répondant de manière ciblée à leurs besoins spécifiques.

Ainsi, l'agent conversationnel pourrait évoluer en un véritable compagnon d'apprentissage, guidant les apprenants à travers des parcours éducatifs individualisés. Ce type de personnalisation améliore non seulement la satisfaction et l'engagement, mais il facilite également la remédiation ciblée en cas de difficultés, tout en maximisant le potentiel d'apprentissage de chaque étudiant.

4. *Évaluation fine des émotions* : L'intégration de technologies avancées, comme l'analyse des expressions faciales ou des signaux physiologiques, permettrait de mesurer précisément l'effet des agents conversationnels sur les émotions des apprenants. Ces données offriraient une meilleure compréhension des dynamiques émotionnelles et favoriseraient des interactions personnalisées, adaptées aux besoins de chaque apprenant pour maximiser leur engagement et leur apprentissage.
5. *Capacités empathiques et adaptatives* : La conception d'un agent conversationnel empathique, capable de détecter les émotions des apprenants en temps réel et d'adapter son interaction en conséquence, constitue une avancée prometteuse dans l'innovation pédagogique. Un tel agent, conçu pour intégrer une base de connaissances, liées au domaine enseigné, pourrait offrir une expérience

éducative plus humaine, engageante et personnalisée. En adaptant son discours à l'humeur et aux besoins spécifiques de chaque apprenant, il pourrait non seulement renforcer l'engagement et la persévérance, mais aussi améliorer les résultats d'apprentissage. Nos travaux vont se poursuivre dans cette direction, notamment en enrichissant les capacités de l'agent par la reconnaissance fine des émotions des apprenants, afin d'optimiser son interaction et de favoriser une expérience d'apprentissage encore plus enrichissante et adaptée.

En conclusion, cette thèse démontre que l'intégration de la RAG dans les agents conversationnels améliore significativement la précision des réponses et leur pertinence pédagogique. Les résultats mettent en évidence un effet positif sur l'acquisition des connaissances et la satisfaction des apprenants, tout en soulignant les défis liés à l'intégration des dimensions émotionnelles. Ces travaux ouvrent des perspectives prometteuses pour le développement d'agents conversationnels en tant que compagnons d'apprentissage, plus empathiques et adaptatifs, capables de répondre aux besoins des apprenants, tout en enrichissant leur expérience d'apprentissage.

ANNEXE A
CERTIFICAT D'ÉTHIQUE



CERTIFICAT D'ÉTHIQUE

2023-10_{modification}

Le comité d'éthique de la recherche de la TÉLUQ certifie avoir examiné
la proposition de recherche soumise par :

Fatma Miladi

Intitulée :

***« Robot conversationnel basé sur l'intelligence artificielle
générative pour les apprenants dans un cours en ligne »***

Et avoir conclu que la recherche proposée est entièrement conforme
aux normes d'éthique de la recherche selon la *Politique d'éthique de la
recherche avec les êtres humains.*

Valide jusqu'au 20 mars 2025

Date : 16 octobre 2024

François Pichette
Président

ANNEXE B

IMPLÉMENTATION

```
def create_context(question):
    # Creating an embedding for the question that's been asked
    product_embedding = get_embedding(question)

    # Finds the relevance of each piece of data in context of the question
    df["similarity"] = df.embedding.apply(lambda x: 1 - cosine_similarity(x, product_embedding))

    # Sorts the text chunks based on how relevant they are to finding the answer to the question
    df2 = df.sort_values("similarity", ascending=False)

    contexte = "\n\n###\n\n".join(c[2:-2] for c in df2["combined"].iloc[0:3])
    return contexte
```

Figure B.1 Exemple de Code Python pour créer le contexte

```
def chat():
    question = request.json['question'] # Receives the question
    derniere_question = str(request.json['lastQuestion'])
    derniere_reponse = str(request.json['lastResponse'])
    if not question:
        return jsonify({'message': 'Veuillez poser une question.'}), 400 # Bad request with error message
    contexte = create_context(question)
    historique = ""
    if len(derniere_question) > 2 and len(derniere_reponse) > 2:
        historique = (
            f"\n\nDans le cas où la question précédente te soit utile pour répondre à la nouvelle question, "
            f"voici la question précédente et la réponse que tu as donnée.\n\n"
            f"Question précédente: {derniere_question}\n\n"
            f"---\n\n"
            f"Réponse précédente: {derniere_reponse}\n\n")
    try:
        response = openai.ChatCompletion.create(
            model="gpt-4",
            messages=[
                {"role": "system",
                 "content": f"Tu es un assistant pour répondre à des questions sur la matière d'un MOOC sur l'intelligence artificielle."
                 "Agis comme un outil d'aide à l'apprentissage. Réponds en français à la question en te basant d'abord sur le contexte ci-dessous.{historique}\n\n"},
                {"role": "user",
                 "content": f"Contexte: {contexte}\n\n---\n\nQuestion: {question}\n\nRéponse:"}
            ],
            temperature=0.7,
            max_tokens=2000)
        final_response = response['choices'][0]['message']['content']
        return Response(final_response, content_type='text/plain; charset=utf-8')
```

Figure B.2 Exemple de Code Python pour générer la réponse

BIBLIOGRAPHIE

- Abdelghani, R., Wang, Y.-H., Yuan, X., Wang, T., Sauz'eon, H. et Oudeyer, P.-Y. (2022). Gpt-3-driven pedagogical agents to train children's curious question-asking skills. *International Journal of Artificial Intelligence in Education*, 1–36.
- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S. et al. (2023). Gpt-4 technical report. *arXiv preprint arXiv :2303.08774*.
- Aimeur, E., Dufort, H., Leibu, D. et Frasson, C. (1997). Some justifications for the learning by disturbing strategy. Dans *Proceedings of the Eighth World Conference on Artificial Intelligence in Education*, 119–126. Citeseer.
- Alkaissi, H. et McFarlane, S. I. (2023). Artificial hallucinations in chatgpt : implications in scientific writing. *Cureus*, 15(2).
- Allouch, M., Azaria, A. et Azoulay, R. (2021). Conversational agents : Goals, technologies, vision and challenges. *Sensors*, 21(24), 8448.
- Arroyo, I., Cooper, D. G., Bureson, W., Woolf, B. P., Muldner, K. et Christopherson, R. (2009). Emotion sensors go to school. Dans *Artificial intelligence in education*, 17–24. Ios Press.
- Arroyo, I., Woolf, B. P., Bureson, W., Muldner, K., Rai, D. et Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education*, 24, 387–426.
- Banchs, R. E. et Li, H. (2012). Iris : a chat-oriented dialogue system based on the vector space model. Dans *Proceedings of the ACL 2012 System Demonstrations*, 37–42.
- Bandyopadhyay, S., Xu, J., Pawar, N. et Touretzky, D. (2022). Interactive visualizations of word embeddings for k-12 students. Dans *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 12713–12720.
- Bangor, A., Kortum, P. et Miller, J. (2009). Determining what individual sus scores mean : Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114–123.
- Beltagy, I., Lo, K. et Cohan, A. (2019). Scibert : A pretrained language model for scientific text. Dans *Conference on Empirical Methods in Natural Language Processing*.
- Bengio, Y., Simard, P. et Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166.
- Bommarito, J., Bommarito, M., Katz, D. M. et Katz, J. (2023). Gpt as knowledge worker : a zero-shot evaluation of (ai) cpa capabilities. *arXiv preprint arXiv :2301.04408*.
- Braun, V. et Clarke, V. (2006). Using thematic analysis in psychology qualitative research in psychology, 3 (2), 77-101. *Online im Internet*, 1–42.
- Brooke, J. (1996). Sus : A quick and dirty usability scale. *Usability Evaluation in Industry*.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901. Récupéré de <https://arxiv.org/abs/2005.14165>

- Bulger, M., Bright, J. et Cobo, C. (2015). The real component of virtual learning : motivations for face-to-face mooc meetings in developing and industrialised countries. *Information, Communication & Society*, 18(10), 1200–1216.
- Ceci, S. J., Williams, W. M. et Barnett, S. M. (2009). Women’s underrepresentation in science : sociocultural and biological considerations. *Psychological bulletin*, 135(2), 218.
- Chan, T.-W. et Baskin, A. B. (1988). Studying with the prince : The computer as a learning companion. Dans *Proceedings of the International Conference on Intelligent Tutoring Systems*, volume 194200.
- Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y. et al. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3), 1–45.
- Chin, J. P., Diehl, V. A. et Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. Dans *Proceedings of the SIGCHI conference on Human factors in computing systems*, 213–218. ACM. <http://dx.doi.org/10.1145/57167.57203>
- Chiu, T. K. et Hew, T. K. (2018). Factors influencing peer learning and performance in mooc asynchronous online discussion forum. *Australasian Journal of Educational Technology*, 34(4).
- Creswell, J. W., Clark, V. L. P., Gutmann, M. L. et Hanson, W. E. (2003). Advanced mixed. *Handbook of mixed methods in social & behavioral research*, 209, 209–240.
- Devlin, J., Chang, M.-W., Lee, K. et Toutanova, K. (2019). Bert : Pre-training of deep bidirectional transformers for language understanding. Dans *North American Chapter of the Association for Computational Linguistics*.
- Dijkstra, R., Genç, Z., Kayal, S., Kamps, J. et al. (2022). Reading comprehension quiz generation using generative pre-trained transformers. Dans *iTextbooks@ AIED*, 4–17.
- Duong, T. et Suppasetserree, S. (2024). The effects of an artificial intelligence voice chatbot on improving vietnamese undergraduate students’ english speaking skills. *International Journal of Learning, Teaching and Educational Research*, 23(3), 293–321.
- Dweck, C. S. (2002). Messages that motivate : How praise molds students’ beliefs, motivation, and performance (in surprising ways). In *Improving academic achievement* 37–60. Elsevier.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211.
- Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, 22(5), 323–327. <http://dx.doi.org/10.1016/j.intcom.2010.04.004>
- Freedman, R., Rosé, C. P., Ringenber, M. A. et VanLehn, K. (2000). Its tools for natural language dialogue : A domain-independent parser and planner. Dans *International Conference on Intelligent Tutoring Systems*, 433–442. Springer.
- Fryer, L. K., Ainley, M., Thompson, A., Gibson, A. et Sherlock, Z. (2017). Stimulating and sustaining interest in a language course : An experimental comparison of chatbot and human task partners. *Computers in human behavior*, 75, 461–468.
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J. et Wang, H. (2023). Retrieval-augmented generation for large language models : A survey. *arXiv preprint arXiv :2312.10997*.

- Gimpel, H., Hall, K., Decker, S., Eymann, T., Lämmermann, L., Mädche, A., Röglinger, M., Ruiner, C., Schoch, M., Schoop, M. *et al.* (2023). *Unlocking the power of generative AI models and systems such as GPT-4 and ChatGPT for higher education : A guide for students and lecturers*. Rapport technique, Hohenheim Discussion Papers in Business, Economics and Social Sciences.
- González-Castro, N., Muñoz-Merino, P. J., Alario-Hoyos, C. et Kloos, C. D. (2021). Adaptive learning module for a conversational agent to support mooc learners. *Australasian Journal of Educational Technology*, 37(2), 24–44.
- Graesser, A. C., Person, N., Harter, D., Group, T. R. *et al.* (2001). Teaching tactics and dialog in autotutor. *International Journal of Artificial Intelligence in Education*, 12(3), 257–279.
- Griol, D. et Callejas, Z. (2013). An architecture to develop multimodal educative applications with chatbots. *International Journal of Advanced Robotic Systems*, 10(3), 175.
- Grudin, J. et Jacques, R. (2019). Chatbots, humbots, and the quest for artificial general intelligence. Dans *Proceedings of the 2019 CHI conference on human factors in computing systems*, 1–11.
- Guo, Y. R., Goh, D. H.-L., Luyt, B., Sin, S.-C. J. et Ang, R. P. (2015). The effectiveness and acceptance of an affective information literacy tutorial. *Computers & Education*, 87, 368–384.
- Hassenzahl, M. (2003). Attrakdiff : A questionnaire to measure perceived hedonic and pragmatic quality. Dans *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 47, 1054–1058. SAGE Publications. <http://dx.doi.org/10.1177/154193120304700616>
- Hattie, J. et Timperley, H. (2007). The power of feedback. review of educational re-search, 77 (1) : 81-112.
- Hebb, D. O. (1949). *The organization of behavior : A neuropsychological theory*.
- Hochreiter, S. et Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hone, K. S. et El Said, G. R. (2016). Exploring the factors affecting mooc retention : A survey study. *Computers & Education*, 98, 157–168.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A. et Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38.
- Johnson, J., Douze, M. et Jégou, H. (2019). Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3), 535–547.
- Johnson, W. L., Rickel, J. W., Lester, J. C. *et al.* (2000). Animated pedagogical agents : Face-to-face interaction in interactive learning environments. *International Journal of Artificial intelligence in education*, 11(1), 47–78.
- Jurafsky, D. (2000). *Speech and language processing*.
- Kasneji, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E. *et al.* (2023). Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103, 102274.
- Kerlyl, A., Hall, P. et Bull, S. (2006). Bringing chatbots into education : Towards natural language negotiation of open learner models. Dans *International conference on innovative techniques and applications of artificial intelligence*, 179–192. Springer.

- Kestin, G., Miller, K., Klales, A., Milbourne, T. et Ponti, G. (2024). Ai tutoring outperforms active learning.
- Kim, Y., Baylor, A. L. et Shen, E. (2007). Pedagogical agents as learning companions : the impact of agent emotion and gender. *Journal of Computer Assisted Learning*, 23(3), 220–234.
- Kim, Y., Jernite, Y., Sontag, D. et Rush, A. M. (2016). Character-aware neural language models. Dans *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2741–2749.
- Kim, Y., Xu, X., McDuff, D., Breazeal, C. et Park, H. W. (2024). Health-llm : Large language models for health prediction via wearable sensor data. *arXiv preprint arXiv :2401.06866*.
- Ko, H.-T., Liu, Y.-K., Tsai, Y.-C. et Suen, S. (2024). Enhancing python learning through retrieval-augmented generation : A theoretical and applied innovation in generative ai education. Dans *International Conference on Innovative Technologies and Learning*, 164–173. Springer.
- Kuratomi, G., Pirozelli, P., Cozman, F. G. et Peres, S. M. (2024). A rag-based institutional assistant. *Anais do XXI Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2024)*.
- Le Cun, Y., Bengio, Y. et Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436–444.
- Lebret, R. P. (2016). *Word embeddings for natural language processing*. Rapport technique, EPFL.
- Levy, D. et Albertos, A. P. (2024). *Teaching Effectively with ChatGPT : A Practical Guide to Creating Better Learning Experiences for Your Students in Less Time*. Kindle Books. Accessed : 2024-01-17. Récupéré de <https://www.amazon.com/Teaching-EYectively-ChatGPT-practical-experiences/dp/B0D8P72M8>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T. et al. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474.
- Li, C. et Xing, W. (2021). Natural language generation using deep learning to support mooc learners. *International Journal of Artificial Intelligence in Education*, 31, 186–214.
- Li, H., Li, C., Xing, W., Baral, S. et Heffernan, N. (2024). Automated feedback for student math responses based on multi-modality and fine-tuning. Dans *Proceedings of the 14th Learning Analytics and Knowledge Conference*, 763–770.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H. et Neubig, G. (2023). Pre-train, prompt, and predict : A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1–35.
- Liu, R., Zenke, C., Liu, C., Holmes, A., Thornton, P. et Malan, D. J. (2024). Teaching cs50 with ai : leveraging generative artificial intelligence in computer science education. Dans *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 750–756.
- Macnamara, B. N. et Burgoyne, A. P. (2023). Do growth mindset interventions impact students' academic achievement? a systematic review and meta-analysis with recommendations for best practices. *Psychological bulletin*, 149(3-4), 133.
- Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S. et al. (2020). Language models are few-shot learners. *arXiv preprint arXiv :2005.14165*, 1.

- Mao, R., Chen, G., Zhang, X., Guerin, F. et Cambria, E. (2023). Gptheval : A survey on assessments of chatgpt and gpt-4. Dans *International Conference on Language Resources and Evaluation*.
- McCulloch, W. S. et Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5, 115–133.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. et Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.
- Miladi, F., Psyché, V., Diattara, A., El Mawas, N. et Lemire, D. (2025a). Design and evaluation of a conversational agent enriched by retrieval-augmented generation : Effect on learners' knowledge acquisition, perceived usability and interaction experience. *International Journal of Technologies in Higher Education*, 22(1).
- Miladi, F., Psyché, V., Diattara, A., El Mawas, N. et Lemire, D. (2025b). Evaluating a gpt-4 and retrieval-augmented generation-based conversational agent to enhance learning experience in a mooc.
- Miladi, F., Psyché, V. et Lemire, D. (2024a). Comparative performance of gpt-4, rag-augmented gpt-4, and students in moocs. Dans *International Conference on Breaking Barriers with Generative Intelligence*, 81–92. Springer.
- Miladi, F., Psyché, V. et Lemire, D. (2024b). Leveraging gpt-4 for accuracy in education : A comparative study on retrieval-augmented generation in moocs. Dans *International Conference on Artificial Intelligence in Education*, 427–434. Springer.
- Minsky, M. et Papert, S. (1969). *Perceptrons : An Introduction to Computational Geometry*. MIT Press.
- Nair, V. et Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. Dans *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807–814.
- Neelakantan, A., Xu, T., Puri, R., Radford, A., Han, J. M., Tworek, J., Yuan, Q., Tezak, N., Kim, J. W., Hallacy, C. et al. (2022). Text and code embeddings by contrastive pre-training. *arXiv preprint arXiv :2201.10005*.
- Newell, A. et Simon, H. A. (1959). The general problem solver : A program that simulates human thought. *Proceedings of the International Conference on Information Processing*, 256–310.
- Nielsen, J. (1994). *Usability Engineering*. San Francisco, CA : Morgan Kaufmann.
- Nilsson, N. J. (1991). Logic and artificial intelligence. *Artificial intelligence*, 47(1-3), 31–56.
- Nkambou, R., Mizoguchi, R. et Bourdeau, J. (2010). *Advances in intelligent tutoring systems*, volume 308. Springer.
- Papi, C., Gérin-Lajoie, S., Czeszak, W. et Tsayem Tchoupou, A. (2022). Accompagnement des étudiants : comment contrer l'isolement en formation à distance ?
- Pekrun, R., Frenzel, A., Goetz, T. et Perry, R. (2007). The control-value theory of achievement emotions : An integrative approach to emotions in educations. *emotion in education*, 13-36.
- Pekrun, R., Goetz, T., Titz, W. et Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement : A program of qualitative and quantitative research. *Educational psychologist*, 37(2), 91–105.

- Pennington, J., Socher, R. et Manning, C. D. (2014). Glove : Global vectors for word representation. Dans *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543.
- Polak, M. P. et Morgan, D. (2024). Extracting accurate materials data from research papers with conversational language models and prompt engineering. *Nature Communications*, 15(1), 1569.
- Psyché, V. (2020). Clom-motsia : Mooc sur l'intelligence artificielle. Accédé le 17 janvier 2024. Récupéré de <https://clom-motsia.telug.ca/>
- Qadir, J. (2023). Engineering education in the era of chatgpt : Promise and pitfalls of generative ai for education. Dans *2023 IEEE Global Engineering Education Conference (EDUCON)*, 1–9. IEEE.
- Radford, A. (2018). Improving language understanding by generative pre-training.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. *et al.* (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. et Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140), 1–67.
- Reimers, N. et Gurevych, I. (2019). Sentence-bert : Sentence embeddings using siamese bert-networks. Dans *Conference on Empirical Methods in Natural Language Processing*.
- Reynolds, L. et McDonell, K. (2021). Prompt programming for large language models : Beyond the few-shot paradigm. Dans *Extended abstracts of the 2021 CHI conference on human factors in computing systems*, 1–7.
- Roos, S. (2018). Chatbots in education : A passing trend or a valuable pedagogical tool?
- Rosenblatt, F. (1958). The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
- Ruan, S., Jiang, L., Xu, J., Tham, B. J.-K., Qiu, Z., Zhu, Y., Murnane, E. L., Brunskill, E. et Landay, J. A. (2019). Quizbot : A dialogue-based adaptive learning system for factual knowledge. Dans *Proceedings of the 2019 CHI conference on human factors in computing systems*, 1–13.
- Rumelhart, D. E., Hinton, G. E. et Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.
- Shawar, B. A. et Atwell, E. (2007). Fostering language learner autonomy through adaptive conversation tutors. Dans *Proceedings of the The fourth Corpus Linguistics conference*, volume 3, 186–193.
- Shortliffe, E. H. (1975). Computer-based medical consultations : Mycin. *New York : Elsevier*.
- Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1), 11–21.
- Su, M.-H., Wu, C.-H., Huang, K.-Y., Hong, Q.-B. et Wang, H.-M. (2017). A chatbot using lstm-based multi-layer embedding for elderly care. Dans *2017 international conference on orange technologies (ICOT)*, 70–74. IEEE.

- Suhaili, S. M., Salim, N. et Jambli, M. N. (2021). Service chatbots : A systematic review. *Expert Systems with Applications*, 184, 115461.
- Sun, D., Boudouaia, A., Zhu, C. et Li, Y. (2024). Would chatgpt-facilitated programming mode impact college students' programming behaviors, performances, and perceptions? an empirical study. *International Journal of Educational Technology in Higher Education*, 21(1), 14.
- Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation*, volume 55 37–76. Elsevier.
- Taneja, K., Maiti, P., Kakar, S., Guruprasad, P., Rao, S. et Goel, A. K. (2024). Jill watson : A virtual teaching assistant powered by chatgpt. Dans *International Conference on Artificial Intelligence in Education*, 324–337. Springer.
- Tegos, S., Demetriadis, S. et Karakostas, A. (2011). Mentorchat : Introducing a configurable conversational agent as a tool for adaptive online collaboration support. Dans *2011 15th panhellenic conference on informatics*, 13–17. IEEE.
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F. et al. (2023). Llama : Open and efficient foundation language models. *arXiv preprint arXiv :2302.13971*.
- Tullis, T. S. et Stetson, J. N. (2004). A comparison of questionnaires for assessing website usability. Dans *Usability professional association conference*, volume 1, 1–12. Minneapolis, USA.
- Vanlehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A. et Wintersgill, M. (2005). The andes physics tutoring system : Lessons learned. *International Journal of Artificial Intelligence in Education*, 15(3), 147–204.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. et Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Vinyals, O. et Le, Q. (2015). A neural conversational model. *arXiv preprint arXiv :1506.05869*.
- Vygotsky, L. S. (1978). *Mind in society : The development of higher psychological processes*, volume 86. Harvard university press.
- Wallace, R. (2003). The elements of aiml style. *Alice AI Foundation*.
- Wang, K., Ramos, J. et Lawrence, R. (2023). Chated : a chatbot leveraging chatgpt for an enhanced learning experience in higher education. *arXiv preprint arXiv :2401.00052*.
- Wei, J., Bosma, M., Zhao, V. Y., Guu, K., Yu, A. W., Lester, B., Du, N., Dai, A. M. et Le, Q. V. (2021). Finetuned language models are zero-shot learners. *arXiv preprint arXiv :2109.01652*.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D. et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35, 24824–24837.
- Weizenbaum, J. (1966). Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45.

- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J. et Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv :2302.11382*.
- Winkler, R. et Söllner, M. (2018). Unleashing the potential of chatbots in education : A state-of-the-art analysis. Dans *Academy of Management Proceedings*, volume 2018, p. 15903. Academy of Management Briarcliff Manor, NY 10510.
- Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M. et Drachslar, H. (2021). Are we there yet?-a systematic literature review on chatbots in education. *Frontiers in artificial intelligence*, 4, 654924.
- Woolf, B. P. (2010). *Building intelligent interactive tutors : Student-centered strategies for revolutionizing e-learning*. Morgan Kaufmann.
- Woolf, B. P., Arroyo, I., Muldner, K., Burleson, W., Cooper, D. G., Dolan, R. et Christopherson, R. M. (2010). The effect of motivational learning companions on low achieving students and students with disabilities. Dans *Intelligent Tutoring Systems : 10th International Conference, ITS 2010, Pittsburgh, PA, USA, June 14-18, 2010, Proceedings, Part I 10*, 327–337. Springer.
- Wu, S., Irsoy, O., Lu, S., Dabrovolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D. et Mann, G. (2023). Bloomberggpt : A large language model for finance. *arXiv preprint arXiv :2303.17564*.
- Wu, Y., Wu, W., Xing, C., Zhou, M. et Li, Z. (2016). Sequential matching network : A new architecture for multi-turn response selection in retrieval-based chatbots. *arXiv preprint arXiv :1612.01627*.
- Xie, Z., Wu, X. et Xie, Y. (2024). Can interaction with generative artificial intelligence enhance learning autonomy ? a longitudinal study from comparative perspectives of virtual companionship and knowledge acquisition preferences. *Journal of Computer Assisted Learning*.
- Xu, A., Liu, Z., Guo, Y., Sinha, V. et Akkiraju, R. (2017). A new chatbot for customer service on social media. Dans *Proceedings of the 2017 CHI conference on human factors in computing systems*, 3506–3510.
- Yan, L., Zhao, L., Echeverria, V., Jin, Y., Alfredo, R., Li, X., Gašević, D. et Martinez-Maldonado, R. (2024). Vizchat : Enhancing learning analytics dashboards with contextualised explanations using multimodal generative ai chatbots. Dans *International Conference on Artificial Intelligence in Education*, 180–193. Springer.
- Yuan, L. et Powell, S. (2013). Moocs and open education : Implications for higher education.
- Zhang, X. et Le Cun, Y. (2015). Character-level convolutional networks for text classification. Dans *Advances in Neural Information Processing Systems*, 649–657.
- Zhao, C., Agrawal, G., Kumarage, T., Tan, Z., Deng, Y., Chen, Y.-C. et Liu, H. (2024). Ontology-aware rag for improved question-answering in cybersecurity education. *arXiv preprint arXiv :2412.14191*.
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z. et al. (2023). A survey of large language models. *arXiv preprint arXiv :2303.18223*.