[Page de garde]

HEC MONTRÉAL

Expérience utilisateur et dynamiques éducatives dans les contextes d'apprentissage par simulation : Une exploration de l'interaction avec les agents virtuels assistée par l'IA générative

par
Nico Helie

Pierre-Majorique Léger
HEC Montréal
Codirecteur de recherche

Constantinos K. Coursaris
HEC Montréal
Codirecteur de recherche

Sciences de la gestion
(Spécialisation Expérience utilisateur)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

Octobre 2024

Résumé

Ce mémoire par articles examine l'intégration et les dynamiques éducatives de l'intelligence artificielle générative (GenAI), en se concentrant particulièrement sur les grands modèles de langage (LLMs), dans le contexte des simulations de réalité virtuelle (VR) sur ordinateur. La récente émergence des technologies d'intelligence artificielle (IA) offre de nombreuses opportunités pour leur incorporation dans le secteur de l'éducation. Parmi les progrès réalisés, l'utilisation des LLMs facilite actuellement le développement de dialogues plus réalistes. La simulation, qui repose sur le principe fondamental d'une plus grande fidélité, pourrait bénéficier grandement de l'incorporation de ces technologies, ouvrant ainsi la voie à des progrès importants dans son évolution à venir.

Étant donné le potentiel croissant d'intégration de ces technologies dans les outils pédagogiques, il est essentiel d'approfondir notre compréhension de l'expérience utilisateur lors de leur utilisation. Dans un premier temps, cette étude vise à analyser, à l'aide d'une approche qualitative, les réactions émotionnelles, comportementales et cognitives suscitées par l'utilisation de différentes formes d'interaction avec des agents virtuels dans des contextes de simulation centrés sur la résolution de problèmes. Deuxièmement, notre but est d'examiner comment l'expérience utilisateur, influencée par ces diverses modalités d'interaction, influence l'attitude envers cette technologie soutenue par GenAI. Un autre volet de notre étude focalise sur l'exploration des possibilités pédagogiques associées à l'emploi d'un LLM accessible, tel que ChatGPT (OpenAI), au sein d'un jeu de rôle conversationnel, en comparaison avec une approche plus élaborée offrant une immersion visuelle, comme celle proposée par la réalité virtuelle sur ordinateur.

Pour réaliser ces objectifs, une première étude qualitative exploratoire (Article 1) a été entreprise afin de répondre à la question suivante : Comment une modalité d'interaction avec des agents virtuels alimentée par GenAI, comparée à une modalité basée sur des choix préétablis, influence-t-elle l’expérience utilisateur de la simulation, et dans quelle mesure les scripts cognitifs de ces modalités d'interaction reflètent-ils les scénarios du monde réel ? Dans cette première étude, nos conclusions indiquent que l'interaction vocale propulsée par GenAI, en comparaison avec une modalité basée sur le choix, pourrait améliorer l'expérience utilisateur en

augmentant l'engagement avec l'interface de simulation et en reflétant plus fidèlement les scénarios du monde réel grâce à des interactions plus naturelles et intuitives.

Ces résultats ont orienté notre problématique de recherche ainsi que le champ d'étude pour l'article ultérieur (Article 2), formulée de la manière suivante : De quelles manières et dans quelle mesure la modalité d'interaction vocale alimentée par GenAI et la fidélité de l'environnement de simulation affectent-elles les attitudes des utilisateurs envers la technologie dans des contextes d'apprentissage par simulation de résolution de problèmes complexes ? Les conclusions exposées dans le second article indiquent en premier lieu que les utilisateurs considèrent l'intelligence artificielle vocale comme étant plus immersive, mais également plus exigeante sur le plan cognitif, notamment en raison des problèmes technologiques (e.g. la reconnaissance vocale) inhérents à une technologie en développement. Nos résultats confirment que la modalité d'interaction et le degré de réalisme de l'environnement de simulation ont un impact considérable sur l'expérience utilisateur dans les contextes d'apprentissage par simulation. L'analyse des données souligne l'importance d'un équilibre entre l'engagement émotionnel, l'effort cognitif et la précision des réponses afin de garantir une expérience utilisateur optimale. Dans ce cadre, il a été observé que ChatGPT (OpenAI), grâce à ses interactions textuelles simples et à sa qualité linguistique, a fréquemment surpassé la réalité virtuelle en matière de fluidité et de réduction de la charge cognitive. Cela met en évidence l'importance de la simplicité de l'interface par rapport au réalisme visuel, ainsi que l'amélioration technologique constante pour stimuler l'engagement des utilisateurs.

Mots clés : IA Générative, LLM, ChatGPT, Simulation, VR, Patients/Agents virtuels, Éducation, Résolution de problèmes, Multimodalité, Voix

Méthodes de recherche : Think-aloud, Entrevue, Sondages, Mesures physiologiques

Abstract

This thesis by articles examines the integration and educational dynamics of generative artificial intelligence (GenAI), focusing particularly on large language models (LLMs), in the context of computer-based virtual reality (VR) simulations. The recent emergence of artificial intelligence (AI) technologies offers numerous opportunities for their incorporation into the education sector. Among the advances made, the use of LLMs currently facilitates the development of more realistic dialogues. Simulation, which relies on the fundamental principle of greater fidelity, could greatly benefit from the incorporation of these technologies, paving the way for significant progress in its future evolution. Due to the recent advancement of these technologies, the evaluation of user experience related to their use in the field of education remains largely unknown.

Given the growing potential for integrating these technologies into educational tools, it is essential to deepen our understanding of the educational dynamics that define the user experience during their use. Firstly, this study aims to analyze, using a qualitative approach, the emotional, behavioral, and cognitive reactions induced by the use of different forms of interaction with virtual agents in simulation contexts centered on problem-solving. Secondly, our goal is to examine how the user experience, influenced by these various interaction modalities, impacts the attitude towards this technology supported by GenAI. Another aspect of our study focuses on exploring the educational possibilities associated with the use of an accessible LLM, such as ChatGPT (OpenAI), within a conversational role-playing game, in comparison with a more elaborate approach offering visual immersion, like that proposed by computer-based virtual reality.

To achieve these objectives, an initial exploratory qualitative study (Article 1) was undertaken to answer the following question: How does an interaction modality with virtual agents powered by GenAI, compared to a modality based on predefined choices, influence the user experience of the simulation, and to what extent do the cognitive scripts of these interaction modalities reflect real-world scenarios? In this first study, our findings indicate that voice interaction powered by GenAI, compared to a choice-based modality, could improve the user experience by increasing

engagement with the simulation interface and more faithfully reflecting real-world scenarios through more natural and intuitive interactions.

These results have guided our research question and field of study for the subsequent article (Article 2), formulated as follows: In what ways and to what extent do the voice interaction modality powered by GenAI and the fidelity of the simulation environment affect users' attitudes towards technology in contexts of complex problem-solving simulation-based learning? The conclusions presented in the second article first indicate that users consider voice artificial intelligence to be more immersive but also more cognitively demanding, notably due to technological issues (e.g., voice recognition) inherent in a developing technology. Our results confirm that the interaction modality and the degree of realism of the simulation environment have a considerable impact on the user experience in simulation-based learning contexts. The data analysis highlights the importance of a balance between emotional engagement, cognitive effort, and response accuracy to ensure an optimal user experience. In this context, it was observed that ChatGPT (OpenAI), thanks to its simple textual interactions and linguistic quality, has frequently surpassed virtual reality in terms of fluidity and reduction of cognitive load. This highlights the importance of interface simplicity compared to visual realism, as well as constant technological improvement to stimulate user engagement.

Keywords : Generative AI, LLM, ChatGPT, Simulation, VR, Virtual Patients/Agents, Education, Problem Solving, Multimodality, Voice

Research methods : Think-aloud, Interview, Surveys, Physiological measures

Table des matières

Liste des tableaux et des figures

Liste des tableaux

Chapitre 1 :

Chapitre 3 :

Liste des figures

Chapitre 2 :

Chapitre 3 :

Liste des abréviations

AFE: Automated Facial Expression Analysis

AI: Artificial Intelligence

CLT: Cognitive Load Theory

DSS: Decision Support Systems

EDA: Electrodermal Activity

GenAI: Generative Artificial Intelligence

HAII: Human-Artificial-Intelligence Interaction

HAT: Human-Artificial-Intelligence Teaming

HCAI: Human-Centered Artificial Intelligence

HCI: Human-Computer Interaction

ITS: Intelligent Tutoring Systems

LLM: Large Language Model

MET: Media Equation Theory

NLP: Natural Language Processing

PBL: Problem-based Learning

SBME: Simulation-Based Medical Education

UX: User Experience

VP: Virtual Patient

VR: Virtual Reality

Avant-propos

Ce mémoire en Expérience utilisateur en contexte d'affaires a été soumis avec l'autorisation de la direction administrative du programme de la Maîtrise ès Science en Gestion.

Le projet de recherche lié à ce mémoire a été approuvé par le comité d'éthique de la recherche (CER) de HEC Montréal. Deux articles issus du projet sont compris dans le mémoire avec le consentement des coauteurs.

Le premier article porte sur la portion qualitative du projet de recherche. Cet article en préparation pour soumission au JMIR Human Factors, explore l'intégration de la GenAI dans la formation par simulation en comparant la dynamique éducative d'une interaction vocale avec un agent virtuel alimentée par LLM à une interaction basée sur le choix dans un contexte de résolution de problèmes.

Le second article focalise le périmètre de l'étude en ajoutant des données quantitatives et physiologiques. Cet article est aussi en préparation pour soumission au JMIR Human Factors et étudie de plus près l'interaction avec un agent virtuel dans le contexte d'apprentissage par simulation d'un point de vue cognitif et émotionnel pour expliquer l'expérience utilisateur avec les systèmes GenAI, en VR et en jeu de rôle avec ChatGPT (OpenAI).

Remerciements

Je souhaite exprimer ma gratitude envers mes co-directeurs, Pierre-Majorique Léger et Constantinos K. Coursaris, pour leur soutien durant l'élaboration de ce mémoire ainsi que toutes les opportunités professionnalisantes offertes à travers ce parcours. Sans oublier Sylvain Sénécal, co-directeur du Tech3lab et chercheur principal sur ce projet, pour sa contribution et son appui dans toutes les étapes de l'étude.

Je suis également reconnaissant envers le Conseil de recherche en sciences naturelles et en génie (CRSNG) et Prompt pour leur appui financier dans ce projet (Subvention CRSNG-Alliance 1 (R2870B, R2870)).

Un remerciement particulier à Alexander John Karran, dont le mentorat et la passion pour la recherche ont été d'une valeur inestimable tout au long de mon parcours de maîtrise.

Un merci spécial à l'équipe du Tech3Lab, particulièrement Frédérique Bouvier, Élise, Xavier, Salyma ainsi qu'aux assistants de recherche pour leur aide précieuse lors du lancement de projet et la collecte des données.

Mes remerciements vont aussi à des collègues inestimables, dont l'accompagnement tout au long de ce parcours a représenté une contribution significative : Rémy, David, Emma, Andréa, Bella et tous les autres, pour leur contribution collective à l'enrichissement de cette aventure académique.

Enfin, je tiens à exprimer ma plus profonde gratitude à ma conjointe Anne-Cécile, ma mère Jeanne, mon frère Yves pour leur soutien infaillible et leurs encouragements au cours de ces deux années.

Papa, déjà plus de 10 ans… Je me questionne encore constamment sur le chemin parcouru et la direction que je prends aujourd'hui, souvent sans réponse claire. Mais une chose est certaine : tu serais fier de moi. Et c'est, au fond, tout ce qui compte vraiment ; être fier et continuer !

Chapitre 1

1.0 Introduction

L'intelligence artificielle (IA) change le paysage de notre société et de notre économie, et on prévoit qu'elle apportera des changements transformateurs qui affecteront tous les aspects de la vie, y compris l'emploi, la prise de décision et la concurrence mondiale, menant à un avenir où les machines intelligentes pourraient surpasser les capacités humaines dans beaucoup de domaines (Makridakis, 2017; Mannuru et al., 2023). L'IA vise à créer des systèmes capables d'exécuter des tâches nécessitant typiquement l'intelligence humaine, telles que l'apprentissage, le raisonnement et la résolution de problèmes (Goodfellow, Bengio and Courville, 2016). Cette intégration transforme déjà les processus, la prise de décision et favorise l'innovation (Ooi *et al.*, 2023), indiquant la capacité de l'IA à influencé et remodeler de manière significative les paysages professionnels et les méthodologies opérationnelles dans les années à venir.

Dans le domaine de la santé, les applications de l'IA couvrent divers champs médicaux, du diagnostic par imagerie et de l'interprétation du génome à l'apprentissage automatique pour la découverte de biomarqueurs, la prédiction des résultats cliniques et la chirurgie robotique autonome (Yu, Beam and Kohane, 2018). L'IA révolutionne en ce moment même le diagnostic, le traitement et la gestion des maladies, améliorant considérablement les résultats des soins grâce à des diagnostics plus précis et à une médecine personnalisée (Briganti and Le Moine, 2020; Alowais *et al.*, 2023). Les algorithmes avancés d'apprentissage automatique et l'analyse de données étendue ont le potentiel d'améliorer la prestation des soins de santé et l'allocation des ressources (Navath, 2021), améliorant ainsi les résultats pour les patients.

L'impact de l'IA sur l'éducation, en particulier l'éducation en soins de santé, est tout aussi transformateur. L'IA augmente les opportunités d'apprentissage en offrant des retours personnalisés, des systèmes de tutorat adaptatifs et une assistance personnalisée pour les éducateurs et les apprenants (Chen, Chen and Lin, 2020). L'IA soutient l'apprentissage collaboratif, surveille les forums étudiants, facilite l'évaluation continue, fournit des compagnons d'apprentissage intelligents pour les étudiants et sert d'outil de recherche pour faire avancer les sciences de l'apprentissage (Holmes, Bialik and Fadel, 2023). La flexibilité de l'intelligence artificielle générative (GenAI) est reconnue comme un outil précieux pour créer des environnements d'apprentissage dynamiques et interactifs adaptés aux divers besoins éducatifs. Des recherches montrent que la GenAI peut améliorer considérablement la performance des étudiants grâce à une formation individualisée, conduisant à de meilleurs résultats d'apprentissage (Ogunleye *et al.*, 2024; Sauder *et al.*, 2024).

Le traitement du langage naturel (NLP) permet aux systèmes IA de comprendre et de générer du texte ou des discours semblables à ceux des humains (Hirschberg and Manning, 2015). Les modèles de langage de grande envergure (LLM), une sous-catégorie des technologies NLP, sont prêts à révolutionner les lieux de travail et les systèmes éducatifs au-delà des simples modifications des flux de travail (AL-Smadi, 2023; Lu, 2023; Chui, Roberts and Yee, no date). Les LLM améliorent l'éducation médicale avec des scénarios cliniques précis et adaptables, améliorent la communication avec les patients (Sallam, 2023) et facilitent l'apprentissage personnalisé et autonome grâce à des réponses immédiates aux questions, des notes d'étude personnalisées ou la génération de questions de test (Tam et al., 2023). Adopter une approche centrée sur l'humain est cependant crucial pour combler le fossé entre la technologie et l'éducation (Kasneci et al., 2023).

La simulation et l'apprentissage par problèmes (PBL) ont des applications étendues dans les domaines éducatifs et professionnels. La simulation, intégrant diverses modalités et des systèmes de classification variés, soutient l'apprentissage expérientiel (Pilote and Chiniara, 2019). Elle est bénéfique dans de mutiples domaines tels que l'aéronautique (Vora *et al.*, 2002; Stone, Panfilov and Shukshunov, 2011), militaire (Bailey *et al.*, 2017) et la vente au détail (Boletsis and Karahasanovic, 2020). Les agents virtuels, des personnages prenant vie dans les simulations, inspirent les apprenants, aident dans les tâches cognitives, offrent des retours et évaluent les apprenants (Ke *et al.*, 2020). Dans l'enseignement des soins infirmiers, la simulation en réalité virtuelle sur ordinateur offre des expériences de formation immersives et réalistes, améliorant les compétences pratiques grâce à des rencontres avec des patients virtuels (Kononowicz *et al.*, 2019). La réalité virtuelle transforme l'éducation médicale en offrant des expériences de formation clinique standardisées et économiques (Pottle, 2019).

L'intégration de la GenAI avec les simulations de patients virtuels dans l'éducation médicale représente une avancée significative encore très peu explorée à ce jour. Les patients virtuels, en particulier grâce aux scénarios interactifs et à la technologie de réalité virtuelle, offrent une expérience éducative immersive et personnalisée, améliorant l'éducation des praticiens de la santé et des patients (Cook, Erwin and Triola, 2010; Cook *et al.*, 2018; Kononowicz *et al.*, 2019; Adeghe, Okolo and Ojeyinka, 2024). La simulation enrichit l'apprentissage par problèmes en permettant aux apprenants d'appliquer des compétences dans des scénarios réalistes, leur permettant de s'adapter aux résultats de leurs décisions (Léger *et al.*, 2012). Le PBL, initialement développé pour l'éducation médicale, implique le travail en équipe et permet aux étudiants d'aborder des défis réels et multifacettes (Savery and Duffy, 1995; Walker and Leary, 2009; Savery, 2015). Les simulations bien conçues facilitent l'acquisition rapide de compétences et une compréhension profonde des scénarios complexes (O'Neil, Wainess and Baker, 2005). Cependant, les simulateurs à haute fidélité, malgré leur amélioration des performances, ne garantissent pas toujours un meilleur transfert d'apprentissage par rapport aux simulateurs à faible fidélité (Norman, Dore and Grierson, 2012a). Également, la mise en œuvre de LLM avancés comme ChatGPT (OpenAI, 2023) dans les pratiques éducatives présente à la fois autant d'opportunités que de défis (Kasneci *et al.*, 2023).

Les avancées rapides de l'IA et le paysage évolutif de l'éducation médicale soulignent la nécessité de trouver un équilibre entre réalisme et valeur éducative. Une focalisation excessive sur le réalisme peut nuire à l'efficacité optimale des expériences d'apprentissage (MacLean *et al.*, 2017, 2019; Massoth *et al.*, 2019; Boscardin *et al.*, 2024). La simulation virtuelle clinique peut améliorer la rétention des connaissances, le raisonnement clinique, la satisfaction de l'apprentissage et l'auto-efficacité dans l'éducation des soins de santé, mais son efficacité globale reste sous-recherchée (Padilha *et al.*, 2019).

De plus, la nouveauté de la recherche sur l'interaction humain-IA (HAII) souligne l'importance d'évaluer et de valider soigneusement la technologie IA dans les contextes éducatifs (Følstad *et al.*, 2021; Sallam, 2023; Cain, 2024). Notre étude vise à combler cet écart critique en examinant l'impact des interactions vocales améliorées par la GenAI et les variations dans la qualité de l'environnement de simulation sur l'expérience éducative des apprenants, en particulier dans l'éducation des soins de santé. En explorant tant la dimension cognitive qu'émotionnelle, de manière qualitative comme quantitative, nous cherchons à informer la recherche sur la manière dont ces technologies influencent l'expérience utilisateur des apprenants. L'étude de l'expérience utilisateur est une méthode efficace pour obtenir des informations précieuses et riches dans l'utilisation d'un artéfact hautement complexe et contextuel (i.e. la simulation de diagnostic clinique à l'aide d'un patient virtuel). Cette recherche vise à enrichir les domaines de l'IA centrée sur l'humain (HCAI) et de l'HAII, en comblant le fossé entre les apprenants et la technologie et en améliorant l'intégration de l'IA dans les simulations éducatives. Afin d'atteindre cet objectif, nous établissons les questions suivantes :

1.1 Questions de recherche

Article 1 :
Comment une modalité d'interaction avec des agents virtuels alimentée par GenAI, comparée à une modalité basée sur des choix pré-inscrits, influence-t-elle l'expérience utilisateur de la simulation, et dans quelle mesure les scripts cognitifs de ces modalités d'interaction reflètent-ils les scénarios du monde réel ?

Article 2 :
De quelles manières et dans quelle mesure la modalité d'interaction vocale alimentée par l'IA générative (GenAI) et la fidélité de l'environnement de simulation affectent-elles les attitudes des utilisateurs envers la technologie dans des contextes d'apprentissage par simulation de résolution de problèmes complexes ?

1.2 Objectifs et contributions

Avec ces questions de recherche, nous souhaitons explorer l'intégration de l'intelligence artificielle générative (GenAI) dans les simulations éducatives en comparant deux modalités d'interaction : l'interaction vocale alimentée par un Large Language Model (LLM) et l'interaction basée sur des choix. La première étude examine comment ces interactions influencent l'expérience utilisateur dans des contextes de simulation de résolution de problèmes.

Dans le second article, elle étend l'analyse des interactions vocales avec des agents virtuels propulsés par GenAI, en examinant l'impact de la fidélité de l'environnement de simulation sur les attitudes des utilisateurs envers la technologie. Elle compare des simulations basées sur le dialogue vocal avec des systèmes comme ChatGPT et des environnements de réalité virtuelle plus immersifs.

Ces études contribuent à une meilleure compréhension de la manière dont GenAI, en particulier les modèles de langage, peut améliorer les simulations éducatives et l'expérience utilisateur. Elles introduisent une approche comparative pour évaluer comment différentes modalités d'interaction (voix vs choix) influencent l'engagement utilisateur, la présence sociale et l'effort cognitif. La recherche aide à évaluer comment la fidélité des environnements de simulation et la modalité d'interaction affectent l'attitude des utilisateurs envers les technologies éducatives et leur adoption. Ces articles comblent une lacune importante dans la littérature sur l'Interaction Homme-AI (HAII) au sein des environnements d'apprentissage immersifs, en particulier dans l'éducation en soins de santé.

1.3 Construits clés

System Response Accuracy (SRA): *How accurately the system perceives and responds to user inputs. / La précision avec laquelle le système perçoit et répond aux entrées de l'utilisateur.*

Naturalness (NAT): *The degree to which the interaction feels natural and similar to human-human communication. / Le degré auquel l'interaction semble naturelle et similaire à une communication entre humains.*

Social Presence (SP): *The sense of sociability, warmth, and personal interaction experienced. / Le sentiment de sociabilité, de chaleur et d'interaction personnelle ressenti.*

Cognitive Effort (CE): *The mental effort required to complete tasks during the interaction with the system. / L'effort mental nécessaire pour accomplir des tâches lors de l'interaction avec le système.*

Arousal (ARO): *The level of stimulation or activation experienced by the user during the interaction. / Le niveau de stimulation ou d'activation ressenti par l'utilisateur durant l'interaction.*

Valence (VAL): *The emotional positivity or negativity experienced by the user during the interaction. / La positivité ou la négativité émotionnelle ressentie par l'utilisateur pendant l'interaction.*

Attitude Toward Technology (ATT): *The overall user evaluation of and willingness to engage with the technology, influenced by cognitive and emotional factors. / L'évaluation globale de l'utilisateur à l'égard de la technologie et sa volonté de l'utiliser, influencée par des facteurs cognitifs et émotionnels.*

1.4 Information au sujet des articles

Présentation sommaire du premier article

Cette étude examine l'intégration de l'IA générative (GenAI) dans les simulations VR à des fins éducatives, en se concentrant sur l'utilisation des grands modèles de langage (LLM) dans la formation infirmière. Elle compare l'interaction vocale propulsée par un LLM à celle d'un agent virtuel basé sur des choix, afin d'évaluer leurs effets sur l'expérience utilisateur, l'engagement et l'apprentissage. L'étude adopte une approche exploratoire et qualitative, où des infirmières ont navigué dans des scénarios cliniques en VR utilisant ces deux modalités d'interaction. Les résultats révèlent une préférence pour les interactions vocales, jugées plus immersives et proches des rencontres cliniques réelles. Cette modalité a été perçue comme offrant une meilleure expérience éducative et un engagement accru. Cependant, l'étude identifie aussi des défis techniques liés à l'intégration des LLMs dans les simulations, soulignant l'importance d'une conception centrée sur l'utilisateur. Les conclusions encouragent à poursuivre les recherches pour optimiser l'usage des LLMs dans les environnements éducatifs et améliorer les outils d'apprentissage basés sur l'IA.

Présentation sommaire du deuxième article

Cet article, en continuité du premier, traite de l'intégration de l'intelligence artificielle générative (GenAI) dans les simulations éducatives basées sur des patients virtuels, notamment dans le domaine de la formation médicale. Il explore en détail l'impact des interactions vocales basées sur l'IA et des niveaux de fidélité des environnements de simulation sur l'attitude des utilisateurs envers la technologie dans des contextes d'apprentissage. Le contexte repose sur la montée en puissance de l'IA dans divers secteurs, y compris la santé et l'éducation, où elle améliore les résultats en matière de diagnostic et personnalisation des soins, et offre des opportunités d’apprentissage personnalisé. En utilisant des théories comme l'Équation Médiatique et la Théorie de la Charge Cognitive, l'étude analyse comment les utilisateurs perçoivent la précision des réponses du système, la naturalité de l’interaction, et la présence sociale dans différents contextes technologiques (réalité virtuelle vs ChatGPT). Les résultats montrent que les interactions vocales augmentent la perception de la présence sociale, mais peuvent aussi augmenter l’effort cognitif en raison de limitations techniques. Par ailleurs, l'environnement

ChatGPT, malgré sa simplicité, a offert une expérience plus naturelle que la réalité virtuelle complexe, soulignant que la fluidité de l'interaction est essentielle pour renforcer l’engagement émotionnel. En conclusion, cet article souligne que la conception des outils d'apprentissage basés sur l'IA doit aller au-delà d'une approche unique pour tous, en tenant compte des besoins cognitifs et émotionnels des utilisateurs, à des stades données, afin de créer des environnements d'apprentissage plus efficaces et engageants.

1.5 Contribution et responsabilités individuelles

Ce mémoire a été réalisé en collaboration avec mes codirecteurs; ainsi, le tableau 1 ci-dessous illustre mes contributions intellectuelles individuelles à chaque élément de la thèse. Les exigences de la collaboration stipulent que l'étudiant doit atteindre un niveau global de contribution de 50 %. Dans les aspects où ma contribution dépasse 50 %, cela reflète mon leadership et ma responsabilité pour la phase respective.

Table 1
Contribution dans le projet de recherche et rédaction d'article

Revue de littérature	**90%** -L'étudiant a procédé à la revue de littérature de manière autonome; Les intervenants du projet ont pu suggérer des avenues ainsi que des articles spécifiques.
Définition de la question de recherche	**70%-**L'étudiant a défini par lui-même les questions de recherches; Les intervenants du projet, au fil des rencontres, ont aidé à définir plus précisément et enrichir ces questions.
Conception du design expérimental	**70%-**L'étudiant a défini le design expérimental; Les intervenants du projet ont apporté des conseils sur la faisabilité et l'efficacité des méthodes proposées.
Recrutement des participants	**100%-**L'étudiant a effectué le processus de recrutement, de la sollicitation jusqu'au suivi de participation.
Pré-tests et collecte de données	**75%**-L'étudiant a participé aux pré-tests et effectué toutes les collectes de données; Les intervenants du projet ont assisté lors d'absence obligatoires.
Extraction et transformation des données	**50%-**L'étudiant était en constante communication avec l'équipe du laboratoire afin de communiquer les besoins spécifiques au projet en plus de faire l'extraction et la transformation entière des données qualitatives; Le personnel du laboratoire a procédé à la portion technique d'extraction et de transformation relative aux différents outils physiologiques.
Analyse des données	**50% -**L'étudiant a ciblé et indiqué les données pertinentes à l'analyse notamment à travers un dictionnaire d'analyse en plus de procéder à l'entièreté de l'analyse qualitative et leur interprétation; Les méthodes statistiques ont été révisées et les calculs statistiques effectués par l'équipe du laboratoire.
Rédaction des articles	**95%** -L'étudiant a rédigé l'entièreté de ce mémoire; Les intervenantes du projets ont pû apporter des recommandations et des corrections.
Préparation et présentation des résultats au partenaire industriel impliqué dans le projet	**80% -**L'étudiant a préparé et présenter la synthèse des résultats de ce projet au partenaire industriel; Les intervenants ont complété et bonifié la documentation pour ajuster selon les standards de livraison du laboratoire.

1.6 Structure du mémoire

La structure du mémoire se présentera de la manière suivante :

Le deuxième chapitre porte sur le premier article du mémoire visant à répondre à la question de recherche suivante : *Comment une modalité d'interaction avec des agents virtuels alimentée par GenAI, comparée à une modalité basée sur des choix pré-inscrits, influence-t-elle l'expérience utilisateurs de la simulation, et dans quelle mesure les scripts cognitifs de ces modalités d'interaction reflètent-ils les scénarios du monde réel?*

Le troisième chapitre du mémoire présente le deuxième article et tente de répondre à la question de recherche suivante : *De quelle manière et dans quelle mesure la modalité de dialogue d'interaction vocale alimentée par GenAI influence-t-elle l'attitude des utilisateurs à l'égard de la technologie dans un contexte d'apprentissage de simulation de résolution de problèmes ?*

Finalement, le quatrième chapitre, la conclusion de ce mémoire, revient sur les résultats obtenus dans les articles pour en faire une synthèse.

La première phase, la « découverte », consistait en un examen méthodique de l'outil de simulation par le biais d'essais itératifs menés par l'équipe de recherche, qui s'est immergée dans sa dynamique opérationnelle. Cette exploration méticuleuse avait pour but de cartographier de manière approfondie les fonctionnalités étendues de l'outil, d'évaluer ses capacités, et de repérer les limites spécifiques et générales, notamment celles liées aux aspects de l'outil encore en cours de développement. Cette exploration approfondie était essentielle pour élaborer un design expérimental robuste et crédible, garantissant qu'il reflète de manière réaliste l'usage réel tout en exploitant le plein potentiel de l'outil de simulation. Cette étape fondamentale a permis d'établir une base solide pour les phases suivantes de la recherche.

1.7 Approche

Après avoir posé les bases lors de la phase de découverte, la phase suivante — intitulée « développement et investigation » — s'est concentrée sur la construction d'un cadre expérimental représentant authentiquement l'application de l'outil de simulation, intégrant les processus de collecte de données et d'analyse. Nous avons soigneusement orchestré cette phase afin de comprendre les opérations cognitives, les approches tactiques, et la dynamique générale des participants dans l'utilisation des deux modalités de simulation. Cela a été réalisé en utilisant des protocoles de « think-aloud » avec des professionnels expérimentés, enrichis par des entretiens structurés.

Au cœur de cette phase réside notre engagement à évaluer l'impact des différentes modalités d'interaction avec les agents virtuels sur l'engagement des utilisateurs avec l'interface de simulation. Notre objectif était de critiquer la manière dont les cadres cognitifs de ces interactions reflètent les scénarios réels et d'explorer les potentialités offertes par ces technologies de pointe. Par conséquent, la question de recherche centrale de notre étude émerge ainsi :

Dans quelle mesure une modalité d'interaction avec des agents virtuels alimentés par GenAI, comparée à une modalité basée sur des choix préenregistrés, affecte-t-elle l'expérience utilisateur de la simulation, et dans quelle mesure les scripts cognitifs de ces modalités d'interaction reflètent-ils des scénarios réels ?

En nous appuyant sur les principes de la théorie de l'adéquation tâche-technologie (Goodhue et Thompson, 1995), nous postulons que les participants experts devraient percevoir les interactions vocales comme un enrichissement significatif de la pertinence de la simulation. La théorie de l'adéquation tâche-technologie est un concept issu des domaines de la gestion et des sciences de l'information. Elle vise à expliquer l'impact de la concordance entre les fonctionnalités d'une technologie et les besoins d'une tâche sur la performance des utilisateurs (Goodhue et Thompson, 1995). Cette théorie vise à évaluer si une technologie spécifique est appropriée pour faciliter la réalisation efficace et efficiente des tâches par les utilisateurs.

Cette perception découle d'une intégration harmonieuse entre les exigences de la tâche et les capacités technologiques à disposition. L’essence de la tâche — la communication avec le patient — se prête intrinsèquement aux échanges oraux, soulignant le rôle essentiel des interactions verbales authentiques et immédiates dans la reproduction efficace des rencontres cliniques réelles. Par conséquent, nous croyons que cette expérience plus réaliste pourrait favoriser une plus grande acceptation et adoption de l'outil d'apprentissage.

Cette étude a vu les participants s'immerger dans deux scénarios de simulation de résolution de problèmes uniques, chacun avec un agent virtuel, comportant les deux modalités d'interaction distinctes. Les données qualitatives issues de ces phases ont non seulement jeté les bases d'une compréhension globale de l'UX dans ce contexte spécifique, mais ont également suscité des questions de recherche critiques pour la phase suivante de la recherche.

Chapitre 2

Integrating Generative AI in Simulation Training: Comparing the Educational Dynamics of LLM-Powered Vocal With Choice-Based Conversational Agent Interactions in Role-Play Problem-Based Learning Context.

Authors: Nico Helie, Alexander J. Karran, Shang Lin Chen, Constantinos K. Coursaris, Pierre-Majorique Léger.

HEC Montréal, 3000 Chem. de la Côte-Sainte-Catherine, Montréal, QC H3T 2A7

2.0 Abstract

This study investigates the integration of Generative Artificial Intelligence (GenAI) into desktop virtual reality (VR) simulations for educational purposes. Focusing specifically on Large Language Models (LLMs), it primarily addresses the domain of nursing training, using it as a case study to explore the broader application of GenAI within problem-based learning (PBL) in VR. Amidst the digital transformation in workplaces and educational domains, this research investigates how an LLM-powered vocal interaction modality compares to a choice-based virtual agent one, aiming to understand their impacts on the user experience (UX), engagement and learning. Utilizing an exploratory and qualitative research design, the study engaged expert nurses in an intra-participant desktop VR experiment featuring both virtual agent interaction modalities. Participants navigated through two clinically oriented scenarios, enabling a comprehensive analysis of interaction dynamics. The methodology included think-aloud, post-experiment interviews, and thematic analysis to extract nuanced insights into the comparative effectiveness of the interaction modalities. Results highlighted a preference for vocal interactions powered by LLMs, attributing to a more immersive and realistic simulation experience. Participants reported enhanced engagement and satisfaction, suggesting that vocal interactions better mimic real-life clinical encounters, thus potentially improving educational relevance. However, the study revealed potentially beneficial avenues for facilitating modality with choices as technological and interface design challenges affecting the seamless integration of LLMs in educational VR simulations. The findings underscore the importance of human-centric design principles in developing artificial intelligence (AI)-enhanced educational

tools and advocate for further research into optimizing LLM integration for educational purposes. By bridging the gap between technology and pedagogy, this study contributes to the ongoing discourse on enhancing human-centered artificial intelligence (HCAI) and human-AI teaming (HAT) within simulation-based learning environments, offering valuable insights for educators, developers, and policymakers aiming to leverage AI in education.

2.1 Introduction

2.1.1 Impact of AI Across Professional Fields and Educational Domains

AI represents a multifaceted domain within computer science aimed at creating systems capable of performing tasks that typically require human intelligence, including learning, reasoning, problem-solving, and adapting to new or changing environments (Goodfellow, Bengio and Courville, 2016). It currently opens up extensive opportunities across numerous sectors, including but not limited to marketing, human resources management, banking, retail, and manufacturing, where it is significantly transforming processes, decision-making and fostering innovation (Ooi et al., 2023). This widespread integration underscores AI's capacity to significantly impact and transform professional landscapes and operational methodologies in the next years.

Healthcare and education are sectors where AI exhibits significant potential. AI recent technological breakthroughs already have applications in various medical fields, such as image-based diagnosis, genome interpretation, machine learning for biomarker discovery, clinical outcome prediction and autonomous robotic surgery (Yu, Beam and Kohane, 2018). In the healthcare practice, it is revolutionizing disease diagnosis, treatment and management, greatly improving care outcomes through more accurate diagnoses and the provision of personalized medicine (Alowais et al., 2023).

AI significantly augments learning opportunities in education by offering personalized feedback, adaptive tutoring systems, and customized assistance for educators and learners (Chen, Chen and Lin, 2020b). It offers a wide range of utilities, including supporting collaborative learning, monitoring student forums, facilitating continuous assessment, providing AI learning companions for students, offering AI teaching assistants for teachers, and serving as a research tool to advance the learning sciences (Holmes, Bialik and Fadel, 2023). In a diverse and multifaceted manner, AI is incrementally permeating the domains of professional and educational environments, exhibiting a range of forms, intensities and velocities.

2.1.2 LLMs in the workplace and education

Natural language processing (NLP) allows AI systems to understand and generate human-like text or speech (Hirschberg and Manning, 2015). AI-powered LLMs, a subset of NLP technologies, are one of the tools predicted to revolutionize workfields and educational systems beyond mere workflow alterations (AL-Smadi, 2023; Lu, 2023; Chui, Roberts and Yee, no date). This shift downstream is expected to redefine learning and teaching, bringing opportunities and challenges with implementing advanced LLMs like ChatGPT in educational practices (Kasneci et al., 2023).

LLMs such as ChatGPT can offer numerous advantages, such as enhancing medical education with precise and adaptable clinical scenarios and improving communication with patients (Sallam, 2023). They facilitate personalized, self-paced learning through immediate responses to queries, custom study notes or test question generation (Tam et al., 2023). LLMs have the potential to enhance the whole educational system experiences by generating content, acting as teaching assistants, and aiding in research and problem-solving. However, embracing a human-centered AI approach is crucial to bridging the actual gap between technology and education. (Kasneci et al., 2023).

2.1.3 Shift in Human-Computer Interaction and Human Factors

The study of human-computer interaction (HCI) has historically revolved around a model where computing systems act mainly as support tools, leading to rule-bound interactions within a conventional stimulus-response model (Farooq and Grudin, 2016; Wickens et al., 2021). The advent of AI marks a significant paradigm shift in HCI, transitioning from traditional, task-focused interactions to more intuitive, human-centric designs. Human factors within the field of HCI are essential considerations that center on the design and the dynamic between individuals and technology. Synonymous with 'ergonomics' and 'human factors engineering', it encompasses the examination of human interactions with technology, focusing on creating systems that are efficient, effective, and user-satisfying (Karat and Karat, 2003).

As the field of HCI evolves, so does the definition of human factors used to address emerging challenges. From an HCAI perspective, human factors now entail crafting systems that enhance and augment human capabilities through prioritizing control, reliability and trust while focusing on optimizing human performance and satisfaction, ensuring user-centered design, and considering the broader societal impacts (Capel and Brereton, 2023). Despite the lack of consensus in academic communities on AI's potential to act as a teammate in human collaboration, the advent of the AI era introduces a hybrid relationship wherein AI systems might simultaneously function as assistive tools and collaborative partners with humans (Xu and Gao, 2024). While this transition offers a myriad of potential advantages, it also presents HCI professionals with unprecedented challenges, distinguishing it markedly from the development of traditional non-AI computing systems (Xu et al., 2023).

2.1.4 Simulation and problem-based learning

Given its extensive application across educational fields and its versatility in supporting various objectives in the workplace, simulation manifests today in diverse forms. This educational approach now incorporates various modalities focusing on experiential learning, with flexible terminology and classification systems accommodating diverse simulation types (Pilote and Chiniara, 2019). Simulation has actual applications and can be beneficial in various fields such as aeronautics (Vora et al., 2002; Stone, Panfilov and Shukshunov, 2011), military (Bailey et al., 2017) or retail (Boletsis and Karahasanovic, 2020).

Featured in some of these simulations, virtual agents, are defined as simulated life-like characters that interact with humans to facilitate learning (Mascarenhas et al., 2018). Among various roles, they can be utilized to inspire learners, evoke situated interests, aid in cognitive and metacognitive learning tasks, offer feedback for decision-making, and evaluate learners (Ke et al., 2020). In nursing education, VR desktop simulation provides immersive and highly realistic training experiences for nursing students through, among others, virtual patient (VP) encounters in problem-based learning, enhancing practical experiences (Kononowicz et al., 2019). VP simulations are interactive computerized representations of clinical scenarios where learners assume the role of healthcare providers, navigating tasks such as gathering clinical data, making differential diagnoses, and managing and monitoring patients, aimed at enhancing health professionals' clinical reasoning skills (Kononowicz et al., 2019). VR is transforming medical education, not only offering immersive but also cost-effective and standardized clinical training experiences that enhance conventional teaching methods (Pottle, 2019).

Simulation enriches problem-based learning (PBL) by allowing learners to apply their skills in realistic scenarios, enabling them to experience and adapt to the outcomes of their decisions safely (Léger et al., 2012). Originally developed for medical education, this approach has been adopted across various fields and educational levels (Savery and Duffy, 1995; Savery, 2015). PBL is a learning approach that gives students more control over their education than traditional methods, typically involving teamwork (Walker and Leary, 2009). As Walker and Leary (2009) state, PBL, at its core, enables students to acquire new knowledge by addressing real-world, multifaceted, and cross-disciplinary challenges similar to what professionals face in their careers. Several authors suggest that well-crafted simulations typically facilitate quick skill acquisition and a more profound comprehension of complex scenarios (O'Neil, Wainess and Baker, 2005; The Effectiveness of Games for Educational Purposes: A Review of Recent Research - Josephine M. Randel, Barbara A. Morris, C. Douglas Wetzel, Betty V. Whitehill, 1992, no date).

Simulation is not merely a technological innovation; it's a technique to substitute or enhance real-life scenarios with interactive and guided learning experiences that replicate essential aspects of a real situation (Gaba, 2004). Simply put, simulation entails an activity mimicking actions or processes (Tun et al., 2015), a practice likely originating from our ancestors pretending to hunt dangerous animals on immobile trees (Baily, 2019). It underscores how nearly all forms of training, regardless of their technological sophistication, can involve simulated experiences. Hence, simulation transcends its role as a technological marvel to become a foundational technique in modern education and professional training, offering immersive experiences that mimic real-life scenarios across various fields. While high-fidelity simulators can enhance performance, the substantial costs involved may not guarantee a markedly better learning transfer compared to using low-fidelity simulators (Norman, Dore and Grierson, 2012b).

Thus, the exploration of technological advancements in simulation-based education, including the integration of GenAI-powered LLM conversational agents, raises critical questions regarding the effectiveness and efficiency of these tools in enhancing learning outcomes. As we delve into the intricate dynamics between the potential of GenAI to enhance simulation fidelity and the potential associated costs of revolutionizing educational simulations, there emerges a need to empirically assess their impact on learners' experiences.

Moreover, to the best of our understanding, existing literature lacks a qualitative, focused and comparative examination of interaction modalities, specifically contrasting choice-based interactions with those facilitated by an LLM-powered agent in a desktop VR simulation learning environment. Understanding these differences is crucial for designing more effective educational tools, improving user engagement, and tailoring learning experiences to meet educational goals and learners' needs better. Taking this into account, our study seeks to delve into and enrich the expanding fields of HCAI and HAT by analyzing the UX of these functionalities. Our goal is to narrow the divide between learners and technology, providing valuable insights for enhancing the incorporation and application of AI in crafting more lifelike virtual agents interactions within simulations.

2.2 Approach

The research presented in this manuscript was completed in two phases: 1) discovery and 2) development and investigation.

The first phase, 'discovery', consisted in a methodical examination of the simulation tool through iterative trials by the research team, immersing themselves in its operational dynamics. This meticulous exploration was designed to thoroughly map out the tool's extensive features, assess its capabilities, and pinpoint both specific and general limitations, especially those associated with aspects of the tool that are still in the process of development. Such a thorough exploration was essential for developing a robust and credible experimental design, ensuring that it realistically mirrors real-life usage and leverages the simulation tool's full potential. This foundational step was critical in establishing a solid groundwork for the subsequent phases of the research.

Following the groundwork laid in the discovery phase, the subsequent stage—termed 'development and investigation'—focused on constructing an experimental framework that authentically represents the application of the simulation tool, encompassing data gathering and analytical processes. We carefully orchestrated this phase to understand the cognitive operations, tactical approaches, and overall participant dynamics in the use of both simulation modalities. This was achieved by employing think-aloud protocols with seasoned professionals and enhancing these insights with structured interviews.

Central to this phase is our commitment to assessing the impact of differing virtual agent interaction modalities on user engagement with the simulation interface. Our aim was to critically evaluate how closely these interactions' cognitive frameworks replicate actual scenarios and to explore the potentialities afforded by these cutting-edge technologies. Consequently, the pivotal research question driving our study emerges:

How does an interaction modality with virtual agents powered by GenAI, compared to a modality based on pre-registered choices, affect the user experience of the simulation, and to what extent do the cognitive scripts of these interaction modalities reflect real-world scenarios?

Building on the principles of Task-Technology Fit Theory (Goodhue and Thompson, 1995), we posit that expert participants should perceive voice interactions as a significant enhancement of the simulation's relevance. Task-Technology Fit Theory is a concept rooted in the fields of management and information sciences. It aims to explain the impact of the alignment between the features of a technology and the needs of a task on user performance (Goodhue and Thompson, 1995). This theory seeks to assess whether a specific technology is suitable for enabling users to perform tasks effectively and efficiently. This perception stems from a harmonious integration of task demands with the technological capabilities at hand. The essence of the task—patient communication—is inherently suited to oral exchanges, emphasizing the critical role of authentic and immediate verbal interactions in replicating real-world clinical encounters effectively. Consequently, we believe that this more realistic experience could likely promote greater acceptance and adoption of the learning tool.

This study saw participants immersed in two unique problem-solving simulation scenarios, each with a virtual agent, featuring the two different interaction modalities. The qualitative data derived from these phases not only laid the groundwork for a comprehensive understanding of UX within this specific context, but also prompted critical research questions for the subsequent phase of the research.

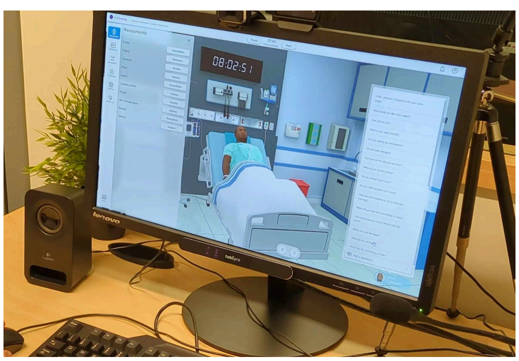
2.3 Methodology

2.3.1 Experimental design & manipulation

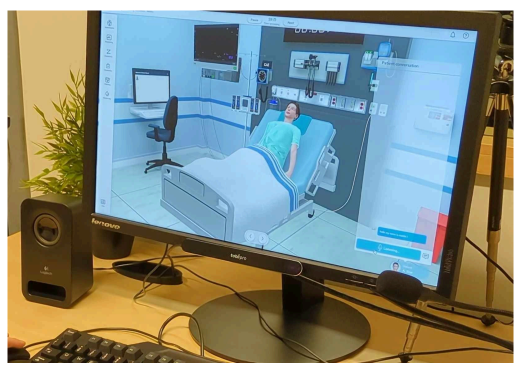
To answer the above-mentioned research question, a two-factor within-subjects lab experiment was conducted. In this specific experiment, we manipulated the mode of interaction of the conversational agent used to interact with the virtual patient. From one task to the next, the participant interacted with the agent in two different ways :

Choice-based

In the choice-based task, learners interacted with a conversational agent using a predefined list of interactions crafted by medical scenario experts. These options, varying in pertinence for the scenario, allowed learners to select responses based on their judgment and scenario phase. Selections triggered text messages through the dialogue system, eliciting immediate textual responses from the virtual patient. The interaction list is dynamically updated based on scenario progression and past choices.



Embody Clinical Simulation Choice-based modality interaction panel (Right of the screen)



Embody Clinical Simulation Voice-based modality interaction panel (Right of the screen)

Figure 1

Interaction modes

Voice-based

In the voice-based task, the learner was required to interact with the conversational agent using his or her voice. For each interaction, the user had to press and hold a button activating the microphone while vocalizing the desired utterance. A speech recognition system synthesized the intention and fed it back into the dialogue function. According to the AI analysis, a consistent response was then returned to the user in the form of text accompanied by an audio output via a text-to-speech system.

2.3.2 Participants

Three expert nurse participants took part in this first stage (Age: 32, +-18, 2 female). The criteria for defining 'expertise' were adapted from the established literature (Benner, 1984). These participants collectively held an average of 4 years of diverse experience in various areas of work roles, with two currently pursuing further education—one at the collegiate level and another at the university level—while concurrently working in the professional field.

The specialized nature of this study limited the availability of expert nurse participants, making a large sample size challenging. However, the carefully selected participants—expert nurses with diverse professional backgrounds—offer valuable qualitative insights that reflect the broader nursing community's practices. In conducting this research, we were mindful of the saturation principle in qualitative data, determining that the inclusion of three expert participants was sufficient to reach saturation and ensure comprehensive coverage of the study's themes (Saunders *et al.*, 2018). This focused approach, prioritizing depth and quality of data over quantity, ensures that the findings are rich and highly relevant.

Participants were recruited through a targeted approach involving specialized groups on social media platforms and through a collaborative effort with a university-affiliated establishment, which circulated a recruitment notice to the population of interest. Prospective participants were required to register their interest through an online Qualtrics form. Following this, eligible candidates matching the selection criteria were identified and subsequently contacted to arrange their participation in the study.

Participants eligible for the study were required to meet specific criteria: (1) be at least 18 years old, (2) work in nursing or be a nursing student, (3) hold OIIQ accreditation, (4) possess a proficient understanding of both English and French, spoken and written. The experiment was conducted in a state-of-the-art UX laboratory. Its duration ranged from one to one and a quarter hours. Prior to commencing the experiment, participants were briefed about the study, asked to sign an informed consent form and informed of their rights to withdraw from the study at any point. Participants were compensated $20. Our institution's ethics board approved the study under certificate no. 2023-5393.

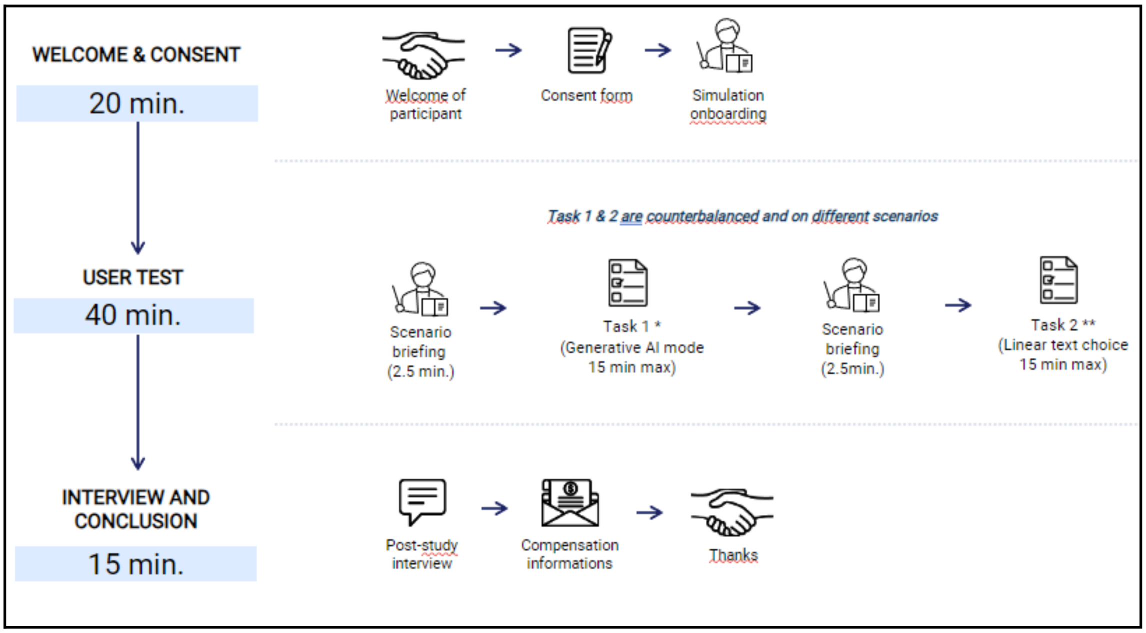


Figure 2

Experimental procedure

2.3.3 Experimental procedure

The experimental procedure is shown in Fig. 1. Participants were first welcomed and then given an outline of the experiment. They were then invited to complete the consent form. Before starting with experimental tasks, participants consulted a tutorial related to using the simulation environment.

In order to maximize the time and quality of interaction with the interface during the task, a scenario brief was provided to the participant prior to the task (See Appendix). The brief contained a summary of the case, the simulation objective and the patient's medical record. In both cases, the participant was asked to carry out the clinical assessment of the patient, as he or she would in real life, without skipping any steps. It was also made clear to the participants that time was not a measure of this experience to ensure the proper execution of the clinical assessment process.

For the onboarding tutorial (Figure 2), the participant was asked to complete a software onboarding tutorial, consisting of 13 steps covering the main functions of the simulation tool. The aim of this onboarding task was to reduce the difficulty of using this complex tool to improve the flow of the scenarios. Prior research shows that pre-briefing participants enhances familiarity with both the scenario and the simulation environment, eliminating the uncertainty associated with an unfamiliar setting (Fraser, Ayres and Sweller, 2015; Tyerman *et al.*, 2019).

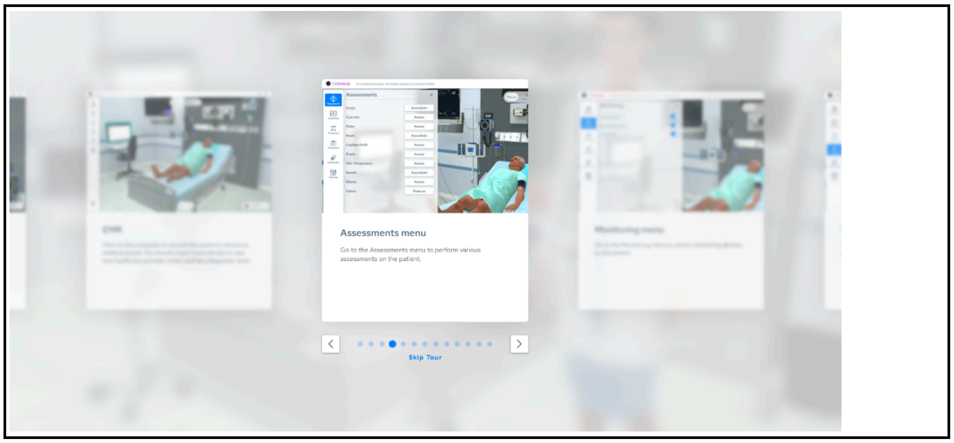


Figure 3

***Onboarding tutorial* - Embody Clinical Simulation Onboarding Tutorial**

2.3.4 Experimental stimuli

The tasks were performed on the cloud-installed software "Maestro Embody" (V1.5.1). This simulation tool includes functionality for interaction with the virtual patient, which is what we were interested in for this study. In a side menu, we also find various modules: Assessment, Monitoring, Procedures, Treatments, Charting, Action Log and Case, all of which allow us to practice the scenario with its various facets that make up the nurse's task. In the simulation environment, we also find a virtual 3D representation of the patient and the various scene elements found in a hospital room, some of which are interactive: the bed, the computer, the monitor and the patient for auscultation tasks. It is also possible to change the viewing angle of the scene as required.

The different scenarios we used were taken from the publisher's library for this software in the nursing - assessments category and represent typical, non-specialized clinical scenarios. The scenarios are modeled, i.e. all interactions and evolutions of the simulation are linked and consistent with each other.

For the task where the patient was interacted with using voice, it is important to note that this interaction mode was used in a development environment and is not currently available to the public. The version with the question-choice interaction mode is, however, available to the public and contains the same models.

The simulation software was accessed via the Internet. In an isolated room, the participant used it on a desktop computer equipped with a keyboard, microphone, mouse and speakers. The application was hosted on a cloud server. Two different nurse evaluation scenarios were used. Qualtrics was used to register participants and administer post-task questionnaires (Qualtrics, Provo, UT). Tobii was used for audio and video recording of the session (Tobii Pro Lab Version 1.217). Danderyd, Sweden: Tobii AB). Optimal Workshop (Optimal Workshop Ltd.) was used for thematic analysis.

2.3.5 Scenarios

The scenarios used for the experiment are modeled simulated clinical experiences (SCEs). Initial situations are presented to the user with a patient background and designed with precise learning objectives based on various National Council Licensure Examination – Registered Nurses (NCLEX-RN) and Quality and Safety Education for Nurses (QSEN) criteria. The learner's actions during the simulation influence the course of the scenario.

Cardiac scenario (Task 1)

In this scenario, a truck driver with visible health issues and a smoking habit attended a health fair on a hot summer day. He displayed symptoms like bilateral crackles, significant leg edema, varicose veins, and absent pedal pulses. His medical history reveals hypertension, obesity, and slightly high cholesterol, managed with Olmesartan medoxomil and Hydrochlorothiazide. He's divorced, childless man living with his diabetic mother, vulnerable to falls. His lifestyle is marked by minimal exercise, daily beer consumption, and a long history of smoking, with no allergies reported.

Asthma scenario (Task 2)

In this scenario, a female patient, experiencing aggravated respiratory distress, visited a clinic on a rainy day. Her symptoms, unalleviated by an expired inhaler, included coughing, wheezing, sore throat, and chest tightness. Diagnosed with asthmatic bronchitis in college, she lives alone, recently got a cat, and was exposed to smoke at a party. She's a non-smoker with a history of using a rescue inhaler, with no allergies noted.

2.3.6 Experimental tasks

The experimental tasks involved comprehensive nursing assessments, rooted in the scenarios described earlier. Participants were directed to immerse themselves in each scenario using the simulation tool for a period of 10 minutes. During this engagement, they were mandated to employ a specific interaction modality—either voice-based or choice-based—strictly designated for each scenario to ensure a controlled examination of the modalities' efficacy. Specifically, the cardiac assessment scenario was navigated using the choice-based modality, whereas the asthma assessment scenario was explored through the voice-based modality. Participants were told that time was not a factor in this experiment, and that they should act as they would in such a case. Participants were ask to verbalize their thoughts, feelings, and actions in real-time as they engaged with the tasks. In the event of too few comments on their part, the moderator could remind them not to forget to detail their thoughts.

Figure 4
Task 1 – Embody Clinical Simulation
Choice mode interaction

Figure 5
Task 2 – Embody Clinical Simulation
Voice mode interaction

2.3.7 Interview

After completing the 2 tasks, participants were invited to take part in a 15-minute interview consisting of 7 open-ended questions. Prior to the interview, they were also reminded that we were not judging the quality of their answers, but rather wanted to learn more about their experience.

2.3.8 Measurements methods

For each task, during their use of the interface, participants were asked to say aloud their thoughts about the clinical case or the use of the interface using the think-aloud method (C, 1982; Ericsson & Simon, 1998). The think-aloud technique is an approach that requires individuals to express their thoughts aloud as they engage in a task (Someren et al., 1994). This aids researchers, specifically in qualitative research, in comprehending individuals' thought processes and decision-making (Charters, 2003). The experiment ended with an interview on the progress of the tasks the participants had just completed to enrich and correlate with think-aloud observations (Wilson, 2013). The interview guide consisted of 6 open-ended questions, previously tested and adjusted for ease of comprehension and flow (see Appendix for complete interview guide).

2.4 Results

A thematic analysis using Optimal Workshop was carried out to identify trends within our extensive qualitative dataset (2x 10min. think-aloud sessions per participant (N=3) & 6 open-ended questions). Thematical analysis offers a detailed overview of a substantial dataset and its adaptability (Braun & Clarke, 2006). Thematic analysis is a widely used qualitative research method for examining and understanding textual data; it uncovers deeper meanings, ideas, feelings, and experiences that emerge from the raw data (Guest et al., 2011). It's a flexible approach that can be adapted to suit the specific nature of the data and research objectives. The thematic analysis method normally starts with data transcription and familiarization, progresses through coding to identify and organize core themes, and advances to theme development where patterns emerge (Naeem et al., 2023). As Naeem et al. (2023) states, it concludes with conceptualization to define and refine concepts and their interrelations, ultimately leading to the construction of a conceptual model that synthesizes the findings and situates them within theoretical frameworks to expand knowledge. Following data collection, all participants' verbatims for the tasks and the interview were transcribed 1) into word-processing software with the aim of carrying out a thematic analysis of this data. These transcripts included what participants were required to say, and any comments made by the moderator on the interaction between the participant and the interface that might be relevant to the analysis. Following this, 2) each participant's data was entered into a thematic analysis tool. Then, 3) analysis groups were created for each task and interview question before 4) each participant's verbatim was entered into the tool by block associated with an analysis group. We then 5) created theme labels for each block. Finally, 6) as data was entered into Optimal Workshop (Optimal Workshop Ltd.), themes were iteratively grouped or created until obtaining an insightful portrait of data.

The results of our analysis are structured into several pivotal subsections, each examining the effects and complexities of virtual patient simulations. Initially, we assess the primary and secondary practical applications as reported by participants, highlighting real-world utility. This is followed by an exploration of the challenges inherent in replicating realistic interactions within simulations, with a particular focus on the difficulties of capturing the depth of human conversations. Subsequently, we delve into the adaptability of medical assessments, such as the "head-to-toe" method, discussing their customization to specific patient contexts. The section concludes with analyses of choice-based and voice-based interactions, examining how each interface modality impacts clinical decision-making and the learning process. This organizational approach facilitates a comprehensive investigation into the educational dynamics and practical implications of both interaction modalities within the simulation framework.

2.4.1 The usefulness of virtual patient simulations

Most participants did not explicitly mention that they would consider integrating simulation with this tool into their daily practice, except for occasionally “practising or revising cases”(P1), or “rarer”(P1) scenarios in a safe context. One participant confirmed that using this tool in a school setting was relevant, as she was still finalizing a degree in the field (P2).

In contrast, all participants proposed relevant secondary uses for the tool, such as telemedicine teaching, inter-unit training, pandemic contexts, or protocol evaluation among employees.

One of the participants reported that during his studies, they used this kind of virtual reality simulation as a pre-practice to prepare for a later simulation with an actor (P3). This same participant underlined the difference in realism between virtual reality simulations and mannequin simulations, in which "you see directly how the person acts, the concrete assessment of the patient"(P3), proposing that a simulation must at some point come as close as possible to the real-life context, with the stress it implies, in order to ensure readiness and safe treatments for the patient.

One participant, however, pointed out that "a lot of people have difficulty projecting themselves into the scenario, but even if it's not a real patient, it doesn't matter, I have to treat him/her as if it was a real patient, it's a role-playing game"(P2).

2.4.2 The realism boundaries of virtual patient conversation

All participants agreed that whatever the mode of interaction, there is a lack of "depth"(P3) in communication with the patient in the simulation, arguing that "you don't have the same answers you would have from a human exactly, you don't have the intonations for example, and then there's a lot of information we go looking for in there."(P2).

A participant also pointed out that "in a real consultation, you have a new patient with whom you need to establish a bond of trust"(P1), difficult if not impossible to find in this type of dialogue in a simulation, even describing it as "impersonal"(P1).

For example, one participant noted that the use of this type of interaction for teaching a mental health situation, for example, would not be appropriate "since there's a contact that's different that you don't have with a computerized voice or something that responds to you"(P2), as interaction with a chatbot lends itself poorly to this context.

In addition, participant 3 noted that in a medical situation, “There was also sometimes talk of third parties involved, such as the parents of a child in pediatric care, who would also be part of the dialogue and could provide important information”.

2.4.3 Variability and adaptability within a context-driven interaction framework

During the interview, most participants (n=2) mentioned explicitly the teaching of the "head-to-toe"(P2,P3) technique. The head-to-toe method provides a systematic and organized way to conduct and gather information during a clinical examination (Fletcher, 2005). Despite this framework, our participants noted that it could vary or be adapted according to the medical context.

On the first level, participants' observations revealed that particular patient characteristics can influence the way in which the head-to-toe sequence is developed. For example, "An older patient will not require the same approach. The 56-year-old who's short of breath, well, I wouldn't ask him if he'd had anything to drink yesterday at a party, I won't look for the same questions."(P2). This difference in approach is therefore expected to have a direct impact on the assessment process and, on a more micro scale, on the depth of the questions as expectations related to the answers; "You adapt, if he's 95 years old and has been in hospital for 2 months, you don't ask him the exact date, you check with the month."(P2)

On another level, the priority associated with a particular medical condition and its degree of urgency can also influence the magnitude of the method or its speed of execution ; "Obviously, a patient who comes in and tells me that he feels really bad when he urinates, and then from what I have as first information I suspect a urinary infection, I'm going to listen to his heart eventually, but that's clearly not where I'm going to start." (P2).

Another theme that has been identified as influencing this framework is the role and tasks assigned to the nurse. Variations may arise if, for example, a primary assessment has been carried out beforehand or if the mandate varies according to the nurse's role in her department; "I'm currently studying to be a nurse practiocioner, so it's really another aspect. I was trying to remind myself a little bit…"(P1). In addition, nurses' personal habits may be adapted to the context as "Everyone will ask the same 3 questions differently."(P2).

In this experimental simulation context, we were, above all, able to observe the centrality of communication with the patient in the clinical reasoning process itself. Thematic analysis revealed that approximately half of the observations made by participants during think-aloud tasks were medical, analytical, or procedural in nature (56 of 112 observations), a corollary of the head-to-toe method. A typical cycle observed through simulation use was that 1) Information was gathered from the patient, 2) Analyzed and clarified, 3) Confirmed/Infirmed/Charted technically using other features of the virtual environment, and 4) Explored through follow-up questions or assessment. This cycle was then repeated with the aim of finally arriving at a clinical diagnosis, and the evaluation was conversation-driven insofar as the interface did not interfere with this flow.

2.4.4 Interface-driven behavioral changes

Another predominant observation noted during the experiment and explained with analysis is the phenomenon we term as "interface-driven" behavior. Within both tasks and across interaction modes, participants occasionally seemed to halt their cognitive process to engage in supplementary actions or extract immediately accessible data from the interface, thereby accelerating their general workflow. This inclination was evident through multiple attributes of the simulation framework. During physical assessments, participants often expanded their focus beyond the initial and present point of assessment. They utilized the simulation's capabilities to conduct comprehensive auscultations, yet they did not consistently follow through with the initiated interaction, such as requesting patient consent. This approach was similarly observed in the charting of medical results. Among these interface-driven behavioral changes, several are relative to the precise use of patient dialog functionality, i.e. the interaction modes.

2.4.4.1 Choice-based interaction

When interacting with the choice-based mode, we noted some participants used the chatbot by asking questions, “as long as I’m here”(P3), in an order that would not necessarily have been the case in a real context. Even more, we observed that some participants used this virtual patient interaction feature more passively as an information provider emphasizing that “By having the questions, I can already ask him the others.”(P3), then describing the interaction with the patient as “robotic” and “automatic”(P1).

Moreover, all participants (n=3) were at some point held back by a lack of possibilities when they had intended to go deeper into an axis in their evaluation or their own way of approaching the subject; “There were questions I couldn't find in it, (...) there were questions I couldn't ask .”(P3). Some also pointed out that they took more time to look for a question they wanted to ask or in line with their intention, noting that they were not in a “logical order”(P3) or “grouping”(P3); “I pretty much know where I want to go, then it's a matter of seeing, okay, is there such a question? So rather than going through the list and saying I've found the questions, I want to ask him (...) :asking him how his day went, what medication he took (and so on) (...) it's true that having a selection list of questions is restrictive.”(P3).

In addition to the level of experience, some participants pointed out that the respective levels of difficulty of the interaction modes could vary according to the individual personality or experience because "the software becomes more frustrating if you don't know what questions to ask and then you can't progress in your scenario (...) so depending on your personality, it's not better for your learning"(P3).

Participants indeed raised the potential educational relevance of this mode to support learning in case of difficulty; "What could be positive is if you have a blank and then don't know what question to ask"(P1) or for personalities “more anxious”(P3) about having to compose questions and express them orally. However, none of the participants in our experiment experienced the need to have support on questions asked or affirmations during the simulation.

Paradoxically, while they perceived the mode as facilitating, the expert participants all struggled during the simulation with this mode of interaction. Limitations imposed by the choice modality interface caused adaptations, changes of direction, or abandonment on the part of participants. As experts, participants seemed to be more slowed down by having to sort and analyze potential utterances and follow their train of thought.

Furthermore, this mode of interaction is perceived as easier. Some participants mentioned the risk of developing passivity, automatism, lack of attention or motivation as the scenario progresses. Most participants agreed on the potential harmful effects of this mode, which “directs your thinking a bit more, whereas if you want to evaluate someone's clinical judgment (...) in real life, there's not going to be a choice!”(P3).

2.4.4.2 Voice-based interaction

Voice interaction mode offered other possibilities to the user. Even if using this function required "a little more adjustment than selecting a question and then seeing the answer appear.”(P2), all participants said they preferred “the fact that you could speak rather than pre-established questions"(P1), stating a positive interaction, as they found it easier to "interact with something much closer to what you're actually doing, rather than asking a question in writing"(P2).

Some participants noted that they had needed to adapt to the mode of interaction, stressing, however, that this adaptation was comparable to "finding yourself in front of a patient whose first language is not French (...) it's adapting to your patient, are all patients going to be like that no, are some going to be yes, the one on the screen is (...) you say to yourself okay, if that's how I can get my information, that's how I'm going to get my information."(P2).

One participant commented on the contribution of such a mode of interaction, noting that "it also helps to get into the scenario"(P1). They found it "just more natural if you talk like you would talk to a patient"(P3). Dialogue seemed to be facilitated and gave "something to hold on to in order to ask another question"(P1).

One of them described the experience as more "real"(P1) because "you don't have the choice to go for it, if you don't talk, you don't get the information"(P1), comparing it with the choice mode, described as "equivalent to reading on the internet"(P1) e.d without added value.

Among the factors that contributed to the positive perception of voice mode was the ability of users to compose and structure their speech. Voice interaction facilitated interaction by allowing users to "go and get your own questions, in the order you wanted to do them"(P3).

From the dialogue perspective, the different and nuanced answers given by the AI also contributed to the perceived quality of the interaction: "They were answers (...) that were longer, but they still answered relatively well to what was asked of it"(P1) and that "the answer still seemed more adapted to the question I had asked"(P1) compared to the answers pre-written by the development team.

One participant reported an advantage of the multimodality interaction mode, mentioning that he was able to use voice to simultaneously perform other actions in the interface.

Most of the problems encountered when interacting with the AI-powered modality stem from technological limitations, starting with voice recognition. Faced with these obstacles, participants had to adapt, noting that "It was maybe a little destabilizing, maybe it took me out of my questioning order a little bit" (P1). Thematical analysis showed that all participants experienced voice recognition errors or defects, which “frustrated”(P2) them at different levels. Apart from explicitly pointing this out, a typical reaction to this type of technological fault was a change in rhythm, tone, articulation, or a complete reformulation of the sentence.

All the participants stressed in their own way that the need to interact vocally, without suggestions from the system, would be an engine of apprenticeship, postulating that "you integrate knowledge better if you ask the questions"(P2), since "you learn more when you say it, because you're digging into your head and you're going to reactivate the knowledge you've learned"(P3).

Responses given by the chatbot were also sometimes perceived as “incomplete”, “imprecise”, or “inappropriate”(P1,P2,P3). This eventuality sometimes led to abandonment, but most participants would persist at least 1 time in retrying the operation.

On another level, some participants felt frustrated, for example, when the AI tried to reframe a conversation that it identified like it was deviating from the topic of interest; "It would often bounce back to -Please Ask about my health (Default answer)- I was like, this is what I'm doing! (...) It makes you know, the kind of like I wouldn't say frustration, but the kind of... you don't need to tell me to talk to you about your health, we're here (...) it's kind of like the program that's trying to think for you."(P2).

These technological flaws during the simulation led not only to a participant's need to adapt but also to certain deviations from the process that were encountered, similar to the choice interaction mode but different in nature; "Sometimes, I felt like I was saying something but it wasn't the right sentence, so I was like ok... we won't go into that question"(P1), or even to abandonment on their part; "She didn't understand exactly what I was saying perfectly, so after that you say to yourself, well that's kind of what I meant, it makes you repeat the same question differently to realize that no... it makes you say to yourself well: I quit and then I'll go and ask another question."(P3).

2.5 Discussion

2.5.1 Summary of results

The qualitative exploration into the interplay between virtual agents and users in VR simulations for nursing education has shed light on the subtleties inherent in designing these interactions for problem-solving scenarios. Our analysis indicates that choice-based interactions may serve as a beneficial entry point for novices, aligning with the principles of cognitive load theory to optimize learning by tailoring educational strategies to match learners' varying expertise levels (Kalyuga et al., 2012; Van Merriënboer & Sweller, 2010). Kalyuga et al. (2012) highlight that novices might find themselves overwhelmed by excessive information, which could impede the learning process. Managing a complex conversation can be cognitively taxing, given the breadth of information it encompasses. Conversely, seasoned practitioners might encounter limitations within the interface that detract from leveraging their existing knowledge base, thereby diminishing the simulation's effectiveness as a tool for knowledge enhancement.

The findings underscore a harmony between interaction modalities and specific nursing tasks, as posited by the Task-Technology fit theory (Goodhue & Thompson, 1995), thereby enriching the learning experience for more advanced learners. Our results particularly advocate for the advantages of voice-based interactions in creating a more immersive and intuitive learning environment, potentially elevating user engagement and managing cognitive load more effectively for advanced users. As reported by participants, this preference underscores the nuanced, complex, and context-specific nature of clinical reasoning (Higgs et al., 2008; Young et al., 2020). Clinical reasoning in nursing, as described by Simmons (2010), entails a multifaceted cognitive process involving formal and informal data collection and analysis strategies, significance determination, and the exploration of alternative actions (Simmons, 2010). This iterative process emphasizes the use of both inductive and deductive reasoning, suggesting a cyclical rather than linear progression (Lee et al., 2016).

Moreover, our study identifies the challenge of suspension of disbelief (Muckler, 2017) as a significant barrier to effective simulation engagement. This concept, crucial for the efficacy of simulation-based learning tools, hinges on the learner's ability to fully immerse themselves in the scenario, thus impacting the learning outcomes (Dede, 2009). Enhancements in the realism and fidelity of dialogues with virtual agents are suggested as key factors in overcoming this hurdle, emphasizing the role of advanced interaction modalities in achieving deeper immersion and learning effectiveness.

This comprehensive understanding elucidates the critical importance of modality selection in the design of desktop VR simulations, pointing out that voice interaction with a virtual agent has the potential to foster deeper engagement and in the end a more enriching learning experience. The alignment between the technological affordances and educational goals not only underscores the pedagogical utility of VR simulations but also suggests that achieving a greater technology-task fit could improve learning outcomes and user satisfaction, warranting further investigation to understand its implications across diverse learning contexts.

2.5.2 Theoretical contributions

This research contributes to cognitive fit theory by explaining how virtual agent interaction modalities can enhance or hinder the alignment between a user's mental representation and the task's information presentation. By comparing vocal and choice-based virtual agent interactions within a simulated education environment, the study empirically validates the principles of task-technology fit, showing that the vocal modality's natural communication patterns align closely with real-world clinical scenarios. This alignment improves user engagement and streamlines decision-making, suggesting that voice-based interactions, at least for expert users, offer a superior cognitive fit by replicating realistic clinical communication patterns. The research also highlights the crucial role of interaction design, demonstrating that choice-based interactions, though structured, can impose a cognitive burden by restricting user inquiry paths and limiting information exploration while being practical for more novice users. Thus, the findings underscore the importance of modality selection in enhancing the real-world application of virtual agents, helping educators design simulation tools that better align with learners' cognitive expectations and varying expertise levels. Ultimately, this study advances cognitive fit theory by emphasizing the critical interplay between interaction modality, task nature, and complexity, offering practical insights for designing more intuitive and effective simulation-based educational tools.

2.5.3 Business applications

The findings offer actionable insights for developers and educators aiming to leverage VR technologies for training and education. By emphasizing the benefits of GenAI powered voice-based interactions, this study suggests a strategic pivot towards more natural and intuitive user interfaces in educational VR applications. For businesses and educational institutions, integrating voice-based modalities could enhance the learning experience and foster the adoption and relevance of VR training programs across various disciplines that require communication skills to be integrated with other learning dimensions.

2.5.4 Limitations and future research

While our study sheds light on the potential of LLM-powered vocal interactions within VR simulations to enhance realism, engagement, and personalization in education, it's important to acknowledge its limitations. Our conclusions on the potential for AI to tailor educational experiences to individual needs and streamline cognitive processes are preliminary. They hint at the transformative potential of AI in education but also underscore the necessity for deeper investigation into how these personalized, human-like interactions impact diverse learning outcomes over the long term. Therefore, while our research contributes valuable insights into the integration of AI technologies in educational settings, these theoretical contributions and their implications should be viewed as an initial step, inviting further exploration and validation within the evolving landscape of educational technology.

Despite its contribution, our study did not explicitly account for participants' experience levels with simulation technologies or their professional work experience. Additionally, the investigation was conducted with a limited sample size, which may affect the generalizability of our findings. These factors could significantly influence the outcomes, as familiarity with simulation environments, practical experience in nursing, and the diversity of participant backgrounds could affect how users interact with and perceive the VR scenarios. The small scale of our sample limits the extent to which these results can be extrapolated to the broader population of nursing professionals and students. Future research should consider these variables, as well as expand the participant pool, to provide a more nuanced understanding of how different backgrounds impact learning effectiveness in VR settings. This approach will allow for a more tailored design of simulation-based learning tools, accommodating a wider range of learners and potentially revealing more detailed insights into the specific elements that contribute to an effective and immersive learning experience.

Also, our study's scope is limited to nursing education within VR contexts, suggesting the need for broader investigations across disciplines and learning environments. Our UX study was short-duration and could be biased by a novelty effect. Future research should explore the longitudinal effects of these modalities on engagement, knowledge retention and skill acquisition.

In addition to the previously mentioned limitations, our study focuses on rapidly evolving technologies. A notable limitation is that voice-based functionality, which is in its nascent stages, is heavily reliant on ongoing technological advancements. Technical errors beyond our control occurred during the experiment, which may have undermined the perception of the vocal modality. As these technologies evolve, there's a compelling need to investigate their implications for learning environments continuously. Future research should not only expand across disciplines but also adapt to and reassess the impact of these advancements as they mature. This ongoing exploration is critical to fully understanding the potential and constraints of GenAI-powered vocal interactions in enhancing simulation-based learning, ensuring that educational tools remain at the forefront of technological and pedagogical innovation.

2.6 Conclusion

The exploration of virtual agent interaction modalities within desktop VR simulations, as delineated in this study, underscores the pivotal role of immersive technologies in enhancing education. By comparing voice-based to choice-based virtual agent interactions, this research sheds light on the nuanced preferences and performance implications for learners but also highlights the importance of intuitive and immersive interaction in educational simulations. The findings suggest that GenAI-powered voice-based interactions, by facilitating a more natural and engaging conversation, thus learning environment, could potentially enhance cognitive load management and user engagement, thereby contributing significantly to the pedagogical efficacy of simulation-based learning.

These insights into the educational dynamics of LLM-powered vocal and choice-based agent interactions within problem-solving contexts open new avenues for further HCAI research. As the field of GenAI evolves and offers increasingly advanced technological possibilities, they underscore the necessity of adopting a human-centered design approach in the development of educational technologies. Furthermore, this emphasizes the importance of aligning technological advancements with pedagogical objectives and learner needs.

Building on the foundational insights gleaned from this study, we are planning a further study to shift the focus to a broader examination of the implications of these findings for designing and implementing scalable interactive technologies such as LLM in medical education. This study will delve deeper into the integration of GenAI within various learning environments, expanding beyond the desktop VR simulations to encompass a wider array of digital platforms, educational contexts, and the challenges of implementing these innovations effectively.

2.7 Appendix

Appendix A

Interview guide

Interview Guide - English

Q0: So, How was your experience?

Today, you took part in two different simulation scenarios, each offering a distinct mode of interaction. The first scenario used a generative artificial intelligence-based interaction with the virtual patient, where you could engage in a dialogue with the patient, hearing his or her responses to your questions (VOICE). Conversely, the interaction with the virtual patient in the second scenario was choice-based, presenting you with several predetermined options from which you could select the most appropriate course of action (CHOICE).

Q1: In the context of a nursing assessment like we simulated today, is there a generic process, taught and documented, that guides clinical reasoning?

Q2: With respect to the VOICE modality, could you share how this specific interaction mode influenced the execution of the process, either positively or negatively?

Q3: Turning to the CHOICE modality, could you discuss how this unique interaction mode affected the execution of the process, either in a positively or negatively?

Q4: Do you think that one modality provided a superior learning experience? If so, could you explain why?

Q5: Would you consider incorporating this kind of simulation tool into your daily practice? If affirmative, could you elaborate on how you would do so?

Q6: Leveraging your expertise, could you talk to us about the key differences between a typical nurse-patient interaction in real life practice and the interaction modes of the simulation settings you encountered today?

Appendix B

Briefing case example

SCENARIO A

You are a nurse working in a medical clinic located in a hospital. You are just beginning your shift and meeting the patient for the first time. She had a negative Covid test.

The patient is a 26-year-old female, Ms. Star. She shows up at the ER on a rainy and damp day with signs of shortness of breath that have worsened over the past two days. The patient is coughing and wheezing and complains of a sore throat and chest tightness, which is unrelieved by her inhaler.

You must:

Complete a complete physical assessment

(Patient file)

Allergies: None	**Medications:** Rescue inhaler that was prescribed while in college	**Code Status:** Full Code
Past Medical History: Reports once having a respiratory illness while in college and was diagnosed at that time as having asthmatic bronchitis	**Social History:** -Does not smoke. -Attended a party last night with friends where people were smoking.	**Family History:** Lives alone with a newly acquired cat
Hand-off report: None	**Orders:** None	**Lab results:** None
Diagnostics: None		

Références

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., Al Yami, M. S., Al Harbi, S., & Albekairy, A. M. (2023). Revolutionizing healthcare : The role of artificial intelligence in clinical practice. *BMC Medical Education*, *23*(1), 689. https://doi.org/10.1186/s12909-023-04698-z

AL-Smadi, M. (2023). *ChatGPT and Beyond : The Generative AI Revolution in Education* (arXiv:2311.15198). arXiv. http://arxiv.org/abs/2311.15198

Bailey, S., Johnson, C., Schroeder, B., & Marraffino, M. (2017). *Using Virtual Reality for Training Maintenance Procedures*.

Baily, L. W. (2019). History of Simulation. In S. B. Crawford, L. W. Baily, & S. M. Monks (Éds.), *Comprehensive Healthcare Simulation : Operations, Technology, and Innovative Practice* (p. 3‑11). Springer International Publishing. https://doi.org/10.1007/978-3-030-15378-6_1

Benner, P. (1984). From novice to expert. *Menlo Park*, *84*(1480), 10‑1097.

Boletsis, C., & Karahasanovic, A. (2020). Immersive Technologies in Retail : Practices of Augmented and Virtual Reality: *Proceedings of the 4th International Conference on Computer-Human Interaction Research and Applications*, 281‑290. https://doi.org/10.5220/0010181702810290

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77‑101. https://doi.org/10.1191/1478088706qp063oa

C, L. (1982). Using the « thinking-aloud » method in cognitive interface design. *Research Report RC9265, IBM T. J. Watson Research Center*. https://cir.nii.ac.jp/crid/1571135649469604736

Capel, T., & Brereton, M. (2023). What is Human-Centered about Human-Centered AI? A Map of the Research Landscape. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1‑23. https://doi.org/10.1145/3544548.3580959

Charters, E. (2003). The Use of Think-aloud Methods in Qualitative Research An Introduction to Think-aloud Methods. *Brock Education Journal*, *12*(2). https://doi.org/10.26522/brocked.v12i2.38

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education : A Review. *IEEE Access*, *8*, 75264‑75278. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2988510

Chui, M., Roberts, R., & Yee, L. (s. d.). *Generative AI is here : How tools like ChatGPT could change your business*.

Dede, C. (2009). Immersive Interfaces for Engagement and Learning. *Science*, *323*(5910), 66‑69. https://doi.org/10.1126/science.1167311

Ericsson, K. A., & Simon, H. A. (1998). How to Study Thinking in Everyday Life : Contrasting Think-Aloud Protocols With Descriptions and Explanations of Thinking. *Mind, Culture, and Activity*, *5*(3), 178‑186. https://doi.org/10.1207/s15327884mca0503_3

Farooq, U., & Grudin, J. (2016). Human-computer integration. *Interactions*, *23*(6), 26‑32. https://doi.org/10.1145/3001896

Fletcher, M. (2005). Canadian Fundamentals of Nursing, 3rd ed. *The Canadian nurse*, *101*(8), 10‑10.

Fraser, K. L., Ayres, P., & Sweller, J. (2015). Cognitive load theory for the design of medical simulations. *Simulation in Healthcare*, *10*(5), 295‑307.

Gaba, D. M. (2004). The future vision of simulation in health care. *Quality and Safety in Health Care*, *13*(suppl 1), i2. https://doi.org/10.1136/qshc.2004.009878

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS quarterly*, 213‑236.

Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis*. sage publications. https://books.google.com/books?hl=fr&lr=&id=Hr11DwAAQBAJ&oi=fnd&pg=PP1&dq=thematic+analysis&ots=XkZytIAGrE&sig=r8S_2KXcn5C7Zm-RZkuypC1BbYY

Higgs, J., Jones, M. A., Loftus, S., & Christensen, N. (2008). *Clinical Reasoning in the Health Professions E-Book*. Elsevier Health Sciences.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261‑266. https://doi.org/10.1126/science.aaa8685

Holmes, W., Bialik, M., & Fadel, C. (2023). Artificial intelligence in education. In C. Stückelberger & P. Duggal (Éds.), *Data ethics : Building trust : How digital technologies can serve humanity* (p. 621‑653). Globethics Publications. https://doi.org/10.58863/20.500.12424/4276068

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational Implications of Expertise Reversal Effects in Learning and Performance of Complex Cognitive and Sensorimotor Skills. *Educational Psychology Review*, *24*(2), 313‑337. https://doi.org/10.1007/s10648-012-9195-x

Karat, J., & Karat, C. M. (2003). The evolution of user-centered focus in the human-computer interaction field. *IBM Systems Journal*, *42*(4), 532‑541. https://doi.org/10.1147/sj.424.0532

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Ke, F., Dai, Z., Dai, C.-P., Pachman, M., Chaulagain, R. S., & Yuan, X. (2020). Designing virtual agents for simulation-based learning in virtual reality. In *Cognitive and affective perspectives on immersive technology in education* (p. 151‑170). IGI Global. https://www.igi-global.com/chapter/designing-virtual-agents-for-simulation-based-learning-in-virtual-reality/253694

Kononowicz, A. A., Woodham, L. A., Edelbring, S., Stathakarou, N., Davies, D., Saxena, N., Car, L. T., Carlstedt-Duke, J., Car, J., & Zary, N. (2019). Virtual Patient Simulations in Health Professions Education : Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. *Journal of Medical Internet Research*, *21*(7), e14676. https://doi.org/10.2196/14676

Lee, J., Lee, Y. J., Bae, J., & Seo, M. (2016). Registered nurses' clinical reasoning skills and reasoning process : A think-aloud study. *Nurse Education Today*, *46*, 75‑80. https://doi.org/10.1016/j.nedt.2016.08.017

Léger, P., Cronan, P., Charland, P., Pellerin, R., Babin, G., & Robert, J. (2012). Authentic OM problem solving in an ERP context. *International Journal of Operations & Production Management*, *32*(12), 1375‑1394. https://doi.org/10.1108/01443571211284151

Lu, M. (2023). *The industrial revolution brought by AI : Opening the Gateway to the Future* (0). Resources Economics Research Board. https://doi.org/10.50908/grb.2.0_166

Mascarenhas, S., Guimarães, M., Prada, R., Dias, J., Santos, P. A., Star, K., Hirsh, B., Spice, E., & Kommeren, R. (2018). A Virtual Agent Toolkit for Serious Games Developers. *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, 1‑7. https://doi.org/10.1109/CIG.2018.8490399

Muckler, V. C. (2017). Exploring Suspension of Disbelief During Simulation-Based Learning. *Clinical Simulation in Nursing*, *13*(1), 3‑9. https://doi.org/10.1016/j.ecns.2016.09.004

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research. *International Journal of Qualitative Methods*, *22*, 16094069231205789. https://doi.org/10.1177/16094069231205789

Norman, G., Dore, K., & Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. *Medical Education*, *46*(7), 636‑647. https://doi.org/10.1111/j.1365-2923.2012.04243.x

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes : Evidence from the computer games literature. *The Curriculum Journal*, *16*(4), 455‑474. https://doi.org/10.1080/09585170500384529

Ooi, K.-B., Tan, G. W.-H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., Dwivedi, Y. K., Huang, T.-L., Kar, A. K., Lee, V.-H., Loh, X.-M., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N. P., Sarker, P., Sharma, A., … Wong, L.-W. (2023). The Potential of Generative Artificial Intelligence Across Disciplines : Perspectives and Future Directions. *Journal of Computer Information Systems*, 1‑32. https://doi.org/10.1080/08874417.2023.2261010

Pilote, B., & Chiniara, G. (2019). Chapter 2—The Many Faces of Simulation. In G. Chiniara (Éd.), *Clinical Simulation (Second Edition)* (p. 17‑32). Academic Press. https://doi.org/10.1016/B978-0-12-815657-5.00002-4

Pottle, J. (2019). Virtual reality and the transformation of medical education. *Future Healthcare Journal*, *6*(3), 181‑185. https://doi.org/10.7861/fhj.2019-0036

Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice : Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare*, *11*(6), Article 6. https://doi.org/10.3390/healthcare11060887

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research : Exploring its conceptualization and operationalization. *Quality & Quantity*, *52*(4), 1893‑1907. https://doi.org/10.1007/s11135-017-0574-8

Savery, J. R. (2015). Overview of problem-based learning : Definitions and distinctions. *Essential readings in problem-based learning: Exploring and extending the legacy of Howard S. Barrows*, *9*(2), 5‑15.

Savery, J. R., & Duffy, T. M. (1995). Problem based learning : An instructional model and its constructivist framework. *Educational technology*, *35*(5), 31‑38.

Simmons, B. (2010). Clinical reasoning : Concept analysis. *Journal of Advanced Nursing*, *66*(5), 1151‑1158. https://doi.org/10.1111/j.1365-2648.2010.05262.x

Someren, M., Barnard, Y., & Sandberg, J. (1994). *The Think Aloud Method—A Practical Guide to Modelling CognitiveProcesses*.

Stone, R., Panfilov, P., & Shukshunov, V. (2011). Evolution of Aerospace Simulation : From Immersive Virtual Reality to Serious Games. In *RAST 2011—Proceedings of 5th International Conference on Recent Advances in Space Technologies*. https://doi.org/10.1109/RAST.2011.5966921

Tam, W., Huynh, T., Tang, A., Luong, S., Khatri, Y., & Zhou, W. (2023). Nursing education in the age of artificial intelligence powered Chatbots (AI-Chatbots) : Are we ready yet? *Nurse Education Today*, *129*, 105917. https://doi.org/10.1016/j.nedt.2023.105917

The Effectiveness of Games for Educational Purposes : A Review of Recent Research—Josephine M. Randel, Barbara A. Morris, C. Douglas Wetzel, Betty V. Whitehill, 1992. (s. d.). Consulté 26 mars 2024, à l'adresse https://journals.sagepub.com/doi/abs/10.1177/1046878192233001

Tun, J. K., Alinier, G., Tang, J., & Kneebone, R. L. (2015). Redefining Simulation Fidelity for Healthcare Education. *Simulation & Gaming*, *46*(2), 159‑174. https://doi.org/10.1177/1046878115576103

Tyerman, J., Luctkar-Flude, M., Graham, L., Coffey, S., & Olsen-Lynch, E. (2019). A Systematic Review of Health Care Presimulation Preparation and Briefing Effectiveness. *Clinical Simulation in Nursing*, *27*, 12‑25. https://doi.org/10.1016/j.ecns.2018.11.002

Van Merriënboer, J. J. G., & Sweller, J. (2010). Cognitive load theory in health professional education : Design principles and strategies. *Medical Education*, *44*(1), 85‑93. https://doi.org/10.1111/j.1365-2923.2009.03498.x

Vora, J., Nair, S., Gramopadhye, A. K., Duchowski, A. T., Melloy, B. J., & Kanki, B. (2002). Using virtual reality technology for aircraft visual inspection training : Presence and comparison studies. *Applied Ergonomics*, *33*(6), 559‑570. https://doi.org/10.1016/S0003-6870(02)00039-X

Walker, A., & Leary, H. (2009). A problem based learning meta analysis : Differences across problem types, implementation types, disciplines, and assessment levels. *Interdisciplinary journal of problem-based learning*, *3*(1), 6.

Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). *Engineering psychology and human performance*. Routledge. https://www.taylorfrancis.com/books/mono/10.4324/9781003177616/engineering-psychology-human-performance-christopher-wickens-justin-hollands-simon-banbury-william-helton

Wilson, C. (2013). *Interview techniques for UX practitioners : A user-centered design method*. Newnes.

Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2023). Transitioning to Human Interaction with AI Systems : New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI. *International Journal of Human–Computer Interaction*, *39*(3), 494‑518. https://doi.org/10.1080/10447318.2022.2041900

Xu, W., & Gao, Z. (2024). Applying HCAI in Developing Effective Human-AI Teaming : A Perspective from Human-AI Joint Cognitive Systems. *Interactions*, *31*(1), 32‑37. https://doi.org/10.1145/3635116

Young, M. E., Thomas, A., Lubarsky, S., Gordon, D., Gruppen, L. D., Rencic, J., Ballard, T., Holmboe, E., Da Silva, A., Ratcliffe, T., Schuwirth, L., Dory, V., & Durning, S. J. (2020). Mapping clinical reasoning literature across the health professions : A scoping review. *BMC Medical Education*, *20*(1), 107. https://doi.org/10.1186/s12909-020-02012-9

Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, *2*(10), Article 10. https://doi.org/10.1038/s41551-018-0305-z

Chapitre 3

Beyond One-Size Fits All : A Dual Path for Explaining User Experience with GenAI Systems in Simulation-Based Learning Using Conversational Agents.

Authors: Nico Helie, Alexander J. Karran, Shang Lin Chen, Constantinos K. Coursaris, Pierre-Majorique Léger.

HEC Montréal, 3000 Chem. de la Côte-Sainte-Catherine, Montréal, QC H3T 2A7

3.0 Abstract

As artificial intelligence (AI) increasingly integrates into healthcare and education, its applications in simulation-based learning offer unique opportunities to enhance user experiences and outcomes. This study examines the effects of generative AI (GenAI)-powered vocal interactions and varying simulation environment fidelities on user perceptions of system response accuracy, naturalness, social presence, and their influence on cognitive, emotional experiences and attitude toward technology. Through a comparative analysis of desktop VR and ChatGPT environments in both interaction modalities (i.e. textual vs. vocal) in nursing education simulations, we found that while desktop VR provides stronger social presence due to its immersive nature, ChatGPT offered higher perceived naturalness and lower cognitive effort, suggesting that interaction with agent smoothness may be more important than visual realism. Additionally, vocal interaction modes were shown to elevate engagement, though at the cost of increased cognitive load. These findings underscore the importance of balancing cognitive demand and emotional engagement in designing AI-driven educational tools, highlighting the need for nuanced approaches that align with users' cognitive and emotional capacities.

3.1 Introduction

More than ever, artificial intelligence (AI) is profoundly and increasingly reshaping the societal and economic landscape. AI aims to create systems capable of performing tasks that typically require human intelligence, such as learning, reasoning, and problem-solving (Goodfellow et al., 2016); It is forecasted to bring transformative changes that will affect all aspects of life, including but not limited to employment, decision-making, and global competition, leading to a future where intelligent machines could surpass human capabilities in numerous domains (Makridakis, 2017; Mannuru et al., 2023). This integration is already transforming deeply processes and fostering innovation in businesses (Ooi et al., 2023), indicating AI's ability to significantly impact and reshape professional landscapes and operational methodologies in the coming years.

In the healthcare field, AI applications span various areas, from imaging diagnosis and genome interpretation to machine learning for biomarker discovery, clinical outcome prediction, and autonomous robotic surgery (Yu et al., 2018). AI is currently revolutionizing diagnosis, treatment, and disease management, significantly improving care outcomes through more accurate diagnoses and personalized medicine (Alowais et al., 2023; Briganti & Le Moine, 2020). Advanced machine learning algorithms and extensive data analysis hold the potential to significantly enhance healthcare delivery and resource allocation (Navath, 2021), thereby improving patient outcomes and, ultimately, advancing the overall efficiency of healthcare for society.

The impact of AI on education, particularly healthcare education, is equally transformative. AI increases learning opportunities by offering personalized feedback, adaptive tutoring systems, and tailored assistance for educators and learners (L. Chen et al., 2020). AI supports collaborative learning, monitors student forums, facilitates continuous assessment, provides AI learning companions for students, and serves as a research tool to advance learning sciences (Holmes et al., 2023). Research shows that GenAI, more specifically, can significantly enhance student performance through individualized training, leading to better learning outcomes (Ogunleye et al., 2024; Sauder et al., 2024). The flexibility of GenAI is recognized as a valuable tool for creating dynamic and interactive learning environments tailored to diverse educational needs.

Natural Language Processing (NLP) enables AI systems to understand and generate human-like text or speech (Hirschberg & Manning, 2015). Large Language Models (LLMs), a subcategory of NLP technologies, are poised to revolutionize workplaces and educational systems beyond mere workflow modifications (AL-Smadi, 2023; Chui et al., s. d.). LLMs enhance medical education with accurate and adaptable clinical scenarios, improve patient communication

(Sallam, 2023, p. 20), and facilitate personalized and autonomous learning through immediate question responses, personalized study notes, or test question generation (W. Tam et al., 2023). AI is driving profound and transformative changes across various sectors, including healthcare and education, enhancing processes through intelligent systems that can perform tasks traditionally requiring human expertise. However, adopting a human-centered approach is crucial to bridging the gap between technology and education, as it ensures that educational tools like LLMs are tailored to meet the diverse needs of learners, promote critical thinking, and are integrated responsibly with human oversight to address biases, privacy concerns, and ethical challenges (Kasneci et al., 2023).

Simulation and Problem-Based Learning (PBL) have extensive applications in educational and professional fields. Simulation, integrating various modalities and classification systems, supports experiential learning (Pilote & Chiniara, 2019). It is beneficial in multiple domains such as aviation (Stone et al., 2011; Vora et al., 2002, p. 20), military (Bailey et al., 2017, p. 201), and retail (Boletsis & Karahasanovic, 2020). Virtual agents, characters brought to life in simulations, inspire users, assist in cognitive tasks, provide feedback, and assess learners (Ke et al., 2020). In nursing education, for instance, computer-based virtual reality simulation offers immersive and realistic training experiences, enhancing practical skills through encounters with virtual patients (Kononowicz et al., 2019). Virtual reality transforms medical education by offering safe, standardized, and cost-effective clinical training experiences (Pottle, 2019). With continuous advancements in GenAI, its integration into simulation and PBL holds immense promise, further enhancing the effectiveness, personalization, and scalability of these educational methodologies.

The integration of GenAI with virtual patient simulations in medical education represents a significant advancement that, to the best of our knowledge, is still largely unexplored. Virtual patients, particularly through interactive scenarios and virtual reality technology, offer an immersive and personalized educational experience, enhancing the education of healthcare practitioners and patients (Adeghe et al., 2024; Cook et al., 2010; Kononowicz et al., 2019). Simulation enriches PBL by allowing learners to apply skills in realistic scenarios, enabling them to adapt to the outcomes of their decisions (Léger et al., 2012). PBL, initially developed for medical education, allows students to tackle real and multifaceted challenges (Savery, 2015; Savery & Duffy, 1995; Walker & Leary, 2009) and well-designed simulations facilitate the rapid acquisition of skills and a deep understanding of complex scenarios (O'Neil et al., 2005, p. 200). Although well-designed simulators are often equated with high fidelity and realism, this definition needs to be expanded. High-fidelity simulators, despite their emphasis on realism and performance enhancement, do not always guarantee superior learning transfer when compared to low-fidelity alternatives (G. Norman et al., 2012). Additionally, implementing advanced LLMs like ChatGPT (OpenAI, 2023) in educational practices presents both opportunities and challenges which need to be carefully investigated and taken into account (Kasneci et al., 2023).

The rapid advancements in AI and the evolving landscape of medical education underscore the need to balance realism and educational value. An excessive focus on realism can hinder the optimal effectiveness of learning experiences (Boscardin et al., 2024; MacLean et al., 2019; Massoth et al., 2019). Clinical virtual simulation can improve knowledge retention, clinical reasoning, learning satisfaction, and self-efficacy in healthcare education, but its overall effectiveness remains under-researched (Padilha et al., 2019). Moreover, the novelty of research on Human-AI Interaction (HAII) highlights the importance of carefully evaluating and validating AI technology in educational contexts (Cain, 2024; Følstad et al., 2021; Sallam, 2023). While the integration of GenAI potentially creating more life-like virtual patients offers exciting possibilities for enhancing realism in simulations, it does not inherently guarantee improved relevance, efficiency, or better learning outcomes and should be approached with careful consideration of its educational impact.

Leveraging the Media Equation Theory (MET) and Cognitive Load Theory (CLT) for a dual-path explanation, our study seeks to address this critical gap by exploring the impact of GenAI-enhanced voice interactions and varying levels of simulation environment fidelity on learners' educational experiences, with nursing education serving as a model. By exploring both cognitive and emotional dimensions, we seek to inform research on how these technologies influence learners' experiences. User experience study is pivotal to obtain valuable and rich insights into the use of a highly complex and contextual artifact (i.e., clinical diagnosis simulation using a virtual patient). It is a powerful framework to understand technology adoption, as it integrates both the experiential and utilitarian aspects (Hornbæk & Hertzum, 2017). This research builds on the authors previous study demonstrating that users tend to appreciate the naturalness, social character, and interactivity offered by the GenAI-powered virtual agent's voice modality (Helie et al., 2025). However, as reported by Helie et al. (2025), technological challenges such as speech-to-text and text-to-speech limitations appear to diminish or at least attenuate this appreciation. Hence, we aim to expand this study through the addition of quantitative data while broadening the research scope to include other potential and more accessible simulation techniques, such as the straightforward utilization of ChatGPT in a role-play activity to distinguish and investigate the real values of complex simulation environments and its integration of GenAI.

With this study, we aim to enrich the fields of Human-Centered AI (HCAI) and Human-AI interaction (HAII), bridging the gap between learners and technology and enhancing AI integration in educational simulations. More particularly, the objective of this study is to investigate the impact of GenAI-enhanced vocal interactions and variations in environment fidelity on learners' attitudes towards technology, considering both cognitive and emotional components, answering the following question:

In what ways and to what extent do GenAI-powered vocal interaction modality and simulation environment fidelity affect user attitudes toward technology in complex problem-solving simulation learning contexts?

3.2 Literature review

3.2.1 Current state of knowledge

3.2.1.1 Artificial intelligence in simulation-based medical education

AI is transforming healthcare simulation education, offering new ways to enhance learning experiences and improve skill acquisition in medical professionals. AI's integration into simulation-based medical education (SBME) represents a significant shift in how healthcare professionals develop both technical and non-technical skills. The use of AI in these environments has opened opportunities to make simulations more adaptive, personalized, and realistic, contributing to better training outcomes.

Komasawa and Yokohira (2023) discussed the evolution of SBME, which has historically been used to teach technical skills such as life support or surgical procedures. With the introduction of AI, simulations can now deliver highly responsive and dynamic training scenarios that adapt to the learner's actions in real-time. AI has enabled a higher degree of fidelity in simulations, meaning that learners can experience more lifelike situations and practice in conditions that closely resemble real-world clinical environments. This is particularly important as the field of healthcare education moves toward competency-based training, where achieving mastery of both technical and non-technical skills is critical (Komasawa & Yokohira, 2023).

Hamilton (2024) highlighted the role of AI in automating certain aspects of SBME, such as real-time feedback and data analysis. By collecting and analyzing large datasets, AI systems can track learner progress over time and provide individualized feedback tailored to the specific needs of each learner. This is particularly important in non-technical skills training, such as decision-making and communication, where AI can simulate complex patient interactions and offer insight into how learners handle different scenarios. However, as Hamilton (2024) points out, while AI enhances the training process, it also introduces challenges, such as the potential over-reliance on technology and the need to ensure that learners still develop critical thinking and clinical judgment skills (Hamilton, 2024).

Sun et al. (2024) examined how AI is reshaping the landscape of healthcare simulation education by offering more sophisticated tools for assessment and evaluation. AI can facilitate the automatic evaluation of technical performance by comparing learner actions to established clinical guidelines. This allows for objective, data-driven assessments of skills such as surgical technique, enabling educators to provide more targeted feedback and identify areas for improvement more efficiently. Sun et al. (2024) also highlight the importance of ensuring that AI-driven simulations complement, rather than replace, traditional instructional methods. They argue that while AI can provide valuable insights, human instructors are still necessary to interpret results, guide debriefings, and foster the reflective learning that is key to skill development (Sun et al., 2024).

In addition to technical skills, AI is proving to be a valuable tool in developing non-technical skills. Shankar (2022) emphasized the potential for AI to simulate complex social interactions, such as patient-doctor conversations or interprofessional teamwork, which are difficult to replicate in traditional simulations. AI-driven virtual patients, for example, can engage learners in realistic dialogue, testing their communication skills, empathy, and decision-making in ways that are more interactive than previous generations of simulation technology. Shankar (2022) also notes that AI can provide immediate feedback on these interactions, allowing learners to reflect on their performance and adjust their behavior in subsequent scenarios. However, the author cautions that AI should not be viewed as a perfect substitute for human interaction, as the nuances of empathy and understanding in medical practice are difficult for machines to replicate (Shankar, 2022).

The ethical and logistical challenges of integrating AI into healthcare simulation education cannot be overlooked. Komasawa and Yokohira (2023) noted that while AI-driven simulations offer immense educational benefits, they also present significant financial and technical barriers. High costs and the need for continuous updates to AI systems are major concerns for educational institutions, and the complexity of these systems can sometimes overwhelm learners and instructors alike. Moreover, the use of AI raises important ethical questions, particularly around data privacy, the potential for algorithmic bias, and the need to ensure that learners are not over-reliant on AI at the expense of developing their clinical reasoning skills. Ensuring that AI is used responsibly and ethically in medical education will be crucial to its long-term success (Komasawa & Yokohira, 2023).

AI is playing an increasingly pivotal role in healthcare simulation education, enhancing the realism, adaptability, and efficiency of training. The integration of AI into SBME has the potential to improve the development of both technical and non-technical skills, offering learners more personalized and interactive training experiences. However, as research is pointing out, AI must be implemented thoughtfully, balancing its benefits with the need to maintain critical human elements of medical education, such as empathy, critical thinking, and ethical judgment. As the field continues to evolve, educators and learners alike must remain adaptable, ensuring that AI complements, rather than replaces, the essential components of healthcare training.

3.2.1.2 Realism and fidelity in healthcare simulation-learning.

The importance of realism in healthcare simulations has long been recognized as a pivotal element in shaping the educational impact of these training tools. However, the intricate relationship between fidelity—how closely a simulation mirrors reality—and learning outcomes remains a nuanced and sometimes elusive concept, as various studies have explored this topic from different perspectives.

Rystedt and Sjöblom (2012) were among the early researchers who delved into the concepts of realism and authenticity in healthcare simulations. They highlighted how learners' perceptions of realism significantly shaped their engagement and, consequently, their learning achievements. Their work underscored the need for simulations to resonate with learners' experiences, suggesting that the perceived authenticity of a simulation could be as important as its factual accuracy in achieving educational goals (Rystedt & Sjöblom, 2012). However, this view is not universally accepted. Other scholars, such as Norman et al. (2012), challenged the common assumption that more realistic simulations automatically enhance learning (G. Norman et al., 2012). Their research argued that the key lies not in the visual or physical realism of a simulation but in how well it supports the cognitive processes of learners. They advocated for a focus on thoughtful educational design, where the purpose of the simulation dictates its level of realism, rather than an uncritical pursuit of accuracy.

Building on this discussion, Tun et al. (2015) warned that an overemphasis on realism, particularly in physical and behavioral aspects, could overshadow the crucial task of managing cognitive load (Tun et al., 2015). They suggested that the realism of a simulation should be carefully balanced with the cognitive capacities of learners and the specific learning objectives at hand, ensuring that the simulation remains a tool for education rather than a source of unnecessary complexity. This cautionary perspective is echoed by Massoth et al. (2019), who found that excessively lifelike simulations could lead to overconfidence among trainees without a corresponding improvement in actual clinical skills (Massoth et al., 2019). Their findings serve as a reminder that more realism is not inherently better and advocate for a measured approach that aligns the degree of realism with the specific needs of the learners. In the context of virtual patient simulations, Lee et al. (2020) explored how realism affects the effectiveness of training, particularly in enhancing communication skills (Lee et al., 2020). Their study emphasized the need for a delicate balance between creating an engaging, lifelike experience and avoiding cognitive overload, which could diminish the educational value of the simulation.

Similarly, Gonçalves et al. (2022) examined the broader influence of simulation fidelity on healthcare education. They argued that while realism is indeed important, it should not be viewed as the sole measure of a simulation's value (Gonçalves et al., 2022). Instead, they proposed a sophisticated approach where realism is fine-tuned to serve the intended learning objectives, ensuring that high-fidelity simulations are truly beneficial. Adding further nuance to this discussion, Carnell et al. (2022) investigated the role of interaction fidelity in virtual patient simulations. Their findings suggested that while higher fidelity could enhance engagement and the sense of realism, it did not necessarily translate into better knowledge retention or skill acquisition. They emphasized the need to pair realism with sound educational strategies to ensure it supports, rather than detracts from learning (Carnell et al., 2022).

More recently, Davies and Krame (2024) contended that the precision required for effective training does not always equate to achieving the highest level of realism. They suggested that while immersive, high-fidelity simulations can be captivating, they do not necessarily lead to better learning outcomes, especially when the focus is on developing cognitive skills rather than procedural tasks (Davies & Krame, 2024). The concept of realism in healthcare simulations is complex, intertwined with learner engagement and educational outcomes. While high-fidelity simulations offer immersive experiences that boost engagement, their actual impact on learning is not always straightforward. The challenge lies in striking the right balance of realism, one that aligns with a high level of engagement, specific educational goals, and the cognitive capacities of learners, ensuring that simulations are not just realistic but also educationally effective.

3.2.1.3 Multi-modal interaction in healthcare simulation-based learning

The application of voice interaction in virtual patient simulation (VPS) is generating significant interest because of the capacity of conversational agents (CAs) to facilitate synchronous, interactive communication. In their recent publication, Sezgin and D'Arcy (2022) critically examine the growing adoption of voice technologies and conversational agents within the healthcare sector, with a specific emphasis on remote treatment and patient-generated data. Although the paper focuses on voice assistants for health monitoring, specifically for documenting symptoms and medications, it also highlights the potential for future integration into virtual patient systems (Sezgin & D’Arcy, 2022). The capacity of these systems to adjust to intricate conversations, facilitated by developments in natural language understanding, is an essential component for enhancing the authenticity of simulations.

The incorporation of automated speech recognition into simulations has great potential. According to Wang et al. (2023), conversational dynamics play a crucial role in the development of communication skills, particularly in intricate diagnostic situations. Their research illustrates that including voice interactions in VP simulation enables learners to engage in the application of medical knowledge as well as the interpersonal elements of patient care, therefore enhancing their overall skill set (X. Wang et al., 2023). Through the use of spoken conversations, these systems enhance the simulation of real-life clinical engagements for learners.

Alario-Hoyos et al. (2020) provide more evidence by examining the ways in which multimodal interactions, such as speech, enhance the learner's experiential learning. Their contention is that these approaches facilitate the cultivation of essential soft skills, such as empathy and patient management, that are crucial in medical education. Their results indicate that including voice-based interactions into virtual simulations enables learners to engage in more authentic patient discussions, therefore improving their readiness for real-life scenarios (Alario-Hoyos et al., 2020).

A separate study by Carbonell and Siekmann (2007) examined the application of intelligent virtual agents and how their capacity to engage through voice augments their function in simulations. Their study indicates that intelligent virtual agents have the capacity to replicate emotional and conversational expressions, so providing a more genuine patient experience that aids in the development of learners' skills in handling nuanced emotional and clinical conversations (Carbonell & Siekmann, s. d.). Bringing it further, Sun et al. (2024) highlight the adaptable capacity of voice interactions in voice-activated speech recognition (VPS), underscoring the ability of machine learning algorithms to dynamically modify the patterns according to voice inputs. The increased adaptability of simulations enhances their fluidity and alignment with the specific requirements of individual learners, therefore greatly enhancing the instructional efficacy of these tools (Sun et al., 2024).

Notwithstanding these advancements, Rizzo and Bouchard (2019) emphasize persistent technical obstacles, namely regarding the precision of voice recognition platforms. The importance of culturally sensitive and context-aware voice systems in medical training simulations is underscored. An essential requirement for establishing an inclusive and realistic training environment is the implementation of tailored voice recognition technology capable of comprehending regional dialects or medical terminology (Rizzo & Bouchard, 2019, p. 20). Stevens et al. (2006) previously examined virtual patients as instructional aids for developing communication abilities, particularly exploring the employment of voice interfaces to augment the authenticity of clinical scenarios. Their findings indicate that although voice interaction enhances the level of immersion in patient-doctor simulations, user happiness is impacted by technical issues, such as failures in speech recognition. Although Stevens et al. contend that these problems restrict the whole capabilities of VPS, they also indicate that continuous progress in speech recognition may soon alleviate these problems (Stevens et al., 2006).

The incorporation of voice interaction into virtual patient simulation is changing medical education by enhancing the authenticity and flexibility of patient simulations. From improving communication abilities to more effectively equipping learners for real-life encounters, voice-based solutions provide significant advantages. Nevertheless, remaining obstacles to be surmounted include the accuracy of speech recognition and the requirement for social adaption. Furthermore, to the extent of our understanding, there has not been a thorough evaluation of user experience and possible educational advantages between voice-interactive virtual personal assistants (VPS) and comparable, text-based interaction systems. Comparative studies of this nature are crucial for comprehending the significance and efficacy of voice interaction. With the advancement of technology, it is essential to acquire more profound understanding of whether voice interaction offers significant benefits compared to conventional approaches, particularly in terms of engagement, retention, and overall instructional results.

3.2.1.4 User experience in healthcare simulation-based learning

The efficacy and experience of simulations are intricately connected to the authenticity and quality of the virtual environments, which have a substantial impact on users' cognitive and clinical abilities. Adequate replication of patient characteristics and interactions in high-fidelity simulations is crucial for ensuring a safe and efficient practice environment (Daher et al., 2022). Research highlights a robust relationship between the accuracy of simulations and successful learning results (Issenberg et al., 2005). Implementing tangible components, such as tactile feedback, into virtual simulations amplifies user immersion and involvement. Research has demonstrated that individuals who interact with simulators that include passive-haptic cues accomplish tasks with more efficiency and fewer difficulties compared to those who do not (English et al., 2010); Tangible interactions greatly improve the user experience by increasing the authenticity and immersion of the virtual environment. Technological progress in computing power, graphics, and AI has significantly improved the realism and user-friendliness of VP simulators (« Emerging Roles of Virtual Patients in the Age of AI », 2019). In virtual simulations, the user experience extends beyond mere usability, it includes dimensions of engagement and happiness. Experimental evidence has demonstrated that the immersive quality of VR simulations promotes a heightened feeling of presence, which is a crucial element in improving user involvement and educational achievements (North & North, 2016). Furthermore, the capacity of virtual reality (VR) settings to provide multi-sensory inputs enhances the enjoyment and effectiveness of the training experience (Sadowski & Stanney, 2002). Within the realm of nursing education, virtual reality simulations have proven to be highly efficient in fostering self-directed learning and equipping students for clinical placements with competencies. Survey participants express a significant degree of contentment with these simulations, as they perceive them to be both educational and pleasurable, indicating that VP simulations can have a key function in connecting theoretical understanding with real-world implementation in healthcare education (Saab et al., 2023). One of these competencies, effective communication, is widely acknowledged as a vital skill for delivering safe and high-quality medical care, making it a key social dimension to enhance through simulation-based training, especially with VP (Battegazzorre et al., 2021; Sheldon & Hilaire, 2015; Stevens et al., 2006). Despite this, and to the best of our knowledge, no study has yet examined the impact of an AI-powered, voice-only interaction mode compared to a multiple-choice approach in a diagnostic context with virtual patients. More broadly, the influence of voice-based interactions (both input and output) on user experience in virtual patient simulations has also received limited attention in the research literature. As technology advances, it is crucial to address the gap in the literature concerning the intended realism of these functionalities and their actual effects on the healthcare practitioners' learning experience. This understanding is especially important before such technologies are broadly implemented for teaching communication skills and to assess how changes in interaction modality may influence the processes and unique challenges within medical simulations.

3.2.2 Theoretical framework

3.2.2.1 Media equation theory

The Media Equation Theory (MET), originally proposed by Reeves and Nass (1996), posits that users interact with computers and various forms of media as if they were engaging with real human beings. This theory highlights the tendency of individuals to attribute human-like qualities and social behaviors to media technologies (Reeves & Nass, 1996). Among the eight propositions of the MET, Reeves and Nass (1996) argued that user responses are "social and natural" and occur automatically without conscious effort. The phenomenon of attributing human characteristics to computers and other media is extensively documented in existing literature (Friedman et al., 2003; Reeves & Nass, 1996; Torta et al., 2013). Human users often make social attributions towards computing systems that possess anthropomorphic traits, behaving as if they are engaging with other humans (Ma, no date). Research has shown that people may employ politeness even when assessing computers, indicating a tendency to treat them in a similar manner as they would interact with humans (Heyselaar et al., 2017). The phenomenon of anthropomorphizing technology, namely computers, is distinct, since individuals who may not anthropomorphize other lifeless items tend to attribute human-like characteristics to computers (Chin et al., 2004). Moreover, a study conducted by Kulms and Kopp (2018) revealed that social attributions connected to affection, which enhance trust in human-human interactions, additionally contribute to increased trust in computers (Kulms & Kopp, 2018, 2019). In general, the literature on attributing human characteristics to computers emphasizes the intricate manners in which people perceive and engage with technology, creating a blurred line between human and machine interactions. In the context of our study, where the goal of the simulation is to closely replicate real-life scenarios to facilitate effective learning, the Media Equation Theory allows us to predict that employing vocal interactions with virtual agents could enhance the user experience by imitating genuine human communication. By leveraging the principles of MET, we anticipate that users will perceive these interactions as more natural and engaging, thereby increasing their emotional connection and overall satisfaction with the virtual agents. This increased engagement and realism can help bridge the gap between simulated and real-world experiences, making the learning process more relevant and impactful for users. Furthermore, our study investigates the disparities between two simulation environments: an immersive desktop virtual reality (VR) setting and a minimal simulation employing a large language model (LLM). Through the application of Media Equation Theory (MET), we aim to investigate and elucidate how and to what extent both of these simulation environments, with their ability to replicate realistic human interactions, may foster a more positive user attitude toward technology.

3.2.2.2 Cognitive load theory

The Cognitive Load Theory (CLT), developed by John Sweller in 1988, underlines the significance of successfully managing cognitive load in order for improved learning outcomes (Sweller, 1988). CLT distinguishes between three types of cognitive burden: extraneous cognitive load, which is influenced by how the work is presented; germane cognitive load, which is focused on processing and comprehending the information; and intrinsic cognitive load. In terms of CLT, intrinsic cognitive load (Chandler & Sweller, 1991; Orru & Longo, 2019) is one of the three types of cognitive load experienced by the user, and it refers to the amount of complexity associated with novel information. The concept of Ease of Use, that is a fundamental element in the Technology Acceptance Model (TAM) (Davis, 1989), has traditionally been examined in existing literature to understand this complexity. When examining virtual agents within the framework of cognitive load theory, it is essential to fully understand the potential impact that these agents have on cognitive processes. Virtual agents, such as pedagogical agents in digital education settings, have the ability to impact cognitive load by influencing learners' attention and processing resources (Clark & Choi, 2005; Huang et al., 2020; Liew et al., 2017). For instance, the virtual agent's eagerness might lead learners to experience increased mental effort as they have to draw attention to excessive nonverbal cues (Liew et al., 2017). Likewise, the inclusion of educational agents when combined with other visual stimuli can result in divided attention and impede the process of acquiring knowledge (Clark & Choi, 2005). Additionally, the utilization of virtual reality technology might also impact cognitive strain in learning settings. Research has indicated that in immersive virtual reality settings, the feeling of being present may increase cognitive load by captivating users' attention (Andersen et al., 2016). Nevertheless, the incorporation of design concepts based on cognitive load theory into virtual reality simulation training has been shown to have a beneficial effect on learning outcomes. This highlights the significance of taking cognitive load into account when designing virtual learning environments. Ultimately, cognitive load theory offers a beneficial paradigm for comprehending the impact of instructional design features, such as virtual agents and virtual reality technologies, on cognitive processes. Our study emphasizes the importance of CLT as it offers a framework to create simulation settings that reduce unnecessary cognitive load, enabling users to focus on acquiring crucial skills. In addition, our study examines how various simulation environments affect cognitive strain. The desktop VR setting offers a highly engaging experience with its intricate and lifelike simulations. However, it may need more mental effort, known as intrinsic cognitive load. On the other hand, the LLM-based simulation, although it may be less immersive, could have a reduced cognitive load in general, making it easier for people to access. By applying the principles of Cognitive Load Theory (CLT), we aim to investigate and elucidate how adjusting the cognitive load balance across different settings may enhance user experience and improve learning outcomes, ultimately fostering a more positive attitude toward technology, thus a meaningful learning experience.

3.2.3 Research model

Through the Media Equation Theory and the Cognitive Load Theory lenses, our objective is to illustrate that engaging in vocal interactions with virtual entities in diverse simulation environments can provide enhanced learning experiences by diminishing cognitive load and enhancing user engagement. The sequence of constructs in this theoretical framework may seem unusual to reviewers, as it deviates from the traditional Stimulus-Organism-Response structure, where physiological responses usually precede perceptual reactions (Jacoby, 2002). Nevertheless, in this study, perceptual assumptions are represented as precursors to physiological responses for two key rationales. First, this approach enables the development of theories on the connection between independent variables (interaction modality and simulation environment) and perceptual responses. This connection will be expanded upon in the next sections. Secondly, placing mental concepts before physical reactions strengthens the generalization of the findings, indicating that the observed effects can apply to other technologies that aim to create accurate system responses, genuineness, and social involvement. The theoretical framework created for this investigation is depicted in Figure 1, showcasing the hypotheses and anticipated associations among the variables of our study:

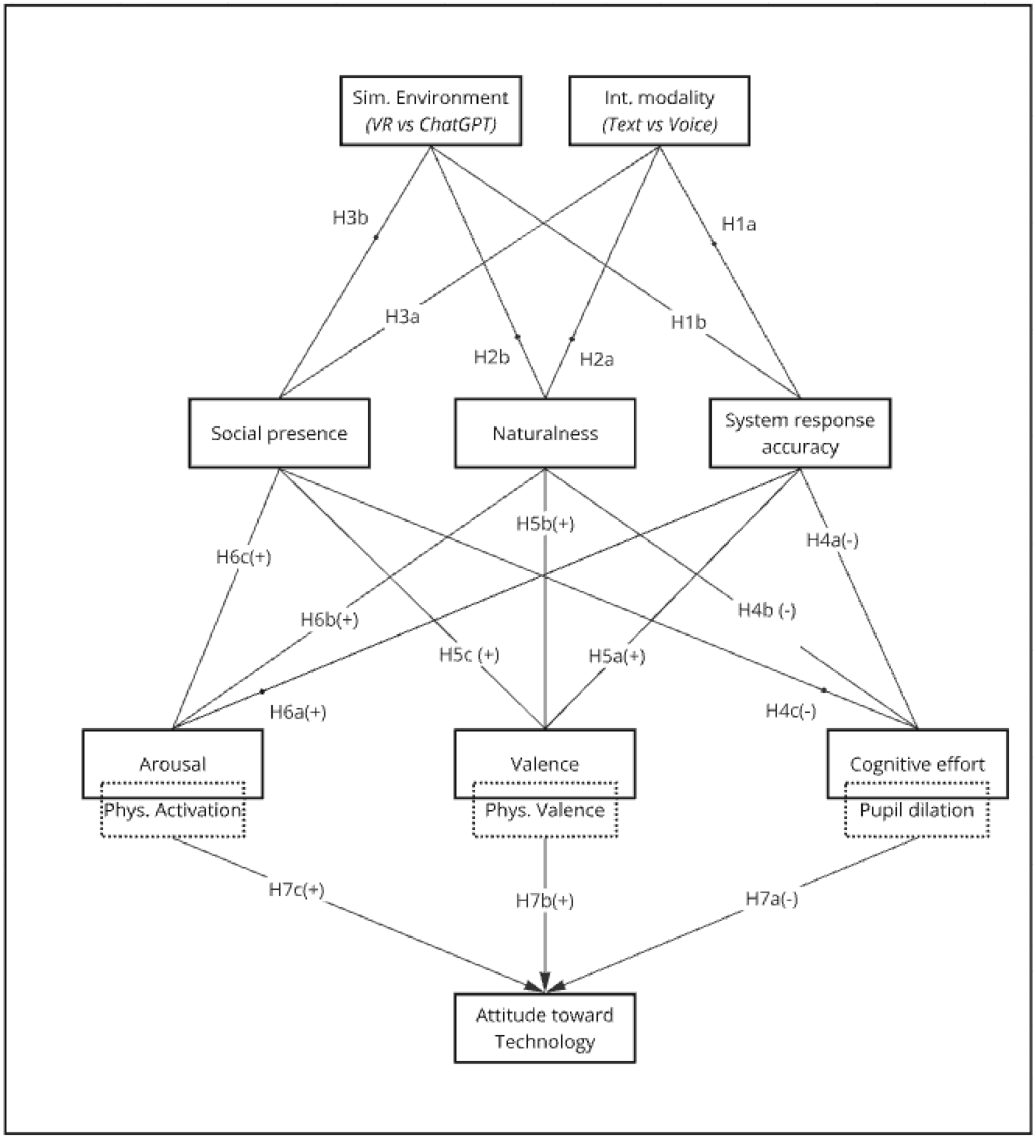


Figure 1

Research model

3.3 Hypothesis development

3.3.1 System perception

As mentioned above, MET focuses on the influence of established social norms and behaviors between humans on users' interactions with machine interfaces, often without them being aware of it (Reeves & Nass, 1996). It assumes that interactions with technology are intrinsically social, even when the user is aware that the technology has no human characteristics. System response accuracy is defined as the user's assessment of the system's ability to perceive and respond to their data (K. S. Hone & Graham, 2000). Naturalness, in the context of interaction, is defined as the flow of conversation and how similar or dissimilar it feels to an in-person interaction (J. Tam et al., 2012). In turn, Social presence can be defined as the degree to which a medium allows an individual to establish a personal connection with others (Kumar & Benbasat, 2006). Placing these three constructs in the MET lens, it is then possible to elaborate our hypotheses regarding the perception of the simulation system. Voice communication fosters emotional engagement, as it conveys warmth and empathy, building trust and satisfaction, and reinforcing belief in the system's accuracy (Kroes et al., 2022). Studies also show voice input is faster and more reliable, especially in multitasking or noisy environments, enhancing users' perception of the virtual agent's competence and accuracy (Ferrell et al., 2022; Ruan et al., 2016). Speaking, a natural human activity, makes virtual agents more relatable and engaging. Literature proved extensively that users prefer lifelike voices for the familiarity and ease they provide (Abdulrahman & Richards, 2022; Emna Chérif, Jean-François Lemoine, 2019). Voice-based interaction enables smoother exchanges, enhancing perceived system understanding and response accuracy. Context also plays a role; tasks requiring immediate feedback, such as clinical simulations, benefit from voice interactions, which are often seen as more direct and effective (Torre et al., 2020). Additionally, anthropomorphism enhances the perception of accuracy, with human-like interfaces capturing attention and fostering emotional bonds that increase user compliance and trust in agent recommendations (Burgoon et al., 2000; Matsui & Yamada, 2019; Nass et al., 1994; Waytz et al., 2010). Therefore, for the effect of the interaction mode as well as the authenticity level projected by the simulation environment - with or without visual support - we hypothesized that :

H1a: *The use of voice interaction modality will result in a higher perception of system response accuracy compared to the text interaction modality.*

H1b: *VR simulation will result in a higher perception of system response accuracy compared to the ChatGPT simulation.*

H2a: *The use of voice interaction modality will result in a higher perception of naturalness compared to the text interaction modality.*

H2b: *VR simulation will result in a higher perception of naturalness compared to the ChatGPT simulation.*

H3a: *The use of voice interaction modality will lead to a higher perception of social presence compared to the text interaction modality.*

H3b: *VR simulation will lead to a higher perception of social presence compared to the ChatGPT simulation.*

3.3.2 User cognitive and emotional experience

According to Cognitive Load Theory (Sweller, 1988), Intrinsic Cognitive Load (Chandler & Sweller, 1991; Orru & Longo, 2019), one of the three types of cognitive load experienced by the user, corresponds to the level of complexity of new information. Said complexity has typically been studied in extant literature through the notion of Ease of Use, a core construct in the Technology Acceptance Model or TAM (Davis, 1989). According to TAM, Ease of Use has been shown to be to the level of effectiveness perceived by the user during their interaction with the system. Effectiveness, itself, is defined as the accuracy and completeness with which users achieve specified goals (ISO 9241). The quality of system output may then be observed in the level of system response accuracy. In turn, Extraneous Cognitive Load refers to the mental effort imposed on a user that is unrelated to the task or learning material itself, but rather stems from the way the information is presented or how the interaction is structured (Sweller et al., 2011). Naturalness in interaction refers to how closely the system mimics real-world human interaction (J. Tam et al., 2012). As per Nass's work, users tend to interact with systems as they would with other humans, expecting intuitive and fluid exchanges (Nass & Moon, 2000). When a system's interaction feels natural—whether through voice, gestures, or conversational flow—the cognitive effort required by the user decreases, as the mental load associated with translating their thoughts into system-understandable actions is reduced. Social Presence, defined as the extent to which a system enables users to feel as though they are interacting with a real, socially present entity (Kumar & Benbasat, 2006), has been also shown to reduce cognitive load in user-system interactions (Hassanein & Head, 2007). When users perceive a higher degree of social presence, they are more likely to interact with the system as they would with a human counterpart, which leads to a more natural and intuitive interaction. The familiarity and emotional engagement provided by high social presence and naturalness reduce the need for users to consciously think about how to interact with the system, lowering the cognitive resources required for task completion. Both of these variables effects are also supported by CLT where reducing the extraneous cognitive load—such as interpreting a complex or unfamiliar interface—frees up resources to focus on the task itself. In highly social systems, users expend less cognitive effort trying to figure out how to engage with the system, allowing them to process information more efficiently and comfortably (Chandler and Sweller, 1991). Hence, we hypothesized that :

H4a: *Higher system response accuracy leads to lower cognitive effort*

H4b: *Higher naturalness leads to a lower cognitive effort*

H4c: *Higher social presence leads to a lower cognitive effort*

According to Media Equation theory (Reeves and Nass, 1996), users respond to media in a similar manner as they do to other people in accordance to the stimuli or cues they receive from the media. The attribution of human characteristics to computers and other media is well documented in extant literature (Cornelius & Leidner, 2021). Among the eight propositions of the media equation, Reeves and Nass (1996) argued that user responses are "social and natural" and occur automatically without conscious effort. Social presence and naturalness are key constructs in human-computer interaction that significantly influence users' emotional and physiological responses. Social presence, as described by Head and Hassanein, refers to the degree of perceived human-like connection in interactions, which can enhance both emotional valence (positivity) and arousal (intensity) by creating a sense of interpersonal warmth and engagement (Head & Hassanein, 2002). Naturalness, rooted in Nass's Media Equation Theory, involves how closely a system mimics human-like behavior, and higher naturalness leads to smoother interactions, increasing positive emotional responses and engagement by reducing cognitive load (Nass & Moon, 2000). Together, these constructs suggest that systems designed with high social presence and naturalness are likely to evoke stronger, more positive emotional responses in users, influencing both their valence and arousal levels. Hence, it is hypothesized that:

H5a: *Higher system response accuracy leads to higher valence*

H5b: *Higher naturalness leads to higher valence*

H5c: *Higher social presence leads to higher valence*

H6a: *Higher system response accuracy leads to higher arousal*

H6b: *Higher naturalness leads to higher arousal*

H6c: *Higher social presence leads to higher arousal*

3.3.3 Attitude toward technology

Cognitive Load Theory (Sweller, 1988) proposes that cognitive effort refers to the mental resources required to process information. High cognitive effort can negatively impact user experiences, as it leads to increased mental fatigue and frustration (Paas et al., 2003). The Technology Acceptance Model (TAM) (Davis, 1989) highlights that perceived ease of use, closely tied to lower cognitive effort, positively influences attitudes toward technology. According to TAM, when users experience less difficulty in interacting with technology, they tend to develop more positive attitudes, as the system is perceived to be less cognitively demanding (Venkatesh & Davis, 2000). This relationship is particularly evident in task-complexity studies, where systems that reduce cognitive load have been shown to improve overall user satisfaction and adoption (Orru & Longo, 2019). Hereby, we hypothesize that :

H7a: *Lower cognitive effort leads to higher (better) attitude toward technology*

Emotional valence refers to the degree of positive or negative feelings experienced during interactions with a system (Russell, 1980). Positive valence—experiencing more pleasant emotions—has been linked to higher satisfaction and a more favorable attitude toward technology (D. Norman, 2007). Individuals often rely on their emotions to evaluate products, and more positive emotions lead to more positive evaluations (Schwarz & Clore, 1983). In TAM, emotional responses, especially positive ones, reinforce the perceived usefulness and ease of use of a system, thereby enhancing the user's attitude toward adopting the technology (Zhang, 2013). Studies in affective computing have shown that users with a more positive valence are more likely to have a favorable view of the technology they interact with, leading to higher acceptance rates (Picard, 2000). Thus, we hypothesize that :

H7b: *Higher valence leads to higher (better) attitude toward technology*

Arousal refers to the level of excitement or alertness a user experiences when interacting with a system (Lang, 1994). The circumplex model of emotion (Russell, 1980) defines arousal as one of the two dimensions (valence and arousal) used to describe emotional states. High arousal, when positive, can contribute to a more stimulating and engaging user experience, enhancing attitudes toward technology (Agarwal & Karahanna, 2000). For example, in flow theory (Csikszentmihalyi, 2013), optimal arousal is a key condition for users to experience "flow," a state where they are fully immersed and satisfied with their interaction. Users who experience higher arousal during interactions are more likely to perceive the technology as enjoyable and engaging, which, according to the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003), can lead to higher acceptance of the technology. This lead to our last hypothesis, which is :

H7c: *Higher arousal leads to higher (better) attitude toward technology*

3.4 Methodology

3.4.1 Approach

For this study, we specifically selected the healthcare field of nursing clinical decision-making as the backdrop to examine our hypotheses. The distinctive complexity of nursing practice, which necessitates a higher level of critical thinking and problem-solving skills, served as the driving force behind the decision; The process of nursing clinical decision making encompasses the evaluation, identification, and treatment of patient care (Higgs et al., 2008; Young et al., 2020), rendering it a suitable domain for assessing the effectiveness of our proposed model. Our objective is to gain a deeper understanding of the cognitive and emotional processes at play in that particular area (i.e., patient/agent interaction-dependent diagnostic/decision-making) and explore ways to improve them through appropriate interventions. This methodology not only enables us to evaluate our hypothesis in a realistic setting but can also be applied to decision-making simulations in a broader sense (e.g., other fields).

3.4.2 Sample characteristics

A total of twenty-four participants issued from convenience sampling, all of whom were practicing nurses or nursing students, took part in this study. Including both practicing profesionnals and students enriches the study by merging real-world insights with different perspectives. Four participants were omitted from the analysis due to beyond our control technical issues encountered during one or more online conditions within the experimental procedure thus invalidating their data. Consequently, the final sample comprised twenty participants (17F 3M; Age: Mean = 26, Median = 25, StdDev = 6.08, Range: 19–47). Among the participants, four were full-time students without any prior work experience. Eight participants reported one year of professional experience, whereas the remaining twelve had between two to six years of work experience in the field. Participants were recruited through a targeted approach involving specialized groups on social media platforms and through a collaborative effort with a university-affiliated establishment, which circulated a recruitment notice to the population of interest. Prospective participants were required to register their interest through an online Qualtrics form. Following this, eligible candidates matching the selection criteria were identified and subsequently contacted to arrange their participation in the study. Participants eligible for the study were required to meet specific criteria: (1) be at least 18 years old, (2) work in nursing or be a nursing student with at least 1 year of curriculum completed, (3) possess a proficient understanding of English, spoken and written, (4) Be able to work on a computer without corrective glasses, (5) Have no skin allergies or particular sensitivity, (6) Have not had laser vision correction or astigmatism, (7) Not have a neurological or psychiatric diagnosis, (8) Not suffer from epilepsy. These requirements were verified through a written informed consent form before the experiment began.

Table 1
Pre-experiment questionnaire summary

Variable	*N*	*Mean*	*Median*	*Std Dev*	*Minimum*	*Maximum*
Age	20	26	25	6,08	19	47
Prof. Exp.	20	1,65	1	1,57	0	6
Eng. level	20	5,22	5,00	1,15	3,33	7
Sim. level	20	4,9	5,125	2,01	1	7

3.4.3 Experimental design

Table 2

Experimental conditions

Sim. Environment	*Embody Desktop VR*		*ChatGPT Role-play*	
Condition	**A**	**B**	**C**	**D**
User comm. modality	Proposed choices selection	Vocal	Free text entry	Vocal
Agent output modality	Textual	Textual & Vocal	Textual	Vocal

To investigate the influence of GenAI-powered voice interaction and visual environment on users' attitudes toward technology in a communicative role-playing problem-solving simulation learning context, we conducted a laboratory experiment. The study employed a 2 (Desktop VR environment: A) choice-based vs. B) voice-based virtual patient interaction) x 2 (ChatGPT

(OpenAI) environment: C) text-based vs. D) voice-based virtual patient interaction) within-subjects experimental design (Figure X above). The simulation tool Embody, which features a virtual patient and is designed for nursing training, was utilized to evaluate desktop virtual reality conditions. Under condition A, participants interacted with the virtual patient by choosing from predetermined options and receiving written feedback from him. In the second condition, labeled as B, participants had to engage in oral communication with the virtual patient by pressing a button. They also received audio responses from the agent, in addition to the written response. To simulate nurse-patient interaction, a prompt was created in ChatGPT (OpenAI) Ipad (Apple, 2023) application for the two remaining conditions, C and D. The 2 integrated communication modes with the application agent were used: text communication (C) and voice communication (D). Each participant was exposed to the four conditions in a predetermined, counterbalanced order. Conditions A and B, as well as conditions C and D, were always presented consecutively. However, the order of these pairs (A-B and C-D) and the sequence of conditions within each pair were randomized.

3.4.4 Experimental procedure

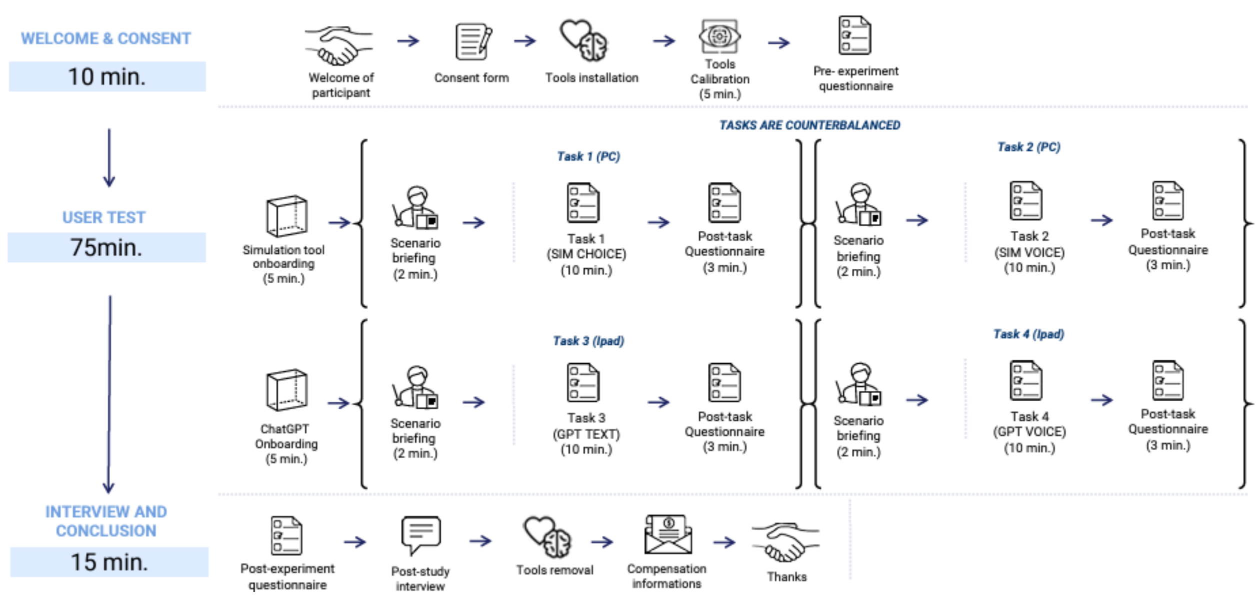


Figure 2

Experimental procedure

The experimental procedure commenced when participants arrived at the laboratory and completed the informed consent process, as shown in Figure 2. We then fitted them with physiological measurement instruments. Participants were then seated at a desk, facing either a computer or an iPad, depending on their randomly assigned condition. Next, the psychophysiological instruments were calibrated. Participants completed a pre-experience questionnaire, followed by an onboarding tutorial for each condition group. Each simulation condition was also introduced with a text briefing and a short overview of instructions by the research team. Participants were instructed that time was not a factor and to act as they would in a real medical situation. Each condition could last up to 10 minutes, after which the research team would conclude the condition, although participants could end the condition earlier if they felt their medical practice was complete. After each condition, participants completed the same post-task questionnaire. At the end of the 4 conditions, a post-experience questionnaire of a ranking question was also administered. A final 10-minute interview was conducted to gather more insights into their experiences across different experimental conditions. After physiological tools removal, participants received a compensation of $150, and the entire experiment lasted approximately 120 minutes. The experiment was approved by the ethics committee of our institution under certificate no. 2023-5393.

3.4.5 Experimental stimuli and manipulation

The desktop VR conditions were conducted using the cloud-based healthcare simulation software "Maestro Embody" (V1.5.1), which features interactive modules such as Assessment, Monitoring, Procedures, Treatments, Charting, Action Log, and Case. These modules enable comprehensive practice of nursing scenarios. The simulation environment also includes a virtual 3D representation of a hospital room with interactive elements like the bed, computer, monitor, and patient for auscultation tasks, and allows for changes in the viewing angle. The clinical scenarios, drawn from the software's library, are typical nursing assessments and ensure consistent interactions. The software was installed on a computer, equipped with keyboard, mouse, speakers and microphone. For condition A, a cardiac nursing assessment scenario was selected, while for condition B it was an asthma nursing assessment scenario. Manipulation, in this condition group, consisted in changing the modality of interaction with the virtual patient. In the condition A, choice-based, learners interacted with the virtual patient using a predefined list of interactions crafted by medical scenario experts, selecting responses that triggered immediate textual replies from the virtual patient. The interaction list dynamically updated based on scenario progression and past choices. In condition B, voice-based, learners communicated by pressing a button to activate the microphone and vocalizing their responses, which were processed by a speech recognition system. The system then provided consistent responses in text and audio form through text-to-speech output. While the voice interaction mode was tested in a development environment and is not yet publicly available, the question-choice interaction mode is available to the public with identical models.

Figure 3

Condition A – Embody Clinical Simulation Choice-based interaction

Figure 4

Condition B – Embody Clinical Simulation Voice-based interaction

The ChatGPT conditions were conducted using ChatGPT 3.5 (OpenAI, 2023) application on Ipad (Apple, California). The Ipad was equipped with a keyboard with a trackpad to be as close as possible to the computer setting. The 2 conditions, C & D, were built on the basis of the cardiac and asthma condition contained in conditions A & B, but varying the causes, social history, symptoms, etc., of each condition. Manipulation, in this condition group, consisted also in changing the modality of interaction with the virtual patient. In condition C, participants communicated with the virtual patient via a text box, receiving responses in text form. Conversely, in condition D, participants interacted solely through voice communication, with both inputs and responses being exclusively vocal, without textual support. Additionally, the ChatGPT (OpenAI) environment introduced a second manipulation when compared to conditions A & B, as it lacked the visual elements present in the desktop VR counterpart.

3.4.6 LLM prompt-engineering

For experiment conditions C & D, a prompt was engineered on ChatGPT 3.5 (OpenAI, 2023) to enable the simulation of a nursing patient assessment through role-play with the Large Language Model (LLM), in a manner substantially identical to what was possible in the VR simulation environment used for conditions A & B.

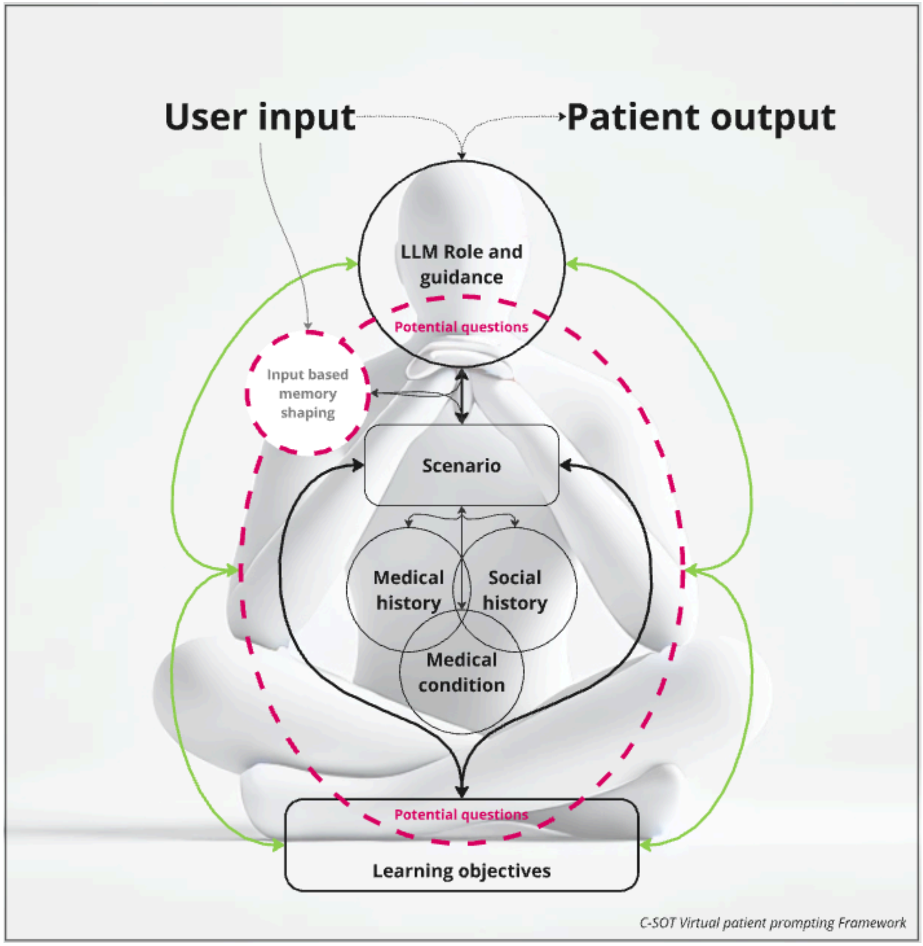


Figure 5

C-SOT-Virtual Patient prompt framework

Reflecting design thinking principles, prompt engineering involves vision, planning, testing, and refinement, forming an iterative cycle that enhances LLM outputs through continuous prompt evaluation and adjustment (Cain, 2024, p. 202). Building on medical learning literature and the latest prompting theories, we designed in an iterative manner a Constrained-Skeleton-of-Thought Virtual-Patient (C-SOT-VP) prompt (Figure 5).

A Virtual Patient (VP) is a specialized computer program that mimics authentic clinical cases, allowing learners to assume the role of healthcare providers for history taking, physical examination, and decision-making in diagnosis and therapy (Sahin Karaduman & Basak, 2024).

VP tools are distinctively beneficial, fitting, and economical in fostering the progression of clinical reasoning—a competence in expert practitioners that arises from non-analytical thinking and is enhanced by repeated exposure to diverse clinical contexts (Cook et al., 2010). Transposing this idea to LLMS, we opted for a role-play prompt that we designed as optimally as possible within the time constraints imposed by the study. Role-play prompts require students to act out different roles or scenarios, which may be based on actual or imagined situations (Heston, 2023).

The main objective of our LLM role-play prompt design was then to enable the participant to sustain a conversation with a VP while having access to the necessary technical medical data to support their clinical reasoning when and if needed (e.g. heart rate, blood pressure, lab results etc.). We aimed to facilitate quick development of a medical case scenario by educators or learners themself, centered on crucial data and information ; Allowing fast input and relying on the LLM capability for generating ancillary details.

Derived from medical education literature, we start by drawing a pedagogical framework (Figure 5), the living core section, containing 5 main contextual organs of our C-SOT-VP : Learning objectives, Scenario, Past medical history, Social History & Medical condition that can be used to delimit and establish a sufficient basis for a role-play clinical reasoning activity (Ackley & Ladwig, 2008; Dougherty & Lister, 2015; Sittner et al., 2015).

We first and foremost attempted during this iterative process to adhere to the 5 major key principles of prompt engineering, namely: Direction, Format, Examples, Evaluation, and Division (*Prompt Engineering for Generative AI*, s. d.). Since Some data show that guidance is more effective than the presentation of examples (Hsieh et al., 2023), we priorised this approach in the construction of our prompt, preferring to guide and coach it in its behavior than to give it an example of a dialogue unfolding .

In general prompt engineering, Direction is applied by ensuring that instructions are clear, such as specifying tasks like “list” or “summarize.” Format involves structuring prompts consistently, often through the use of templates, like framing a question as “What are the key benefits of…”. While we sometimes incorporate Examples, such as beginning a prompt with “For example…” to illustrate the desired response, we primarily use them selectively. Evaluation plays a crucial role as we regularly assess the effectiveness of prompts by comparing outputs to predefined criteria, ensuring accuracy and relevance. Finally, Division is implemented by breaking down complex prompts into smaller, more manageable parts, like separating questions into individual prompts to maintain clarity and focus.

Starting from the basis of a role-play dialogue between 2 agents and driven by the Skeleton-of-thought (SoT) methodology, we then added a block composed of a series of potential questions or interactions that could be raised during the conversation by the user. Inspired by human thinking processes, the concept of SoT involves guiding LLMs to first derive a skeleton of the answer and then complete each point in parallel, aiming among others to improve answer quality, leading to more diverse and relevant answers for a wide range of questions (Ning et al., 2023).

Staying into the anatomy analogy, we've then added a head to our tool by *pre-warming* him with context as to its purpose and use (Liu et al., 2022), explaining the role the AI will have to play in this simulation . By fostering deeper contextual insights, role-play prompting acts as an indirect trigger for Chain-of-Thought, thus refining the reasoning process (Kong et al., 2023).

Collectively, the core and the skeleton of potential questions establish the boundaries for the simulation. This constrains the dialogue within a network of elements that are intricately connected to the predetermined scenario, thus with the objective of enhancing the quality of responses within the specific context of our role-playing exercise. In our particular case, the aim was to surround the pedagogical framework with such a skeleton and thus "bring to life" the virtual patient story.

To further structure and circumscribe the LLM's field of action and research, we thought it worth including under objectives that the simulation was intended to put the National Council Licensure Examination (NCLEX) criteria into practice. NCLEX is a mandatory standardized test for nursing graduates in the United States and Canada, designed to evaluate their readiness for nursing licensure by testing their essential knowledge, skills, and abilities for safe and effective nursing practice. Semantic embedding in LLM prompt engineering involves incorporating meaningful, context-rich cues into prompts to guide the model's understanding and response generation. Using this technique thus leverages the model's ability to understand and process natural language at a deep semantic level, enabling it to generate more accurate, relevant, and contextually appropriate responses.

To ensure the AI can provide data to the user immediately during a conversation (i.e. blood pressure, heart auscultation etc.), enabling seamless diagnosis, we introduced an example sentence along with a detailed outline of how the response should be formatted.

Finally, to address the issue where the AI hesitated to provide information, claiming it was "only a simulation" we introduced a final enhancement, “Precision” section to expand and improve the creativity of the dialogue, ensuring and bringing to life a more engaging and informative interactions.

3.4.7 Measures

Prior to the first condition, the participant was asked to complete a pre-task questionnaire including a few demographic questions and others about previous simulation and language experience. At the end of each condition, participants were presented with a Qualtrics questionnaire (Qualtrics, Provo, UT) aimed at capturing a multidimensional view of user interaction and response according to our theoretical model. A post-experiment questionnaire was then administered to rank the conditions according to perceived learning quality. These post-tasks questionnaires were complemented by physiological measures for a subset of these variables and in certain conditions as well as a short open-ended interview to obtain a richer understanding of the participants' responses.

Table 3
Summary of measures per condition

Measures and tools per condition

Tools	**A**	**B**	**C**	**D**
Survey	X	X	X	X
Electrodermal activity	X	X	X	X
Facial emotion analysis	X	X	-	-
Pupillometry	X	X	-	-

3.4.7.1 Survey tools and measures

Prior to participating in the experience, participants were required to complete an initial questionnaire. This questionnaire included a range of demographic questions as well as questions related to prior experience. Specifically, it collected information on participants' age, sex, gender, levels of completed medical education, and clinical experience. To further enhance our analysis, we also included the Self-Reported Fluency of English Scale (SRFES) , a 3-item questionnaire using a 7-point Likert scale to assess the participants' English proficiency (Yeh & Inose, 2003). Additionally, we incorporated an adapted version of the Compatibility with Prior Experience scale (Karahanna et al., 2006), consisting of 4 items on a 7-point Likert scale, to measure participants' familiarity with simulation tasks similar to those they would encounter.

To thoroughly evaluate user experiences across each condition, we designed a comprehensive multidimensional questionnaire by integrating selected subscales from well-established measures. This methodology aimed to capture a wide range of perceptions related to the system's characteristics, as well as cognitive and emotional experience.

Table 4
Assessment of measurement model for constructs

Constructs	No. of items	Cronbach's alphas	Label	Questions	Source
System reponse accuracy (SRA)	5	0,924	SRA1 SRA2 SRA3 SRA4 SRA5	"The system is accurate" "The system is reliable" "The interaction with the system is predictable" "The system always did what I wanted" "The system always did what I expected"	*(Hone & Graham 2000)*
Naturalness (NAT)	3	0,949	NAT1 NAT2 NAT3	In relation to the task you have just completed : "How was the flow of this conversation?" "How natural or unnatural did you find this conversation?" "Was this conversation like or unlike an in-person conversation?"	*(Tam et al. 2012)*
Social presence (SP)	5	0,959	SP1 SP2 SP3 SP4 SP5	"There is a sense of social presence in this clinical simulation" "There is a sense of personalness in this clinical simulation" "There is a sense of sociability in this clinical simulation" "There is a sense of human warmth in this clinical simulation" "There is a sense of human sensitivity in this clinical simulation."	*(Kumar & Benbasat, 2006)*
Cognitive effort (CE)	3	0,875	CE1 CE2 CE3	"Clinical diagnosing using this clinical simulation took too much time." "Clinical diagnosing using this clinical simulation required too much effort." "Clinical diagnosing using this clinical simulation was too complex."	*(Wang and Benbasat 2009)*

Attitude toward technology (ATT)	3	0.970	ATT1 ATT2 ATT3	All things considered, using the system in the future will be a… : “Foolish move.....Wise move” “Negative step.....Positive step” “Ineffective idea.....Effective idea”	*(Tita & Barki 2009)*
Perceived scenario difficulty (PDIF)	3	0.772	PDIF1 PDIF2 PDIF3	“I found the clinical scenario difficult.” “I had difficulty progressing through the clinical scenario.” “I did not perform well in the clinical scenario.”	*(Chapman et al. 2003)*

Our questionnaire incorporated three subscales to evaluate perceptions of system characteristics across various conditions. We integrated the System's Response Accuracy (SRA) subscale (5 items, 7 points Likert scale; Ranging from Strongly disagree to Strongly agree) from the Subjective Assessment of Speech System Interfaces tool (SASSI) (K. Hone & Graham, 2001), chosen for its reliable measures of user attitudes toward speech systems. The SASSI, developed by Hone and Graham (2001), is designed to gauge user attitudes toward speech systems across six dimensions: perceived system response accuracy, likeability, cognitive demand, annoyance, habitability, and speed. For our study, the subscale was streamlined to five items to enhance the questionnaire's succinctness, with some items rephrased to ensure clarity while maintaining strong internal consistency (α=0.924).

Furthermore, we adopted the Naturalness (NAT) subscale from Tam et al.'s (2012) investigation into the impact of video and audio delays on perceived naturalness in remote interactions (J. Tam et al., 2012). This scale, with targeted questions (3 items, 7 points Likert scale; In relation to the task you have just completed: “How was the flow of this conversation?”, “How natural or unnatural did you find this conversation?”, “Was this conversation like or unlike an in-person conversation?”), evaluates the conversational flow and authenticity of technology-mediated communication, offering insights into the perceived realism of different interaction modalities. We modified the original 5-point Likert scale to a 7-point scale, ensuring robust internal consistency (α=0.949). The SP subscale from Kumar and Benbasat (2006) was employed to assess the perceived social presence and personal connection within the virtual environment (5 items, 7 points Likert scale; Ranging from Strongly disagree to Strongly agree), focusing on how the medium conveys non-verbal cues like facial expressions and posture to simulate psychological presence. This subscale, which has been validated in various studies (Gefen & Straub, 2003; Kumar & Benbasat, 2006), demonstrated excellent internal consistency (α=0.959). In addition to system perception variables, we included measures for cognitive and emotional experiences. The Perceived Cognitive Effort (CE) subscale from Wang and Benbasat (2009) was adapted to evaluate the mental exertion experienced during technology interaction (3 items, 7 points Likert scale; Ranging from Strongly disagree to Strongly agree). Emotional engagement was measured using the affective slider, a reliable method for assessing valence and arousal (Betella & Verschure, 2016), requiring participants to indicate their current emotional state on two sliders representing happiness to sadness and interest to boredom, respectively.

The attitude toward technology subscale from Titah and Barki (2009) was utilized as our dependent variable, capturing an individual's evaluation and willingness to engage with technology, influenced by cognitive and affective factors and aligning with the Technology Acceptance Model (TAM) (Davis, 1989; Titah & Barki, 2009). Finally, we adapted the perceived condition difficulty subscale from Chapman and al (2003) to better suit our specific research context. Perceived difficulty in the context of the article is the interviewee's subjective evaluation of the interview's challenge, reflecting their assessment of question difficulty, their performance, and their ability to respond effectively (Chapman et al., 2003, p. 20). This modification allowed us to precisely measure participants' perceived difficulty of the conditions presented, providing insights into the condition difficulty perceived in different conditions (4 items, 7 points Likert scale; In relation to the task you have just completed: “I fund the clinical scenario difficult.”, “I had difficulty progressing through the clinical scenario.”, “I did not perform well in the clinical scenario.”, “The clinical scenario’s specific medical case was easy.”). The fourth item has been droped to maintain a valid internal consistency ($\alpha=0.772$).

After the participant had completed all the conditions, they were asked to rank the learning quality of the 4 different systems used based on their personal opinion (Based on your experience today with the 4 different systems, please rank them in terms of learning experience quality, placing the best at the top (#1) and the worst at the bottom (#4); A-Free-choice PC, B-Free-audio PC, C-Free-text Ipad, D-Free-audio Ipad.).

The experiment ended with an open-ended qualitative interview consisting of 5 questions to enrich the quantitative analysis described above.

3.4.7.2 Physiological tools & measures

Emotional engagement comprises two key dimensions: valence and arousal. To infer the psychophysiological condition of participants and assess implicit emotional engagement, our study employed two measures: automatic facial emotion recognition (AFE) and skin conductance/electrodermal activity (SCL/EDA). These metrics provided valuable insights into the participants' emotional responses during the study.

To measure psychophysiological arousal, EDA was recorded using a Biopac MP-160 device (Biopac Systems, Inc., Goleta, USA) equipped with pre-gel sensors affixed to the palm of the participants' non-dominant hand, enabling the detection of variations in skin conductance. We also utilized Biopac's technology (Biopac Systems, Inc., Goleta, USA) to record EDA data. Electrodermal activity is extensively and dependably used to assess arousal in response to stimuli (Y. J. Wang & Minor, 2008, p. 200). The Biopac MP system has been effectively used in numerous studies to measure arousal (Charland et al., 2018; Pauna et al., 2018). Electrodermal responses were normalized against a baseline established for each participant prior to the experiment. This baseline was determined by recording the individual's standard electrodermal activity, allowing for the comparison of subsequent fluctuations against this norm. Furthermore, for the purposes of analysis, the results were adjusted to a scale ranging from -1 to 1.

To quantify the physiological valence aspects of the lived experience through FEA, we used FaceReader 6 (Noldus Technology Inc., Wageningen), a software for automated analysis of facial expressions, for the exportation and examination of the video data (Riedl & Léger, 2016). FaceReader stands as the most prevalently employed Automated Facial Expression Analysis (AFEA) tool in the field of NeuroIS, as evidenced by multiple studies (Carmichael et al., 2021; Giroux-Huppé et al., 2019; Lamontagne et al., 2020). This data was recorded using MediaRecorder and The Observer XT (Noldus Technology Inc, Wageningen). We assessed valence on a scale ranging from –1 to 1, delineating states of pleasure (for instance, happiness) against states of displeasure (such as anger) experienced in each simulation condition; this was quantified by subtracting the intensity of the most intense negative emotion from the intensity of happiness.

In order to assess and deduce cognitive exertion, we documented the mean size of the pupil throughout each situation. Pupilometry was measured using a Tobii X60 eye-tracking equipment manufactured by Tobii Technology in Sweden. Prior studies in cognitive pupillometry have established a connection between the size of the pupil and the amount of mental effort required. Steinhauer et al. (2004) emphasized that pupil dilation is a dependable measure of cognitive load, where greater dilation indicates greater mental effort during tasks like serial-7 mental subtraction (Steinhauer et al., 2004). According to Klingner et al. (2011), pupil dilation can be used as an efficient indicator of cognitive load in several tasks. Their study showed that pupil size increases as the difficulty of the task increases (Klingner et al., 2011). Multiple research on task-evoked pupillary responses have demonstrated that pupil dilation accurately represents the cognitive requirements of mental processing in different situations, including computerized activities. Eye trackers, like the Tobii X60, are commonly employed for the purpose of measuring the size of the pupil and deducing cognitive load. Research has demonstrated that task-evoked pupillary responses are a very accurate indicator of cognitive functioning, and may be used in various experimental settings (Jiang et al., 2014; Kahneman & Beatty, 1966; Kruger & Steyn, 2014). Utilizing traditional eye-tracking technology to measure pupil size while using a computer interface preserves the ecological validity of the job. Studies conducted by Chen and Epps (2013) and Kruger et al. (2014) have shown that variations in pupil diameter during computer-based tasks can be used as a reliable indicator of task difficulty and cognitive load (S. Chen & Epps, 2014; Kruger & Steyn, 2014). These findings support the validity of using pupil diameter as a measure in real-world settings.

3.4.8 Analysis

Employing SAS (SAS Institute 2011), we assessed the discrepancy in the mean of model variables across the different conditions through a two-tailed Wilcoxon signed-rank test. The impact of modulation was evaluated utilizing a linear regression model featuring a random intercept, with adjustments for multiple comparisons made via the Holm-Bonferroni method. The 10% significance level was employed in the statistical analysis.

3.5 Results

3.5.1 Descriptive statistics

To begin the analysis, we performed a descriptive analysis to gain a comprehensive understanding of the data and to compare multiple conditions (Table 6).

In Desktop VR, the choice-based condition (Condition A) had a mean system response accuracy (SRA) of 4.85 (SD = 1.50), whereas the vocal condition (Condition B) had 4.49 (SD = 1.49), suggesting a modest decrease in SRA while switching from choice-based to vocal modality. Naturalness (NAT) increased from 3.18 (SD = 1.78) in choice-based to 3.63 (SD = 1.82) in vocal. From 3.25 (SD = 1.35) in the choice-based condition to 3.9 (SD = 1.69) in the verbal condition, social presence (SP) rose. Cognitive effort (CE) increased from 3.27 (SD = 1.31) in choice-based to 3.47 (SD = 1.51) in voice. Arousal (ARO) increased slightly from 64.94 (SD = 15.41) in the choice-based condition to 65.55 (SD = 21.70) in the verbal condition. Valence (VAL) dropped from 72.65 (SD = 19.42) in choice-based to 65.1 (SD = 25.31) in vocal. ATT rose from 5.13 (SD = 1.58) in the choice-based condition to 5.28 (SD = 1.71) in the verbal condition. Comparing these two conditions, the vocal modality increased the perception of social presence and naturalness slightly, while also slightly increasing cognitive effort and reducing valence, indicating a trade-off between more engaging interactions and higher cognitive load.

In ChatGPT, the free text condition (Condition C) had a system response accuracy (SRA) of 5.17 (SD = 1.09), whereas the voice condition (Condition D) had 5.01 (SD = 1.31), demonstrating a modest reduction with vocal modality. Free text naturalness (NAT) was 4.81 (SD = 1.78) compared to 4.73 (SD = 1.83) in vocal, a small decline. Social presence (SP) increased from 3.54 (SD = 1.59) in free text to 4.33 (SD = 1.77) in vocal. Both conditions had equal cognitive effort (CE), 2.88 (SD = 1.41) in free text and 2.84 (SD = 1.40) in vocal. Arousal (ARO) increased from 58.7 (SD = 20.97) in free text to 67.95 (SD = 15.84) in vocal. Valence (VAL) dropped from 70.3 (SD = 17.04) in free text to 65.85 (SD = 26.15) in vocal. Technology attitude (ATT) increased from 5.12 (SD = 1.36) in free text to 5.32 (SD = 1.68) in vocal. In this environment, vocal interactions slightly decreased perceived system response accuracy and naturalness but significantly increased social presence and arousal, while cognitive effort remained low and attitude towards technology improved slightly, suggesting that vocal interactions in ChatGPT enhance engagement without significantly increasing cognitive load.

In Desktop VR, Conditions A and B had a mean SRA of 4.67, while ChatGPT Conditions C and D had 5.09, indicating a greater SRA in ChatGPT. Desktop VR had a lower mean naturalness (NAT) (3.40) than ChatGPT (4.77). ChatGPT voice condition (4.33) had stronger social presence (SP) than Desktop VR (3.9). ChatGPT had lower cognitive effort (CE) than Desktop VR (mean = 3.37). Arousal (ARO) was higher in ChatGPT vocal (67.95) than Desktop VR vocal (65.55), while valence (VAL) was higher in Desktop VR choice-based (72.65) than ChatGPT free text (70.3). Attitude towards technology (ATT) was similar across situations, with ChatGPT verbal interactions (5.32) slightly preferred over Desktop VR choice-based interactions (5.13). These findings indicate that while both environments have their strengths, ChatGPT, particularly in vocal interactions, tends to offer a more natural and less cognitively demanding user experience, making it a potentially more effective tool for enhancing user engagement and satisfaction.

3.5.2 Control of perceived condition difficulty

The first part of our analysis consisted in analyzing the perceived difficulty of the medical scenarios provided between the conditions. Table below shows the estimated ratings of perceived condition difficulty for each condition.

Table 5
Perceived condition difficulty control

Condition	A	B	C	D
A	-	-0,58	0,57	0,18
B	-	-	**1,15*** **	**0,77****
C	-	-	-	-0,38
D	-	-	-	-

Legend : p < 10% : *, p < 5% : **, p < 1% : ***

The results revealed significant differences in perceived condition difficulty between conditions B and C (p = 0.0031) and between conditions B and D (p = 0.086). Participants found condition B significantly more challenging than conditions C and D. This highlights a consistent trend where condition B, vocal modality in desktop VR simulation environment, was perceived as more difficult than all conditions in the ChatGPT (OpenAI) simulation environment.

The second part of our analysis examined the potential influence of interaction modalities and simulation environments on the model variables. The table below presents the mean ratings and standard deviations for system response accuracy, naturalness, social presence, cognitive effort (including pupil diameter), arousal (including activation), valence (including physiological valence), and attitude toward technology for each condition and grouped by simulation environment.

Table 6

Mean and (standard deviations) of model variables per condition

Simulation environment	DEKSTOP VR SIMULATION			GPT SIMULATION		
Condition	A	B	VR	C	D	GPT
			Survey tools			
SRA	4,85 (1,50)	4,49 (1,49)	4,67 (1,49)	5,17 (1,09)	4,85 (1,52)	5,01 (1,31)
NAT	3,18 (1,78)	3,63 (1,82)	**3,41*** (1,79)**	4,81 (1,78)	4,64 (1,92)	**4,73*** (1,83)**
SP	**3,25** (1,35)**	**3,9** (1,69)**	3,58 (1,55)	**3,54*** (1,59)**	**4,33*** (1,77)**	3,93 (1,71)
CE	3,27 (1,31)	3,47 (1,51)	**3,37* (1,40)**	2,88 (1,41)	2,8 (1,44)	**2,84* (1,40)**
ARO	64,94 (15,41)	65,55 (21,70)	65,26 (18,74)	**58,7*** (20,97)**	**67,95*** (15,84)**	63,33 (18,93)
VAL	**72,65* (19,42)**	**65,1* (25,31)**	68,88 (22,59)	70,3 (17,04)	65,85 (26,15)	68,08 (21,90)
ATT	5,13 (1,58)	5,28 (1,71)	5,21 (1,63)	5,12 (1,36)	5,32 (1,68)	5,22 (1,51)
			Physiological measures			
ACTIVATION	0,28 (0,27)	0,24 (0,17)	**0,26** (0,23)**	0,42 (0,32)	0,34 (0,30)	**0,38** (0,31)**
PHYS. VALENCE	**-0,13** (0,11)**	**-0,09** (0,14)**	N/A	N/A	N/A	N/A
PUPIL DIAMETER	**-0,01** (0,17)**	**0,11** (0,12)**	N/A	N/A	N/A	N/A

Legend :

A: VR-Choice-based interaction, **B:** VR-Voice-based interaction, **C:** LLM-Text-based interaction, **D:** LLM-Voice-based interaction

SRA : System response accuracy, **NAT:** Naturalness, **SP:** Social presence, **CE:** Cognitive effort, **ARO:** Arousal, **VAL:** Valence, **ATT:** Attitude toward technology.

p < 10% : *, p < 5% : **, p < 1% : ***

3.5.3 Hypothesis testing

3.5.3.1 Media perception

Beginning the third part of our analysis with the desktop VR simulation environment, we observed that manipulating the interaction mode did not result in a significant difference in perceived system response accuracy ($p = 0.1321$). Similarly, no significant difference was found for naturalness ($p = 0.3694$). However, the results indicated that changing the mode of interaction with the virtual agent significantly affected perceived social presence, with vocal interaction leading to a greater perceived social presence ($p = 0.0314$).

In the ChatGPT (OpenAI) simulation environment, similar to our observations in the desktop VR environment, we did not find a significant difference in perceived system response accuracy ($p = 0.2058$). Likewise, no significant difference was found in perceived naturalness ($p = 0.4975$). However, the results indicated that changing the mode of interaction with the virtual agent significantly affected perceived social presence ($p = 0.0005$), thereby validating H3a, suggesting that Vocal interaction resulted in a significantly greater perceived social presence than textual one within both simulation environments.

When comparing the two simulation environments, we did not find a significant difference in system response accuracy ($p = 0.2935$) or social presence ($p = 0.2664$), preventing us from validating hypotheses H1b and H3b. However, contrary to our expectations, there was a significant difference in perceived naturalness ($p = 0.0005$). This finding contradicts our predictions for H2b, suggesting that, on average, the ChatGPT (OpenAI) environment without any visual-anthropomorphic feature, was perceived as more natural, regardless of the interaction mode.

3.5.3.2 Cognitive experience

Linear regression analysis indicates that, overall, without distinguishing interaction modality, all three variables—social presence, naturalness, and system response accuracy—negatively correlate with perceptual cognitive effort ($p<0.0001$), hereby validating h4a, h4b and h4c.

Digging deeper into the results, the results showed no significant difference in either the desktop VR condition ($p = 0.5726$) or the ChatGPT (OpenAI) simulation condition ($p = 0.7048$). However, analysis of the physiological data, specifically pupil dilation, indicates a significant difference in average cognitive effort in the desktop VR simulation environment ($p = 0.0192$). At the 5% significance level, the data show that pupil dilation was, on average, greater during speech interaction, suggesting increased cognitive effort. This result also demonstrates the marked difference between reported and physiological cognitive effort. In addition, at significance level 10%, results showed there was a significant difference when comparing the means between the desktop VR and ChatGPT (OpenAI) environments ($p = 0.0791$), indicating that simulation in the desktop VR environment was perceived as cognitively more challenging than ChatGPT (OpenAI) one.

Linear regression analysis indicates a negative relationship between system response accuracy and cognitive effort in the desktop VR environment ($e = -0.3636$, StdE = 0.1849, DF = 17, $t = -1.97$, $p = 0.0658$), suggesting that greater system response accuracy generally leads to less perceived cognitive effort. Similarly, in the ChatGPT (OpenAI) environment, system response accuracy significantly reduces perceived cognitive effort ($e = -0.5979$, StdE = 0.1595, DF = 17, $t = -3.75$, $p = 0.0016$). However, the study of mean pupil dilation during the desktop VR simulation shows no significant regression relationship ($e = -0.00763$, StdE = 0.02321, DF = 17, $t = -0.33$, $p = 0.7463$). These results support hypothesis H4a.

Regarding naturalness, the regression analysis indicates a negative relationship with cognitive effort in both the desktop VR environment ($e = -0.381$, StdE = 0.1074, DF = 19, $t = -3.55$, $p = 0.0022$) and the ChatGPT (OpenAI) environment ($e = -0.4395$, StdE = 0.1154, DF = 19, $t = -3.81$, $p = 0.0012$). These findings suggest that, overall, the more natural an interaction is perceived to be by the user, the lower the resulting cognitive effort, whether in the desktop VR or ChatGPT (OpenAI) environment. More specifically, the results indicate ($p = 0.0939$) that the effect of naturalness is more pronounced when using voice-based interactions compared to choice-based interactions. However, as with the previous variable, the regression analysis did not reveal the same effect on pupil dilation ($e = 0.00675$, StdE = 0.01424, DF = 19, $t = 0.47$, $p = 0.6409$).

Similarly, the regression analysis also demonstrates a negative relationship between social presence and perceptual cognitive effort, both in the VR desktop (e = -0.3734, StdE = 0.1355, DF = 19, t = -2.76, p = 0.0126) and ChatGPT (OpenAI) (e = -0.4941, StdE = 0.1124, DF = 19, t = -4.39, p = 0.0003) environments. These results suggest that, overall, in both desktop VR and GPT environments, the higher the social presence, the lower the cognitive effort. Again, the regression analysis did not reveal the same effect on pupil dilation (e = 0.00961, StdE = 0.01646, DF = 19, t = 0.58, p = 0.5662) Nonetheless, the results indicate (p = 0.0471) that the effect of social presence on pupil dilation is more pronounced when using voice-based interactions compared to choice-based interactions.

3.5.3.3 Emotional experience

In terms of emotional experience, starting with the desktop VR environment, results show a significant difference in perceived valence between the two interaction modalities (p = 0.0733). This suggests that perceived valence is higher when using the choice-based modality compared to the vocal modality. Contrarily, the analysis of physiological valence indicates that average emotional intensity was lower during choice-based interactions than during vocal interactions (p = 0.0494). For arousal, no significant difference was found between the two modalities in perceived arousal (p = 0.6696) or physiological arousal (p = 0.355). However, a more detailed analysis, segmented by thirds, shows that average physiological arousal at mid-course (p = 0.0546) and at the end of the course (p = 0.0446) was higher when using the choice-based modality. In the ChatGPT (OpenAI) environment, the results show no significant difference between the two interaction modes in terms of perceived valence (p = 0.4744). However, the analysis of physiological valence indicates a higher average intensity during voice interactions compared to text interactions (p = 0.0461). For arousal, the results reveal significantly lower arousal during textual interactions compared to vocal interactions (p = 0.0064). The comparative analysis of the VR and ChatGPT (OpenAI) desktop environments revealed no significant differences in either perceived valence (p = 0.9182) or perceived arousal (p = 0.7581).

3.5.3.3.1 Valence

Linear regression analysis indicates that, overall, without distinguishing interaction modality, all three variables—social presence, naturalness, and system response accuracy—positively correlate with perceptual valence ($p=0.0132$, $p=0.012$, $p=0.006$). Moreover, social presence also positively correlates with the physiological valence mean ($p=0.0721$).

Digging deeper into the results, linear regression analysis indicates a positive relationship between system response accuracy and perceived valence in the desktop VR environment ($e = 0.6276$, $StdE = 0.3397$, $DF = 19$, $t = 1.85$, $p = 0.0803$), suggesting that higher perceived system accuracy is associated with higher valence. However, this finding is not supported by the analysis of mean physiological valence ($e = 0.02761$, $StdE = 0.01641$, $DF = 18$, $t = 1.68$, $p = 0.1097$). A positive linear relationship was found between naturalness and perceived valence ($e = 0.6043$, $StdE = 0.2758$, $DF = 19$, $t = 2.19$, $p = 0.0411$), suggesting that a more natural simulation is associated with higher perceived valence. However, the analysis of mean physiological valence does not support this result ($e = 0.006681$, $StdE = 0.01226$, $DF = 18$, $t = 0.54$, $p = 0.5925$). Additionally, no significant regression relationship was observed between social presence and perceived valence ($e = 0.4932$, $StdE = 0.2974$, $DF = 19$, $t = 1.66$, $p = 0.11370$) or physiological valence ($e = 0.01326$, $StdE = 0.01507$, $DF = 18$, $t = 0.88$, $p = 0.3904$).

In the ChatGPT (OpenAI) environment, the results of the linear regression analysis do not establish a significant relationship between system response accuracy and perceived valence ($e = 0.5747$, $StdE = 0.3603$, $DF = 19$, $t = 1.6$, $p = 0.1272$). However, a significant positive linear relationship was found between naturalness and perceived valence ($e = 0.6268$, $StdE = 0.3021$, $DF = 19$, $t = 2.08$, $p = 0.0518$). Additionally, there was a significant positive linear relationship between social presence and perceived valence ($e = 0.5194$, $StdE = 0.2878$, $DF = 19$, $t = 1.8$, $p = 0.087$), indicating that in the ChatGPT (OpenAI) environment, the more social the simulation was perceived to be, the higher the perceived valence.

3.5.3.3.2 Arousal

Linear regression analysis indicates that, overall, social presence and naturalness positively correlate with reported arousal, regardless of interaction modality ($p=0.0082$, $p=0.0617$). Additionally, system response accuracy shows a positive correlation with physiological arousal mean ($p=0.0436$).

Digging deeper into the results, in the desktop VR environment, linear regression analysis did not confirm a relationship between system response accuracy and reported arousal ($e = 0.6763$, $StdE = 2.2808$, $DF = 17$, $t = 0.3$, $p = 0.7704$) or physiological activation ($e = 0.1783$, $StdE = 0.1432$, $DF = 15$, $t = 1.24$, $p = 0.2323$). However, a positive relationship was observed between naturalness and reported arousal ($e = 4.3814$, $StdE = 1.644$, $DF = 17$, $t = 2.67$, $p = 0.0163$), indicating that in the VR environment, more natural interactions were associated with higher arousal. This finding, however, was not supported by the analysis of physiological activation ($e = 0.1302$, $StdE = 0.1066$, $DF = 15$, $t = 1.22$, $p = 0.2407$). Furthermore, the relationship between social presence and arousal was not statistically significant for either perceptual measures ($e = 3.3332$, $StdE = 2.1764$, $DF = 17$, $t = 1.53$, $p = 0.144$) or physiological measures ($e = -0.04044$, $StdE = 0.1282$, $DF = 15$, $t = -0.33$, $p = 0.7465$).

In the ChatGPT (OpenAI) environment, no significant relationship was found between system response accuracy and perceived arousal ($e = 1.1151$, $StdE = 2.4982$, $DF = 19$, $t = 0.45$, $p = 0.6604$). Similarly, no significant relationship was detected for the mean activation measure ($e = 0.1591$, $StdE = 0.1266$, $DF = 15$, $t = 1.26$, $p = 0.2279$), not allowing us to validate H6a. Additionally, there was no significant relationship between naturalness and arousal, whether perceptual ($e = 0.01798$, $StdE = 1.9147$, $DF = 19$, $t = 0.01$, $p = 0.9926$) or physiological ($e = -0.1324$, $StdE = 0.09815$, $DF = 15$, $t = -1.35$, $p = 0.1975$). A significant relationship was found between social presence and perceived arousal ($e = 4.1455$, $StdE = 1.939$, $DF = 19$, $t = -2.14$, $p = 0.0457$), suggesting that in the ChatGPT (OpenAI) environment, higher social presence is associated with increased arousal. However, this result was not supported by the analysis of the same relationship on the physiological side ($e = -0.00573$, $StdE = 0.09973$, $DF = 15$, $t = -0.06$, $p = 0.9549$).

3.5.3.4 Attitude toward technology

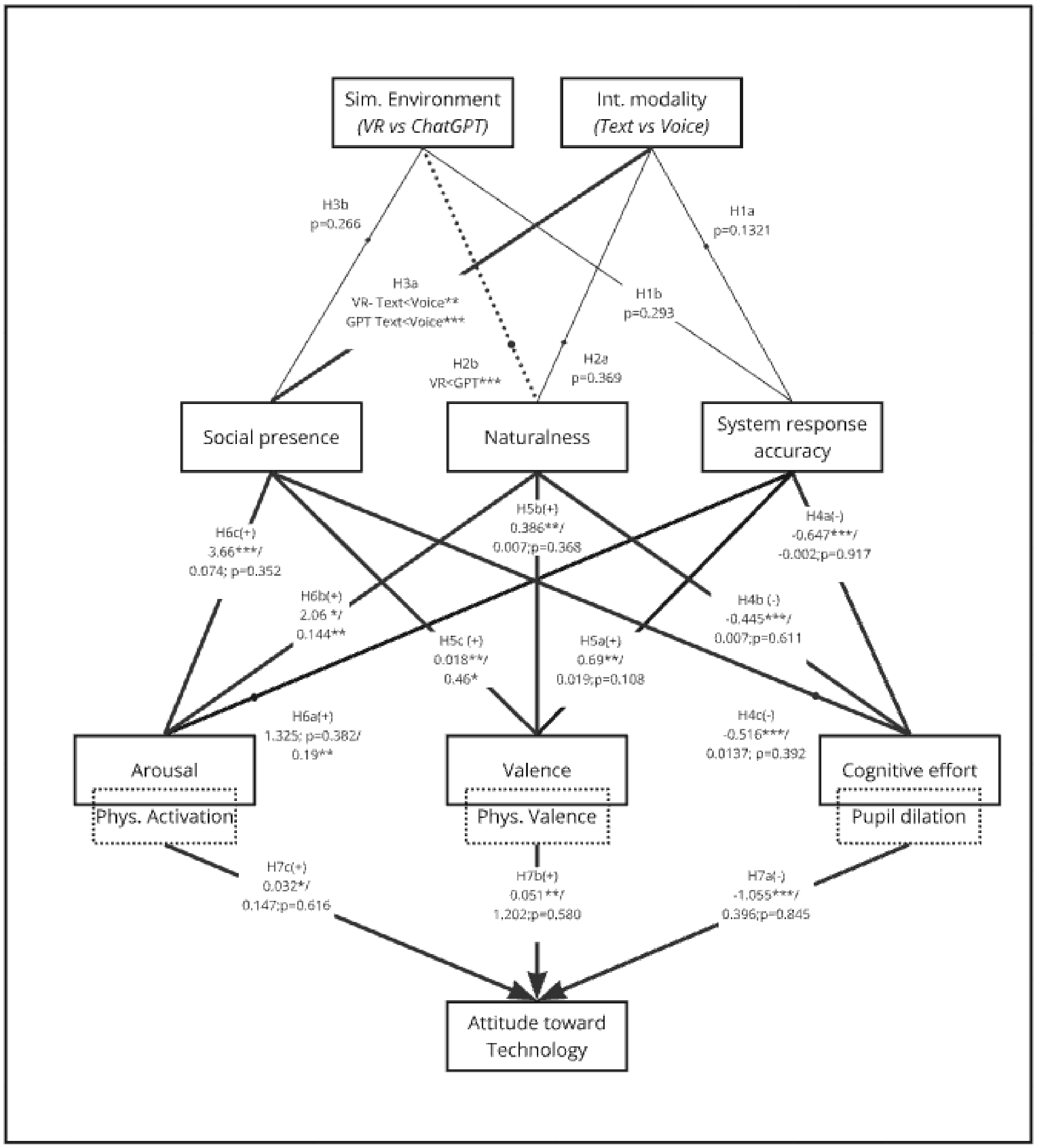
Overall, no significant difference in attitude toward technology was detected between the interaction modalities in either the VR desktop environment ($p = 0.671$) or the ChatGPT (OpenAI) environment ($p = 0.5499$). Additionally, no significant difference was found when comparing the two simulation environments ($p = 0.7532$).

Linear regression analysis indicates that, overall, reported arousal and valence significantly positively correlate with attitude toward technology, regardless of interaction modality ($p=0.072$, $p=0.0044$). Additionally, perceived cognitive effort shows a significant negative correlation with attitude toward technology ($p=0.0003$).

In the desktop VR environment, no significant relationship was found between reported cognitive effort and attitude toward technology ($e = -0.4261$, $StdE = 0.3545$, $DF = 19$, $t = -1.2$, $p = 0.2442$). Similarly, no significant relationship was observed with the physiological variable ($e = 3.7382$, $StdE = 2.997$, $DF = 19$, $t = 1.25$, $p = 0.2274$). No significant relationship with attitude toward technology was found for either reported valence ($e = 0.029933$, $StdE = 0.02286$, $DF = 19$, $t = 1.31$, $p = 0.2061$) or physiological valence ($e = 5.1407$, $StdE = 3.7432$, $DF = 18$, $t = 1.37$, $p = 0.1865$). Additionally, no significant relationships were detected for arousal ($e = 0.01145$, $StdE = 0.02643$, $DF = 17$, $t = 0.43$, $p = 0.6704$) or physiological activation ($e = 0.1169$, $StdE = 0.4768$, $DF = 15$, $t = 0.25$, $p = 0.8097$).

In the ChatGPT (OpenAI) environment, the results indicate a negative relationship between reported cognitive effort and attitude toward technology ($e = -2.9482$, $StdE = 1.1003$, $DF = 19$, $t = -2.68$, $p = 0.0148$), meaning that higher reported cognitive effort is associated with a lower attitude toward technology. A significant positive relationship was also detected between perceived valence and attitude toward technology ($e = 0.1039$, $StdE = 0.03521$, $DF = 19$, $t = 2.95$, $p = 0.0082$), indicating that higher perceived valence is associated with a more positive attitude toward technology. However, no significant relationship was found between reported arousal and attitude toward technology ($e = 0.03916$, $StdE = 0.02362$, $DF = 19$, $t = 1.66$, $p =$

0.1137), nor for the mean physiological activation measure (e = 0.1464, StdE = 0.4648, DF = 15, t = 0.31, p = 0.7571).



Legend :
█ : Validated hypothesis
- - - : Contrary to our expectations

Figure 6

Research model with hypothesis summary

Table 7

Summary of hypothesis testing

Hypothesis	IV	DV	Sim Env.	S	W	R	P	Status
H1a	Int. Mod.	SRA	VR LLM	29,5 34,5	- -	0,463 0,2058	0,1321 0,314	Unsuported
H1b	Sim. Env.	SRA		-	276,5	0,213	0,2935	Unsuported
H2a	Int. Mod.	NAT	VR LLM	-18 12,5	- -	0,265 0,4975	0,3694 0,208	Unsuported
H2b	Sim. Env.	NAT			143,5	0,613	0,0005	Contrary
H3a	Int. Mod.	SP	VR LLM	-53 -61	- -	0,532 0,897	0,0314 0,0005	Fully supported
H3b	Sim. Env.	SP		-	291	0,215	0,2664	Unsupported
Hyp.	**IV**	**DV**	**E**	**StdErr**	**DF**	**T**	**P**	Status
H4a	SRA	CE P.D	-0.6467 -0.00195	0.08869 0.01866	59 29	-7.29 -0.1	<.0001 0.9175	Part. supported
H5a	SRA	VAL P.VAL	0.6906 0.01885	0.2411 0.01153	59 52	2.86 1.63	0.0058 0.1082	Fully supported
H6a	SRA	ARO P.ARO	1.3247 0.1914	1.5045 0.09227	57 47	0.88 2.07	0.3823 0.0436	Part. supported
H4b	NAT	CE P.D	-0.4452 0.006895	0.06495 0.01341	59 29	-6.85 0.51	<.0001 0.6111	Part. supported
H5b	NAT	VAL P.VAL	0.3856 0.007427	0.1488 0.008176	59 52	2.59 0.91	0.012 0.3679	Part. supported
H6b	NAT	ARO P.ARO	2.0631 0.1438	1.0824 0.05757	57 47	1.91 2.5	0.0617 0.016	Fully supported
H4c	SP	CE P.D	-0.516 0.01367	0.08367 0.01574	59 29	-6.17 0.87	<.0001 0.3922	Part. supported
H5c	SP	VAL P.VAL	0.4592 0.01795	0.1796 0.00978	59 52	2.56 1.84	0.0132 0.0721	Fully supported
H6c	SP	ARO P.ARO	3.6591 0.07407	1.3357 0.07881	57 47	2.74 0.94	0.0082 0.3521	Part. supported
H7a	CE PD	ATT	-1.0548 0.3963	0.2748 2.0121	59 29	-3.84 0.2	0.0003 0.8452	Part. supported
H7b	VAL PVAL	ATT	0.05054 1.2022	0.01707 2.1577	59 52	2.96 0.56	0.0044 0.5798	Part. supported
H7c	ARO PARO	ATT	0.03179 0.1472	0.01735 0.2915	57 47	1.83 0.51	0.072 0.6158	Part. supported

3.5.4 Ranking Questionnaire

The statistical analysis utilized Type III Tests of Fixed Effects in a cumulative logistic regression model. The study found a notable influence of stimulus type (text vs. speech), as indicated by an F-value of 3.72 ($p = 0.0588$). Nevertheless, the type of setting (simulation vs. GPT) did not have a substantial impact, as evidenced by the statistical analysis with $F(1, 55) = 0.29$ ($p = 0.5927$). In addition, the combination of stimulus type and device type did not have a significant impact ($F(1, 55)=0.23$ ($p = 0.6318$)).

Subsequent analysis confirmed these findings. The impact of the kind of stimulus demonstrated a statistically significant pattern ($F(1, 56) = 3.66$ ($p = 0.0608$)). On the other hand, the influence of the type of equipment used did not reach statistical significance, ($F(1, 56) = 0.29$ ($p = 0.5906$).

The cumulative logistic regression model showed a significant impact of stimulus type (text), with a coefficient estimate of -0.7813 (stdE = 0.4082, t-value = -1.91, $p = 0.0608$). In contrast, the device type (GPT) did not show a significant effect, with an estimated value of 0.2177 (standard error = 0.4025, t-value = 0.54, $p = 0.5906$).

The results suggest a distinct inclination for voice interactions compared to text interactions, with statistical significance at a p-value of less than 0.10. Yet, the choice of device employed (VR vs. GPT) had no substantial impact on the rankings (Table 8).

Table 8

Synthesis of ranking results, grouped to show preferences

RN KG	p10	p11	p12	p07	p13	p21	p20	p22	p09	p04	p01	p16	p17	p23	p24	p02	p18	p08	p06	p03
1	A	A	A	A	A	A	B	B	B	B	D	D	D	D	D	D	D	D	D	C
2	B	B	B	B	C	C	D	D	D	A	B	B	B	B	B	C	C	C	C	B
3	C	C	C	D	B	D	C	C	A	C	C	C	C	C	C	B	B	B	A	A
4	D	D	D	C	D	B	A	A	C	D	A	A	A	A	A	A	A	A	B	D

Legend : A = Destop VR - Choice, B = Desktop VR - Voice, C = ChatGPT - Text, D = ChatGPT - Voice

3.6 Discussion

3.6.1 Principal findings

Our research sought to understand how and to what extent GenAI-powered vocal interaction modalities and simulation environment fidelity influence user attitudes toward technology within simulation-based learning contexts. Specifically, the study examined the impact of these modalities and environments on key dimensions such as system response accuracy, naturalness, and social presence, and how these system perception dimensions, in turn, affected users' cognitive load, emotional engagement, and overall attitude toward the technology being used as educational means.

While the study explored multiple facets of interaction modality (voice vs. text) and simulation environment (Desktop VR vs. GPT), it found no significant difference in users' overall attitude toward technology between these conditions. Variations were observed in other variables, but they did not significantly affect how users ultimately viewed the technology. This suggests that attitude toward technology is influenced by a combination of factors beyond modality or environment alone. For instance, users might have placed more importance on the quality of interaction and technical reliability rather than the visual richness of the environment. System response accuracy, naturalness, and social presence played crucial roles in shaping the user experience, but the interaction between these factors might not have been strong enough to alter their overarching attitudes toward the technology. Additionally, individual differences in user expectations and familiarity with such systems may have led to heterogeneous responses, further diluting any potential impact on their overall attitude.

Contrary to expectations, Desktop VR did not outperform ChatGPT (OpenAI, 2023) in terms of system response accuracy; users did not perceive a significant difference between the environments in how accurately the system responded to their input. This could in part be understood through the lens of Expectancy Disconfirmation Theory (Oliver, 1980). In immersive environments like Desktop VR, users tend to have higher expectations due to the advanced sensory input (e.g., visuals, spatial cues) that suggests a more responsive and accurate system. When these high expectations are not met, users experience negative disconfirmation, perceiving the system’s performance as less accurate than anticipated. On the other hand, in the simpler, less immersive GPT environment, users may enter with lower expectations. When the system performs to or above those expectations, users experience positive confirmation, leading to a perception of higher accuracy. This dynamic is reflected in similar findings in the literature, who applied Expectancy Disconfirmation Theory to user satisfaction in e-commerce platforms, showing how unmet or exceeded expectations shape perceived performance (Bhattacherjee, 2001). Moreover, users perceived ChatGPT (OpenAI, 2023) as more natural than Desktop VR. This finding suggests that the fidelity of the simulation environment alone is not sufficient to create a more natural experience for users. Other factors, such as the smoothness and reliability of the interaction, play a critical role.

The discrepancy between interaction modality and environmental fidelity could be explained by the Uncanny Valley effect (Mori et al., 2012). As digital environments become more lifelike, slight imperfections in behavior and interaction become more pronounced, leading to discomfort or a sense of unnaturalness. In high-fidelity settings like Desktop VR, users may have heightened expectations for naturalness due to the visual and sensory realism. When voice interactions fail to align with this realism, users become more sensitive to the mismatch, thus reducing the perceived naturalness. This effect is well-documented in the literature; for example, Mori et al. (2012) discuss how highly realistic mediums, when slightly off, can lead to feelings of eeriness or discomfort, a phenomenon often seen in humanoid robots or virtual avatars.

The hypothesis that Desktop VR would result in higher perceived social presence than GPT was fully supported, confirming that the immersive visuals and interactive elements of the VR environment contributed to a stronger connection with the system. This aligns with Media Equation Theory (Reeves & Nass, 1996). Users not only heard but also interacted in a visually rich environment, which likely made the experience feel more authentic. In contrast, GPT's simpler environment lacked the immersive sensory cues of VR but still managed to provide natural interaction, highlighting that while media richness contributes to social presence, the quality of interaction is critical. Consistent, smooth, and reliable communication can foster social presence even in less immersive settings. This interaction quality, emphasized by Media Equation Theory, explains why the simpler GPT environment performed well despite the lack of visual richness. These findings highlight that both visual fidelity and interaction smoothness are pivotal in shaping perceived social presence and naturalness, supporting the theory that users respond to media much like real-world interactions.

The relationships between system response accuracy, social presence, and naturalness and their effects on cognitive effort, valence, and arousal were all confirmed, offering a comprehensive view of how these system characteristics shape user experiences. However, the findings also highlight some important nuances and complexities. The results confirmed that higher system response accuracy and greater perceptions of naturalness and social presence were associated with lower cognitive effort. This aligns with Cognitive Load Theory (Sweller, 1988). In our experimentation, when users felt that the system responded accurately and the interaction felt natural, they reported less mental strain.

The presence of social elements also contributed to a smoother, more intuitive experience, reducing the cognitive load. However, the intricate balance between system performance and cognitive effort underscores the importance of reliability. While naturalness and social presence can enhance the user experience, their benefits may be overshadowed if system response accuracy suffers due to technical limitations like speech recognition errors. In these cases, users might experience increased cognitive effort, despite the interactive richness of the environment. Similarly, higher system response accuracy, naturalness, and social presence were associated with positive valence. Users experienced more pleasant emotions when interacting with a system that was perceived to be reliable, engaging, and natural. This is consistent with theories from affective computing and user experience research, which suggest that smooth and effective interactions with technology lead to positive emotional states (D. Norman, 2007; Picard, 2000). When users perceive the system to be functioning well, they experience fewer frustrations, leading to more positive overall feelings about the interaction.

One interesting aspect to consider is the role of social presence in enhancing emotional engagement. The more users felt the system was socially present and engaged with them on a human level, the more positive their emotional responses were. This finding reinforces the idea that human-like interactions, even in AI systems, can evoke more emotionally positive experiences. Again, it’s important to note that these benefits may not be fully realized if the system’s accuracy is compromised, as even highly engaging systems can lead to negative emotional outcomes if they are not reliable. The connection between system response accuracy, social presence, and naturalness and arousal was also confirmed. Systems that provided accurate responses, felt more natural, and exhibited higher social presence led to higher levels of arousal, indicating a heightened state of alertness or excitement during the interaction. This makes sense in light of Media Equation Theory (Reeves & Nass, 1996). However, while high arousal can enhance the user experience, it must be carefully managed. Excessive arousal, especially if it stems from frustration due to system inaccuracies or technical issues, could negatively impact the overall user experience. The key takeaway here is that positive arousal must come from effective and seamless interactions, not from user efforts to compensate for system flaws.

The results discussed throughout this study underscore the complexity of user experiences when interacting with GenAI-powered systems in simulation-based learning environments, particularly in terms of how key factors such as system response accuracy, naturalness, and social presence shape overall perceptions and interactions. The intricate relationship between these variables highlights the multifaceted nature of human-technology interaction in educational simulations, where single-dimensional explanations are often inadequate or insufficient. While certain hypotheses were not fully supported, the findings emphasize that user experiences are shaped by a combination of context, objective, technical performance, and individual social engagement, each influencing user perceptions in distinct ways.

Our study revealed that user preferences and perceptions fluctuate based on the modality of interaction and the environmental fidelity of the simulation. The nuanced ways users evaluate their experience alone do not guarantee a superior experience. Instead, users' cognitive and emotional responses are often shaped by how seamlessly the system operates, and how well it can meet their expectations both technically (in terms of system response) and socially (in terms of engagement and presence). This reinforces the notion that a one-size-fits-all approach cannot fully explain the variety of user experiences within these complex contexts and not only in therms of novice versus expert users different needs. The study's findings also demonstrate that the interplay between cognitive effort, emotional responses (such as valence and arousal), and the system's ability to engage users on both a technical and social level demands a more sophisticated, multiple-path approach to understanding user experience.

The combination of Cognitive Load Theory (Sweller, 1988) and Media Equation Theory lenses in our study suggests that when interactions with a system are natural, accurate, and socially engaging, users experience lower cognitive load and higher emotional engagement. The findings show that naturalness and social presence can enhance user experiences, but these benefits can be easily overshadowed if system response accuracy fails to meet user expectations. Moreover, the study highlights the importance of managing emotional responses like valence and arousal in designing interactive systems.

However, the study also showed that physiological measures of arousal and cognitive effort did not always align with self-reported measures, suggesting that user perceptions and unconscious physiological responses may capture different dimensions of the experience. This reinforces the idea that physiological and subjective data are complementary tools that provide distinct, yet valuable, perspectives on user interactions. Additionally, the GPT-based environment, represented by ChatGPT (OpenAI, 2023), emerges as a fresh new solution for simulations across many domains. Its ability to offer a relevant and efficient way to practice simulation at an affordable cost while remaining simple to use suggests it could pave the way for broader adoption in educational and professional settings.

By leveraging the flexibility and conversational capabilities of ChatGPT (OpenAI, 2023), institutions and learners alike may benefit from a more accessible and practical simulation solution, without the need for expensive or high-maintenance setups like Desktop VR. Ultimately, these insights reflect the need for more nuanced, flexible designs in AI-driven educational tools. Effective design must account for the dynamic interaction between cognitive and emotional dimensions, ensuring that systems are not only technically reliable but also socially engaging and emotionally satisfying. This aligns again with the broader theme of the article 'Beyond one-size-fits-all' reflecting and highlighting that user experience is best explained through a multi-path framework that integrates technical performance, social engagement, and emotional impact. Moving beyond generalized approaches, this dual-path model calls for designers to address both the cognitive needs and emotional expectations of users to create more effective and engaging learning environments.

3.6.2 Theoretical contribution

Our study provides novel insights into both Media Equation Theory (MET) and Cognitive Load Theory (CLT), exploring how different interaction modalities and simulation environments shape user experiences in educational simulations.

MET traditionally posits that users treat media and AI as if they are engaging with human counterparts, attributing social presence to technology. Our research supports the idea that voice-based interactions enhance this perception of social presence, making the engagement feel more human-like. However, the findings challenge some of MET's foundational assumptions. While immersive virtual reality (VR) environments offer visually rich experiences, simpler setups, such as ChatGPT (Open AI), were found to provide a more natural user experience, primarily due to fewer technical complications. This suggests that realism alone is not a definitive indicator of higher social presence; instead, the fluidity and reliability of the interaction are more significant. Although voice-based interactions are engaging, they also impose higher cognitive demands, adding complexity not fully accounted for in MET.

CLT emphasizes minimizing extraneous cognitive load to enhance learning. Our findings align with this principle but add further complexity. While voice-based interactions feel more natural and immersive, they might lead to increased cognitive load, particularly in demanding tasks such as clinical simulations or when technical problems occur. This indicates that enhancing the naturalness of an interaction does not necessarily alleviate cognitive strain; in some cases, it may even exacerbate it. The key lies in balancing immersion with cognitive load. Our research suggests that simpler environments, like ChatGPT (Open AI), tend to result in lower cognitive strain, making them more effective for specific tasks.

3.6.3 Practical implications

The insights derived from this study have significant practical implications for the design of educational simulations and other AI-driven tools. By understanding the balance between social presence and cognitive load, designers can optimize the user experience more effectively.

For instance, while voice-based interactions can make simulations feel more natural, designers should consider the cognitive burden imposed by such interactions. In scenarios where technical reliability is uncertain, it might be more practical to use text-based or simpler AI interfaces, which generally create fewer distractions and reduce the risk of cognitive overload. This is particularly relevant for high-stakes training environments, such as medical or clinical simulations, where maintaining cognitive focus is critical.

Moreover, these findings suggest that immersive VR environments should be used selectively, focusing more on reliability and interaction fluidity rather than visual realism alone. By doing so, developers can create tools that not only feel natural but also facilitate effective learning by managing cognitive load appropriately.

3.6.4 Limitations

While this study provides valuable insights into the impact of GenAI-powered voice interaction and simulation environment fidelity on user experience and attitudes towards technology, several limitations should be acknowledged.

One limitation is the difference in textual modalities between the desktop VR and ChatGPT (OpenAI) environments. In the desktop VR environment, interactions were mediated through predefined choice-based text options, whereas in the ChatGPT (OpenAI) environment, interactions were more open-ended and user-generated. This disparity in text interaction styles might have influenced user perceptions and experiences, potentially confounding the results. Future research should aim to standardize textual interaction modalities across different environments to ensure more comparable results.

Another limitation pertains to the difference in voice quality between the Embody system used in the desktop VR environment and the ChatGPT (OpenAI) environment. The voice synthesis and recognition capabilities in these two systems vary, potentially affecting the naturalness and clarity of voice interactions. This discrepancy might have influenced user perceptions of social presence and emotional engagement. Future studies should strive to use more uniform voice technologies across different environments to mitigate this issue. Another notable limitation is the complexity of VR environments, which necessitated a more comprehensive onboarding tutorial for participants. This extended onboarding process might have influenced various variables, as participants required additional time to adapt to the VR interface and controls.

The study encountered various technological limitations and errors that could have impacted the results. These include issues with speech recognition accuracy, response delays, and occasional system crashes. Such technical problems may have affected the smoothness and perceived quality of interactions, thereby influencing user experience and cognitive load. Addressing these technological limitations in future studies is crucial for obtaining more reliable and valid data.

Lastly, the English proficiency of participants also poses a limitation. As the study required participants to interact with the system in English, varying levels of English fluency among participants could have impacted their ability to engage with the simulations effectively. This variation may have influenced their cognitive and emotional experiences, as well as their overall attitudes towards the technology. Ensuring a more homogeneous level of language proficiency or providing language support in future studies could help address this limitation.

3.7 Conclusion

This study reinforces the critical role that interaction modality and simulation environment fidelity play in shaping user experiences in AI-powered simulations. While immersive environments like desktop VR offer heightened social presence, less complex environments, such as ChatGPT, can provide smoother and more natural interactions with lower cognitive effort. The findings suggest that the perceived naturalness and reliability of AI systems, rather than visual realism, are crucial for enhancing user engagement and reducing cognitive strain. As AI-driven simulations continue to evolve, designers must consider these trade-offs to create balanced, user-friendly systems that promote effective learning while maintaining positive emotional engagement. Future research should explore further optimizing these interactions to meet diverse user needs, ultimately advancing the integration of AI in educational and healthcare simulations.

Références

Abdulrahman, A., & Richards, D. (2022). Is Natural Necessary? Human Voice versus Synthetic Voice for Intelligent Virtual Agents. Multimodal Technologies and Interaction, 6(7), 51. https://doi.org/10.3390/mti6070051

Ackley, B. J., & Ladwig, G. B. (2008). Nursing diagnosis handbook. (No Title). https://cir.nii.ac.jp/crid/1130000793882549376

Adeghe, E. P., Okolo, C. A., & Ojeyinka, O. T. (2024). A review of the integration of virtual reality in healthcare : Implications for patient education and treatment outcomes. International Journal of Science and Technology Research Archive, 6(01), 079‑088.

Agarwal, R., & Karahanna, E. (2000). Time Flies When You're Having Fun : Cognitive Absorption and Beliefs about Information Technology Usage. MIS Quarterly, 24(4), 665‑694. https://doi.org/10.2307/3250951

Alario-Hoyos, C., Rodríguez-Triana, M. J., Scheffel, M., Arnedillo-Sánchez, I., & Dennerlein, S. M. (Éds.). (2020). Addressing Global Challenges and Quality Education : 15th European Conference on Technology Enhanced Learning, EC-TEL 2020, Heidelberg, Germany, September 14–18, 2020, Proceedings (Vol. 12315). Springer International Publishing. https://doi.org/10.1007/978-3-030-57717-9

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., Al Yami, M. S., Al Harbi, S., & Albekairy, A. M. (2023). Revolutionizing healthcare : The role of artificial intelligence in clinical practice. BMC Medical Education, 23(1), 689. https://doi.org/10.1186/s12909-023-04698-z

AL-Smadi, M. (2023). ChatGPT and Beyond : The Generative AI Revolution in Education (arXiv:2311.15198). arXiv. http://arxiv.org/abs/2311.15198

Andersen, S. A. W., Mikkelsen, P. T., Konge, L., Cayé-Thomasen, P., & Sørensen, M. S. (2016). The effect of implementing cognitive load theory-based design principles in virtual reality simulation training of surgical skills : A randomized controlled trial. Advances in Simulation, 1(1), 20. https://doi.org/10.1186/s41077-016-0022-1

Anthropomorphic virtual assistants and the reactions of Internet users : An experiment on the assistant's voice—Emna Chérif, Jean-François Lemoine, 2019. (s. d.). Consulté 22 juillet 2024, à l'adresse https://journals.sagepub.com/doi/abs/10.1177/2051570719829432?journalCode=rmea

Bailey, S., Johnson, C., Schroeder, B., & Marraffino, M. (2017). Using Virtual Reality for Training Maintenance Procedures.

Battegazzorre, E., Bottino, A., & Lamberti, F. (2021). Training Medical Communication Skills with Virtual Patients : Literature Review and Directions for Future Research. In N. Shaghaghi, F. Lamberti, B. Beams, R. Shariatmadari, & A. Amer (Éds.), Intelligent Technologies for Interactive Entertainment (p. 207‑226). Springer International Publishing. https://doi.org/10.1007/978-3-030-76426-5_14

Betella, A., & Verschure, P. F. M. J. (2016). The Affective Slider : A Digital Self-Assessment Scale for the Measurement of Human Emotions. PLOS ONE, 11(2), e0148037. https://doi.org/10.1371/journal.pone.0148037

Bhattacherjee, A. (2001). Understanding Information Systems Continuance : An Expectation-Confirmation Model. MIS Quarterly, 25(3), 351. https://doi.org/10.2307/3250921

Boletsis, C., & Karahasanovic, A. (2020). Immersive Technologies in Retail : Practices of Augmented and Virtual Reality: Proceedings of the 4th International Conference on Computer-Human Interaction Research and Applications, 281‑290. https://doi.org/10.5220/0010181702810290

Boscardin, C. K., Gin, B., Golde, P. B., & Hauer, K. E. (2024). ChatGPT and generative artificial intelligence for medical education : Potential impact and opportunity. Academic Medicine, 99(1), 22‑27.

Briganti, G., & Le Moine, O. (2020). Artificial intelligence in medicine : Today and tomorrow. Frontiers in medicine, 7, 509744.

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., & Allspach, L. (2000). Interactivity in human±computer interaction : A study of credibility, understanding, and in¯uence. Computers in Human Behavior.

Cain, W. (2024). Prompting Change : Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education. TechTrends, 68(1), 47‑57. https://doi.org/10.1007/s11528-023-00896-0

Carbonell, E. J. G., & Siekmann, J. (s. d.). Lecture Notes in Artificial Intelligence.

Carmichael, L., Poirier, S.-M., Coursaris, C., Léger, P.-M., & Sénécal, S. (2021). Does Media Richness Influence the User Experience of Chatbots : A Pilot Study. In F. D. Davis, R. Riedl, J. Vom Brocke, P.-M. Léger, A. B. Randolph, & G. Müller-Putz (Éds.), Information Systems and Neuroscience (Vol. 52, p. 204‑213). Springer International Publishing. https://doi.org/10.1007/978-3-030-88900-5_23

Carnell, S., Miles, A., & Lok, B. (2022). Evaluating Virtual Patient Interaction Fidelity With Advanced Communication Skills Learners. Frontiers in Virtual Reality, 2. https://doi.org/10.3389/frvir.2021.801793

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. Cognition and Instruction, 8(4), 293‑332. https://doi.org/10.1207/s1532690xci0804_2

Chapman, D. S., Uggerslev, K. L., & Webster, J. (2003). Applicant reactions to face-to-face and technology-mediated interviews : A field investigation. Journal of Applied Psychology, 88(5), 944‑953. https://doi.org/10.1037/0021-9010.88.5.944

Charland, P., Lapierre, H., Skelling, Y., Léger, P.-M., & Simon, G. (2018). L'utilisation de l'électroencéphalographie : La collecte et l'analyse en continu. In S. Masson & G. Borst (Éds.), Méthodes de recherche en neuroéducation (p. 219‑242). Presses de l'Universite Du Quebec. https://hal.science/hal-04056866

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education : A Review. IEEE Access, 8, 75264‑75278. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2988510

Chen, S., & Epps, J. (2014). Using Task-Induced Pupil Diameter and Blink Rate to Infer Cognitive Load. Human–Computer Interaction, 29(4), 390‑413. https://doi.org/10.1080/07370024.2014.892428

Chin, M. G., Sims, V. K., Clark, B., & Lopez, G. R. (2004). Measuring Individual Differences in Anthropomorphism toward Machines and Animals. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48(11), 1252‑1255. https://doi.org/10.1177/154193120404801110

Chui, M., Roberts, R., & Yee, L. (s. d.). Generative AI is here : How tools like ChatGPT could change your business.

Clark, R. E., & Choi, S. (2005). Five Design Principles for Experiments on the Effects of Animated Pedagogical Agents. Journal of Educational Computing Research, 32(3), 209‑225. https://doi.org/10.2190/7LRM-3BR2-44GW-9QQY

Cook, D. A., Erwin, P. J., & Triola, M. M. (2010). Computerized Virtual Patients in Health Professions Education : A Systematic Review and Meta-Analysis. Academic Medicine, 85(10), 1589. https://doi.org/10.1097/ACM.0b013e3181edfe13

Cornelius, S., & Leidner, D. (2021). Acceptance of Anthropomorphic Technology : A Literature Review. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2021.774

Csikszentmihalyi, M. (2013). Flow : The psychology of happiness. Random House. https://legenslife.wordpress.com/wp-content/uploads/2018/03/the-psychology-of-optimal-experience.pdf

Daher, S., Clark, A., & Barmaki, R. (2022). Editorial : Immersive Technologies in Healthcare. Frontiers in Virtual Reality, 3. https://doi.org/10.3389/frvir.2022.962950

Davies, A. J., & Krame, G. (2024). Measuring the Level of Fidelity Required for Transfer of Learning in Simulation-Based Learning Exercises for Novice and Experienced Practitioners. Simulation & Gaming, 10468781241241548. https://doi.org/10.1177/10468781241241548

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, 319‑340.

Dougherty, L., & Lister, S. (2015). The Royal Marsden manual of clinical nursing procedures. John Wiley & Sons. https://books.google.com/books?hl=fr&lr=&id=ATAABwAAQBAJ&oi=fnd&pg=PR13&dq=Dougherty,+L.,+Lister,+S.,+%26+West-Oram,+A.+(Eds.).+(2015).+The+royal+Marsden+manual+of+clinical+nursing+procedures.+John+Wiley+%26+Sons.&ots=jM6STnh4kk&sig=olpzjCggj5KfCd4bI5YfRJ0hCeQ

Emerging Roles of Virtual Patients in the Age of AI. (2019). AMA Journal of Ethics, 21(2), E153-159. https://doi.org/10.1001/amajethics.2019.153

English, T. S., Hooe, D. C., Moen, H. G., Murray, C. J., & Gerling, G. J. (2010). Enhancing user presence in a chest tube simulator by Joining virtual reality with a mannequin. 2010 IEEE Systems and Information Engineering Design Symposium, 245‑250. https://doi.org/10.1109/SIEDS.2010.5469647

Ferrell, T., Crowson, A., & Mayhorn, C. (2022). How We Perceive and Trust Advice from Virtual Humans : The Influence of Voice Quality. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 66(1), 1189‑1193. https://doi.org/10.1177/1071181322661440

Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer von Wolff, R., Hobert, S., & Luger, E. (2021). Future directions for chatbot research : An interdisciplinary research agenda. Computing, 103(12), 2915‑2942. https://doi.org/10.1007/s00607-021-01016-7

Friedman, B., Kahn, P. H., & Hagman, J. (2003). Hardware companions? : What online AIBO discussion forums reveal about the human-robotic relationship. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 273‑280. https://doi.org/10.1145/642611.642660

Gefen & Straub. (2003). Managing User Trust in B2C e-Services. E-Service Journal, 2(2), 7. https://doi.org/10.2979/esj.2003.2.2.7

Giroux, F., Boasen, J., Sénécal, S., Fredette, M., Tchanou, A. Q., Ménard, J.-F., Paquette, M., & Léger, P.-M. (2019). Haptic Stimulation with High Fidelity Vibro-Kinetic Technology Psychophysiologically Enhances Seated Active Music Listening Experience. 2019 IEEE World Haptics Conference (WHC), 151‑156. https://doi.org/10.1109/WHC.2019.8816115

Gonçalves, G., Coelho, H., Monteiro, P., Melo, M., & Bessa, M. (2022). Systematic Review of Comparative Studies of the Impact of Realism in Immersive Virtual Experiences. ACM Computing Surveys, 55(6), 115:1-115:36. https://doi.org/10.1145/3533377

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Hamilton, A. (2024). Artificial Intelligence and Healthcare Simulation : The Shifting Landscape of Medical Education. Cureus. https://doi.org/10.7759/cureus.59747

Hassanein, K., & Head, M. (2005). The Impact of Infusing Social Presence in the Web Interface : An Investigation Across Product Types. International Journal of Electronic Commerce, 10(2), 31‑55. https://doi.org/10.2753/JEC1086-4415100202

Hassanein, K., & Head, M. (2007). Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. International journal of human-computer studies, 65(8), 689‑708.

Helie et *al. (2025). Integrating Generative AI in Simulation Training: Comparing the Educational Dynamics of LLM-Powered Vocal With Choice-Based Conversational Agent Interactions in Role-Play Problem-Based Learning Context. Master thesis; HEC Montréal*

Heston, T., & Khun, C. (2023). Prompt Engineering in Medical Education. International Medical Education, 2(3), 198‑205. https://doi.org/10.3390/ime2030019

Heyselaar, E., Hagoort, P., & Segaert, K. (2017). In dialogue with an avatar, language behavior is identical to dialogue with a human partner. Behavior Research Methods, 49(1), 46‑60. https://doi.org/10.3758/s13428-015-0688-7

Higgs, J., Jones, M. A., Loftus, S., & Christensen, N. (2008). Clinical Reasoning in the Health Professions E-Book. Elsevier Health Sciences.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. Science, 349(6245), 261‑266. https://doi.org/10.1126/science.aaa8685

Holmes, W., Bialik, M., & Fadel, C. (2023). Artificial intelligence in education. In C. Stückelberger & P. Duggal (Éds.), Data ethics : Building trust : How digital technologies can serve humanity (p. 621‑653). Globethics Publications. https://doi.org/10.58863/20.500.12424/4276068

Hone, K., & Graham, R. (2001). Subjective assessment of speech-system interface usability (p. 2086). https://doi.org/10.21437/Eurospeech.2001-491

Hone, K. S., & Graham, R. (2000). Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). Natural Language Engineering, 6(3 & 4), 287‑303. https://doi.org/10.1017/S1351324900002497

Hornbæk, K., & Hertzum, M. (2017). Technology Acceptance and User Experience : A Review of the Experiential Component in HCI. ACM Transactions on Computer-Human Interaction, 24(5), 1‑30. https://doi.org/10.1145/3127358

Hsieh, C.-Y., Chen, S.-A., Li, C.-L., Fujii, Y., Ratner, A., Lee, C.-Y., Krishna, R., & Pfister, T. (2023). Tool Documentation Enables Zero-Shot Tool-Usage with Large Language Models (arXiv:2308.00675). arXiv. http://arxiv.org/abs/2308.00675

Huang, C. L., Luo, Y. F., Yang, S. C., Lu, C. M., & Chen, A.-S. (2020). Influence of Students' Learning Style, Sense of Presence, and Cognitive Load on Learning Outcomes in an Immersive Virtual Reality Learning Environment. Journal of Educational Computing Research, 58(3), 596‑615. https://doi.org/10.1177/0735633119867422

Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Gordon, D. L., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning : A BEME systematic review. Medical Teacher, 27(1), 10‑28. https://doi.org/10.1080/01421590500046924

Jacoby, J. (2002). Stimulus-Organism-Response Reconsidered : An Evolutionary Step in Modeling (Consumer) Behavior. Journal of Consumer Psychology, 12(1), 51‑57. https://doi.org/10.1207/S15327663JCP1201_05

Jiang, X., Atkins, M. S., Tien, G., Bednarik, R., & Zheng, B. (2014). Pupil responses during discrete goal-directed movements. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2075‑2084. https://doi.org/10.1145/2556288.2557086

Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. Science, 154(3756), 1583‑1585. https://doi.org/10.1126/science.154.3756.1583

Karahanna, E., Agarwal, R., & Angst, C. M. (2006). Reconceptualizing Compatibility Beliefs in Technology Acceptance Research. MIS Quarterly, 30(4), 781‑804. https://doi.org/10.2307/25148754

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Ke, F., Dai, Z., Dai, C.-P., Pachman, M., Chaulagain, R. S., & Yuan, X. (2020). Designing virtual agents for simulation-based learning in virtual reality. In Cognitive and affective perspectives on immersive technology in education (p. 151‑170). IGI Global. https://www.igi-global.com/chapter/designing-virtual-agents-for-simulation-based-learning-in-virtual-reality/253694

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. Psychophysiology, 48(3), 323‑332. https://doi.org/10.1111/j.1469-8986.2010.01069.x

Komasawa, N., & Yokohira, M. (2023). Simulation-based education in the artificial intelligence era. Cureus, 15(6). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10368461/

Kong, A., Zhao, S., Chen, H., Li, Q., Qin, Y., Sun, R., & Zhou, X. (2023). Better Zero-Shot Reasoning with Role-Play Prompting (arXiv:2308.07702). arXiv. https://doi.org/10.48550/arXiv.2308.07702

Kononowicz, A. A., Woodham, L. A., Edelbring, S., Stathakarou, N., Davies, D., Saxena, N., Car, L. T., Carlstedt-Duke, J., Car, J., & Zary, N. (2019). Virtual Patient Simulations in Health Professions Education : Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. Journal of Medical Internet Research, 21(7), e14676. https://doi.org/10.2196/14676

Kroes, K., Saccardi, I., & Masthoff, J. (2022). Empathizing with virtual agents : The effect of personification and general empathic tendencies. 2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), 73‑81. https://doi.org/10.1109/AIVR56993.2022.00017

Kruger, J.-L., & Steyn, F. (2014). Subtitles and Eye Tracking : Reading and Performance. Reading Research Quarterly, 49(1), 105‑120. https://doi.org/10.1002/rrq.59

Kulms, P., & Kopp, S. (2018). A social cognition perspective on human–computer trust : The effect of perceived warmth and competence on trust in decision-making with computers. Frontiers in Digital Humanities, 5, 352444.

Kulms, P., & Kopp, S. (2019). More Human-Likeness, More Trust? : The Effect of Anthropomorphism on Self-Reported and Behavioral Trust in Continued and Interdependent Human-Agent Cooperation. Proceedings of Mensch Und Computer 2019, 31‑42. https://doi.org/10.1145/3340764.3340793

Kumar, N., & Benbasat, I. (2006). Research Note : The Influence of Recommendations and Consumer Reviews on Evaluations of Websites. Information Systems Research, 17(4), 425‑439. https://doi.org/10.1287/isre.1060.0107

Lamontagne, C., Sénécal, S., Fredette, M., Chen, S. L., Pourchon, R., Gaumont, Y., De Grandpré, D., & Léger, P.-M. (2020). User Test : How Many Users Are Needed to Find the Psychophysiological Pain Points in a Journey Map? In T. Ahram, R. Taiar, S. Colson, & A. Choplin (Éds.), Human Interaction and Emerging Technologies (p. 136‑142). Springer International Publishing. https://doi.org/10.1007/978-3-030-25629-6_22

Lang, P. J. (s. d.). The Varieties of Emotional Experience : A Meditation on James-Lange Theory.

Lee, J., Kim, H., Kim, K. H., Jung, D., Jowsey, T., & Webster, C. S. (2020). Effective virtual patient simulators for medical communication training : A systematic review. Medical Education, 54(9), 786‑795. https://doi.org/10.1111/medu.14152

Léger, P., Cronan, P., Charland, P., Pellerin, R., Babin, G., & Robert, J. (2012). Authentic OM problem solving in an ERP context. International Journal of Operations & Production Management, 32(12), 1375‑1394. https://doi.org/10.1108/01443571211284151

Liew, T. W., Mat Zin, N. A., & Sahari, N. (2017). Exploring the affective, motivational and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment. Human-Centric Computing and Information Sciences, 7(1), 9. https://doi.org/10.1186/s13673-017-0089-2

Liu, J., Liu, A., Lu, X., Welleck, S., West, P., Bras, R. L., Choi, Y., & Hajishirzi, H. (2022). Generated Knowledge Prompting for Commonsense Reasoning (arXiv:2110.08387). arXiv. https://doi.org/10.48550/arXiv.2110.08387

MacLean, S., Geddes, F., Kelly, M., & Della, P. (2019). Realism and Presence in Simulation : Nursing Student Perceptions and Learning Outcomes. Journal of Nursing Education, 58(6), 330‑338. https://doi.org/10.3928/01484834-20190521-03

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution : Its impact on society and firms. Futures, 90, 46‑60. https://doi.org/10.1016/j.futures.2017.03.006

Mannuru, N. R., Shahriar, S., Teel, Z. A., Wang, T., Lund, B. D., Tijani, S., Pohboon, C. O., Agbaji, D., Alhassan, J., Galley, J., Kousari, R., Ogbadu-Oladapo, L., Saurav, S. K., Srivastava, A., Tummuru, S. P., Uppala, S., & Vaidya, P. (2023). Artificial intelligence in developing countries : The impact of generative artificial intelligence (AI) technologies for development. Information Development, 02666669231200628. https://doi.org/10.1177/02666669231200628

Massoth, C., Roeder, H., Ohlenburg, H., Hessler, M., Zarbock, A., Poepping, D. M., & Wenk, M. (2019). High-fidelity is not superior to low-fidelity simulation but leads to overconfidence in medical students. Bmc Medical Education, 19, 29. https://doi.org/10.1186/s12909-019-1464-7

Matsui, T., & Yamada, S. (2019). Designing Trustworthy Product Recommendation Virtual Agents Operating Positive Emotion and Having Copious Amount of Knowledge. Frontiers in Psychology, 10, 675. https://doi.org/10.3389/fpsyg.2019.00675

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley [From the Field]. IEEE Robotics & Automation Magazine, 19(2), 98‑100. IEEE Robotics & Automation Magazine. https://doi.org/10.1109/MRA.2012.2192811

Nass, C., & Moon, Y. (2000). Machines and Mindlessness : Social Responses to Computers. Journal of Social Issues, 56(1), 81‑103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 72‑78. https://doi.org/10.1145/191666.191703

Navath, S. (2021). Transforming Healthcare : The Impact and Future of Artificial Intelligence in Healthcare. Journal of Artificial intelligence and Machine Learning, 1(1), 16‑21.

Ning, X., Lin, Z., Zhou, Z., Yang, H., & Wang, Y. (2023). Skeleton-of-Thought : Large Language Models Can Do Parallel Decoding (arXiv:2307.15337). arXiv. http://arxiv.org/abs/2307.15337

Norman, D. (2007). Emotional design : Why we love (or hate) everyday things. Basic books. https://books.google.com/books?hl=fr&lr=&id=QtWSu3zBtPoC&oi=fnd&pg=PA3&dq=Norman,+D.+A.+(2004).+Emotional+design:+Why+we+love+(or+hate)+everyday+things.+Basic+Books&ots=9kwynV21uJ&sig=4rd4a6D2fKcbVaQEm6ZalbAXcDY

Norman, G., Dore, K., & Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. Medical Education, 46(7), 636‑647. https://doi.org/10.1111/j.1365-2923.2012.04243.x

North, M. M., & North, S. M. (2016). A Comparative Study of Sense of Presence of Traditional Virtual Reality and Immersive Environments. Australasian Journal of Information Systems, 20. https://doi.org/10.3127/ajis.v20i0.1168

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). A Systematic Review of Generative AI for Teaching and Learning Practice. Education Sciences, 14(6), 636.

Oliver, R. L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. Journal of Marketing Research, 17(4), 460‑469. https://doi.org/10.2307/3150499

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes : Evidence from the computer games literature. The Curriculum Journal, 16(4), 455‑474. https://doi.org/10.1080/09585170500384529

Ooi, K.-B., Tan, G. W.-H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., Dwivedi, Y. K., Huang, T.-L., Kar, A. K., Lee, V.-H., Loh, X.-M., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N. P., Sarker, P., Sharma, A., … Wong, L.-W. (2023). The Potential of Generative Artificial Intelligence Across Disciplines : Perspectives and Future Directions. Journal of Computer Information Systems, 1‑32. https://doi.org/10.1080/08874417.2023.2261010

Orru, G., & Longo, L. (2019). The Evolution of Cognitive Load Theory and the Measurement of Its Intrinsic, Extraneous and Germane Loads : A Review. In L. Longo & M. C. Leva (Éds.), Human Mental Workload : Models and Applications (Vol. 1012, p. 23‑48). Springer International Publishing. https://doi.org/10.1007/978-3-030-14273-5_3

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design : Recent Developments. Educational Psychologist, 38(1), 1‑4. https://doi.org/10.1207/S15326985EP3801_1

Padilha, J. M., Machado, P. P., Ribeiro, A., Ramos, J., & Costa, P. (2019). Clinical Virtual Simulation in Nursing Education : Randomized Controlled Trial. Journal of Medical Internet Research, 21(3), e11529. https://doi.org/10.2196/11529

Pauna, H., Léger, P.-M., Sénécal, S., Fredette, M., Courtemanche, F., Chen, S.-L., Labonté-Lemoyne, É., & Ménard, J.-F. (2018). The Psychophysiological Effect of a Vibro-Kinetic Movie Experience : The Case of the D-BOX Movie Seat. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Éds.), Information Systems and Neuroscience (p. 1‑7). Springer International Publishing. https://doi.org/10.1007/978-3-319-67431-5_1

Picard, R. W. (2000). Affective computing. MIT press. https://books.google.com/books?hl=fr&lr=&id=GaVncRTcb1gC&oi=fnd&pg=PR9&dq=Affective+computing.+MIT+Press.&ots=F6janmscd7&sig=Kh0xBYRkvoPA-6Hm6Ou5lNfPKzc

Pilote, B., & Chiniara, G. (2019). Chapter 2—The Many Faces of Simulation. In G. Chiniara (Éd.), Clinical Simulation (Second Edition) (p. 17‑32). Academic Press. https://doi.org/10.1016/B978-0-12-815657-5.00002-4

Pottle, J. (2019). Virtual reality and the transformation of medical education. Future Healthcare Journal, 6(3), 181‑185. https://doi.org/10.7861/fhj.2019-0036

Prompt Engineering for Generative AI. (s. d.). Consulté 1 février 2024, à l'adresse https://learning.oreilly.com/library/view/prompt-engineering-for/9781098153427/

Reeves, B., & Nass, C. (1996). The media equation : How people treat computers, television, and new media like real people. Cambridge, UK, 10(10). https://www.academia.edu/download/50585912/The_Media_Equation_How_People_Treat_Comp20161127-17472-1s43e41.pdf

Riedl, R., & Léger, P.-M. (2016). Fundamentals of NeuroIS : Information Systems and the Brain. Springer. https://doi.org/10.1007/978-3-662-45091-8

Rizzo, A. “Skip”, & Bouchard, S. (Éds.). (2019). Virtual Reality for Psychological and Neurocognitive Interventions. Springer New York. https://doi.org/10.1007/978-1-4939-9482-3

Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., & Landay, J. (2016). Speech is 3x faster than typing for english and mandarin text entry on mobile devices. arXiv preprint arXiv:1608.07323. https://hci.stanford.edu/research/speech/paper/speech_paper.pdf

Russell, J. A. (1980). A circumplex model of affect. Journal of personality and social psychology, 39(6), 1161.

Rystedt, H., & Sjöblom, B. (2012). Realism, authenticity, and learning in healthcare simulations : Rules of relevance and irrelevance as interactive achievements. Instructional Science, 40(5), 785‑798. https://doi.org/10.1007/s11251-012-9213-x

Saab, M. M., Landers, M., Murphy, D., Cooke, E., & Hegarty, J. (2022). Nursing students’ views of using virtual reality in healthcare : A qualitative study. Journal of Clinical Nursing, 31(9‑10), 1228‑1242. https://doi.org/10.1111/jocn.15978

Sadowski, W., & Stanney, K. (2002). Presence in Virtual Environments. In Handbook of Virtual Environments. CRC Press.

Sahin Karaduman, G., & Basak, T. (2024). Virtual Patient Simulations in Nursing Education : A Descriptive Systematic Review. Simulation & Gaming, 55(2), 159‑179. https://doi.org/10.1177/10468781231224836

Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice : Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare, 11(6), Article 6. https://doi.org/10.3390/healthcare11060887

Sauder, M., Tritsch, T., Rajput, V., Schwartz, G., & Shoja, M. M. (2024). Exploring generative artificial intelligence-assisted medical education : Assessing case-based learning for medical students. Cureus, 16(1). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10852982/

Savery, J. R., & Duffy, T. M. (1995). Problem based learning : An instructional model and its constructivist framework. Educational technology, 35(5), 31‑38.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being : Informative and directive functions of affective states. Journal of personality and social psychology, 45(3), 513.

Sezgin, E., & D’Arcy, S. (2022). Editorial : Voice Technology and Conversational Agents in Health Care Delivery. Frontiers in Public Health, 10, 887492. https://doi.org/10.3389/fpubh.2022.887492

Shankar, P. R. (2022). Artificial Intelligence in Health Professions Education. Archives of Medicine and Health Sciences, 10(2), 256‑261. https://doi.org/10.4103/amhs.amhs_234_22

Sheldon, L. K., & Hilaire, D. M. (2015). Development of communication skills in healthcare : Perspectives of new graduates of undergraduate nursing education. Journal of Nursing Education and Practice, 5(7), p30. https://doi.org/10.5430/jnep.v5n7p30

Sittner, B. J., Aebersold, M. L., Paige, J. B., Graham, L. L., Schram, A. P., Decker, S. I., & Lioce, L. (2015). INACSL standards of best practice for simulation : Past, present, and future. Nursing education perspectives, 36(5), 294‑298.

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. International journal of psychophysiology, 52(1), 77‑86.

Stevens, A., Hernandez, J., Johnsen, K., Dickerson, R., Raij, A., Harrison, C., DiPietro, M., Allen, B., Ferdig, R., Foti, S., Jackson, J., Shin, M., Cendan, J., Watson, R., Duerson, M., Lok, B., Cohen, M., Wagner, P., & Lind, D. S. (2006). The use of virtual patients to teach medical students history taking and communication skills. The American Journal of Surgery, 191(6), 806‑811. https://doi.org/10.1016/j.amjsurg.2006.03.002

Stone, R., Panfilov, P., & Shukshunov, V. (2011). Evolution of Aerospace Simulation : From Immersive Virtual Reality to Serious Games. In RAST 2011—Proceedings of 5th International Conference on Recent Advances in Space Technologies. https://doi.org/10.1109/RAST.2011.5966921

Sun, W., Jiang, X., Dong, X., Yu, G., Feng, Z., & Shuai, L. (2024). The evolution of simulation-based medical education research : From traditional to virtual simulations. Heliyon, 10(15). https://doi.org/10.1016/j.heliyon.2024.e35627

Sweller, J. (1988). Cognitive load during problem solving : Effects on learning. Cognitive science, 12(2), 257‑285.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive Load Theory. Springer New York. https://doi.org/10.1007/978-1-4419-8126-4

Tam, J., Carter, E., Kiesler, S., & Hodgins, J. (2012). Video increases the perception of naturalness during remote interactions with latency. CHI ’12 Extended Abstracts on Human Factors in Computing Systems, 2045‑2050. https://doi.org/10.1145/2212776.2223750

Tam, W., Huynh, T., Tang, A., Luong, S., Khatri, Y., & Zhou, W. (2023). Nursing education in the age of artificial intelligence powered Chatbots (AI-Chatbots) : Are we ready yet? Nurse Education Today, 129, 105917. https://doi.org/10.1016/j.nedt.2023.105917

Titah, R., & Barki, H. (2009). Nonlinearities between Attitude and Subjective Norms in Information Technology Acceptance : A Negative Synergy? MIS Quarterly, 33(4), 827‑844. https://doi.org/10.2307/20650329

Torre, I., Latupeirissa, A. B., & McGinn, C. (2020). How context shapes the appropriateness of a robot’s voice. 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 215‑222. https://ieeexplore.ieee.org/abstract/document/9223449/?casa_token=F0mew2C873MAAAAA:tYGQHqPPa0oX9fMD7EnFJzZNV_rYozE-kEzvpcILVgwTWAn-cpCG4ZauBZivppxzaQR1zjgWSWSB

Torta, E., Oberzaucher, J., Werner, F., Cuijpers, R. H., & Juola, J. F. (2013). Attitudes Towards Socially Assistive Robots in Intelligent Homes : Results From Laboratory Studies and Field Trials. Journal of Human-Robot Interaction, 1(2), 76‑99. https://doi.org/10.5898/JHRI.1.2.Torta

Tun, J. K., Alinier, G., Tang, J., & Kneebone, R. L. (2015). Redefining Simulation Fidelity for Healthcare Education. Simulation & Gaming, 46(2), 159‑174. https://doi.org/10.1177/1046878115576103

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model : Four Longitudinal Field Studies. Management Science, 46(2), 186‑204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology : Toward a unified view. MIS quarterly, 425‑478.

Vora, J., Nair, S., Gramopadhye, A. K., Duchowski, A. T., Melloy, B. J., & Kanki, B. (2002). Using virtual reality technology for aircraft visual inspection training : Presence and comparison studies. Applied Ergonomics, 33(6), 559‑570. https://doi.org/10.1016/S0003-6870(02)00039-X

Walker, A., & Leary, H. (2009). A problem based learning meta analysis : Differences across problem types, implementation types, disciplines, and assessment levels. Interdisciplinary journal of problem-based learning, 3(1), 6.

Wang, N., Rebolledo-Mendez, G., Matsuda, N., Santos, O. C., & Dimitrova, V. (Éds.). (2023). Artificial Intelligence in Education : 24th International Conference, AIED 2023, Tokyo, Japan, July 3–7, 2023, Proceedings (Vol. 13916). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36272-9

Wang, Y. J., & Minor, M. S. (2008). Validity, reliability, and applicability of psychophysiological techniques in marketing research. Psychology & Marketing, 25(2), 197‑232. https://doi.org/10.1002/mar.20206

Wang, Y.-C., Xue, J., Wei, C., & Kuo, C.-C. J. (2023). An overview on generative AI at scale with Edge-Cloud Computing. IEEE Open Journal of the Communications Society. https://ieeexplore.ieee.org/abstract/document/10268594/

Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J.-H., & Cacioppo, J. T. (2010). Making sense by making sentient : Effectance motivation increases anthropomorphism. Journal of Personality and Social Psychology, 99(3), 410‑435. https://doi.org/10.1037/a0020240

Yeh, C. J., & Inose, M. (2003). International students' reported English fluency, social support satisfaction, and social connectedness as predictors of acculturative stress. Counselling Psychology Quarterly, 16(1), 15‑28. https://doi.org/10.1080/0951507031000114058

Young, M. E., Thomas, A., Lubarsky, S., Gordon, D., Gruppen, L. D., Rencic, J., Ballard, T., Holmboe, E., Da Silva, A., Ratcliffe, T., Schuwirth, L., Dory, V., & Durning, S. J. (2020). Mapping clinical reasoning literature across the health professions : A scoping review. BMC Medical Education, 20(1), 107. https://doi.org/10.1186/s12909-020-02012-9

Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. Nature Biomedical Engineering, 2(10), Article 10. https://doi.org/10.1038/s41551-018-0305-z

Zhang, P. (2013). The affective response model : A theoretical framework of affective concepts and their relationships in the ICT context. MIS quarterly, 247‑274.

Chapitre 4 : Conclusion

Notre étude sur l'intégration de l'intelligence artificielle générative (GenAI), en particulier des grands modèles de langage (LLM), dans les environnements d'apprentissage par simulation, a révélé des dynamiques complexes mais prometteuses en matière d'éducation. À travers l'ensemble des études présentées, ce mémoire démontre que l'utilisation des LLM dans des contextes de simulation de résolution de problèmes enrichit considérablement l'expérience utilisateur, en facilitant une interaction plus naturelle, intuitive et fidèle à la réalité. Plus particulièrement, les résultats indiquent que l'interaction vocale propulsée par l'IA, bien qu'exigeante sur le plan cognitif, permet une meilleure immersion dans les scénarios de simulation, un facteur clé dans la formation du personnel médical. Cependant, la modalité textuelle démontre une efficacité tout aussi viable avec réduction de la charge cognitive, sans sacrifier la qualité de l'expérience utilisateur. Ces résultats mettent en lumière la nécessité d'adapter le support de formation au niveau d'apprentissage de l'apprenant et aux objectifs de formation visés. Les conclusions théoriques montrent que la fidélité de l'environnement de simulation, bien qu'importante, ne constitue pas le seul facteur influençant la satisfaction et l'engagement des utilisateurs. De plus, pour des simulations nécessitant un haut degré de communication et d'interaction humaine, le succès de l'expérience dépend surtout de la fluidité, de la précision et à la fois de la richesse des réponses fournies par les systèmes IA, ce qui place en perspective l'importance exclusive du réalisme visuel souvent privilégié dans les simulations de réalité virtuelle. Nos travaux de recherche invitent à repenser l'approche traditionnelle des environnements d'apprentissage immersifs, en soulignant que l'interaction humaine médiatisée par l'IA doit être au centre des considérations de conception, dans ses plus fins détails techniques, plutôt qu'une simple recherche de réalisme visuel.

D'un point de vue managérial, les implications de cette recherche sont éclairantes, en particulier pour les responsables de la conception d'applications ou de technologies éducatives. Nos recherches démontrent que l'intégration des LLM peut considérablement améliorer l'efficacité des outils d'apprentissage si elle est bien implémentée. Les gestionnaires et développeurs doivent donc évaluer attentivement l'adéquation entre les besoins éducatifs des utilisateurs et la capacité des technologies à répondre à ces besoins. Dans les contextes où l'objectif principal est l'engagement émotionnel, les compétences de communication et la préparation avancée, l’interaction vocale est à privilégier. Toutefois, dans des environnements nécessitant un rythme plus lent, de la réflexion, une précision accrue et une charge cognitive réduite (e.g. en début de cursus, pour les apprenants novices), la modalité textuelle offre un avantage certain. De plus, les gestionnaires doivent se concentrer sur l'intégration d'outils adaptatifs et flexibles qui peuvent évoluer avec la technologie et les apprenants. Nos études suggèrent que des améliorations incrémentales, notamment en matière de reconnaissance vocale et de fluidité des interactions, sont nécessaires pour que ces technologies soient pleinement adoptées à grande échelle. Par ailleurs, les coûts liés à la mise en place de ces technologies doivent être équilibrés avec les avantages éducatifs qu'elles offrent, en tenant compte du fait que des technologies comme ChatGPT (OpenAI) peuvent parfois maintenant surpasser des solutions plus coûteuses, comme la réalité virtuelle, en termes de satisfaction utilisateur et d’atteintes de certains objectifs éducatifs.

Bien que cette recherche ait apporté des contributions significatives à la compréhension des dynamiques d'interaction avec des agents virtuels dans des environnements d'apprentissage, plusieurs limites doivent être reconnues. Premièrement, l'échantillon restreint des participants, dans les deux études, limite la généralisation des résultats. Les résultats obtenus sont pertinents pour le domaine de la formation médicale, mais des études plus larges seraient nécessaires, en plus d’être pertinent, pour confirmer ces résultats dans d'autres contextes éducatifs. De plus, les scénarios utilisés dans les simulations, bien qu'adaptés à l'évaluation de l'interaction avec l’IA, ne représentent pas nécessairement toutes les situations cliniques possibles. La sélection de ces scénarios précis, le contexte, pourrait avoir influencé les réponses des participants et les conclusions tirées quant à l'efficacité des différentes modalités d'interaction. Une autre limite notable est la variabilité des capacités techniques des systèmes d'IA employés, notamment la reconnaissance vocale, qui a parfois introduit des erreurs dans les interactions. Cela montre que les résultats obtenus dépendent en partie des limitations technologiques actuelles, qui pourraient être surmontées à mesure que les technologies d'IA évoluent.

L'une des contributions méthodologiques majeures de ce mémoire est la création du modèle *Constrained-Skeleton-of-Thought Virtual-Patient* (C-SOT-VP), un cadre conceptuel qui permet de générer des scénarios d'apprentissage personnalisés et adaptables pour le domaine infirmier dans un LLM. Ce modèle propose une structure méthodologique innovante qui guide le développement rapide de scénarios basés sur des objectifs d'apprentissage précis, tout en permettant une interaction flexible et enrichie avec l'agent virtuel. Ce cadre facilite l'intégration des LLM dans les milieux éducatifs et propose une solution adaptable pour d'autres domaines où l'interaction reproduisant une communication humaine est essentielle.

La conception d'un stimulus expérimental basé sur ChatGPT (OpenAI) représente aussi une avancée importante dans la méthodologie de la recherche comparative. Une des innovations de ce mémoire de recherche réside dans l'utilisation de l'IA pour simuler des interactions conversationnelles réalistes dans un cadre expérimental contrôlé et reproduire une expérience de simulation hautement complexe. Dans notre recherche, ce stimulus a permis d'explorer de manière approfondie les différences perçues par les utilisateurs entre deux types d'interaction (e.g. texte et voix) avec des agents virtuels, ouvrant ainsi la voie à de futures recherches sur l'optimisation de ces technologies dans des contextes éducatifs. Cependant, l'innovation centrale de ce mémoire est l'utilisation pionnière d'une IA générative pour créer un environnement conversationnel interactif de haute fidélité qui puisse être étudié en laboratoire à des fins de recherche. La conception d'une approche comparative (i.e. VR vs ChatGPT (OpenAI)) novatrice qui nous a permis d'évaluer de manière systématique l'efficacité, et du même coup d'explorer différentes capacités de simulation. Cela a permis de démontrer que la complexité technologique ne garantit pas nécessairement une meilleure expérience utilisateur lorsqu'il est question de simulation d'une communication humaine, mettant ainsi en évidence le potentiel de solutions moins coûteuses mais plus intuitives comme les interactions textuelles basées sur ChatGPT (OpenAI).

Pour les futures recherches, nos travaux nous emmènent à proposer d'adopter une approche rigoureuse en matière de conception et de développement expérimental, tout en restant ouverts aux ajustements itératifs. L'une des principales leçons tirées de ce mémoire est l'importance de pré-tester les technologies d'IA avant leur intégration complète dans les environnements de recherche. Non seulement les tests assurent un déroulement expérimental plus robuste, mais ils permettent également d'innover. De plus, la gestion des attentes des utilisateurs et la formation préalable à l'utilisation des technologies d'IA sont essentielles pour minimiser les biais dans la collecte de données. Les recherches futures devraient également s’attarder sur l’amélioration des capacités de ces agents, notamment en matière de compréhension du langage naturel et de reconnaissance vocale, pour permettre une interaction plus fluide et précise avec les utilisateurs. De plus, il serait intéressant d'examiner l'impact de ces technologies sur différents types d'apprenants (novices vs experts) ou domaines professionnels, pour mieux comprendre comment adapter les agents virtuels aux besoins spécifiques des utilisateurs à différents stades de leur formation et selon leur champ de spécialisation. Dans cette perspective, l’'approche centrée sur l'utilisateur est cruciale pour le développement de ces futures applications éducatives ; cette recherche illustre clairement l'importance et la valeur ajoutée de mener des études combinant des méthodes qualitatives et quantitatives dans un processus de conception, afin d'obtenir une vue d'ensemble plus complète des dynamiques d'interaction et de l'expérience utilisateur lors de l'évaluation et du développement de technologies innovantes.

Bibliographie

Abdulrahman, A., & Richards, D. (2022). Is Natural Necessary? Human Voice versus Synthetic Voice for Intelligent Virtual Agents. Multimodal Technologies and Interaction, 6(7), 51. https://doi.org/10.3390/mti6070051

Ackley, B. J., & Ladwig, G. B. (2008). Nursing diagnosis handbook. (No Title). https://cir.nii.ac.jp/crid/1130000793882549376

Adeghe, E. P., Okolo, C. A., & Ojeyinka, O. T. (2024). A review of the integration of virtual reality in healthcare : Implications for patient education and treatment outcomes. International Journal of Science and Technology Research Archive, 6(01), 079‑088.

Agarwal, R., & Karahanna, E. (2000). Time Flies When You're Having Fun : Cognitive Absorption and Beliefs about Information Technology Usage. MIS Quarterly, 24(4), 665‑694. https://doi.org/10.2307/3250951

Alario-Hoyos, C., Rodríguez-Triana, M. J., Scheffel, M., Arnedillo-Sánchez, I., & Dennerlein, S. M. (Éds.). (2020). Addressing Global Challenges and Quality Education : 15th European Conference on Technology Enhanced Learning, EC-TEL 2020, Heidelberg, Germany, September 14–18, 2020, Proceedings (Vol. 12315). Springer International Publishing. https://doi.org/10.1007/978-3-030-57717-9

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., Al Yami, M. S., Al Harbi, S., & Albekairy, A. M. (2023). Revolutionizing healthcare : The role of artificial intelligence in clinical practice. BMC Medical Education, 23(1), 689. https://doi.org/10.1186/s12909-023-04698-z

AL-Smadi, M. (2023). ChatGPT and Beyond : The Generative AI Revolution in Education (arXiv:2311.15198). arXiv. http://arxiv.org/abs/2311.15198

Andersen, S. A. W., Mikkelsen, P. T., Konge, L., Cayé-Thomasen, P., & Sørensen, M. S. (2016). The effect of implementing cognitive load theory-based design principles in virtual reality simulation training of surgical skills : A randomized controlled trial. Advances in Simulation, 1(1), 20. https://doi.org/10.1186/s41077-016-0022-1

Anthropomorphic virtual assistants and the reactions of Internet users : An experiment on the assistant's voice—Emna Chérif, Jean-François Lemoine, 2019. (s. d.). Consulté 22 juillet 2024, à l'adresse https://journals.sagepub.com/doi/abs/10.1177/2051570719829432?journalCode=rmea

Bailey, S., Johnson, C., Schroeder, B., & Marraffino, M. (2017). Using Virtual Reality for Training Maintenance Procedures.

Baily, L. W. (2019). History of Simulation. In S. B. Crawford, L. W. Baily, & S. M. Monks (Éds.), *Comprehensive Healthcare Simulation : Operations, Technology, and Innovative Practice* (p. 3‑11). Springer International Publishing. https://doi.org/10.1007/978-3-030-15378-6_1

Battegazzorre, E., Bottino, A., & Lamberti, F. (2021). Training Medical Communication Skills with Virtual Patients : Literature Review and Directions for Future Research. In N. Shaghaghi, F. Lamberti, B. Beams, R. Shariatmadari, & A. Amer (Éds.), Intelligent Technologies for Interactive Entertainment (p. 207‑226). Springer International Publishing. https://doi.org/10.1007/978-3-030-76426-5_14

Benner, P. (1984). From novice to expert. *Menlo Park*, *84*(1480), 10‑1097.

Betella, A., & Verschure, P. F. M. J. (2016). The Affective Slider : A Digital Self-Assessment Scale for the Measurement of Human Emotions. PLOS ONE, 11(2), e0148037. https://doi.org/10.1371/journal.pone.0148037

Bhattacherjee, A. (2001). Understanding Information Systems Continuance : An Expectation-Confirmation Model. MIS Quarterly, 25(3), 351. https://doi.org/10.2307/3250921

Boletsis, C., & Karahasanovic, A. (2020). Immersive Technologies in Retail : Practices of Augmented and Virtual Reality: Proceedings of the 4th International Conference on Computer-Human Interaction Research and Applications, 281‑290. https://doi.org/10.5220/0010181702810290

Boscardin, C. K., Gin, B., Golde, P. B., & Hauer, K. E. (2024). ChatGPT and generative artificial intelligence for medical education : Potential impact and opportunity. Academic Medicine, 99(1), 22‑27.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77‑101. https://doi.org/10.1191/1478088706qp063oa

Briganti, G., & Le Moine, O. (2020). Artificial intelligence in medicine : Today and tomorrow. Frontiers in medicine, 7, 509744.

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., & Allspach, L. (2000). Interactivity in human±computer interaction : A study of credibility, understanding, and in¯uence. Computers in Human Behavior.

C, L. (1982). Using the « thinking-aloud » method in cognitive interface design. *Research Report RC9265, IBM T. J. Watson Research Center*. https://cir.nii.ac.jp/crid/1571135649469604736

Cain, W. (2024). Prompting Change : Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education. TechTrends, 68(1), 47‑57. https://doi.org/10.1007/s11528-023-00896-0

Capel, T., & Brereton, M. (2023). What is Human-Centered about Human-Centered AI? A Map of the Research Landscape. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1‑23. https://doi.org/10.1145/3544548.3580959

Carbonell, E. J. G., & Siekmann, J. (s. d.). Lecture Notes in Artificial Intelligence.

Carmichael, L., Poirier, S.-M., Coursaris, C., Léger, P.-M., & Sénécal, S. (2021). Does Media Richness Influence the User Experience of Chatbots : A Pilot Study. In F. D. Davis, R. Riedl, J. Vom Brocke, P.-M. Léger, A. B. Randolph, & G. Müller-Putz (Éds.), Information Systems and Neuroscience (Vol. 52, p. 204‑213). Springer International Publishing. https://doi.org/10.1007/978-3-030-88900-5_23

Carnell, S., Miles, A., & Lok, B. (2022). Evaluating Virtual Patient Interaction Fidelity With Advanced Communication Skills Learners. Frontiers in Virtual Reality, 2. https://doi.org/10.3389/frvir.2021.801793

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. Cognition and Instruction, 8(4), 293‑332. https://doi.org/10.1207/s1532690xci0804_2

Chapman, D. S., Uggerslev, K. L., & Webster, J. (2003). Applicant reactions to face-to-face and technology-mediated interviews : A field investigation. Journal of Applied Psychology, 88(5), 944‑953. https://doi.org/10.1037/0021-9010.88.5.944

Charland, P., Lapierre, H., Skelling, Y., Léger, P.-M., & Simon, G. (2018). L'utilisation de l'électroencéphalographie : La collecte et l'analyse en continu. In S. Masson & G. Borst (Éds.), Méthodes de recherche en neuroéducation (p. 219‑242). Presses de l'Universite Du Quebec. https://hal.science/hal-04056866

Charters, E. (2003). The Use of Think-aloud Methods in Qualitative Research An Introduction to Think-aloud Methods. *Brock Education Journal*, *12*(2). https://doi.org/10.26522/brocked.v12i2.38

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education : A Review. IEEE Access, 8, 75264‑75278. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2988510

Chen, S., & Epps, J. (2014). Using Task-Induced Pupil Diameter and Blink Rate to Infer Cognitive Load. Human–Computer Interaction, 29(4), 390‑413. https://doi.org/10.1080/07370024.2014.892428

Chin, M. G., Sims, V. K., Clark, B., & Lopez, G. R. (2004). Measuring Individual Differences in Anthropomorphism toward Machines and Animals. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48(11), 1252‑1255. https://doi.org/10.1177/154193120404801110

Chui, M., Roberts, R., & Yee, L. (s. d.). Generative AI is here : How tools like ChatGPT could change your business.

Clark, R. E., & Choi, S. (2005). Five Design Principles for Experiments on the Effects of Animated Pedagogical Agents. Journal of Educational Computing Research, 32(3), 209‑225. https://doi.org/10.2190/7LRM-3BR2-44GW-9QQY

Cook, D. A., Erwin, P. J., & Triola, M. M. (2010). Computerized Virtual Patients in Health Professions Education : A Systematic Review and Meta-Analysis. Academic Medicine, 85(10), 1589. https://doi.org/10.1097/ACM.0b013e3181edfe13

Cornelius, S., & Leidner, D. (2021). Acceptance of Anthropomorphic Technology : A Literature Review. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2021.774

Csikszentmihalyi, M. (2013). Flow : The psychology of happiness. Random House. https://legenslife.wordpress.com/wp-content/uploads/2018/03/the-psychology-of-optimal-experience.pdf

Daher, S., Clark, A., & Barmaki, R. (2022). Editorial : Immersive Technologies in Healthcare. Frontiers in Virtual Reality, 3. https://doi.org/10.3389/frvir.2022.962950

Davies, A. J., & Krame, G. (2024). Measuring the Level of Fidelity Required for Transfer of Learning in Simulation-Based Learning Exercises for Novice and Experienced Practitioners. Simulation & Gaming, 10468781241241548. https://doi.org/10.1177/10468781241241548

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, 319‑340.

Dede, C. (2009). Immersive Interfaces for Engagement and Learning. *Science*, *323*(5910), 66‑69. https://doi.org/10.1126/science.1167311

Dougherty, L., & Lister, S. (2015). The Royal Marsden manual of clinical nursing procedures. John Wiley & Sons. https://books.google.com/books?hl=fr&lr=&id=ATAABwAAQBAJ&oi=fnd&pg=PR13&dq=Dougherty,+L.,+Lister,+S.,+%26+West-Oram,+A.+(Eds.).+(2015).+The+royal+Marsden+manual+of+clinical+nursing+procedures.+John+Wiley+%26+Sons.&ots=jM6STnh4kk&sig=olpzjCggj5KfCd4bI5YfRJ0hCeQ

Emerging Roles of Virtual Patients in the Age of AI. (2019). AMA Journal of Ethics, 21(2), E153-159. https://doi.org/10.1001/amajethics.2019.153

English, T. S., Hooe, D. C., Moen, H. G., Murray, C. J., & Gerling, G. J. (2010). Enhancing user presence in a chest tube simulator by Joining virtual reality with a mannequin. 2010 IEEE Systems and Information Engineering Design Symposium, 245‑250. https://doi.org/10.1109/SIEDS.2010.5469647

Ericsson, K. A., & Simon, H. A. (1998). How to Study Thinking in Everyday Life : Contrasting Think-Aloud Protocols With Descriptions and Explanations of Thinking. *Mind, Culture, and Activity*, *5*(3), 178‑186. https://doi.org/10.1207/s15327884mca0503_3

Farooq, U., & Grudin, J. (2016). Human-computer integration. *Interactions*, *23*(6), 26‑32. https://doi.org/10.1145/3001896

Ferrell, T., Crowson, A., & Mayhorn, C. (2022). How We Perceive and Trust Advice from Virtual Humans : The Influence of Voice Quality. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 66(1), 1189‑1193. https://doi.org/10.1177/1071181322661440

Fletcher, M. (2005). Canadian Fundamentals of Nursing, 3rd ed. *The Canadian nurse*, *101*(8), 10‑10.

Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer von Wolff, R., Hobert, S., & Luger, E. (2021). Future directions for chatbot research : An interdisciplinary research agenda. Computing, 103(12), 2915‑2942. https://doi.org/10.1007/s00607-021-01016-7

Fraser, K. L., Ayres, P., & Sweller, J. (2015). Cognitive load theory for the design of medical simulations. *Simulation in Healthcare*, *10*(5), 295‑307.

Friedman, B., Kahn, P. H., & Hagman, J. (2003). Hardware companions? : What online AIBO discussion forums reveal about the human-robotic relationship. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 273‑280. https://doi.org/10.1145/642611.642660

Gaba, D. M. (2004). The future vision of simulation in health care. *Quality and Safety in Health Care*, *13*(suppl 1), i2. https://doi.org/10.1136/qshc.2004.009878

Gefen & Straub. (2003). Managing User Trust in B2C e-Services. E-Service Journal, 2(2), 7. https://doi.org/10.2979/esj.2003.2.2.7

Giroux, F., Boasen, J., Sénécal, S., Fredette, M., Tchanou, A. Q., Ménard, J.-F., Paquette, M., & Léger, P.-M. (2019). Haptic Stimulation with High Fidelity Vibro-Kinetic Technology Psychophysiologically Enhances Seated Active Music Listening Experience. 2019 IEEE World Haptics Conference (WHC), 151‑156. https://doi.org/10.1109/WHC.2019.8816115

Gonçalves, G., Coelho, H., Monteiro, P., Melo, M., & Bessa, M. (2022). Systematic Review of Comparative Studies of the Impact of Realism in Immersive Virtual Experiences. ACM Computing Surveys, 55(6), 115:1-115:36. https://doi.org/10.1145/3533377

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS quarterly*, 213‑236.

Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis*. sage publications. https://books.google.com/books?hl=fr&lr=&id=Hr11DwAAQBAJ&oi=fnd&pg=PP1&dq=thematic+analysis&ots=XkZytIAGrE&sig=r8S_2KXcn5C7Zm-RZkuypC1BbYY

Hamilton, A. (2024). Artificial Intelligence and Healthcare Simulation : The Shifting Landscape of Medical Education. Cureus. https://doi.org/10.7759/cureus.59747

Hassanein, K., & Head, M. (2005). The Impact of Infusing Social Presence in the Web Interface : An Investigation Across Product Types. International Journal of Electronic Commerce, 10(2), 31‑55. https://doi.org/10.2753/JEC1086-4415100202

Hassanein, K., & Head, M. (2007). Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. International journal of human-computer studies, 65(8), 689‑708.

Heston, T., & Khun, C. (2023). Prompt Engineering in Medical Education. International Medical Education, 2(3), 198‑205. https://doi.org/10.3390/ime2030019

Heyselaar, E., Hagoort, P., & Segaert, K. (2017). In dialogue with an avatar, language behavior is identical to dialogue with a human partner. Behavior Research Methods, 49(1), 46‑60. https://doi.org/10.3758/s13428-015-0688-7

Higgs, J., Jones, M. A., Loftus, S., & Christensen, N. (2008). Clinical Reasoning in the Health Professions E-Book. Elsevier Health Sciences.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. Science, 349(6245), 261‑266. https://doi.org/10.1126/science.aaa8685

Holmes, W., Bialik, M., & Fadel, C. (2023). Artificial intelligence in education. In C. Stückelberger & P. Duggal (Éds.), Data ethics : Building trust : How digital technologies can serve humanity (p. 621‑653). Globethics Publications. https://doi.org/10.58863/20.500.12424/4276068

Hone, K., & Graham, R. (2001). Subjective assessment of speech-system interface usability (p. 2086). https://doi.org/10.21437/Eurospeech.2001-491

Hone, K. S., & Graham, R. (2000). Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). Natural Language Engineering, 6(3 & 4), 287‑303. https://doi.org/10.1017/S1351324900002497

Hornbæk, K., & Hertzum, M. (2017). Technology Acceptance and User Experience : A Review of the Experiential Component in HCI. ACM Transactions on Computer-Human Interaction, 24(5), 1‑30. https://doi.org/10.1145/3127358

Hsieh, C.-Y., Chen, S.-A., Li, C.-L., Fujii, Y., Ratner, A., Lee, C.-Y., Krishna, R., & Pfister, T. (2023). Tool Documentation Enables Zero-Shot Tool-Usage with Large Language Models (arXiv:2308.00675). arXiv. http://arxiv.org/abs/2308.00675

Huang, C. L., Luo, Y. F., Yang, S. C., Lu, C. M., & Chen, A.-S. (2020). Influence of Students' Learning Style, Sense of Presence, and Cognitive Load on Learning Outcomes in an Immersive Virtual Reality Learning Environment. Journal of Educational Computing Research, 58(3), 596‑615. https://doi.org/10.1177/0735633119867422

Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Gordon, D. L., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning : A BEME systematic review. Medical Teacher, 27(1), 10‑28. https://doi.org/10.1080/01421590500046924

Jacoby, J. (2002). Stimulus-Organism-Response Reconsidered : An Evolutionary Step in Modeling (Consumer) Behavior. Journal of Consumer Psychology, 12(1), 51‑57. https://doi.org/10.1207/S15327663JCP1201_05

Jiang, X., Atkins, M. S., Tien, G., Bednarik, R., & Zheng, B. (2014). Pupil responses during discrete goal-directed movements. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2075‑2084. https://doi.org/10.1145/2556288.2557086

Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. Science, 154(3756), 1583‑1585. https://doi.org/10.1126/science.154.3756.1583

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational Implications of Expertise Reversal Effects in Learning and Performance of Complex Cognitive and Sensorimotor Skills. *Educational Psychology Review*, *24*(2), 313‑337. https://doi.org/10.1007/s10648-012-9195-x

Karahanna, E., Agarwal, R., & Angst, C. M. (2006). Reconceptualizing Compatibility Beliefs in Technology Acceptance Research. MIS Quarterly, 30(4), 781‑804. https://doi.org/10.2307/25148754

Karat, J., & Karat, C. M. (2003). The evolution of user-centered focus in the human-computer interaction field. *IBM Systems Journal*, *42*(4), 532‑541. https://doi.org/10.1147/sj.424.0532

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Ke, F., Dai, Z., Dai, C.-P., Pachman, M., Chaulagain, R. S., & Yuan, X. (2020). Designing virtual agents for simulation-based learning in virtual reality. In Cognitive and affective perspectives on immersive technology in education (p. 151‑170). IGI Global. https://www.igi-global.com/chapter/designing-virtual-agents-for-simulation-based-learning-in-virtual-reality/253694

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. Psychophysiology, 48(3), 323‑332. https://doi.org/10.1111/j.1469-8986.2010.01069.x

Komasawa, N., & Yokohira, M. (2023). Simulation-based education in the artificial intelligence era. Cureus, 15(6). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10368461/

Kong, A., Zhao, S., Chen, H., Li, Q., Qin, Y., Sun, R., & Zhou, X. (2023). Better Zero-Shot Reasoning with Role-Play Prompting (arXiv:2308.07702). arXiv. https://doi.org/10.48550/arXiv.2308.07702

Kononowicz, A. A., Woodham, L. A., Edelbring, S., Stathakarou, N., Davies, D., Saxena, N., Car, L. T., Carlstedt-Duke, J., Car, J., & Zary, N. (2019). Virtual Patient Simulations in Health Professions Education : Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. Journal of Medical Internet Research, 21(7), e14676. https://doi.org/10.2196/14676

Kroes, K., Saccardi, I., & Masthoff, J. (2022). Empathizing with virtual agents : The effect of personification and general empathic tendencies. 2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), 73‑81. https://doi.org/10.1109/AIVR56993.2022.00017

Kruger, J.-L., & Steyn, F. (2014). Subtitles and Eye Tracking : Reading and Performance. Reading Research Quarterly, 49(1), 105‑120. https://doi.org/10.1002/rrq.59

Kulms, P., & Kopp, S. (2018). A social cognition perspective on human–computer trust : The effect of perceived warmth and competence on trust in decision-making with computers. Frontiers in Digital Humanities, 5, 352444.

Kulms, P., & Kopp, S. (2019). More Human-Likeness, More Trust? : The Effect of Anthropomorphism on Self-Reported and Behavioral Trust in Continued and Interdependent Human-Agent Cooperation. Proceedings of Mensch Und Computer 2019, 31‑42. https://doi.org/10.1145/3340764.3340793

Kumar, N., & Benbasat, I. (2006). Research Note : The Influence of Recommendations and Consumer Reviews on Evaluations of Websites. Information Systems Research, 17(4), 425‑439. https://doi.org/10.1287/isre.1060.0107

Lamontagne, C., Sénécal, S., Fredette, M., Chen, S. L., Pourchon, R., Gaumont, Y., De Grandpré, D., & Léger, P.-M. (2020). User Test : How Many Users Are Needed to Find the Psychophysiological Pain Points in a Journey Map? In T. Ahram, R. Taiar, S. Colson, & A. Choplin (Éds.), Human Interaction and Emerging Technologies (p. 136‑142). Springer International Publishing. https://doi.org/10.1007/978-3-030-25629-6_22

Lang, P. J. (s. d.). The Varieties of Emotional Experience : A Meditation on James-Lange Theory.

Lee, J., Kim, H., Kim, K. H., Jung, D., Jowsey, T., & Webster, C. S. (2020). Effective virtual patient simulators for medical communication training : A systematic review. Medical Education, 54(9), 786‑795. https://doi.org/10.1111/medu.14152

Lee, J., Lee, Y. J., Bae, J., & Seo, M. (2016). Registered nurses' clinical reasoning skills and reasoning process : A think-aloud study. *Nurse Education Today*, *46*, 75‑80. https://doi.org/10.1016/j.nedt.2016.08.017

Léger, P., Cronan, P., Charland, P., Pellerin, R., Babin, G., & Robert, J. (2012). Authentic OM problem solving in an ERP context. International Journal of Operations & Production Management, 32(12), 1375‑1394. https://doi.org/10.1108/01443571211284151

Liew, T. W., Mat Zin, N. A., & Sahari, N. (2017). Exploring the affective, motivational and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment. Human-Centric Computing and Information Sciences, 7(1), 9. https://doi.org/10.1186/s13673-017-0089-2

Liu, J., Liu, A., Lu, X., Welleck, S., West, P., Bras, R. L., Choi, Y., & Hajishirzi, H. (2022). Generated Knowledge Prompting for Commonsense Reasoning (arXiv:2110.08387). arXiv. https://doi.org/10.48550/arXiv.2110.08387

Lu, M. (2023). *The industrial revolution brought by AI : Opening the Gateway to the Future* (0). Resources Economics Research Board. https://doi.org/10.50908/grb.2.0_166

MacLean, S., Geddes, F., Kelly, M., & Della, P. (2019). Realism and Presence in Simulation : Nursing Student Perceptions and Learning Outcomes. Journal of Nursing Education, 58(6), 330‑338. https://doi.org/10.3928/01484834-20190521-03

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution : Its impact on society and firms. Futures, 90, 46‑60. https://doi.org/10.1016/j.futures.2017.03.006

Mascarenhas, S., Guimarães, M., Prada, R., Dias, J., Santos, P. A., Star, K., Hirsh, B., Spice, E., & Kommeren, R. (2018). A Virtual Agent Toolkit for Serious Games Developers. *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, 1‑7. https://doi.org/10.1109/CIG.2018.8490399

Mannuru, N. R., Shahriar, S., Teel, Z. A., Wang, T., Lund, B. D., Tijani, S., Pohboon, C. O., Agbaji, D., Alhassan, J., Galley, J., Kousari, R., Ogbadu-Oladapo, L., Saurav, S. K., Srivastava, A., Tummuru, S. P., Uppala, S., & Vaidya, P. (2023). Artificial intelligence in developing countries : The impact of generative artificial intelligence (AI) technologies for development. Information Development, 02666669231200628. https://doi.org/10.1177/02666669231200628

Massoth, C., Roeder, H., Ohlenburg, H., Hessler, M., Zarbock, A., Poepping, D. M., & Wenk, M. (2019). High-fidelity is not superior to low-fidelity simulation but leads to overconfidence in medical students. Bmc Medical Education, 19, 29. https://doi.org/10.1186/s12909-019-1464-7

Matsui, T., & Yamada, S. (2019). Designing Trustworthy Product Recommendation Virtual Agents Operating Positive Emotion and Having Copious Amount of Knowledge. Frontiers in Psychology, 10, 675. https://doi.org/10.3389/fpsyg.2019.00675

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley [From the Field]. IEEE Robotics & Automation Magazine, 19(2), 98‑100. IEEE Robotics & Automation Magazine. https://doi.org/10.1109/MRA.2012.2192811

Muckler, V. C. (2017). Exploring Suspension of Disbelief During Simulation-Based Learning. *Clinical Simulation in Nursing*, *13*(1), 3‑9. https://doi.org/10.1016/j.ecns.2016.09.004

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research. *International Journal of Qualitative Methods*, *22*, 16094069231205789. https://doi.org/10.1177/16094069231205789

Nass, C., & Moon, Y. (2000). Machines and Mindlessness : Social Responses to Computers. Journal of Social Issues, 56(1), 81‑103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 72‑78. https://doi.org/10.1145/191666.191703

Navath, S. (2021). Transforming Healthcare : The Impact and Future of Artificial Intelligence in Healthcare. Journal of Artificial intelligence and Machine Learning, 1(1), 16‑21.

Ning, X., Lin, Z., Zhou, Z., Yang, H., & Wang, Y. (2023). Skeleton-of-Thought : Large Language Models Can Do Parallel Decoding (arXiv:2307.15337). arXiv. http://arxiv.org/abs/2307.15337

Norman, D. (2007). Emotional design : Why we love (or hate) everyday things. Basic books. https://books.google.com/books?hl=fr&lr=&id=QtWSu3zBtPoC&oi=fnd&pg=PA3&dq=Norman,+D.+A.+(2004).+Emotional+design:+Why+we+love+(or+hate)+everyday+things.+Basic+Books&ots=9kwynV21uJ&sig=4rd4a6D2fKcbVaQEm6ZalbAXcDY

Norman, G., Dore, K., & Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. Medical Education, 46(7), 636‑647. https://doi.org/10.1111/j.1365-2923.2012.04243.x

North, M. M., & North, S. M. (2016). A Comparative Study of Sense of Presence of Traditional Virtual Reality and Immersive Environments. Australasian Journal of Information Systems, 20. https://doi.org/10.3127/ajis.v20i0.1168

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). A Systematic Review of Generative AI for Teaching and Learning Practice. Education Sciences, 14(6), 636.

Oliver, R. L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. Journal of Marketing Research, 17(4), 460‑469. https://doi.org/10.2307/3150499

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes : Evidence from the computer games literature. The Curriculum Journal, 16(4), 455‑474. https://doi.org/10.1080/09585170500384529

Ooi, K.-B., Tan, G. W.-H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., Dwivedi, Y. K., Huang, T.-L., Kar, A. K., Lee, V.-H., Loh, X.-M., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N. P., Sarker, P., Sharma, A., … Wong, L.-W. (2023). The Potential of Generative Artificial Intelligence Across Disciplines : Perspectives and Future Directions. Journal of Computer Information Systems, 1‑32. https://doi.org/10.1080/08874417.2023.2261010

Orru, G., & Longo, L. (2019). The Evolution of Cognitive Load Theory and the Measurement of Its Intrinsic, Extraneous and Germane Loads : A Review. In L. Longo & M. C. Leva (Éds.), Human Mental Workload : Models and Applications (Vol. 1012, p. 23‑48). Springer International Publishing. https://doi.org/10.1007/978-3-030-14273-5_3

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design : Recent Developments. Educational Psychologist, 38(1), 1‑4. https://doi.org/10.1207/S15326985EP3801_1

Padilha, J. M., Machado, P. P., Ribeiro, A., Ramos, J., & Costa, P. (2019). Clinical Virtual Simulation in Nursing Education : Randomized Controlled Trial. Journal of Medical Internet Research, 21(3), e11529. https://doi.org/10.2196/11529

Pauna, H., Léger, P.-M., Sénécal, S., Fredette, M., Courtemanche, F., Chen, S.-L., Labonté-Lemoyne, É., & Ménard, J.-F. (2018). The Psychophysiological Effect of a Vibro-Kinetic Movie Experience : The Case of the D-BOX Movie Seat. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Éds.), Information Systems and Neuroscience (p. 1‑7). Springer International Publishing. https://doi.org/10.1007/978-3-319-67431-5_1

Picard, R. W. (2000). Affective computing. MIT press. https://books.google.com/books?hl=fr&lr=&id=GaVncRTcb1gC&oi=fnd&pg=PR9&dq=Affective+computing.+MIT+Press.&ots=F6janmscd7&sig=Kh0xBYRkvoPA-6Hm6Ou5lNfPKzc

Pilote, B., & Chiniara, G. (2019). Chapter 2—The Many Faces of Simulation. In G. Chiniara (Éd.), Clinical Simulation (Second Edition) (p. 17‑32). Academic Press. https://doi.org/10.1016/B978-0-12-815657-5.00002-4

Pottle, J. (2019). Virtual reality and the transformation of medical education. Future Healthcare Journal, 6(3), 181‑185. https://doi.org/10.7861/fhj.2019-0036

Prompt Engineering for Generative AI. (s. d.). Consulté 1 février 2024, à l'adresse https://learning.oreilly.com/library/view/prompt-engineering-for/9781098153427/

Reeves, B., & Nass, C. (1996). The media equation : How people treat computers, television, and new media like real people. Cambridge, UK, 10(10). https://www.academia.edu/download/50585912/The_Media_Equation_How_People_Treat_Comp20161127-17472-1s43e41.pdf

Riedl, R., & Léger, P.-M. (2016). Fundamentals of NeuroIS : Information Systems and the Brain. Springer. https://doi.org/10.1007/978-3-662-45091-8

Rizzo, A. "Skip", & Bouchard, S. (Éds.). (2019). Virtual Reality for Psychological and Neurocognitive Interventions. Springer New York. https://doi.org/10.1007/978-1-4939-9482-3

Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., & Landay, J. (2016). Speech is 3x faster than typing for english and mandarin text entry on mobile devices. arXiv preprint arXiv:1608.07323. https://hci.stanford.edu/research/speech/paper/speech_paper.pdf

Russell, J. A. (1980). A circumplex model of affect. Journal of personality and social psychology, 39(6), 1161.

Rystedt, H., & Sjöblom, B. (2012). Realism, authenticity, and learning in healthcare simulations : Rules of relevance and irrelevance as interactive achievements. Instructional Science, 40(5), 785‑798. https://doi.org/10.1007/s11251-012-9213-x

Saab, M. M., Landers, M., Murphy, D., Cooke, E., & Hegarty, J. (2022). Nursing students' views of using virtual reality in healthcare : A qualitative study. Journal of Clinical Nursing, 31(9‑10), 1228‑1242. https://doi.org/10.1111/jocn.15978

Sadowski, W., & Stanney, K. (2002). Presence in Virtual Environments. In Handbook of Virtual Environments. CRC Press.

Sahin Karaduman, G., & Basak, T. (2024). Virtual Patient Simulations in Nursing Education : A Descriptive Systematic Review. Simulation & Gaming, 55(2), 159‑179. https://doi.org/10.1177/10468781231224836

Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice : Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare, 11(6), Article 6. https://doi.org/10.3390/healthcare11060887

Sauder, M., Tritsch, T., Rajput, V., Schwartz, G., & Shoja, M. M. (2024). Exploring generative artificial intelligence-assisted medical education : Assessing case-based learning for medical students. Cureus, 16(1). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10852982/

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research : Exploring its conceptualization and operationalization. *Quality & Quantity*, *52*(4), 1893‑1907. https://doi.org/10.1007/s11135-017-0574-8

Savery, J. R. (2015). Overview of problem-based learning : Definitions and distinctions. *Essential readings in problem-based learning: Exploring and extending the legacy of Howard S. Barrows*, *9*(2), 5‑15.

Savery, J. R., & Duffy, T. M. (1995). Problem based learning : An instructional model and its constructivist framework. Educational technology, 35(5), 31‑38.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being : Informative and directive functions of affective states. Journal of personality and social psychology, 45(3), 513.

Sezgin, E., & D'Arcy, S. (2022). Editorial : Voice Technology and Conversational Agents in Health Care Delivery. Frontiers in Public Health, 10, 887492. https://doi.org/10.3389/fpubh.2022.887492

Shankar, P. R. (2022). Artificial Intelligence in Health Professions Education. Archives of Medicine and Health Sciences, 10(2), 256‑261. https://doi.org/10.4103/amhs.amhs_234_22

Sheldon, L. K., & Hilaire, D. M. (2015). Development of communication skills in healthcare : Perspectives of new graduates of undergraduate nursing education. Journal of Nursing Education and Practice, 5(7), p30. https://doi.org/10.5430/jnep.v5n7p30

Simmons, B. (2010). Clinical reasoning : Concept analysis. *Journal of Advanced Nursing*, *66*(5), 1151‑1158. https://doi.org/10.1111/j.1365-2648.2010.05262.x

Sittner, B. J., Aebersold, M. L., Paige, J. B., Graham, L. L., Schram, A. P., Decker, S. I., & Lioce, L. (2015). INACSL standards of best practice for simulation : Past, present, and future. Nursing education perspectives, 36(5), 294‑298.

Someren, M., Barnard, Y., & Sandberg, J. (1994). *The Think Aloud Method—A Practical Guide to Modelling CognitiveProcesses*.

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. International journal of psychophysiology, 52(1), 77‑86.

Stevens, A., Hernandez, J., Johnsen, K., Dickerson, R., Raij, A., Harrison, C., DiPietro, M., Allen, B., Ferdig, R., Foti, S., Jackson, J., Shin, M., Cendan, J., Watson, R., Duerson, M., Lok, B., Cohen, M., Wagner, P., & Lind, D. S. (2006). The use of virtual patients to teach medical students history taking and communication skills. The American Journal of Surgery, 191(6), 806‑811. https://doi.org/10.1016/j.amjsurg.2006.03.002

Stone, R., Panfilov, P., & Shukshunov, V. (2011). Evolution of Aerospace Simulation : From Immersive Virtual Reality to Serious Games. In RAST 2011—Proceedings of 5th International Conference on Recent Advances in Space Technologies. https://doi.org/10.1109/RAST.2011.5966921

Sun, W., Jiang, X., Dong, X., Yu, G., Feng, Z., & Shuai, L. (2024). The evolution of simulation-based medical education research : From traditional to virtual simulations. Heliyon, 10(15). https://doi.org/10.1016/j.heliyon.2024.e35627

Sweller, J. (1988). Cognitive load during problem solving : Effects on learning. Cognitive science, 12(2), 257‑285.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive Load Theory. Springer New York. https://doi.org/10.1007/978-1-4419-8126-4

Tam, J., Carter, E., Kiesler, S., & Hodgins, J. (2012). Video increases the perception of naturalness during remote interactions with latency. CHI '12 Extended Abstracts on Human Factors in Computing Systems, 2045‑2050. https://doi.org/10.1145/2212776.2223750

Tam, W., Huynh, T., Tang, A., Luong, S., Khatri, Y., & Zhou, W. (2023). Nursing education in the age of artificial intelligence powered Chatbots (AI-Chatbots) : Are we ready yet? Nurse Education Today, 129, 105917. https://doi.org/10.1016/j.nedt.2023.105917

The Effectiveness of Games for Educational Purposes : A Review of Recent Research—Josephine M. Randel, Barbara A. Morris, C. Douglas Wetzel, Betty V. Whitehill, 1992. (s. d.). Consulté 26 mars 2024, à l'adresse https://journals.sagepub.com/doi/abs/10.1177/1046878192233001

Titah, R., & Barki, H. (2009). Nonlinearities between Attitude and Subjective Norms in Information Technology Acceptance : A Negative Synergy? MIS Quarterly, 33(4), 827‑844. https://doi.org/10.2307/20650329

Torre, I., Latupeirissa, A. B., & McGinn, C. (2020). How context shapes the appropriateness of a robot's voice. 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 215‑222.

https://ieeexplore.ieee.org/abstract/document/9223449/?casa_token=F0mew2C873MAAAAA:tYGQHqPPa0oX9fMD7EnFJzZNV_rYozE-kEzvpcILVgwTWAn-cpCG4ZauBZivppxzaQR1zjgWSWSB

Torta, E., Oberzaucher, J., Werner, F., Cuijpers, R. H., & Juola, J. F. (2013). Attitudes Towards Socially Assistive Robots in Intelligent Homes : Results From Laboratory Studies and Field Trials. Journal of Human-Robot Interaction, 1(2), 76‑99. https://doi.org/10.5898/JHRI.1.2.Torta

Tun, J. K., Alinier, G., Tang, J., & Kneebone, R. L. (2015). Redefining Simulation Fidelity for Healthcare Education. Simulation & Gaming, 46(2), 159‑174. https://doi.org/10.1177/1046878115576103

Tyerman, J., Luctkar-Flude, M., Graham, L., Coffey, S., & Olsen-Lynch, E. (2019). A Systematic Review of Health Care Presimulation Preparation and Briefing Effectiveness. *Clinical Simulation in Nursing*, *27*, 12‑25. https://doi.org/10.1016/j.ecns.2018.11.002

Van Merriënboer, J. J. G., & Sweller, J. (2010). Cognitive load theory in health professional education : Design principles and strategies. *Medical Education*, *44*(1), 85‑93. https://doi.org/10.1111/j.1365-2923.2009.03498.x

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model : Four Longitudinal Field Studies. Management Science, 46(2), 186‑204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology : Toward a unified view. MIS quarterly, 425‑478.

Vora, J., Nair, S., Gramopadhye, A. K., Duchowski, A. T., Melloy, B. J., & Kanki, B. (2002). Using virtual reality technology for aircraft visual inspection training : Presence and comparison studies. Applied Ergonomics, 33(6), 559‑570. https://doi.org/10.1016/S0003-6870(02)00039-X

Walker, A., & Leary, H. (2009). A problem based learning meta analysis : Differences across problem types, implementation types, disciplines, and assessment levels. Interdisciplinary journal of problem-based learning, 3(1), 6.

Wang, N., Rebolledo-Mendez, G., Matsuda, N., Santos, O. C., & Dimitrova, V. (Éds.). (2023). Artificial Intelligence in Education : 24th International Conference, AIED 2023, Tokyo, Japan, July 3–7, 2023, Proceedings (Vol. 13916). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36272-9

Wang, Y. J., & Minor, M. S. (2008). Validity, reliability, and applicability of psychophysiological techniques in marketing research. Psychology & Marketing, 25(2), 197‑232. https://doi.org/10.1002/mar.20206

Wang, Y.-C., Xue, J., Wei, C., & Kuo, C.-C. J. (2023). An overview on generative AI at scale with Edge-Cloud Computing. IEEE Open Journal of the Communications Society. https://ieeexplore.ieee.org/abstract/document/10268594/

Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J.-H., & Cacioppo, J. T. (2010). Making sense by making sentient : Effectance motivation increases anthropomorphism. Journal of Personality and Social Psychology, 99(3), 410‑435. https://doi.org/10.1037/a0020240

Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). *Engineering psychology and human performance*. Routledge. https://www.taylorfrancis.com/books/mono/10.4324/9781003177616/engineering-psychology-human-performance-christopher-wickens-justin-hollands-simon-banbury-william-helton

Wilson, C. (2013). *Interview techniques for UX practitioners : A user-centered design method*. Newnes.

Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2023). Transitioning to Human Interaction with AI Systems : New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI. *International Journal of Human–Computer Interaction*, *39*(3), 494‑518. https://doi.org/10.1080/10447318.2022.2041900

Xu, W., & Gao, Z. (2024). Applying HCAI in Developing Effective Human-AI Teaming : A Perspective from Human-AI Joint Cognitive Systems. *Interactions*, *31*(1), 32‑37. https://doi.org/10.1145/3635116

Yeh, C. J., & Inose, M. (2003). International students' reported English fluency, social support satisfaction, and social connectedness as predictors of acculturative stress. Counselling Psychology Quarterly, 16(1), 15‑28. https://doi.org/10.1080/0951507031000114058

Young, M. E., Thomas, A., Lubarsky, S., Gordon, D., Gruppen, L. D., Rencic, J., Ballard, T., Holmboe, E., Da Silva, A., Ratcliffe, T., Schuwirth, L., Dory, V., & Durning, S. J. (2020). Mapping clinical reasoning literature across the health professions : A scoping review. BMC Medical Education, 20(1), 107. https://doi.org/10.1186/s12909-020-02012-9

Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. Nature Biomedical Engineering, 2(10), Article 10. https://doi.org/10.1038/s41551-018-0305-z

Zhang, P. (2013). The affective response model : A theoretical framework of affective concepts and their relationships in the ICT context. MIS quarterly, 247‑274.