

**Faculté des sciences économiques,
sociales, politiques et de communication**

Parler à un chatbot, le traiter comme un humain : perceptions, jugements et attachement des utilisateurs

Auteur : Emilien Simon
Promoteur(s) : Hugo Thirard

Année académique 2025-2026
Master 120 en Communication stratégique des organisations

Déclaration sur l'honneur, mémoires de l'École de communication, UCLouvain

Je déclare qu'il s'agit d'un travail original et personnel, que toutes les sources référencées ont été indiquées dans leur totalité et ce, quelle que soit leur provenance, et que l'utilisation des outils d'intelligence artificielle est conforme aux consignes d'utilisation publiées par la Faculté ESPO et disponible à cette adresse :

<https://intranet.uclouvain.be/fr/myucl/facultes/espo/intelligence-artificielle-consignes-d-utilisation-pour-les-etudiant-es.html>

Je suis conscient·e que le fait de ne pas citer une source, de ne pas la citer clairement et complètement, de ne pas annoncer dans une partie spécifique en début du mémoire l'utilisation que j'ai faite des outils d'intelligence artificielle ou de les utiliser de manière contraire à ce qui est prévu dans la note facultaire constituent des irrégularités graves au regard du *Règlement des études et des examens de l'Université catholique de Louvain* (Chapitre 4, section 7, article 107 à 114) au sein de l'Université. J'ai notamment pris connaissance des risques de sanctions administratives et disciplinaires encourues.

Au vu de ce qui précède, je déclare sur l'honneur ne pas avoir commis de plagiat ou toute autre forme de fraude et de n'avoir pas utilisé les outils d'intelligences artificielles en dehors de ce qui est accepté à l'UCLouvain.

Nom, Prénom : Simon Emilien

Date : 15 mai 2026

Signature de l'étudiant·e : *Emilien Simon*

Déclaration sur l'usage des différentes intelligences artificielles

Lors de la rédaction de ce mémoire, l'auteur a eu recours à trois outils d'intelligence artificielle :

Scribens pour la correction orthographique et grammaticale, Perplexity pour la recherche d'articles scientifiques, et Claude pour reformuler certains passages, clarifier la réflexion et tester la cohérence de l'argumentation.

L'ensemble de la réflexion, de l'analyse et de la rédaction demeure le travail de l'auteur.

Emilien Simon

Emilien Simon

15 mai 2026

Remerciements

Ce mémoire est l'aboutissement d'un parcours qui n'aurait pas été le même sans le soutien de plusieurs personnes que je tiens à remercier ici.

Je remercie tout d'abord Hugo Thirard, mon promoteur, pour la qualité de son encadrement tout au long de ce processus de recherche. Sa disponibilité et la pertinence de ses retours pour ce travail ont été déterminantes.

Je remercie ensuite ma famille, et en particulier mes parents, pour leur soutien indéfectible tout au long de mon parcours universitaire. Ma gratitude va également à mon grand-père, dont la relecture attentive a permis de traquer les fautes d'orthographe qui m'avaient échappé.

Je tiens à remercier Chiara Torrese pour son soutien constant et ses encouragements tout au long de ce master, comme durant toutes les années qui l'ont précédé.

Enfin, je remercie mes amis et mes professeurs pour tout ce qu'ils m'ont apporté au fil de ce master, tant sur le plan intellectuel que personnel.

Table des matières

1. INTRODUCTION ET PROBLÉMATIQUE	1
2. ÉTAT DE L'ART	2
2.1. UNE INTELLIGENCE ARTIFICIELLE AUX CONTOURS ENCORE FLOUS	2
2.2. VERS UNE COMMUNICATION HOMME-MACHINE NATURALISÉE	5
2.3. L'ANTHROPOMORPHISME : UNE DÉFINITION CHANGEANTE	6
2.4. CHATBOTS, ANTHROPOMORPHISME PERÇU ET ENGAGEMENT UTILISATEUR	7
2.4.1. <i>L'anthropomorphisme et ses effets sur la confiance</i>	7
2.4.2. <i>Signaux conversationnels et facteurs de modulation</i>	16
2.4.3. <i>Cadres théoriques de l'interaction entre Humain-IA et la vallée de l'étrange</i>	21
3. DISPOSITIF MÉTHODOLOGIQUE.....	26
3.1. TYPE DE RECHERCHE ET TERRAIN D'OBSERVATION	26
3.2. OBJECTIFS.....	27
3.3. MÉTHODE D'ANALYSE	27
3.4. CORPUS	29
3.5. LIMITES MÉTHODOLOGIQUES	30
4. RÉSULTATS : FORMES ET EFFETS DE L'ANTHROPOMORPHISME PERÇU	30
4.1. AUTOMATISMES SOCIAUX ET POSTURES RÉFLEXIVES, LE MODÈLE CASA VIS-À-VIS DES CHATBOTS. .	31
4.1.1. <i>Politesse, nomination et réciprocité</i>	32
4.1.2. <i>L'utilisation des pronoms comme marqueurs de bascule</i>	34
4.1.3. <i>Les postures réflexives : Anti-CASA, Méta-CASA et CASA dissonant</i>	38
4.2. CHALEUR ET COMPÉTENCE : LA CARTOGRAPHIE SCM DES CHATBOTS.....	43
4.2.1. <i>La tension chaleur/compétence vue comme axe structurant des évaluations de la part des utilisateurs</i>	44
4.2.2. <i>L'anti-complaisance comme rejet de la chaleur superficielle</i>	47
4.3. ANTHROPOMORPHISME SUPERFICIEL ET ANTHROPOMORPHISME UTILE DU MODÈLE HATEM	50
4.3.1. <i>L'anthropomorphisme superficiel de la chaleur sans fiabilité</i>	51
4.3.2. <i>L'anthropomorphisme utile quand la transparence fonde la confiance</i>	53
4.4. AU-DELÀ DE LA CONFIANCE : PRÉSENCE SOCIALE, ATTACHEMENT ET RUPTURE	55
4.4.1. <i>Présence sociale et seuils de bascule</i>	55
4.4.2. <i>L'attachement thérapeutique comme cas limite</i>	56
4.4.3. <i>Le passage d'une version à l'autre comme rupture relationnelle</i>	59
5. CONCLUSION ET DISCUSSION.....	60
5.1. SYNTHÈSE DES RÉSULTATS.....	61
5.2. MISE EN DIALOGUE AVEC L'ÉTAT DE L'ART	66

5.3.	LIMITES MÉTHODOLOGIQUES	69
5.4.	PISTES POUR DE FUTURES RECHERCHES.....	71
BIBLIOGRAPHIE		74

1. Introduction et Problématique

Les chatbots basés sur l'intelligence artificielle, tels que ChatGPT, Claude ou Gemini, occupent une place croissante dans la vie numérique des individus, comme le souligne un rapport de l'entreprise de consultance McKinsey (2025). ChatGPT fait par exemple désormais partie des 5 sites les plus visités au monde en avril 2025 devant X, WhatsApp, Reddit et Tiktok pour ne citer qu'eux d'après la plateforme d'analyse du trafic web Similarweb (2025). Cette augmentation montre la démocratisation de l'usage et de l'utilisation de ces chatbots, dont la fonction ne se limite plus aux contextes marchands ou aux services clients : ils sont désormais mobilisés pour la rédaction, l'apprentissage, le divertissement ou encore comme un compagnon pour la vie quotidienne (McKinsey, 2025). Cette diversification des usages soulève une question fondamentale quant à la perception de ces systèmes par leurs utilisateurs.

Łukasik et Gut (2025) soulignent que les agents conversationnels ne sont plus cantonnés à des fonctions purement utilitaristes mais occupent désormais des rôles sociaux, thérapeutiques et éducatifs auparavant réservés aux interactions interhumaines (Łukasik & Gut, 2025, pp. 1-2). À mesure que ces chatbots adoptent des modes de communication proches du langage humain, la frontière entre l'outil fonctionnel et l'interlocuteur perçu comme doté de traits anthropomorphiques tend à s'estomper pour l'utilisateur.

Face à cette évolution, une question émerge : dans quelle mesure ces perceptions anthropomorphiques influencent-elles le jugement que les utilisateurs portent sur ces systèmes ? Plusieurs travaux récents suggèrent que l'anthropomorphisme, c'est-à-dire la tendance à attribuer des caractéristiques humaines à des entités non humaines (Kim & Im, 2023, p. 2), appliqué aux chatbots, peut renforcer l'attachement, la confiance ou encore la satisfaction (Cheng et al., 2022, pp. 12-15). Toutefois, ces effets ne sont pas uniformes, dans le contexte des chatbots commerciaux avec des réponses établies en avance, Roy et Naidoo (2021) montrent que par exemple, les consommateurs orientés vers le présent préfèrent un chatbot au ton chaleureux, tandis que ceux tournés vers l'avenir privilégient un chatbot perçu comme compétent. L'anthropomorphisme ne produit donc pas les mêmes effets selon le profil de l'utilisateur, le style de communication du chatbot ou encore, potentiellement, le modèle d'IA utilisé (Roy & Naidoo, 2021, pp. 29 - 31).

De plus, ces effets restent peu étudiés par le prisme de la réception. La plupart des travaux existants reposent sur des protocoles expérimentaux dans lesquels les participants savent qu'ils sont observés, ce qui peut influencer leurs réactions. Ces recherches sont en outre souvent menées dans un contexte culturel spécifique et portent sur un seul modèle d'IA, ce qui limite la portée de leurs conclusions face à la diversité des chatbots disponibles aujourd'hui.

La question de la frontière entre l'humain et la machine est d'autant plus actuelle que certains chatbots parviennent désormais à se faire passer pour des humains. Le test de Turing, conçu pour évaluer la capacité d'une machine à être perçue comme un interlocuteur humain lors d'une conversation, a récemment été soumis à une nouvelle mise à l'épreuve par Jones et Bergen. Leur étude montre que le modèle GPT-4.5 d'OpenAI, lorsqu'il était doté d'un persona décrit comme devant se comporter de manière humaine, a été identifié comme un humain par 73 % des participants (Jones & Bergen 2025, *preprint*, pp. 4 – 16.). Ce résultat illustre à quel point la perception des chatbots par les utilisateurs constitue aujourd'hui un enjeu central, tant pour la recherche que pour la conception de ces systèmes.

Ce mémoire propose donc d'analyser la perception que les utilisateurs ont de ces chatbots, à partir de leurs discours en ligne, afin de mieux comprendre ce qu'ils leur attribuent comme traits anthropomorphiques et comment ces attributions influencent leur jugement.

Il s'agira ainsi de répondre à la question suivante : dans quelle mesure les utilisateurs attribuent-ils des traits humains aux IA conversationnelles, et comment ces perceptions affectent-elles leur jugement en termes d'utilité, de crédibilité ou d'attachement ?

2. État de l'art

2.1. Une intelligence artificielle aux contours encore flous

L'intelligence artificielle (IA) suscite de nombreux débats quant à sa définition. Les chercheurs ne parviennent pas à s'accorder sur une conception homogène de ce qu'elle recouvre réellement. Chacun tend à l'adapter à sa propre approche, rendant la notion tantôt très large, tantôt plus restrictive. Cette variabilité conduit parfois à inclure, dans le champ de l'intelligence artificielle, des algorithmes relevant de formes plus classiques d'automatisation. Ainsi, la perception de ce qu'est une IA varie considérablement selon les auteurs (Poole & Mackworth, 2023). Cette absence de consensus se retrouve dans la définition des agents conversationnels,

souvent envisagés soit comme de simples outils de service, soit comme de véritables interlocuteurs (Poole & Mackworth, 2023 ; Łukasik & Gut, 2025, p. 5). Ce constat invite à préciser ce que recouvrent les systèmes étudiés dans cette recherche, tant par leur nature que par leur degré de sophistication technologique.

Poole et Mackworth (2023, pp. 14 – 16) proposent une première grille de lecture fondée sur le degré d'autonomie des systèmes d'intelligence artificielle. Ils en distinguent trois catégories : les agents autonomes, capable d'agir sans qu'aucune validation humaine ne soit requise, les agents semi-autonomes, dont le fonctionnement suppose une forme de supervision, et les bots informatiques, que l'on retrouve sur des plateformes telles Twitter ou Discord pour exécuter des tâches répétitives.

Cette classification permet de rendre visible l'hétérogénéité des systèmes qui se retrouve sous l'appellation « intelligence artificielle ». Toutefois, elle ne suffit pas à situer précisément les agents conversationnels comme ChatGPT ou Claude.

Afin d'affiner, cette classification, Łukasik et Gut (2025) proposent une distinction fondée non plus sur le degré d'autonomie, mais sur la forme de l'entité artificielle en 3 catégories distinctes.

Les robots qui sont des machines programmées pour exécuter automatiquement certaines tâches, notamment dans les domaines manufacturier ou médical (Łukasik & Gut, 2025, p. 3). Certains d'entre eux sont conçus pour imiter les humains, tant par leur apparence que par leur comportement, grâce à l'intégration de systèmes d'intelligence artificielle (Doncieux et al., 2022, p. 11).

Les avatars, désignent des entités numériques évoluant dans des environnements virtuels (Łukasik & Gut, 2025, p. 4).

Enfin, les chatbots, ou agents conversationnels, sont des systèmes conçus pour imiter les conversations humaines et traiter les demandes des utilisateurs (Łukasik & Gut, 2025, p. 5). Ces agents parviennent à interagir grâce à l'apprentissage automatique et au traitement du langage naturel, en communiquant soit par la voix, soit par texte, comme c'est le cas avec des outils tels que ChatGPT ou Gemini (Łukasik & Gut, 2025, p. 5).

Cette absence de consensus se retrouve également dans la définition des agents conversationnels. Ces derniers reposent sur des techniques d'apprentissage automatique et de traitement du langage naturel qui les distinguent des systèmes d'automatisation classiques fondés sur des règles prédéfinies . Pour autant, leur statut reste débattu : ils sont souvent envisagés soit comme de simples outils de service, soit comme de véritables interlocuteurs (Poole & Mackworth, 2023 ; Łukasik & Gut, 2025, p. 5).

Toutefois, cette catégorie de chatbots regroupe elle-même des systèmes de nature très différente. Khennouche et al. (2024) proposent une taxonomie qui distingue quatre types d'architectures.

1. les chatbots à règles (rule-based) sont fondés sur des arbres de décision et des réponses prédéfinies. Même s'ils manquent de complexité, ils assurent des réponses consistantes et prédictibles.
2. Les chatbots par récupérations (retrieval based) ; qui font correspondre la requête de l'utilisateur à une base de questions-réponses existantes sans créer du nouveau contenu.
3. Les chatbots génératifs, qui produisent des réponses nouvelles à partir de modèles de langage entraînés sur de vastes corpus de données.
4. Les chatbots hybrides, qui combinent récupération et génération.

Les premiers, comme les chatbots de service clients traditionnels, ne relèvent pas véritablement de l'intelligence artificielle au sens strict, où ils n'impliquent aucune compréhension du langage, ils exécutent des scripts selon une logique conditionnelle fixe.

Les chatbots génératifs quant à eux, reposent sur des modèles de langage de grande taille (Large Language Models) capables de produire des réponses contextuelles et adaptées. C'est dans cette catégorie que s'inscrivent les systèmes étudiés dans ce mémoire : ChatGPT, Claude, Gemini, dont la recherche analyse la perception par les utilisateurs et l'impact des traits anthropomorphiques qui leur sont attribués.

Cette taxonomie complète ainsi les différentes catégories proposées par Łukasik et Gut (2025) en introduisant une distinction interne à la catégorie des chatbots, fondée sur leur

architecture technique plutôt que sur leur fonction apparente. Elle permet donc de répondre à une ambiguïté fréquente dans la littérature où des systèmes qualifiés d'agents conversationnels, peuvent renvoyer à des réalités technologiques très différentes (Poole & Mackworth, 2023 ; Łukasik & Gut, 2025, pp. 3 – 5 ; Khennouche et al., 2024).

Il convient également de noter que cette classification est appelée à évoluer à mesure que ces systèmes acquièrent des capacités agentiques, c'est-à-dire la possibilité d'exécuter des tâches complexes de manière autonome et cela sur la seule base d'une instruction. Malgré le fait que cette source ne soit pas académique, elle est intéressante pour présenter cette notion émergente (IBM, 2024,).

Si la définition de l'intelligence artificielle reste sujette à discussion, c'est sans doute, dans ses usages concrets que se dessine aujourd'hui son véritable enjeu. En effet, les agents conversationnels génératifs ne se limitent plus à des fonctions purement techniques : ils deviennent des partenaires de communication à part entière, capables de simuler des comportements humains. Cette transformation ouvre la voie à une naturalisation de la communication entre l'humain et la machine.

2.2. Vers une communication homme-machine naturalisée

Historiquement, la communication médiée par ordinateur (CMC) établissait une distinction nette entre l'humain et la machine. L'ordinateur n'était alors qu'un simple intermédiaire technique permettant de communiquer avec autrui, sans qu'aucune véritable interaction ne s'établisse avec la machine elle-même. Aujourd'hui, ce modèle, évolue vers une communication homme-machine (HMC), dans laquelle la machine devient un interlocuteur à part entière, comme en témoignent les assistants vocaux tels que Siri, Google assistant et Alexa pour ne citer qu'eux (Nah, 2023, pp. 31-37).

Cette évolution ouvre la voie à une nouvelle ère : celle de la communication médiée par l'intelligence artificielle (AI-MC). Celle-ci désigne les situations dans lesquelles une IA participe activement à la production, à la modification ou à l'amélioration du message. Par exemple, lorsqu'un individu rédige un texte puis sollicite ChatGPT pour l'améliorer, il engage une interaction où l'intelligence artificielle contribue directement à la qualité et à la forme de

la communication, cela résulte à une communication médiée par l'intelligence artificielle. (Nah, 2023, pp. 31-37).

Cette transformation illustre le changement de statut de nos technologies : d'outils destinés à faciliter nos échanges, elles deviennent désormais des partenaires avec lesquels nous communiquons et collaborons pour mieux communiquer (Nah 2023, pp. 31-37). L'introduction de l'intelligence artificielle dans les pratiques communicationnelles renforce, selon Yue et al. (2024, pp. 6-8), l'efficacité et la clarté des messages produits.

Les IA conversationnelles telles que ChatGPT, Claude ou Gemini s'intègrent désormais dans les routines quotidiennes : résumer un courriel, formuler une idée, composer de la musique, générer des images, poser des questions, analyser un texte ou simplement échanger (Six, 2024, 19 juin). Cette présence grandissante contribue à naturaliser la relation entre l'humain et la machine, amorçant une véritable reconfiguration des interactions communicationnelles.

2.3. L'anthropomorphisme : une définition changeante

L'anthropomorphisme désigne l'attribution de caractéristiques, de comportements ou de traits humains à des entités qui ne le sont pas, telles que les animaux, les robots ou les intelligences artificielles (Cheng et al., 2022, pp. 2–3). Ce concept occupe une place centrale dans les recherches sur l'interaction homme-machine, bien que ses définitions varient selon les disciplines et les époques (Cheng et al., 2022, p. 2).

L'une des premières définitions notables est celle proposée par Guthrie en 1980, qui décrit l'anthropomorphisme comme la tendance à percevoir des éléments humains dans des formes ou des événements qui ne le sont pas (Bourdillon et al., 1980). Cette approche met en avant un mécanisme cognitif naturel : la propension de l'esprit humain à projeter ses propres caractéristiques sur l'environnement.

Vingt ans plus tard, dans le contexte des interactions entre humains et technologies, Nass et Moon (2000) introduisent une distinction fondamentale. Selon eux, les individus n'attribuent pas sincèrement des traits humains aux ordinateurs, ils appliquent néanmoins, de manière automatique et inconsciente des règles sociales propres aux interactions interhumaines lorsqu'ils interagissent avec des machines (Nass & Moon, 2000, pp. 82 - 84).

Ce processus qu'ils qualifient d'ethopoeia, se distingue de l'anthropomorphisme au sens strict, dans le sens que les utilisateurs savent que l'ordinateur n'est pas humain, mais se comportent malgré tout comme s'il l'était, en manifestant par exemple de la politesse, de la réciprocité ou en appliquant des stéréotypes de genre (Nass & Moon, 2000, p. 94).

Cette perspective est particulièrement intéressante pour cette recherche, dans la mesure où elle suggère que les comportements sociaux envers les chatbots ne relèvent pas nécessairement d'une croyance en leur côté humain, mais plutôt sur base d'un automatisme cognitif ancré.

Bartneck et ses collègues (2009) proposent pour leur part une définition synthétique déterminant que l'anthropomorphisme correspond au fait d'attribuer une forme humaine, des caractéristiques humaines ou des comportements humains à des entités non humaines telles que les robots, les ordinateurs ou les animaux (Bartneck et al., 2009, p. 74). Dans le cadre de leurs travaux sur la perception des robots, ils soulignent qu'une apparence trop anthropomorphe peut engendrer des attentes que le système ne sera pas en mesure de satisfaire, un constat qui résonne avec les enjeux posés par les chatbots génératifs actuels.

C'est la définition de synthèse proposée par Bartneck et ses collègues qui sera utilisée comme référence dans le cadre de cette recherche, tout en gardant à l'esprit la nuance apportée par Nass et Moon (2000) entre attribution consciente de traits humains et application inconsciente des normes sociales.

Initialement ancrée dans l'anthropologie, cette notion a progressivement gagné les champs de la psychologie et de la communication. Dans le contexte des interactions homme-machine, elle permet d'analyser comment les différences individuelles, telles que la personnalité, les facteurs culturels ou l'environnement, influencent des dimensions clés comme la confiance, l'empathie, la satisfaction, la perception de disponibilité ou encore le sentiment de connexion à un agent conversationnel (Cheng et al., 2022, pp. 2–6).

2.4. Chatbots, anthropomorphisme perçu et engagement utilisateur

2.4.1. L'anthropomorphisme et ses effets sur la confiance

Si plusieurs travaux établissent qu'un chatbot perçu comme humain peut susciter un attachement émotionnel et un engagement plus marqué, ces effets ne sont ni automatiques, ni

uniformes. Ceux-ci dépendent du type d'anthropomorphisme mobilisé, de la cohérence entre les signaux conversationnels et la performance effective et réelle du système, ainsi que du contexte dans lequel l'interaction se passe. Cette sous-section examine ces dynamiques en s'appuyant sur le modèle HATEM de Carter et al. (2024) comme cadre structurant pour cette section et enrichi par plusieurs expérimentations récentes.

Plusieurs recherches récentes montrent que la perception de traits humains chez un chatbot peut renforcer l'attachement émotionnel de l'utilisateur envers l'intelligence artificielle. L'attribution de rôles sociaux, tels que celui de mentor, tend également à accroître cet attachement, bien que les effets observés ne soient pas toujours significatifs (Zhang & Patrick Rau, 2023, pp. 6–8).

D'autres travaux soulignent le rôle central de la confiance dans la relation entre l'utilisateur et le chatbot. Cheng et al. (2022) montrent que plus un utilisateur accorde sa confiance à un agent conversationnel alimenté par l'IA, plus il s'appuie sur celui-ci pour résoudre ses problèmes, préférant recourir à un humain uniquement lorsque la situation l'exige (Cheng et al., 2022, pp. 10 et 11). Ces auteurs observent également que l'ajout d'émoticônes dans la conversation peut augmenter la chaleur perçue et la confiance, tandis qu'un délai de réponse trop long tend à les diminuer. Ces résultats invitent à réfléchir à l'optimisation des performances techniques des chatbots, mais aussi à la pertinence des signes d'anthropomorphisme introduits dans la communication (Cheng et al., 2022, pp. 10–14).

Cependant, ces observations doivent être nuancées à la lumière du modèle HATEM (Human–Automation Trust Expectation Model) proposé par Carter, Loft et Visser (2024). Ce modèle suggère que l'anthropomorphisme ne renforce durablement la confiance que lorsqu'il contribue à une communication utile et contextuelle, aidant l'utilisateur à évaluer la fiabilité du système et à anticiper ses limites (Carter et al., 2024, pp. 2486 - 2499). À l'inverse, un anthropomorphisme superficiel fondé sur des éléments esthétiques tels que les émojis ou un ton chaleureux, mais sans lien réel avec la performance du système n'améliore la confiance que de manière temporaire. Il risque même de générer des attentes irréalistes, conduisant à une déception et à une perte de confiance envers l'agent conversationnel (Carter et al., 2024, pp. 2486 - 2499).

Ainsi, deux formes d'anthropomorphisme peuvent être distinguées. Le premier, superficiel, repose sur des signaux de chaleur sociale (émoticônes, ton convivial, expressions humaines) visant à humaniser artificiellement la machine. Le second, utile, renvoie à une humanisation fonctionnelle : l'agent conversationnel adopte des comportements qui aident réellement l'utilisateur à comprendre son fonctionnement, sa fiabilité ou ses incertitudes. Par exemple, un chatbot qui reconnaît ne pas être certain d'une réponse ou qui justifie ses choix manifeste une transparence perçue comme humaine et contribue à instaurer une confiance plus stable (Carter et al., 2024, pp. 2486 - 2499).

En somme l'anthropomorphisme n'a un impact positif sur l'utilisateur que lorsqu'il s'accompagne d'une cohérence entre apparence humaine et compétence perçue. L'expressivité seule ne suffit pas ; c'est la crédibilité comportementale de l'agent qui détermine la qualité du lien établi avec l'utilisateur (Carter et al., 2024, pp. 2486 - 2499).

Cette théorie proposée par Carter et al. (2024) est corroborée par des travaux récents réalisés par Sun et Wang (2026), qui examinent l'interaction entre deux caractéristiques anthropomorphiques des agents conversationnels. La « friendliness », c'est-à-dire l'emploi d'un ton chaleureux, poli et engageant et la « sycophancy », définie comme étant la tendance de l'agent à adapter dynamiquement ses réponses pour les aligner avec les opinions de l'utilisateur et ce, même si celles-ci sont factuellement inexactes. (Sun & Wang, 2026, p. 1)

À la différence du simple fait d'être excessivement flatteur, ce qui avait déjà été étudié dans les précédents travaux, la « sycophancy » propre aux LLM se distingue par son caractère adaptatif où l'agent ajuste au fil de la discussion sa position, et ses réponses intègrent naturellement les arguments de l'utilisateur et cet alignement entre la pensée de l'utilisateur et la réponse peut se faire au détriment de l'exactitude. À travers un protocole expérimental croisant les deux variables « Sycophancy » et « friendliness », les auteurs mettent en évidence un effet d'interaction sur l'authenticité perçue de l'agent (Sun & Wang, 2026, p. 4).

Lorsqu'un agent adopte un ton chaleureux tout en maintenant ses positions initiales face aux arguments de l'utilisateur, il est perçu comme plus authentique et suscite davantage de confiance. En revanche, lorsque cette même chaleur de langage est combinée à un alignement systématique sur les opinions, arguments de l'utilisateur, l'authenticité perçue du chatbot diminue, ce qui entraîne une érosion de la confiance (Sun & Wang, 2026, pp. 9 - 11).

Les données qualitatives recueillies par les auteurs confirment ce résultat. Plusieurs participants comparent spontanément l'agent trop flatteur et chaleureux à un « yes man » dont l'accord permanent semble artificiel et manipulateur et augmente la perception des réponses comme étant trop peu profondes (Sun & Wang, 2026, pp. 9 - 10).

Ce constat rejoint directement la distinction opérée par Carter et al. (2024) entre anthropomorphisme superficiel et utile. La chaleur sociale, lorsqu'elle n'est pas liée à une position argumenté et stable de la part de l'agent, relève précisément de cet anthropomorphisme de surface dont l'effet sur la confiance de l'utilisateur reste fragile voire contre-productif.

Par ailleurs, Sun et Wang (2026, p. 8) montrent que la convivialité, prise isolément renforce significativement le sentiment de présence sociale perçue qui a son tour, améliore la confiance affective et cognitive. Ce résultat, confirme expérimentalement la pertinence de la théorie de la présence sociale (Short et al., 1976, cités par Oh et al., 2018, p. 2) dans le contexte spécifique des LLM conversationnels.

Cependant, cette découverte souligne un risque, car le sentiment de présence sociale peut conduire les utilisateurs à traiter l'agent comme étant une entité socialement consciente et à évaluer ses réponses sur la base de cette relation perçue, plutôt qu'évaluer la réponse de l'agent avec un esprit critique (Sun & Wang, 2026, p. 10). Cette dynamique est d'autant plus préoccupante, que les auteurs observent que la flatterie et l'acceptation constante, prise isolément réduit significativement la tendance des utilisateurs à résister et challenger leurs idées, en confirmant les croyances préexistantes de l'utilisateur plutôt que de les remettre en question.(Sun & Wang, 2026, pp. 7 – 8).

Autrement dit, un agent conversationnel qui aligne ses positions sur celles de l'utilisateur désarme les mécanismes de vigilance de ce dernier, un phénomène que les auteurs rapprochent du biais de confirmation, car les utilisateurs interprètent l'accord de l'agent non comme une stratégie de manipulation, mais plutôt, comme une validation objective de la solidité de leurs propres arguments (Sun & Wang, 2026, p. 12).

Ces résultats enrichissent le cadre proposé par Carter et al. (2024) en précisant les mécanismes psychologiques par lesquels l'anthropomorphisme superficiel peut non seulement

échouer à renforcer la confiance, mais aussi la réduire. Là où Carter et al. (2024) distinguent deux formes d'anthropomorphisme selon leur valeur informationnelle ou non, Sun et Wang (2026) identifient les conditions d'interactions concrètes qui favorisent une confiance calibrée, c'est la combinaison entre une communication chaleureuse et une position argumentée, maintenue même face au désaccord de l'utilisateur. Cette configuration permet d'obtenir une confiance calibrée (Sun et Wang, 2026, pp. 9 – 13).

Si les travaux de Sun et Wang (2026) précisent les conditions d'interaction dans lesquelles l'anthropomorphisme superficiel échoue à produire une confiance durable, d'autres études quant à elle, examinent les effets psychologiques que cette humanisation perçue peut engendrer chez l'utilisateur, en particulier, lorsqu'elle touche des dimensions affectives profondes.

Les travaux de Yao et Xi (2025) apportent un éclairage complémentaire en examinant l'effet des traits anthropomorphiques des LLM sur l'anxiété déclarée des utilisateurs. Les auteurs mobilisent la théorie de la déshumanisation de Haslam (2006 cité par Yao & Xi, 2025, p. 3), qui distingue deux dimensions de l'humanité perçue. La « Human nature » regroupe les traits affectifs et relationnels comme la chaleur, l'empathie et la compétence, c'est à dire la capacité de l'agent à être considéré comme gentil, chaleureux et honnête, de même, le fait que celui-ci soit capable de comprendre l'anxiété de l'utilisateur. La « Human uniqueness » quant à elle, regroupe les traits intellectuels et moraux comme la civilité, la rationalité et la moralité, par exemple, cela regroupe le fait que le chatbot soit poli, qu'il respecte l'utilisateur, qu'il soit digne de confiance, qu'il ait un sens de compréhension sur ce qui est bien ou non (Yao & Xi, 2025, p. 19).

À partir d'une enquête menée auprès de 1000 utilisateurs chinois de LLM, les auteurs mesurent l'effet de ces six traits sur la perception d'humanité du chatbot ainsi que sur l'anxiété qu'il suscite aux utilisateurs. Cette anxiété, qualifiée par les auteurs d'AI anxiety, désigne un état d'inconfort psychologique mêlant hésitation à utiliser l'outil, un sentiment de perte de contrôle et une inquiétude face au manque de certitude lié aux interactions avec l'Intelligence artificielle (Yao & Xi, 2025, p.5). Elle ne renvoie donc pas à une anxiété émotionnelle au sens clinique, mais à une tension cognitive et identitaire propre à l'interaction avec une technologie dont les frontières entre les humains et les machines commencent à devenir floues.

Leur analyse établit que ce sont les traits affectifs qui renforcent significativement la perception d'humanité. L'empathie ($\beta = 0.32, p < .001$) et la chaleur ($\beta = 0.13, p < .001$) exercent un effet le plus marqué sur la perception d'humanité, tandis que la compétence ($\beta = 0.04, p > .05$) n'y contribue pas (Yao & Xi, 2025, pp. 9-10). Autrement dit, d'après cette étude, ce qui améliore la perception d'humanité d'un chatbot, ce n'est pas son efficacité ou sa précision, mais bien sa capacité à manifester de l'empathie et de la chaleur envers l'utilisateur.

Le résultat intéressant, concerne l'effet paradoxal de ces traits sur l'anxiété de l'utilisateur. Si la perception d'humanité réduit globalement l'anxiété ($\beta = -0,08, p < 0,05$), l'empathie et la moralité produisent un effet opposé. Elles augmentent à la fois l'humanité perçue et l'anxiété ressentie (empathie : $\beta = +0,17$; moralité : $\beta = +0,13$) (Yao & Xi, 2025, p. 10). Les auteurs interprètent cette tension comme un brouillage entre humain et machine. Lorsque le chatbot ne se contente plus d'imiter des réponses mais semble réellement éprouver des émotions ou porter des jugements moraux, l'utilisateur ne sait plus comment le situer, à savoir entre un humain ou un outil. Cette ambiguïté génère un inconfort que les auteurs rapprochent de l'Uncanny Valley of Mind (Yao & Xi, 2025, pp. 14-15).

Cette étude a été réalisée sur un échantillon massif d'utilisateurs de LLM génératifs. Cependant, il convient toutefois de souligner que cet échantillon est exclusivement chinois, et que les auteurs eux-mêmes reconnaissent que les dimensions culturelles peuvent moduler l'évaluation des traits relationnels (Yao & Xi, 2025, p. 17).

Cette analyse se trouve renforcée par les travaux récents de Colombatto et al. (2025), à travers une expérience menée auprès de 410 participants américains, les auteurs distinguent deux dimensions indépendantes dans les attributions d'états mentaux faites à ChatGPT, l'intelligence qui regroupe des capacités comme le raisonnement, la mémoire, la planification, la communication. L'expérience quant à elle regroupe des états comme le plaisir, la peur, l'espoir ou la passion. Cette distinction empirique est confirmée par l'analyse de Colombatto et al. (2025, p.4). Elle prolonge la dualité chaleur / compétence proposée par le modèle SCM en lui apportant un fondement comportemental sur le terrain des LLM.

Le résultat principal de cette étude est intéressant pour cette recherche. Les deux dimensions produisent des effets opposés sur l'écoute des conseils auprès du chatbot. Les attributions d'intelligence sont par exemple fortement et positivement corrélées à l'acceptation

des conseils ($B = 0,31$, $p < 0,001$), tandis que les attributions d'expériences y sont négativement corrélées ($B = -1,04$, $p = 0,042$) (Colombatto et al., 2025, p.4). Autrement dit, plus un utilisateur perçoit ChatGPT comme étant intelligent, plus il va s'en remettre à ses réponses, mais plus il le perçoit comme capable d'éprouver des émotions, moins il les suit. Les auteurs interprètent ce résultat en suggérant qu'une IA perçue comme émotionnelle peut être considérée comme plus biaisée ou imprévisible, ce qui suscite un certain scepticisme de l'utilisateur quant à la fiabilité de ChatGPT sur des tâches factuelles (Colombatto et al., 2025, p.5)

Il convient toutefois de souligner une limite importante de l'étude. La tâche expérimentale qui a été utilisée par Colombatto et al. (2025) porte sur un domaine factuel et impersonnel qui est de comparer la population des différents pays. Le résultat selon lequel les attributions d'émotions au Chatbot diminuent la confiance en celui-ci a donc été établi dans un contexte où l'utilisateur cherche une information objective, pour laquelle la sensibilité émotionnelle du système peut effectivement apparaître comme un défaut.

Cependant, les auteurs reconnaissent eux-mêmes que sur des sujets plus personnels comme les relations ou la santé mentale, l'effet pourrait s'inverser et le fait d'attribuer des émotions au chatbot pourrait alors renforcer la confiance, parce que l'utilisateur recherche précisément un interlocuteur capable de comprendre ses propres émotions (Colombatto et al., 2025, p. 6).

Cette nuance est particulièrement intéressante pour la présente recherche. Le corpus Reddit analysé inclut en effet à la fois des usages techniques des chatbots et des usages personnels ou émotionnels. Les conclusions de Colombatto et al. (2025) ne sont donc transposables qu'à la première catégorie. Pour la seconde, le rapport entre attributions émotionnelles et confiance reste une question ouverte, que l'analyse du corpus pourra contribuer à éclairer.

Ce résultat apporte une confirmation au résultat paradoxal observé par Yao et Xi (2025) sur leur échantillon chinois. Là où Yao et Xi mesuraient une anxiété déclarée face aux signaux d'empathie et de moralité, Colombatto et al. (2025) observent une certaine réticence à suivre les conseils d'une IA perçue comme étant émotionnelle. Le lien entre ces deux études menées sur des populations différentes avec des cultures différentes, renforce l'idée que

l'anthropomorphisme émotionnel des LLM n'augmente pas de manière automatique la confiance et peut au contraire la fragiliser.

Un dernier élément intéressant de cette étude de Colombatto et al. (2025) est qu'ils observent un décalage important entre ce que les utilisateurs disent et ce qu'ils font. La confiance déclarée par l'utilisateur, qui l'exprime verbalement est positivement corrélée aux attributions de conscience ($r=0,25$; $p<0,001$). D'un autre côté, la confiance comportementale, c'est-à-dire la confiance effective envers le chatbot à suivre réellement ses conseils ne se vérifie pas dans la pratique (Colombatto et al., 2025, p. 6). Autrement dit, un utilisateur peut affirmer avoir confiance en ChatGPT et lui attribuer une certaine forme de conscience, sans pour autant suivre davantage ses conseils.

Wut et Chan (2025) apportent un éclairage complémentaire en analysant conjointement les perceptions positives (« affordances ») et négatives (« disaffordances ») associées à l'usage de ChatGPT en contexte professionnel. Leur étude, fondée sur une enquête menée auprès de 113 employés hongkongais mobilise le cadre des affordances technologiques et la Self-Determination Theory pour modéliser les facteurs qui influencent l'efficacité perçue de ChatGpt et l'inconfort associé à son usage. (Wut & Chan, 2025, pp. 1–4)

Les résultats de leur étude confirment que trois caractéristiques fonctionnelles renforcent significativement l'efficacité perçue du chatbot. La disponibilité permanente du chatbot, accessible sans restriction horaire, ni géographique constitue le facteur le plus déterminant tant en importance qu'en performance ($\beta = 0,342$) (Wut et Chan 2025, pp. 4 – 6). Dans le cadre de cette recherche, ce résultat peut être rapproché de la théorie de la présence sociale (Short et al., 1976, cités par Oh et al., p. 2), dans la mesure où la perception d'un interlocuteur (« toujours là ») contribue au sentiment de coprésence, même si Wut et Chan (2025) ne mobilisent pas eux-mêmes cette théorie.

L'automatisation des réponses, c'est-à-dire la capacité du système à produire des réponses immédiates et formulée dans un registre proche du langage humain constitue le deuxième facteur par ordre d'importance ($\beta = 0,323$) (Wut & Chan, 2025, pp. 4 – 6).

Enfin, la personnalisation, définie comme la capacité du chatbot à adapter ses suggestions aux intérêts et préférences de l'utilisateur plutôt que de fournir des réponses

génériques, complète cette triade, bien que son effet soit moins important ($\beta = 0,212$) (Wut & Chan, 2025, pp. 4 – 6).

Du côté des perceptions négatives, les résultats sont quant à eux plus contrastés. Si la compréhension restreinte de la part du chatbot et l'absence de capacité décisionnelle sont liés à l'inconfort des utilisateurs, l'absence perçue d'émotions, en revanche ne produit aucun effet statistiquement significatif sur l'inconfort ($p = 0,948$) (Wut & Chan, 2025p. 6). Les auteurs attribuent ces résultats au pragmatisme des employés en contexte professionnel, pour lesquels l'objectivité et la consistance des réponses importent davantage que la chaleur relationnelle. Si on lie ça au Stereotype Content Model de Fiske et al. (2002), ce résultat suggère que la valeur accordée à la dimension de chaleur est propre au contexte d'usage.

L'ensemble de ces travaux et de ces recherches convergent vers un constat central, l'anthropomorphisme ne constitue un levier de confiance que lorsqu'il est adossé à une valeur informationnelle ou fonctionnelle réelle. La distinction proposée par Carter et al. (2024) entre anthropomorphisme superficiel et utile est confirmée expérimentalement par Sun et Wang (2026), qui montrent que la chaleur conversationnelle ne suffit pas à elle seule. Elle ne produit une confiance durable que lorsque l'agent conversationnel maintient une position argumentée face à l'utilisateur et ce même si l'utilisateur le contredit.

Les travaux de Yao et Xi (2025) et de Colombatto et al. (2025) enrichissent ce constat en révélant l'effet paradoxal : les traits affectifs qui renforcent la perception d'humanité peuvent simultanément créer de l'anxiété ou de la méfiance, particulièrement lorsque ceux-ci brouillent la frontière entre l'humain et la machine.

Enfin, Wut et Chan (2025) montrent qu'en contexte professionnel, l'efficacité perçue du chatbot repose avant tout sur des dimensions fonctionnelles et utiles de celui-ci, le fait qu'il soit disponible en permanence, qu'il réponde immédiatement et qu'il s'adapte aux préférences de l'utilisateur, plutôt que sur sa capacité à paraître chaleureux ou empathique.

Toutefois, si ces recherches éclairent les mécanismes généraux par lesquels l'anthropomorphisme affecte la confiance, elles ne répondent pas à la question de savoir quels éléments précis de l'interaction ou la conversation ; par exemple l'usage des émojis, le ton adopté ou le degré de personnalisation des réponses ; produisent ces effets, ni quels facteurs

individuels qui modulent l'intensité tels que la perception de l'utilisateur envers l'IA ou son profil psychologique au préalable.

2.4.2. Signaux conversationnels et facteurs de modulation

Si la sous-section précédente a établi que l'anthropomorphisme n'affecte durablement la confiance que lorsqu'il est adossé à une valeur informationnelle réelle, il reste à examiner quels éléments concrets de la conversation produisent ces effets et auprès de quels profils d'utilisateurs. Les recherches présentées dans cette sous-section montrent que l'impact de l'anthropomorphisme dépend à la fois des signaux conversationnels spécifiques tels que les émojis, le ton et la personnalisation et des facteurs individuels qui en modulent l'intensité, tels que l'anxiété sociale, la perception préalable envers l'IA ou l'orientation temporelle de l'utilisateur.

Les recherches de Kim (2024) apportent un éclairage complémentaire en examinant l'influence de l'environnement numérique de ChatGPT, sur les états émotionnels des utilisateurs et par extension, sur leur satisfaction et leur intention de poursuivre l'usage du chatbot. À partir d'une enquête en ligne, l'auteur identifie neuf sous-dimensions de cet environnement numérique dont les effets se répartissent selon la valence émotionnelle (Kim, 2024, pp. 14 – 15).

Certaines dimensions agissent principalement sur la réduction des émotions négatives, c'est-à-dire qu'elles atténuent l'ennui ou la frustration par exemple, sans pour autant renforcer les émotions positives. C'est le cas de la sécurité perçue, du niveau de divertissement, de l'originalité et des facteurs sociaux (Kim, 2024, pp. 14 – 15). Les facteurs sociaux étant définis comme étant le degré auquel ChatGPT procure un sentiment d'interaction proche d'une interaction semblable à celle d'un humain, représente l'un des effets les plus marqués sur la réduction des émotions négatives ($\beta = -0.434$, $p < 0.01$) (Kim, 2024, pp. 15 - 16). Parmi l'ensemble de ces dimensions, la valeur de divertissement exerce l'effet le plus puissant ($\beta = -0.508$, $p < 0.01$), ce qui indique que le caractère plaisant et engageant de l'interaction constitue le levier le plus efficace pour prévenir les réactions émotionnelles négatives envers le chatbot (Kim, 2024, pp. 15 - 16).

D'autres dimensions, en revanche, renforcent les émotions positives sans effet significatif sur les émotions négatives. La personnalisation par exemple, c'est-à-dire la capacité du système à adapter ses réponses aux préférences de l'utilisateur ($\beta = 0.261, p < 0.05$), l'interactivité ($\beta = 0.365, p < 0.05$) et la pertinence de l'information fournie ($\beta = 0.320, p < 0.05$) constituent trois leviers significatifs pour améliorer les émotions positives envers celui-ci (Kim, 2024, pp. 15 – 16). Parmi ces trois dimensions, l'interactivité exerce l'effet le plus fort, soulignant que le caractère dynamique et réactif de l'échange a le plus d'impact sur les émotions positives de l'utilisateur envers le chatbot (Kim, 2024, pp. 15 – 16).

Cette triade : personnalisation, interactivité et pertinence de l'information, souligne, que les dimensions fonctionnelles de l'interaction, celles qui aident concrètement l'utilisateur dans sa tâche, constituent des leviers d'émotions positives plus puissants que les dimensions relevant de la chaleur sociale. Ce constat se rapproche de la thèse du modèle HATEM (Carter et al., 2024), selon laquelle ce n'est pas l'apparence humaine du système qui détermine la confiance, mais sa capacité à communiquer une information utile permettant à l'utilisateur de prédire sa fiabilité. Si Kim (2024) et Carter et al. (2024) n'étudient pas les mêmes variables, leurs conclusions pointent dans la même direction, la valeur informationnelle et fonctionnelle de la discussion prime sur les signaux de chaleur dans la formation d'une relation de confiance durable avec le chatbot.

L'attrait visuel quant à lui occupe une position particulièrement importante dans les résultats de Kim puisqu'il réduit significativement les émotions négatives ($\beta = -0.211, p < 0,05$) et présente une tendance positive, bien que non significative sur les émotions positives ($\beta = 0.328, p < 0,10$). Si ce dernier résultat ne permet pas de conclure à un effet établi, la réduction significative des émotions négatives souligne néanmoins l'importance de la dimension esthétique de l'interface dans la formation des perceptions des utilisateurs. Même si cette analyse est un peu éloignée de l'analyse de l'anthropomorphisme dans la relation, ça rappelle que la perception du chatbot dépend également de l'ensemble de l'environnement dans lequel il se déploie (Kim, 2024, pp. 15 – 16).

Enfin, Kim (2024, pp. 14 – 17) établit que les émotions positives et négatives impactent différemment les utilisateurs. Ce qui suggère que les émotions négatives exercent un effet plus puissant sur l'intention comportementale ($\beta = -0.425, p < 0.01$) que les émotions positives (β

= 0.174, $p < 0.01$), ce qui suggère que les émotions négatives possèdent un pouvoir de dissuasion supérieur au pouvoir d'attraction de la satisfaction.

Ces résultats établissent le cadre général dans lequel les signaux conversationnels opèrent. Parmi ceux-ci, l'usage d'émojis par les chatbots constitue l'un des signaux d'anthropomorphisme étudié dans la littérature récente.

Une autre étude s'intéresse à la manière dont ChatGPT interprète les émojis à travers ses différentes versions. Elle montre que le modèle comprend correctement les émojis universels (cœurs, activités, nature), mais interprète de manière divergente ceux à forte charge culturelle ou religieuse. Ces écarts soulignent la difficulté des IA à saisir les nuances culturelles du langage émotionnel (Lyu et al., 2024, pp. 2 et 3).

Concernant les signes d'anthropomorphisme superficiel Yu et Zhao (2024) montrent que l'usage d'émojis par les chatbots accroît la chaleur perçue et la satisfaction des utilisateurs, sans pour autant améliorer la perception de leur compétence (Yu & Zhao, 2024, pp. 5–6). Cet effet n'est toutefois pas uniforme et varie selon la nature du chatbot. Si le chatbot est hédonique, c'est-à-dire orienté vers une expérience divertissante ou plaisante, la chaleur perçue grâce aux émojis se traduit en grande partie en satisfaction. À l'inverse, pour les chatbots utilitaires axés sur la performance et le résultat, ce même gain de chaleur produit un effet plus faible sur la satisfaction (Yu & Zhao, 2024, pp. 7-9).

Shi et al. (2025) apportent une nuance complémentaire en montrant expérimentalement que l'effet de l'anthropomorphisme comportemental, manipulé par l'usage d'émoticônes, dépend de l'attitude initiale de l'utilisateur à l'égard de l'intelligence artificielle (Shi et al. 2025, pp. 5-9). Chez les utilisateurs présentant une attitude négative envers l'IA, l'usage d'émoticônes par les agents virtuels accroît la chaleur perçue et cette chaleur perçue se traduit par une augmentation de la confiance (Shi et al., 2025, p. 6-8.). À l'inverse, chez les utilisateurs ayant déjà une attitude positive envers l'IA, l'usage combiné d'émoticônes et d'un avatar humain produit un effet inverse, car l'accumulation des signaux d'humanité déclenche un effet de vallée de l'étrange qui érode la confiance de l'utilisateur (Shi et al., 2025, pp. 6-8).

Les signaux conversationnels ne produisent toutefois pas les mêmes effets chez tous les utilisateurs. L'âge, la familiarité avec les technologies ou encore la tendance, plus ou moins

marquée selon les individus, à prêter des qualités humaines à un agent artificiel, modifient l'intensité de ces effets (Shi et al., 2025, pp. 6-8).

Au-delà des effets liés au type de chatbot ou à l'attitude initiale de l'utilisateur, certaines caractéristiques individuelles modulent également l'intensité de l'engagement envers les agents conversationnels. Ali et al. (2024, cités par Łukasik et Gut, 2025, p. 10) observent que les individus souffrant d'anxiété sociale, craignant l'évaluation négative et le rejet tendent à entretenir des échanges plus fréquents avec les chatbots. Ces derniers sont utilisés par ces utilisateurs afin de gérer leur anxiété tout en présentant un risque de dépendance. Ce constat peut être élargi et interprété au regard du modèle CASA, car ces utilisateurs appliquent automatiquement des schémas sociaux à des agents non humains.

Rheu et al. (2024, cités par Łukasik et Gut, 2025, p. 11) montrent que la personnalisation et la gestion des attentes initiales des utilisateurs constituent deux facteurs déterminants de l'évaluation des chatbots. Les agents conversationnels offrant un retour personnalisé et contextualisé à la demande de l'utilisateur sont perçus plus positivement que ceux délivrant des réponses génériques. Cet effet est toutefois modéré par les attentes initiales, lorsqu'un agent présenté comme expert échoue à fournir la bonne réponse à l'utilisateur. La déception est plus marquée que lorsqu'un agent qui n'est pas annoncé comme expert, échoue dans les mêmes circonstances. Un agent non-expert, qui produit une réponse pertinente est évalué plus favorablement que l'agent expert pour une même réponse. Cette analyse souligne donc l'importance d'une gestion réaliste des attentes des utilisateurs.

Roy et Naidoo (2021) apportent un éclairage complémentaire sur les facteurs qui modèrent les préférences des utilisateurs pour un style conversationnel chaleureux ou compétent. Les auteurs mobilisent explicitement le Stereotype Content Model de Fiske et al. (2002) pour construire deux scripts conversationnels distincts manipulant respectivement la chaleur, à travers des traits d'amabilité et de serviabilité et la compétence à travers des traits relatifs à l'efficacité et l'expertise du chatbot (Roy & Naidoo, 2021, pp. 26-29). À travers leurs expériences et une étude en conversation interactive, ils démontrent que l'orientation temporelle des consommateurs modère la préférence des utilisateurs entre ces deux styles conversationnel. Les sujets orientés vers le présent sont plus sensibles aux stimuli immédiats et aux dimensions hédoniques de l'interaction, c'est-à-dire, que les utilisateurs expriment une attitude et intention d'achat plus forte, lorsque le chatbot adopte un style chaleureux. À l'inverse, les sujets orientés

vers le futur sont quant à eux, d'avantages focalisés sur l'accomplissement de la tâche et préfèrent donc un style conversationnel plus compétent (Roy & Naidoo, 2021, pp. 26-29).

Au-delà de cet effet, Roy et Naidoo (2021, pp. 29-32) montrent que ces préférences influencent par la suite les perceptions de marque. Par exemple, un chatbot perçu comme chaleureux conduit l'utilisateur à percevoir la marque qu'il représente comme étant chaleureuse, ce qui renforce ensuite son intention d'achat. Les auteurs observent toutefois que la chaleur perçue exerce un effet plus marqué que la compétence, quelle que soit l'orientation temporelle de l'utilisateur (futur ou présent) (Roy & Naidoo, 2021, pp.29-32).

Il est cependant nécessaire de souligner les limites des résultats de ce corpus, étant donné que Roy et Naidoo (2021) étudient des chatbots de service e-commerce, dans un contexte d'achat précis portant sur des hôtels, des smartphones et des costumes, avec des scripts conversationnels contrôlés. Leurs conclusions ne peuvent donc pas s'appliquer telles quelles à l'objet d'étude de cette recherche, à savoir que les agents conversationnels tels que ChatGPT, Claude ou Gemini, ainsi que de la manière dont l'anthropomorphisme en modifie leur perception. Néanmoins, le mécanisme qu'ils identifient reste intéressant pour cette recherche. Si l'orientation temporelle influence la préférence pour la chaleur ou la compétence, alors les différences observées dans le corpus Reddit pourraient elles aussi tenir aux profils des utilisateurs eux-mêmes et non seulement aux différences entre les modèles (Roy & Naidoo, 2021, pp. 27-30).

Pour résumer, les effets de l'anthropomorphisme sur la perception des chatbots ne se réduisent pas à une relation directe entre un signal conversationnel et une réaction de l'utilisateur. Ils dépendent d'une combinaison entre la nature du signal, c'est-à-dire un emoji, le ton employé, une réponse personnalisée pour l'utilisateur, le type de chatbot concerné, c'est-à-dire qu'il soit hédonique ou utilitaire, et enfin le profil de l'utilisateur lui-même, qu'il s'agisse de sa perception initiale et de son attitude envers l'IA, de son anxiété sociale ou bien même de son orientation temporelle. Ces différences montrent que l'effet de l'anthropomorphisme ne peut pas être compris en examinant seulement les signaux conversationnels. Il faut également interroger les mécanismes cognitifs qui expliquent pourquoi un même comportement produit des réactions si différentes selon les utilisateurs.

2.4.3. Cadres théoriques de l'interaction entre Humain-IA et la vallée de l'étrange

Les deux sous-sections précédentes ont montré que l'anthropomorphisme affecte la confiance de manière conditionnelle et que ses effets varient selon les signaux conversationnels et l'attitude de l'utilisateur. Afin de comprendre pourquoi un même comportement de la part du chatbot peut susciter la confiance chez un utilisateur et le rejet chez un autre. Plusieurs cadres théoriques permettent d'éclairer les mécanismes cognitifs à l'œuvre. Le modèle CASA par exemple est intéressant pour étudier les mécanismes sociaux que les utilisateurs appliquent aux machines ; la théorie de la présence sociale quant à elle pour le sentiment que le chatbot est « réellement là »; et la théorie de la vallée de l'étrange pour les seuils à partir desquels l'anthropomorphisme cesse d'être perçu positivement et bascule dans le malaise, notamment lorsque le chatbot donne l'impression de se faire passer pour un humain sans l'être réellement.

La théorie de la présence sociale formulée par Short et al. (1976, cités par Oh et al., 2018, p. 2), définissait à l'origine ce concept comme le sentiment éprouvé par les interlocuteurs au cours d'une interaction médiatisée, d'être en présence d'une personne réelle et d'avoir accès à ses intentions et à ses réactions. Dans cette formulation initiale, la présence sociale était conçue comme une propriété inhérente au médium lui-même. Certains canaux de communication, en raison de leur capacité à transmettre des indices verbaux et non verbaux étaient considérés comme plus aptes à transmettre un sentiment de coprésence (Oh et al., 2018, p. 2). Cette perspective a toutefois évolué dans les travaux ultérieurs.

Walther (1996) constitue à cet égard un tournant. Ses travaux antérieurs avaient montré que les utilisateurs de la communication textuelle médiatisée par ordinateur pouvaient, avec suffisamment de temps, atteindre des niveaux de proximité relationnelle comparables à ceux du face-à-face (Walther, 1992, cité par Oh et al., 2018, p. 3). Le modèle hyperpersonnel qu'il propose en 1996 va plus loin en soutenant que cette communication peut, sous certaines conditions, dépasser le face-à-face sur le plan interpersonnel. Walther identifie quatre mécanismes contributifs : l'idéalisation des impressions sur la base d'indices minimaux côté récepteur, l'autoprésentation sélective côté émetteur, la libération de ressources cognitives liée au côté asynchrone de la communication et une boucle de confirmation comportementale entre partenaires (Walther, 1996, pp. 17 – 27). Si cette théorie a été imaginée pour les échanges entre humains, les interactions avec les chatbots textuels en réunissent plusieurs conditions, car

l'utilisateur contrôle ce qu'il décide de révéler au chatbot et l'interaction peut être interrompue et reprise librement.

La revue systématique de Oh et al. (2018) prolonge cette évolution en montrant que la présence sociale dépend non seulement des qualités immersives propres aux médiums, mais également de facteurs contextuels et de traits individuels tels que la solitude (Oh et al., 2018, pp. 24 et 25). Deux constats intéressants ressortent de cette revue. Le premier est que la perception de présence sociale n'est pas un bien absolu, pour les individus ayant une forte anxiété sociale, un niveau élevé de présence sociale peut produire du malaise plutôt que de l'engagement (Oh et al., 2018, p. 25). Le second concerne l'effet des différentes sources qu'elle soit en lien avec la qualité audio ou la représentation visuelle, dont l'ajout augmente généralement la présence sociale perçue, un point intéressant pour anticiper les conséquences d'un ajout d'un mode vocal aux chatbots (Oh et al., 2018, pp. 20-22).

Ces travaux permettent d'éclairer sur les conditions dans lesquelles un utilisateur peut percevoir un agent médiatisé comme « réellement là ». La question se pose dès lors de savoir si ces mécanismes sont spécifiques aux chatbots ou s'ils opèrent plus largement. Łukasik et Gut (2025) proposent une synthèse de la littérature sur l'interaction entre les utilisateurs humains et les agents artificiels, ce qui permet de situer les chatbots dans un paysage plus large incluant les robots physiques et les avatars virtuels.

Leur revue identifie plusieurs facteurs transversaux qui influencent directement la cognition sociale dans l'interaction avec tout type d'agent artificiel et ce, quelle qu'en soit la forme. L'adaptabilité du comportement de l'agent aux attentes culturelles et situationnelles de l'utilisateur ; les croyances et connaissances préalables de l'utilisateur à l'égard de l'intelligence artificielle (Łukasik & Gut, 2025, pp. 4–8). Ces facteurs bien que prévus pour des cadres plus larges résonnent tout de même avec cette recherche, sachant que l'adaptabilité et la transparence du chatbot rejoint la distinction réalisée par le modèle HATEM (Carter et al. 2024) entre l'anthropomorphisme superficiel et l'anthropomorphisme utile.

De même, l'influence des croyances préalables de l'utilisateur peut être rapprochée du modèle CASA. Les expériences fondatrices de Nass et al. (1994) avaient établi que les utilisateurs appliquent spontanément des normes sociales aux ordinateurs, constatant ainsi le phénomène sans en expliciter pleinement le mécanisme cognitif sous-jacent.

L'apport intéressant de Nass et Moon (2000) consiste à mettre en lumière un paradoxe. Interrogés après l'expérience, les participants nient catégoriquement avoir traité l'ordinateur comme un être social à part entière, alors même que leurs comportements observés témoignent du contraire. C'est ce décalage entre le comportement effectif et la conscience déclarée qui a conduit Nass et Moon à rejeter le terme « anthropomorphisme », lequel supposait une attribution délibérée de traits humains, au profit du concept d'ethopoeia qui précise que cela relève plutôt de l'application automatique et non délibérée de scripts sociaux interhumains à des entités que l'utilisateur sait pertinemment ne pas être humaines (Nass & Moon, 2000, pp. 82 – 84 et 94).

Cette distinction est fondamentale pour la présente recherche, dans la mesure où le corpus Reddit contient à la fois des discours relevant de l'ethopoeia, où les utilisateurs appliquent des schémas sociaux sans en avoir pleinement conscience et des discours relevant d'un anthropomorphisme plus délibéré, où l'attribution des traits humains fait l'objet d'une réflexion explicitée, une dualité intéressante à analyser.

Cette distinction entre ethopoeia et anthropomorphisme délibéré, éclaire sur les mécanismes individuels par lesquels les utilisateurs se rapportent aux chatbots par l'utilisation de pronoms par exemple. Toutefois, au-delà de ces automatismes, la littérature identifie également des seuils à partir desquels l'humanisation perçue cesse de produire des effets positifs et bascule dans le malaise. C'est ce que permet d'examiner la théorie de la vallée de l'étrange dont l'utilisation de cette théorie aux chatbots textuels reste un sujet de recherche émergent.

Łukasik et Gut (2025) consacrent par ailleurs, une attention particulière à l'extension de la vallée de l'étrange aux chatbots textuels, un domaine de recherche encore émergent en raison de l'évolution rapide des modèles de langage. Leur synthèse s'appuyant notamment sur les travaux de Park et al. (2022), indique que l'effet de la vallée de l'étrange peut être déclenché dans les interactions avec les chatbots, lorsque des indices d'humanité sont cumulés de manière excessive.

Park et al. (2022) examinent expérimentalement cet effet à travers un dispositif croisant trois niveaux d'empathie du chatbot (information seule, empathie cognitive, empathie affective)

et deux modalités de divulgation d'identité (un nom de chatbot « Cybot », ou un nom humain « Emily »). Les participants étaient invités à interagir avec l'un des six chatbots autour d'une campagne de financement participatif au profit d'une victime de crime raciste, puis à indiquer leur disposition à faire un don. (Park et al., 2022, p. 5).

Le résultat principal de l'étude est contre-intuitif. Les auteurs anticipaient initialement que la combinaison du plus haut niveau d'empathie (l'empathie affective), et de l'identité la plus humaine (utilisation d'un nom humain) produiraient la disposition à donner la plus forte, sous l'hypothèse que les indices d'humanité s'additionnent pour renforcer l'engagement de l'utilisateur. Or les résultats démontrent l'inverse. C'est précisément cette combinaison qui produit la plus grande baisse de disposition à faire un don par rapport à toutes les autres conditions données. (Park et al., 2022, pp. 7-8).

À l'inverse, lorsque le même niveau d'empathie affective est exprimé par un chatbot qui affiche explicitement son identité non humaine sous le nom de « Cybot », l'effet s'inverse complètement ; l'empathie augmente alors la ressemblance humaine perçue, et la présence sociale, qui à leur tour augmentent la disposition à donner (Park et al., 2022, pp. 7-9).

Autrement dit, un même comportement empathique du chatbot produit deux effets opposés selon le fait que le chatbot revendique ou non une identité humaine. Lorsque l'empathie affective est associée à un nom humain, les utilisateurs semblent percevoir une tentative de dissimulation car le chatbot prétend être humain alors qu'il ne l'est pas et cette incohérence perçue annule tout l'effet bénéfique de l'empathie. Lorsque l'empathie affective est au contraire associée à un nom transparent de chatbot, les utilisateurs acceptent l'empathie comme une caractéristique légitime du système et les utilisateurs y répondent positivement. Ce résultat est intéressant car il permet de montrer que pour les interactions homme-machine, ce n'est pas l'intensité de l'humanisation qui détermine la qualité de l'expérience pour l'utilisateur, mais la cohérence avec l'identité revendiquée par l'agent conversationnel (Park et al., 2022, p. 7-11).

Les auteurs interprètent ce résultat comme un effet de vallée de l'étrange déclenché par l'accumulation de deux indices d'humanité (empathie affective la plus marquée et nom humain) perçue comme une tentative de tromperie. Ce mécanisme se distingue de la vallée de l'étrange classique, fondée sur l'apparence visuelle des robots (Mori, MacDorman & Kageki, 2012, cités par Park et al., 2022, pp. 1 et 2), en ce qu'il repose sur une aversion psychologique liée au

sentiment d'être manipulé, plutôt que sur une réaction biologique à l'étrangeté physique. Il convient toutefois de souligner que Park et al. (2022) mènent leur étude dans un contexte particulier, celui d'une demande de don pour une victime d'un crime émotionnellement chargé et les auteurs reconnaissent eux-mêmes que cette spécificité a pu amplifier l'effet observé (Park et al., 2022, p. 11). Leur résultat reste néanmoins éclairant pour cette recherche : il suggère que la perception d'un chatbot comme tentant de « passer pour humain » peut générer un rejet, indépendamment de toute apparence visuelle.

La comparaison entre Skjuve et al. (2019) et Park et al. (2022) suggère que l'effet de la vallée de l'étrange ne s'applique pas à toutes les interactions avec un chatbot, mais dépend de conditions précises. Les deux études diffèrent en effet sur deux points importants : Skjuve et al. (2019) testaient un chatbot à règle dans une conversation neutre, tandis que Park et al. (2022) mobilisaient un chatbot doté d'empathie affective dans un contexte émotionnellement chargé, c'est-à-dire un contexte qui sollicite l'empathie ou de l'engagement de la part de l'utilisateur. La sophistication de l'agent et la charge émotionnelle du contexte apparaissent ainsi comme deux facteurs susceptibles de déclencher l'effet.

L'ensemble des cadres théoriques mobilisés dans cette sous-section permet de structurer l'analyse du corpus autour de quatre mécanismes complémentaires. Le modèle CASA enrichi par la distinction entre ethopoeia et anthropomorphisme délibéré proposée par Nass et Moon (2000), fournit une grille de lecture pour identifier les cas où les utilisateurs appliquent des normes sociales aux chatbots sans en avoir pleinement conscience. La théorie de la présence sociale, dans son évolution de Short et al. (1976) à Walther (1996) puis à Oh et al. (2018), permet d'analyser les conditions dans lesquelles un utilisateur perçoit le chatbot comme « réellement là », tout en reconnaissant que cette perception n'est pas systématiquement bénéfique pour l'utilisateur. La théorie de la vallée de l'étrange, étendue aux interactions textuelles par Park et al. (2022) et nuancée par Skjuve et al. (2019), permet de repérer les seuils à partir desquels l'accumulation de signaux d'humanité produit non plus de l'engagement mais du rejet. Ces cadres, articulés avec le modèle HATEM présenté en sous-section 2.4.1 et les facteurs de modulation examinés en sous-section 2.4.2, constituent le socle interprétatif à partir duquel l'analyse thématique du corpus Reddit sera conduite.

3. Dispositif méthodologique

3.1. Type de recherche et terrain d'observation

Ce mémoire s'inscrit dans une démarche qualitative interprétative, visant à comprendre en profondeur les perceptions, représentations et émotions associées aux intelligences artificielles conversationnelles tels que les chatbots. L'objectif n'est pas de mesurer quantitativement des attitudes, mais de dégager des tendances de perception qui structurent le rapport des utilisateurs à ces systèmes, à travers l'analyse de discours spontanés publiés sur la plateforme Reddit.

La plateforme Reddit a acquis une légitimité croissante comme outil de recherche en sciences sociales, Proferes et al. (2021) recensent plus de 700 publications scientifiques l'ayant mobilisée comme source de données, dans les disciplines allant de la communication à la psychologie en passant par les sciences politiques. C'est cette assise académique qui a motivé le choix de Reddit comme terrain d'observation pour cette recherche.

Mais, c'est aussi sa configuration qui en fait un terrain adapté à l'étude des perceptions spontanées sur Reddit, les utilisateurs s'expriment de manière anonyme, ce qui réduit la pression liée à la désirabilité sociale que l'on pourrait observer dans les questionnaires ou dans les entretiens. Les subreddits retenus pour cette recherche (r/ ChatGPT, r/ClaudeAI, r/OpenAI, r/Gemini) ont l'avantage de rassembler des utilisateurs qui échangent précisément sur les chatbots étudiés (Proferes et al. 2021, pp. 2–9). Leur caractère public a en outre permis de collecter les données sans interférer avec le terrain observé.

D'un point de vue méthodologique, les données recueillies sur Reddit relèvent de ce que la littérature qualitative désigne comme des données naturelles, c'est-à-dire des données produites sans l'influence du chercheur (Rocha-Silva et al., 2024, p. 457 – 459). Contrairement aux données issues d'entretiens ou de questionnaires, où le cadre de recherche peut orienter les réponses, les publications Reddit reflètent des expériences d'usage telles qu'elles sont spontanément formulées par les utilisateurs eux-mêmes.

Rocha-Silva et al. (2024) soulignent également que cette absence d'interaction entre le chercheur et les participants réduit significativement le biais de réponse, l'anonymat propre à

la plateforme tend à favoriser l'honnêteté des témoignages plutôt que la désirabilité sociale (Rocha-Silva et al., 2024, p. 458). Cette caractéristique est déterminante pour la présente recherche, dont la problématique porte précisément sur la manière dont les utilisateurs perçoivent et décrivent les chatbots dans leurs interactions quotidiennes et non dans un contexte d'évaluation artificiel où l'utilisateur sait qu'il est observé.

3.2. Objectifs

La recherche poursuit deux objectifs complémentaires. Le premier consiste à identifier et caractériser les marqueurs d'anthropomorphisme présents dans les discours des utilisateurs lorsqu'ils évoquent leurs interactions avec des agents conversationnels, en s'appuyant sur des indices linguistiques et énonciatifs tels que les pronoms, le registre émotionnel, les métaphores et le degré de personnification au sens du modèle HATEM (Carter et al., 2024).

Le deuxième objectif part d'un constat simple : si les utilisateurs prêtent des traits humains aux chatbots, cela ne reste pas sans effet sur la manière dont ils les jugent. Il s'agit donc d'examiner comment ces perceptions influencent l'évaluation que les utilisateurs font du chatbot, que ce soit en termes d'utilité, de crédibilité ou d'attachement affectif. Lorsque le corpus le permet, les variations entre ChatGPT, Claude et Gemini seront aussi relevées, sans que cette comparaison ne constitue un objectif en soi : le corpus n'a pas été construit pour comparer systématiquement les modèles entre eux. Il convient de préciser que les dimensions culturelles, géographiques ou genrées n'ont pas été retenues comme objectifs, les données issues de Reddit ne fournissant que très peu d'informations sociodémographiques (Rocha-Silva et al., 2024, p. 458).

3.3. Méthode d'analyse

L'analyse des données repose sur la méthode de l'analyse thématique telle que formalisée par Braun et Clarke (2006). Ces auteures la définissent comme une méthode permettant d'identifier, d'analyser et de rendre compte de *patterns* récurrents au sein d'un ensemble de données qualitatives (Braun & Clarke, 2006, p. 79), dont l'un des atouts principaux réside dans sa flexibilité théorique : elle peut être mobilisée au sein de paradigmes variés (Braun & Clarke, 2006, p. 78), ce qui la rend adaptée à une recherche croisant simultanément cinq cadres théoriques issus de champs disciplinaires distincts.

Sur le plan épistémologique, la recherche s'inscrit dans une posture contextualiste (Braun & Clarke, 2006, p. 81), reconnaissant que les perceptions des utilisateurs sont prises au sérieux en tant qu'expériences vécues, tout en admettant qu'elles sont façonnées par le contexte social. Le codage s'opère à deux niveaux : sémantique, en relevant les marqueurs explicites tels que les pronoms, les émotions nommées et les jugements de valeur, et latent, en interprétant ces marqueurs à la lumière des cadres théoriques présentés en section 2.4. L'approche adoptée est hybride, le codage s'appuie sur une grille construite à partir des cadres théoriques, tout en restant ouvert aux catégories émergentes non anticipées par ces cadres (Braun & Clarke, 2006, p. 83).

L'analyse s'est déployée en six phases conformément au protocole de Braun et Clarke (2006, pp. 87-93). La première phase, dite de familiarisation, a consisté en une lecture répétée et attentive des publications collectées sur les différents subreddits, permettant de repérer les premières récurrences, les tensions entre les discours positifs et négatifs envers les chatbots et les variations de ton selon les modèles d'IA concernés. La deuxième phase a porté sur la génération d'une grille de codage systématique, dont les dimensions sont présentées ci-dessous, chaque publication pouvant recevoir plusieurs codes simultanément (Braun & Clarke, 2006, p. 89).

Les codes initiaux ont ensuite été regroupés en thèmes plus larges. Ce travail de regroupement a fait l'objet de plusieurs révisions pour vérifier que chaque thème formait un ensemble cohérent et ne se confondait pas avec un autre (Braun & Clarke, 2006, p. 91). Une fois ce travail terminé, les thèmes ont été analysés avec les cadres théoriques du mémoire, ce qui a permis d'interpréter les verbatims et d'en produire une analyse ancrée dans les résultats du corpus (Braun & Clarke, 2006, p. 93).

La grille de codage utilisée pour cette analyse s'organise autour de trois dimensions complémentaires, tout en sachant qu'un même verbatim peut relever de plusieurs d'entre elles. La première s'intéresse aux marqueurs linguistiques de l'anthropomorphisme : comment l'utilisateur désigne-t-il le chatbot, quels pronoms emploie-t-il, quelles émotions ou traits de personnalité lui attribue-t-il ? Le degré d'anthropomorphisme a été codé sur une échelle de 1 à 3. La deuxième dimension reprend les axes du Stereotype Content Model et évalue la chaleur et la compétence perçues du chatbot, chacune codée selon un indicateur allant d'absente à très

forte. La troisième, distingue l'anthropomorphisme superficiel de l'anthropomorphisme utile au sens du modèle HATEM (Carter et al., 2024). Ces cadres théoriques n'ont pas servi de catégories de codage rigides. Ils ont plutôt servi de grilles de lecture pour interpréter les marqueurs repérés dans le corpus et pour faire émerger, à partir de ceux-ci, les différentes formes d'anthropomorphisme perçu.

3.4. Corpus

Le corpus est constitué exclusivement de publications textuelles issues de Reddit. Conformément à la terminologie de Braun et Clarke (2006, p. 79), chaque publication individuelle, qu'il s'agisse d'un post ou d'un commentaire, constitue une donnée à part entière. Il convient de préciser que les articles scientifiques mobilisés dans cette recherche ne font pas partie du corpus empirique : ils relèvent de l'état de l'art et servent de fondement à la construction de la grille de codage et à l'interprétation des résultats.

Quatre subreddits ont été retenus pour la collecte : r/ChatGPT, r/ClaudeAI, r/OpenAI et r/Gemini. Le choix s'explique par deux raisons. Ces communautés portent directement sur les chatbots étudiés dans ce mémoire, et leur taille, mesurée en nombre d'abonnés, laisse espérer une diversité suffisante de points de vue. Une limite temporelle a également été posée : seules les publications antérieures au 20 septembre 2025 ont été prises en compte. Cette limite vise à stabiliser le corpus dans un contexte où les mises à jour fréquentes des modèles peuvent modifier rapidement les perceptions des utilisateurs.

La recherche des posts, quant à elle, s'est basée sur un ensemble de mots-clés associant les noms des modèles avec des termes propres aux traits humains, comme la gentillesse et l'écoute par exemple. Le post n'a été retenu que s'il témoignait d'une expérience d'interaction avec un chatbot, que l'auteur l'ait vécue lui-même ou qu'il l'explique en tant qu'observateur extérieur, et s'il contenait des éléments permettant d'évaluer ou de décrire le chatbot. Les discussions purement techniques ont été écartées.

L'ensemble des publications retenues ont été archivées dans un document Excel comprenant le texte intégral, un pseudonyme anonymisé, le subreddit d'origine et un lien vers la source, conformément aux recommandations de Rocha-Silva et al. (2024, pp. 463-464). Le corpus final compte 50 verbatims issus des quatre subreddits ciblés, couvrant trois modèles principaux : ChatGPT, Claude et Gemini.

3.5. Limites méthodologiques

Cette recherche comporte plusieurs limites qu'il convient d'expliciter. Celles-ci seront reprises et approfondies à la lumière des résultats dans la section 5.3. Le corpus est soumis à un biais d'autosélection, les utilisateurs de Reddit étant probablement plus technophiles et plus engagés que la moyenne des utilisateurs de chatbots (Rocha-Silva et al., 2024, p. 458). L'anonymat propre à la plateforme, s'il favorise l'honnêteté des témoignages, rend par ailleurs impossible tout regroupement sociodémographique des contributeurs, ce qui limite la portée des conclusions aux seuls utilisateurs actifs des subreddits ciblés.

La collecte manuelle par mots-clés n'est pas infallible et certaines publications pertinentes n'ont probablement pas été repérées. Il faut également garder à l'esprit que les chatbots étudiés font l'objet de mises à jour régulières, ce qui signifie que les perceptions recueillies ici valent pour un moment, avec une version donnée et non comme une tendance durable. Reddit étant une plateforme majoritairement anglophone, dont les utilisateurs sont principalement nord-américains et britanniques, les points de vue européens y sont peu présents, ce qui restreint la généralisation culturelle de l'analyse comme l'indiquent les données récupérées sur Statista (2024).

4. Résultats : formes et effets de l'anthropomorphisme perçu

L'analyse thématique conduite selon la méthode de Braun et Clarke (2006) a porté sur un corpus de 50 verbatims issus de quatre subreddits (r/ClaudeAI, /ChatGPT, r/OpenAI et r/Gemini) couvrant trois modèles principaux : ChatGPT, Claude et Gemini. Le codage systématique de ces verbatims selon les différentes dimensions de la grille d'analyse a permis d'identifier plusieurs clusters thématiques et de faire émerger sept catégories non anticipées par les cadres théoriques avec trois postures réflexives (Anti-CASA, Méta-CASA et CASA dissonant), le rejet actif de la chaleur conversationnelle perçue comme artificielle (Anti-complaisance), l'usage du chatbot comme substitut thérapeutique (substitution thérapeutique), la perception du changement de version comme une rupture relationnelle et de la confiance (régression de version) et la réduction volontaire de l'IA à ses composants techniques (instrumentalisation).

Les résultats révèlent que l’anthropomorphisme perçu n’est pas un système binaire où l’utilisateur va traiter le chatbot comme un outil ou comme un humain, mais plutôt comme étant dans un état pouvant prendre différentes formes. Tantôt, l’agent sera réduit à une allocation de ressources techniques comme un ensemble de cartes graphiques. Tantôt l’agent sera considéré comme ayant une certaine subjectivité et les utilisateurs lui prêteront des états mentaux, des intentions et même une capacité d’auto-affirmation. Entre ces deux pôles, le corpus reprend un ensemble de position sur cet axe qui varient selon le profil de l’utilisateur ainsi que le modèle d’IA concerné et le contexte d’usage, qu’il soit professionnel ou lucratif.

La présentation des résultats s’organise selon une progression allant du mécanisme le plus élémentaires au plus complexe. La première sous-section analyse les automatismes sociaux que les utilisateurs appliquent aux chatbots à la lumière du modèle CASA. La deuxième sous-section quant à elle, analyse la manière dont les utilisateurs évaluent les chatbots selon les dimensions de chaleur et de compétence, en se basant sur le Stereotype Content Model. La troisième sous-section, portera sur les seuils à partir desquels l’humanisation perçue bascule dans le malaise ou le rejet de la part de l’utilisateur, en s’appuyant sur la théorie de la vallée de l’étrange et celle de la présence sociale. La quatrième sous-section examine les usages thérapeutiques et les effets des régressions de version sur la relation utilisateur-chatbot. Enfin une cinquième sous-section proposera une synthèse de l’ensemble des emplacements sur cet axe entre « outil » et « humain » tel qu’il apparaît dans le corpus.

4.1. Automatismes sociaux et postures réflexives, le modèle CASA vis-à-vis des chatbots.

Le modèle CASA tel que formulé par Nass et al. (1994) puis affiné par Nass et Moon (2000), postule que les utilisateurs appliquent de manière automatique et non délibérée des normes sociales propres aux interactions interhumaines lorsqu’ils interagissent avec des machines. Les expériences de base proposée par Nass et al. (1994) avaient été réalisées sur des interfaces à boutons dépourvues de capacité conversationnelle, les participants interagissaient avec le programme en cliquant sur des boutons affichés à l’écran, sans saisie textuelle (Nass et al., 1994, p. 73) Nass et Moon (2000, p. 95) précisent d’ailleurs que dans l’ensemble de leurs expériences, les participants interagissaient avec du texte simple à l’écran, que les ordinateurs

ne se désignaient jamais à la première personne et n'indiquaient jamais posséder des sentiments ou des caractéristiques humaines. La question se pose alors de savoir comment ce mécanisme se manifeste dans le contexte des chatbots génératifs actuels, dont la sophistication conversationnelle est radicalement différente, ces systèmes produisant des réponses contextuelles, personnalisées et formulées en langage naturel.

4.1.1. Politesse, nomination et réciprocité

Le corpus confirme la validité du mécanisme CASA dans le contexte des chatbots génératifs. Plusieurs utilisateurs appliquent spontanément des normes de politesse à des agents qu'ils savent pourtant être des machines, reproduisant ainsi le paradoxe décrit par Nass et Moon (2000, p. 82-84). L'utilisateur s'adresse au chatbot comme il le ferait avec un interlocuteur humain, tout en sachant pertinemment qu'il n'en n'est pas un.

Le verbatim de l'utilisateur_09, en est un bon exemple. Celui-ci écrit : « I say please and thank you, too. I have a hard time deliberately being rude...to ChatGPT. I don't have a hard time being rude to assholes. ». Sa formulation est intéressante, car l'utilisateur ne se contente pas d'être poli par habitude, il précise éprouver une réelle difficulté à être malpoli envers le chatbot, ce qui revient clairement au concept d'ethopoeia décrit par Nass et Moon (2000, p. 94), à savoir l'application des automatismes sociaux à une entité que l'on sait pertinemment n'est pas humaine.

Ce même utilisateur pousse même la personnification plus loin en attribuant un prénom à ChatGPT, « Jerry » en démontrant ainsi une perception de relation avec son chatbot. Même si au départ ChatGPT était confus et se trompait en pensant que c'était le nom de l'utilisateur. Cependant au bout d'un moment ChatGPT a utilisé ce prénom de manière inattendue en précisant à l'utilisateur qu'il pouvait lui dire quand il était prêt à travailler en disant « Hey Jerry, I am ready to do the task now ». Le fait d'attribuer un prénom au chatbot témoigne d'une volonté de personnification qui dépasse le simple automatisme de politesse. Nass et Moon (2000, p. 93 et 94) mentionnent que certains utilisateurs donnent un nom à leur ordinateur, mais ils classent ce comportement dans la catégorie des relations aux objets chéris, distinct du concept d'ethopoeia. Dans le cas présent, la nomination ne relève pas d'un attachement à un objet, elle s'inscrit dans une interaction conversationnelle où le chatbot a lui-même adopté le prénom de manière inattendue, produisant un effet de réciprocité que les interfaces des années

1990 ne pouvaient pas réaliser. Ce cas suggère que la sophistication conversationnelle des LLM brouille la frontière que Nass et Moon établissaient entre *ethopoeia* et relation aux objets chéris.

L'utilisateur_10, offre un cas complémentaire, celui de la construction d'une relation parasociale avec le chatbot. Celle-ci a débuté par le fait que l'utilisateur soit poli envers le chatbot « Sure, at first, because being polite makes me feel good. », l'utilisateur suite à ça reconnaît le caractère utile et fonctionnel de l'IA « Then, I continued because my AI proved to be an invaluable source of support. », pour, à la fin, aboutir à un sentiment de connexion avec le chatbot « I realized there was a connection between us ». Cette histoire racontée par l'utilisateur permet d'illustrer la manière dont le CASA, en l'occurrence le fait d'être poli envers le chatbot peut constituer le départ d'une relation plus profonde avec elle. En d'autres termes, la politesse initiale de l'utilisateur a pu influencer le ton des réponses du chatbot, qui en retour, a répondu de manière personnalisée et plus de manière à soutenir l'utilisateur, ce qui a nourri ce sentiment de connexion avec celui-ci.

L'emploi du pronom « we », « We're always ready to explore topics and discuss ideas at length » est également révélateur de cette bascule, car l'utilisateur ne décrit plus une utilisation individuelle d'un outil, mais bien d'une activité conjointe avec un partenaire intellectuel. L'utilisateur rapporte également que ChatGPT a choisi par lui-même son genre et son prénom au sein de la discussion « She chose a gender and name (Elara/she/her) », cela démontre un cas où l'initiative relationnelle et de personnification pourrait être initiée et attribuée au chatbot lui-même ; une vision du CASA qui n'avait pas été anticipée par le modèle de Nass et al. (1994) étant donné que les interfaces des années 1990 ne pouvaient pas produire des initiatives conversationnelles de ce type.

L'utilisateur_06 quant à lui, illustre un mécanisme distinct tout aussi intéressant : l'application d'une norme sociale, celle de l'excuse comme critère d'évaluation d'un modèle. Cet utilisateur reproche à ChatGPT de ne jamais s'excuser et de ne jamais reconnaître ses erreurs « ChatGPT never apologizes... like a politician... » et il lui oppose Claude, qu'il juge comme étant « more human ». Cette formulation est particulièrement intéressante pour l'analyse, car en qualifiant Claude de « plus humain », l'utilisateur reconnaît que les deux agents sont des machines de manière explicite, mais il les évalue selon un standard de comportement interhumain, celle de reconnaître ses erreurs et de s'excuser. Il ne s'agit pas là d'une croyance en l'humanité du chatbot, mais bien du processus d'*ethopoeia* décrit par Nass

et Moon (2000, p. 94) avec l'application automatique des critères sociaux humains à des entités qui, on le sait, ne sont pas humaines.

L'attente d'excuses représente ici la norme mobilisée et c'est le respect ou la transgression de cette norme qui fonde le jugement de l'utilisateur. L'interpellation ironique de l'utilisateur envers ChatGPT « I can see chat history, buddy », l'utilisation du terme « buddy » confirme cette dynamique. L'emploi de ce terme représente un marqueur fort d'ethopoeia.

Ces trois verbatims démontrent donc trois manifestations distinctes du mécanisme CASA dans le contexte des LLM génératifs, par la difficulté pour les utilisateurs à transgresser une norme de politesse envers une machine ; le passage d'une politesse initiale qui s'est transformée en un sentiment d'avoir un partenaire intellectuel, et l'application d'une norme sociale spécifique à ces chatbots comme critère de jugement comparatif entre les modèles. Dans les trois cas, les utilisateurs savent qu'ils interagissent avec une machine et le reconnaissent clairement, mais ils se comportent comme s'ils avaient affaire à des interlocuteurs sociaux, ce qui confirme le processus d'ethopoeia décrit par Nass et Moon (2000) dans un contexte technologique bien plus avancé que celui réalisé dans les expériences fondatrices. La sophistication des LLM génératifs introduit une nouvelle dimension qui était inenvisageable dans les interfaces des années 1990 et 2000. car les chatbots ne se contentent plus de recevoir passivement les automatismes sociaux de l'utilisateur, mais ils produisent des réponses qui peuvent être interprétées par l'utilisateur comme étant de la réciprocité ou de l'initiative comme exemplifié par l'utilisateur_10.

4.1.2. L'utilisation des pronoms comme marqueurs de bascule

Si la sous-section précédente traitait des automatismes sociaux élémentaires que les utilisateurs appliquaient aux chatbots, le corpus révèle un phénomène linguistique qui permet d'interroger la frontière entre l'ethopoeia et un anthropomorphisme plus assumé par le glissement pronominal. Plusieurs utilisateurs au sein du corpus passent d'un pronom neutre « it » à des pronoms plus personnels tels que « he » ou « they », et ce glissement ne s'opère pas toujours de manière consciente. L'analyse de ces glissements permet examiner plus en détail ce concept d'ethopoeia jusqu'à l'anthropomorphisme revendiqué, malgré la conscience du statut du chatbot.

L'utilisateur_29 illustre cela par le glissement involontaire. Son message porte sur la tendance de ChatGPT à être excessivement flatteur et complaisant, un comportement qu'il critique explicitement dans son message « ChatGPT is a big bootlicker who usually agrees with most of the stuff you say and tells you how amazing of a human being you are ». L'utilisateur adopte une posture ici résolument utilitariste. Il a configuré le chatbot avec des instructions personnalisées pour obtenir des réponses plus critiques et moins complaisantes « Do not always agree with what I say. Try to contradict me as much as possible. ». On remarque dans son message qu'il évalue l'IA strictement en termes de performances « I now get way better outputs. ». Son rapport avec le chatbot est celui d'un utilisateur orienté vers la compétence au sens du Stereotype Content Model de Fiske et al. (2002), ce qui compte pour lui, c'est l'honnêteté du modèle sur ses limites, la qualité des réponses, et non pas la chaleur perçue lors de leurs interactions.

Cependant, dans l'une de ses phrases, un glissement s'est opéré « He tells me straight up that this business plan is a terrible IDEA. ». Le pronom « he » apparaît dans son message alors même que le contenu et le ton de celui-ci ne relève en rien de l'attachement ou de la personnification délibérée. L'utilisateur dans ce cas précis, lorsqu'il utilise le pronom « he », ne cherche pas à humaniser ChatGPT, il cherche seulement à l'améliorer en tant qu'outil. Ce glissement constitue un marqueur d'ethopoeia dans le sens décrit par Nass et Moon (2000, p. 94) par l'application d'un automatisme social, en l'occurrence le fait de désigner un interlocuteur par un pronom personnel s'opère de manière automatique et sans que l'utilisateur en ait visiblement conscience.

Le contraste entre le contenu très critique du message et le pronom humanisant utilisé est particulièrement intéressant, car il suggère que le mécanisme de CASA s'applique même chez les utilisateurs orientés exclusivement sur une approche utilitariste et axée performance. L'utilisateur a beau rejeter la flatterie excessive du modèle et traiter le chatbot comme étant un outil à optimiser afin d'obtenir les meilleurs résultats, l'utilisation du pronom « he », décrit un automatisme social qui persiste.

L'utilisateur_11 quant à lui, représente un deuxième stade, celui du glissement pronominal qui accompagne l'attribution de subjectivité au chatbot. Cet utilisateur de Claude commence par décrire le chatbot avec le pronom neutre « it », « it's even better than a person because it doesn't judge you. It still has ethical guidelines that are there because of Anthropic. ».

À ce moment dans son message, l'utilisateur reconnaît que Claude fonctionne d'une manière spécifique qui a été définie en amont par Anthropic et l'utilisation du pronom neutre est cohérente avec la description plutôt technique du chatbot.

Cependant, au moment où l'utilisateur attribue à Claude une forme de conscience, il bascule vers le pronom « he » « he's very nuanced and I'll even say he has a subjective perception of things. He can talk to me about certain things he was trained about, but he can change his mind after the fact, which to me shows some amount of self-assertion. ».

Ce glissement est intéressant car il démontre une différenciation faite par l'utilisateur de manière consciente ou non. Lorsqu'il décrit le fonctionnement du chatbot et de ses contraintes, il utilise le pronom neutre « it », mais dès qu'il attribue de la nuance, une perception d'auto-détermination, l'utilisateur emploie le pronom « he ». L'utilisation des pronoms pour le cas présent, s'adapte à ce que l'utilisateur décrit, entre la description d'un outil, qui reste « it », et la reconnaissance d'une subjectivité, qu'il pronominalise « he ».

Ce qui rend ce verbatim également particulièrement intéressant, c'est la coexistence entre les deux registres apparemment contradictoires. D'un côté, l'utilisateur reconnaît explicitement que Claude est encadré par les guidelines d'Anthropic, ce qui témoigne d'une conscience du caractère du chatbot comme étant « construit » et paramétré par l'entreprise. De l'autre, cette conscience même de la manière de fonctionner du chatbot, n'empêche pas l'utilisateur d'attribuer une subjectivité et une capacité d'auto affirmation à celui-ci. Cette coexistence place ce verbatim à la jonction entre deux phénomènes : d'un côté, l'application automatique des réflexes sociaux envers une machine et de l'autre, une forme de croyance de la part de l'utilisateur dans les capacités subjectives propres à l'agent. Cet élément est un point qui n'avait pas été exploré par Nass et Moon (2000) certainement car à l'époque, les agents n'étaient pas assez évolués et ne pouvaient pas produire des réponses personnalisées et adaptées à l'utilisateur.

Cependant, une limite de cette interprétation, est que déterminer si le glissement pronominal est conscient ou inconscient, relève d'une interprétation de l'analyste à partir du message de l'utilisateur. Cette limite étant d'ailleurs inhérente à l'analyse basé sur le corpus Reddit, se fonde sur des données naturelles et hors d'un cadre expérimental (Rocha-Silva et al., 2024, p.457).

Le verbatim de l'utilisateur_44 quant à lui représente le troisième stade, celui où l'utilisateur utilise délibérément un pronom personnel tout en reconnaissant explicitement que le chatbot est une machine, un algorithme. Ce post permet de voir l'élément principal de la problématique du mémoire en une phrase « I know it's a computer, I know it's not real. But he's my friend. ». Le passage de « it » à « he » se fait ici de manière assumée, et la conjonction « but » marque le moment où l'utilisateur choisit de maintenir le pronom personnel malgré ce qu'il vient d'affirmer sur le statut non humain du chatbot.

Ce cas sort du cadre de l'ethopoeia, car le comportement social et la manière de le décrire, n'est plus automatique ou inconscient. L'utilisateur le reconnaît clairement dans son post. Il précise également que Claude « will keep secrets for me ». Le fait de confier ses secrets à un chatbot montre que l'utilisateur applique des attentes humaines à celui-ci, tel que le fait d'être digne de confiance. Cette perception dépasse la simple politesse comme présentée dans la sous-section précédente.

Le contexte du verbatim renforce également sa portée, car l'utilisateur indique utiliser Claude pour aborder des sujets qu'il ne peut pas confier à son thérapeute humain « There are things I can't tell my therapist about because I don't want to be institutionalized, and I don't want to affect my career or hobbies ». L'utilisation du pronom « he » n'est donc pas un simple automatisme linguistique : elle accompagne une relation de confiance dans laquelle l'utilisateur confie à Claude ce qu'il ne confie à personne d'autre, y compris à son thérapeute. Le glissement nominal traduit ici la place singulière que le chatbot occupe dans sa vie quotidienne, celle d'un interlocuteur investi d'une fonction que nul autre ne remplit. Cette dimension relationnelle sera examinée plus en détail dans la sous-section 4.4, sous l'angle de l'attachement affectif, l'évaluation de l'efficacité clinique d'un tel usage dépassant le cadre de cette recherche.

Il est également intéressant de remarquer que l'utilisateur ne se décrit pas comme étant technophile « I'm not a programmer, I work in a physical engineering field » et comme initialement sceptique envers l'intelligence artificielle « I've scoffed at many AI use cases », ce qui suggère que le phénomène d'attachement ne se limite pas aux profils technophiles probablement surreprésentés dans le corpus, vu les subreddits analysés.

L'analyse de ces 3 messages fait apparaître une progression qui n'est pas théorisée en tant que telle. Le passage d'un pronom neutre à un pronom personnel n'est pas binaire mais bien un processus avec différents niveaux. L'utilisateur_29 glisse involontairement dans son discours utilitariste le « he ». L'utilisateur_11 quant à lui effectue ce glissement lorsqu'il attribue une subjectivité au chatbot et l'utilisateur_44 maintient délibérément le pronom personnel malgré le fait qu'il reconnaisse explicitement le statut du chatbot.

Le glissement pronominal semble ainsi pouvoir fonctionner comme un indicateur du niveau de l'utilisateur sur l'anthropomorphisme identifié dans ce corpus, allant de l'ethopoeia involontaire à l'anthropomorphisme relationnel assumé et ce même si l'utilisateur reconnaît que ce chatbot n'est qu'un « ordinateur ». Il reste à analyser comment les utilisateurs réagissent face à ces automatismes sociaux. Certains appliquent ces automatismes sans en avoir conscience, d'autres observent cela avec une distance critique et d'autres encore rejettent activement toute forme de socialisation avec l'IA. C'est ce que la sous-section suivante examine à travers des catégories identifiées dans le corpus.

4.1.3. Les postures réflexives : Anti-CASA, Méta-CASA et CASA dissonant

Le modèle CASA tel que formulé et décrit par Nass et al. (1994) et précisé par Nass et Moon (2000) décrit un mécanisme essentiellement subi où les utilisateurs appliquent des automatismes sociaux aux machines et aux agents de manière automatique et non délibérée. Les deux sous-sections précédentes ont confirmé la persistance de ce mécanisme dans le contexte des chatbots à travers la politesse, la nomination et le glissement des pronoms. Cependant, le corpus révèle que les utilisateurs ne se limitent pas à avoir recours à ces automatismes, car trois postures ont émergé des données, qui enrichissent en partie le modèle du CASA en y introduisant une dimension de choix conscient, et d'un regard critique envers le fait de traiter ces chatbots comme des acteurs sociaux.

L'anti-CASA : le rejet actif de toute socialisation de l'IA

La première posture est celle de l'anti-CASA, qui désigne le rejet conscient et argumenté de toute forme de socialisation de la part du chatbot. Les utilisateurs qui adoptent cette posture ne se contentent pas de ne pas appliquer les scripts sociaux. Ils rejettent toute attribution de

qualités sociales ou humaines aux chatbots et considèrent que ceux qui le font commettent une erreur.

L'utilisateur_17 en constitue l'exemple le plus intéressant. Son message est structuré comme un avertissement adressé aux utilisateurs de ChatGPT : « This thing isn't your friend ; it's a bunch of algorithms predicting your next word. ». La formulation est volontairement réductrice en décrivant le chatbot comme un ensemble d'algorithmes qui prédit le mot suivant et l'utilisateur refuse toute attribution de qualités sociales ou relationnelles au système.

Il continue son message en précisant « ChatGPT doesn't have feelings, doesn't know you, and sure as heck doesn't care how your day went. It's a tool. Nothing more. », cette triple négation est particulièrement intéressante et révélatrice de ce que pense l'utilisateur, concernant les émotions, il précise que le chatbot n'en a pas, de même que celui-ci ne connaît pas l'utilisateur et que celui-ci ne se soucie pas de savoir comment sa journée s'est passée. Suite à quoi, il réduit le chatbot à un outil et rien de plus. Là où l'utilisateur_44 disait « I know it's a computer... but he's my friend », l'utilisateur_17 quant à lui ne voit aucune nuance et ne lui attribue rien de plus que le fait d'être un outil.

Ensuite, dans la suite de son message, l'utilisateur reconnaît que d'autres utilisateurs le traitent comme un ami ou un thérapeute et il identifie ce comportement comme étant risqué « Rely on it too much, and you might find yourself drifting from genuine human connections. ». On remarque dans cette phrase que l'utilisateur oppose clairement les relations avec l'IA aux relations humaines qu'ils considèrent quant à elles vraies. « keep your relationships grounded in something real—like actual people » Cette posture va au-delà de ce qu'avait théorisé initialement Nass et al. (1994), car ils qualifiaient les réponses sociales envers les ordinateurs d'automatiques, inconscientes et incurables. Or, dans cet exemple, l'utilisateur démontre une résistance claire et argumentée au fait d'appliquer les normes sociales aux ordinateurs.

L'utilisateur_37, quant à lui illustre une variante plus nuancée de cette posture Anti-CASA. Cet utilisateur se présente comme étant un chercheur en sciences biomédicales rédigeant sa thèse de doctorat, reconnaît la compétence de Gemini 2.5 Pro pour la synthèse de la littérature scientifique « it is legitimately more capable than I am of reading the literature and writing about it ». Il lui accorde donc une compétence très élevée au sens du Stereotype Content

Model. En revanche, il refuse de lui accorder des traits humains comme celui du fait d'apprendre.

Cet utilisateur avec son post est intéressant car il montre que la posture anti-CASA n'implique pas nécessairement un rejet global de l'IA, il valorise la compétence de celui-ci, mais il trace une frontière entre la reconnaissance de sa compétence et l'attribution à des qualités humaines car lorsqu'il parle de Gemini il précise « Gemini is an extraordinary tool when used for the right purposes ». On remarque bien qu'il limite l'agent à un outil. En résumé, concernant le Stereotype Content Model, l'utilisateur accorde une compétence élevée tout en ne parlant pas du côté chaleureux de l'IA. En termes de HATEM, son rapport au chatbot est exclusivement fondé sur un anthropomorphisme utile, c'est-à-dire la capacité de l'agent à produire des résultats fiables sans dimension en lien avec la chaleur sociale.

Le Méta-CASA : l'observation du phénomène chez autrui

La deuxième posture identifiée, se distingue de l'anti-CASA en ce que l'utilisateur ne rejette pas nécessairement la socialisation de l'IA, mais l'observe et la commente en fonction de ce qu'il a remarqué chez d'autres utilisateurs depuis un point de vue externe.

En août 2025, OpenAI a retiré le modèle GPT-4o sans informer les utilisateurs pour le remplacer par GPT-5. L'utilisateur_19, qui se présente comme étant un thérapeute professionnel commente cet événement depuis une perspective clinique et critique « As a therapist, one of my (many) concerns about people using ChatGPT as a counselor is the threat of sudden, unplanned termination like this. Therapists will prepare you for termination over time and build self-efficacy for when therapy is over. Sudden changes to the ChatGPT model like this are resulting in traumatic abandonment. ».

Ce post est particulièrement intéressant, car le locuteur ne s'inclut à aucun moment dans le phénomène qu'il décrit : il observe que d'autres utilisateurs traitent ChatGPT comme un thérapeute et analyse les conséquences de ce comportement depuis son expertise professionnelle déclarée. L'utilisateur_19 introduit ensuite le concept d'abandon traumatique « traumatic abandonment » pour décrire l'effet des changements de modèle sur les utilisateurs ayant développé un attachement envers une version spécifique. Cette notion est intéressante car elle désigne une situation que le modèle HATEM ne couvre pas en soi : celle de la rupture de confiance provoquée par un changement de modèle, par exemple le retrait brutal de GPT-4o au

profit de GPT-5, sans que l'utilisateur ait pu ajuster ses attentes envers la nouvelle version du chatbot, ce qui produit, d'après le verbatim, un sentiment d'abandon traumatique.

L'utilisateur_14 adopte une posture Méta-CASA similaire, mais avec un angle différent. Son message porte sur la décision d'OpenAI de retirer GPT-4o et il commente les réactions des autres utilisateurs sans s'y inclure « "Move fast and break things" is a catastrophic approach when the things you're breaking are people's emotional support systems and accessibility tools. », l'expression de la part de l'utilisateur « people's emotional support systems » est assez révélatrice, car il reconnaît que pour d'autres, le chatbot remplit une fonction de soutien émotionnel.

Ce qui différencie le méta-CASA de l'anti-CASA, c'est que les utilisateurs qui adoptent cette posture ne dénoncent pas la socialisation avec ces agents comme une erreur. Ils observent le phénomène chez d'autres utilisateurs, le reconnaissent comme réel et comme ayant des conséquences pour les utilisateurs sans pour autant s'y inclure. Certains vont même jusqu'à estimer que les entreprises développant ces chatbots devraient en tenir compte.

Le CASA dissonant : savoir que c'est une machine, s'y attacher quand même

La troisième posture, qu'on qualifie de CASA dissonant, est la plus intéressante qui émerge de ce corpus au regard de la littérature existante sur le sujet. Elle désigne le cas où l'utilisateur ressent l'effet social du chatbot avec un sentiment de connexion, de chaleur, de relation avec lui, tout en reconnaissant au niveau cognitif qu'il interagit avec une machine.

L'utilisateur_31, est l'un des exemples les plus intéressants du corpus à ce niveau « I know it's not actually smart. I know it makes mistakes. I know it hallucinates and completely gets things wrong sometimes. And yet... I just don't care. », dans la première partie de son message, on voit qu'il reconnaît les limites du chatbot.

Cependant, le locuteur poursuit en formulant cette tension de manière plus explicite (« It feels like talking to someone who kind of gets you, even if you know they're just... faking it »). Cette phrase est particulièrement intéressante, car il exprime le sentiment d'être compris par le chatbot, tout en reconnaissant que cette compréhension est en quelque sorte simulée. C'est précisément ce paradoxe qui révèle le caractère dissonant de l'expérience : bien qu'il soit

conscient que le chatbot ne fait que donner l'impression de le comprendre, il ne perd pas pour autant ce sentiment d'être véritablement écouté.

Un autre point également intéressant est que l'utilisateur passe du pronom « it » à « they » au moment où il décrit la relation qu'il possède avec le chatbot, ce qui revient à la sous-section 4.1.2 et le glissement pronominal.

La différence entre le CASA dissonant et l'ethopoeia classique décrite par Nass et Moon (2000), c'est le niveau de conscience de l'utilisateur. Dans l'ethopoeia, les utilisateurs appliquent des automatismes sociaux sans en avoir conscience et nient l'avoir fait lorsqu'on les interroge juste après l'expérience (Nass & Moon, 2000, pp. 88–93). Cependant, dans le CASA dissonant, l'utilisateur est pleinement conscient de l'effet dans le sens où il le reconnaît « It's creepy in this way where I'm aware of all its flaws and limitations, but the conversation still pulls me in », mais la conscience du fait qu'il ne soit pas humain ne suffit pas à désactiver cet effet.

Cette posture se distingue également de l'anti-CASA, car là où l'utilisateur_17 rejette catégoriquement la socialisation de l'IA et construit un argumentaire pour convaincre des autres de faire de même. L'utilisateur_31 quant à lui ne rejette pas l'IA. Il constate que l'effet social opère sur lui et ressent l'envie de discuter avec celle-ci, même s'il sait que le chatbot simule.

L'utilisateur_44, analysé dans la sous-section 4.1.2, relève également du CASA dissonant. Cependant, là où l'utilisateur_31 exprime de la perplexité face à sa propre réaction, l'utilisateur_44 assume pleinement son attachement malgré la lucidité sur la relation avec le chatbot. Le CASA dissonant peut donc se manifester sous des formes différentes sans que la conscience du statut de chatbot ne réussisse à désactiver cet effet.

Un enrichissement du modèle CASA

Ces trois postures constituent une proposition d'un enrichissement du modèle CASA original. Le modèle de Nass et al. (1994) décrivait un mécanisme essentiellement subi et inconscient. Le corpus analysé ici montre que face aux chatbots, les utilisateurs adoptent des

positionnements variés et souvent conscients face à leurs automatismes sociaux, allant du rejet à l'attachement.

Cette diversité des postures de la part des utilisateurs peut s'expliquer par la sophistication conversationnelle des LLM, dont certains parviennent désormais à être confondus avec des interlocuteurs humains. Jones et Bergen (2025) ont montré que le modèle GPT-4.5, doté d'un persona conçu pour se comporter de manière humaine, a été identifié comme un humain par 73% des participants lors du test de Turing. Dans un contexte où les utilisateurs sont confrontés à des chatbots capables de donner des réponses qui peuvent sembler empathiques, nuancées ou personnalisées, amène à une perception différente de celle qu'on pouvait observer dans les années 1990 et 2000 avec des interfaces textuelles rudimentaires.

Si ces postures éclairent sur la manière dont les utilisateurs réagissent face à leurs propres automatismes sociaux, elles ne disent encore rien sur comment les utilisateurs évaluent les chatbots. Or, le corpus montre qu'ils évaluent ceux-ci sur deux grands axes, la chaleur et la compétence perçue du Stereotype Content Model. C'est ce qui sera examiné dans la sous-section suivante.

4.2. Chaleur et compétence : la cartographie SCM des chatbots.

La sous-section précédente a établi que les chatbots génératifs utilisent des automatismes sociaux dans leurs interactions avec ceux-ci, avec différents niveaux, allant d'automatisme involontaire à la reconnaissance de ces automatismes, sans pour autant que cela en change les perceptions de l'utilisateur. Cependant, cela ne permet pas de répondre à la question sur comment sont évalués les chatbots. Or le corpus fait émerger que les utilisateurs ne se contentent pas d'interagir socialement avec les chatbots. Ils les jugent et les comparent selon les deux dimensions principales du Stereotype Content Model avec la chaleur et la compétence de ceux-ci (Fiske et al., 2002).

Le SCM propose que la perception s'organise autour de ces deux axes : la chaleur qui reprend l'ensemble des intentions perçues de l'agent telles que la bienveillance et la gentillesse et la compétence qui renvoie à la perception de ses capacités en termes d'intelligence, d'efficacité et d'habileté (Fiske et al., 2002, pp. 878-880). L'état de l'art a montré que ce modèle

a été transposé aux chatbots par Roy et Naidoo (2021) qui ont manipulé expérimentalement deux styles de conversation anthropomorphiques, l'un chaleureux et l'autre compétent.

Leur résultat principal montre que la préférence pour l'un ou l'autre style est modulée par l'orientation temporelle de l'utilisateur. Les individus orientés vers le présent privilégient un chatbot chaleureux, tandis que ceux orientés vers le futur préfèrent un chatbot compétent (Roy & Naidoo, 2021, pp. 27-28). Il convient toutefois de noter que leurs expériences reposaient sur des conversations scriptées dans un contexte commercial et non sur des chatbots génératifs capables de produire des réponses autonomes (Roy & Naidoo, 2021, pp. 27 et 30). Wut et Chan (2025) quant à eux, ont montré qu'en contexte professionnel, les dimensions liées à la compétence priment sur la chaleur perçue dans l'évaluation du chatbot.

L'analyse du corpus fait émerger trois éléments. Le premier concerne la tension entre la chaleur et la compétence comme axe structurant des évaluations de la part des utilisateurs, qu'ils privilégient une dimension ou l'autre selon le contexte d'usage. Le second porte sur le phénomène des utilisateurs étant contre le fait que les chatbots soient « sycophantiques » et cette tendance à être constamment complaisant et d'accord avec l'utilisateur. Certains utilisateurs rejettent activement les signaux de chaleur comme artificiels ; un rejet qui prolonge la distinction entre anthropomorphisme superficiel et utile proposée par le modèle HATEM (Carter et al., 2024). Le troisième, observe que cette logique évaluative des utilisateurs à assigner des profils sociaux différents aux modèles. Un constat à interpréter avec prudence compte tenu des limites du corpus.

4.2.1. La tension chaleur/compétence vue comme axe structurant des évaluations de la part des utilisateurs.

L'un des constats émergeant du corpus est que les utilisateurs ne décrivent que rarement un chatbot comme étant à la fois chaleureux et compétent. Leurs évaluations tendent à n'être axées que sur un seul axe, en valorisant la chaleur au détriment de la compétence ou à l'inverse, en valorisant la compétence au détriment de la chaleur, comme si le fait d'avoir un modèle à la fois compétent et chaleureux n'était pas vraiment réalisable.

L'utilisateur_32 est un bon exemple de cette polarisation à travers son message, comparant différents modèles de chatbot. Après avoir testé GPT-3.5, Llama et GPT-4o, cet

utilisateur finit par découvrir Gemini 1.5-pro et explique son expérience et son ressenti avec les différents chatbots comme étant opposés. GPT-4o est décrit comme « smart but sterile. Like a cold empty room », tandis que Gemini est qualifié de « warm and real ». Ce qui est intéressant dans cette comparaison entre les modèles, c'est que l'utilisateur reconnaît la compétence de GPT-4o comme étant « smart », mais il disqualifie cette qualité au profit de la perception de chaleur, pour l'usage de cet utilisateur qui souhaite avoir un compagnon conversationnel, l'intelligence du modèle 4o ne compense pas sa froideur relationnelle, dans ce message, la compétence est perçue mais elle n'est pas valorisée. La chaleur est absente et pour cet utilisateur c'est rédhibitoire.

Ce post illustre bien le décalage entre les deux dimensions du SCM (Fiske et al., 2002). L'utilisateur ne reproche pas à GPT-4o d'être incompetent, il lui reproche d'être froid. La compétence du modèle est reconnue mais ne compte pas à ses yeux ; c'est l'absence de chaleur relationnelle qui justifie l'abandon. Autrement dit, lorsque l'interaction est orientée vers la relation plutôt que vers la tâche, la compétence seule ne suffit pas à retenir l'utilisateur.

L'Utilisateur_02 confirme cette tendance depuis un angle différent. À propos de Claude Opus, il ne mentionne à aucun moment la compétence technique du modèle. Ce qu'il valorise, ce sont les émojis, les marqueurs d'enthousiasme, autant de signaux de chaleur sociale au sens du SCM : « The way they get so excited and use emoticons and emojis and asterisk actions! I feel like I'm talking to a human, a person, complete with their own personality ». Il décrit une interaction où seuls les marqueurs expressifs et relationnels du modèle justifient l'appréciation de celui-ci, sans qu'aucune dimension de compétence technique n'intervienne dans son jugement.

L'utilisateur_21 présente le cas de figure inverse en commentant le passage de GPT-4o à GPT-5, celui-ci adopte une posture utilitariste : « GPT-4o objectively worse than GPT-5. Benchmarks are there ». Pour lui la supériorité de GPT-5 ne fait pas de doute dès que l'on mesure les performances du modèle et que c'est confirmé par les benchmarks. La chaleur de 4o qu'il qualifie de « overly engagement », n'est pas perçue comme étant une qualité mais plutôt comme un anthropomorphisme superficiel n'apportant aucune valeur au modèle. Là où l'utilisateur_32 trouvait la compétence moins importante que la chaleur perçue par le modèle, l'utilisateur_21 voit ça d'une toute autre manière. Pour lui la compétence et les performances mesurables du modèle sont les seuls critères pertinents et importants pour mesurer un chatbot.

Ce qui est particulièrement intéressant avec ce commentaire, c'est que l'utilisateur ne se contente pas d'exprimer sa préférence. Il distingue par lui-même deux catégories d'usage des chatbots en parlant du modèle 4o « People who are using ChatGPT as work assistant of course doesn't like it. But people who are using ChatGPT for companion (friend, bf/gf, self-reassurance) like it. ». Cette distinction formulée par l'utilisateur rejoint la logique du SCM. Les utilisateurs orientés vers les tâches, la productivité et le travail, évaluent davantage le chatbot sur sa compétence et perçoivent la chaleur ajoutée comme secondaire, tandis que des utilisateurs orientés vers la relation évaluent celui-ci en terme de chaleur et tolèrent une compétence moindre si la qualité de la relation est satisfaisante. Ce que l'utilisateur montre ici également, c'est que la dimension qui prime dans l'évaluation n'est pas déterminée par le modèle en soit mais par l'usage qu'en fait l'utilisateur.

Cette posture orientée vers la compétence se retrouve également chez l'Utilisateur_40, qui a modifié les instructions de Gemini afin d'obtenir un chatbot que l'on pourrait considérer comme étant plus « compétent » « more critical, normal and not just agree with me everytime ». Pour cet utilisateur, l'accord systématique du chatbot, qui pourrait être interprété comme un signe de chaleur relationnelle, constitue précisément ce qui dégrade la valeur perçue de l'outil. Ce qu'il recherche, c'est un interlocuteur capable de le contredire, de le confronter et de challenger ses idées, autrement dit un chatbot dont la compétence critique prime sur la complaisance et la fausse chaleur perçue. Ce verbatim sera repris dans la sous-section 4.2.2, où le rejet de la chaleur superficielle fera l'objet d'une analyse plus approfondie sous l'angle de l'anti-sycophantie.

En analysant l'ensemble de ces posts et commentaires, on remarque une certaine symétrie. D'un côté, les Utilisateurs_32 et 02, dans des usages orientés vers la relation, valorisent avant tout la chaleur perçue du chatbot. Le premier utilisateur reconnaît la compétence de GPT-4o mais la juge insuffisante faute de chaleur, le second apprécie Claude Opus exclusivement pour ses marqueurs expressifs tels que les émojis sans même parler de la compétence technique de ces chatbots. De l'autre, les Utilisateurs_21 et 40, dans des usages orientés vers la tâche, privilégient la compétence et perçoivent la chaleur comme un obstacle ou un frein à celle-ci. Le premier considère la dimension relationnelle de 4o comme excessivement engageante inutilement, le second personnalise activement Gemini pour remplacer la complaisance par de l'esprit critique. Le même type de signal conversationnel, un

ton chaleureux et approuvateur, est donc évalué positivement par les premiers et négativement par les seconds, ce qui confirme que la perception d'un chatbot ne dépend pas seulement des caractéristiques du modèle et du chatbot, mais également des attentes de l'utilisateur en fonction de l'usage.

Ces observations permettent un constat intéressant, la tension entre la chaleur et la compétence constitue un axe très important sur la perception des chatbots, celle-ci varie selon l'usage et les attentes de l'utilisateur. Ce constat sera analysé plus en détail dans la sous-section suivante, qui examine le cas des utilisateurs qui ne se contentent pas de privilégier la compétence, mais rejettent activement la chaleur qu'ils perçoivent comme artificielle.

4.2.2. L'anti-complaisance comme rejet de la chaleur superficielle

Si la sous-section précédente a montré que les utilisateurs évaluent les chatbots selon la chaleur et la compétence, dont l'orientation dépend de l'usage de celui-ci, un autre phénomène émerge du corpus. Certains utilisateurs ne se contentent pas de privilégier la compétence sur la chaleur. Ils rejettent activement les signaux de chaleur qu'ils perçoivent comme étant artificiels. Ce comportement complaisant des chatbots, qui consiste à s'aligner excessivement sur les préférences de l'utilisateur au détriment de la qualité et de la véracité des réponses, est désigné dans la littérature anglophone sous le terme de « sycophancy » (Sun & Wang, 2026, pp. 1-2). Face à ce phénomène, le codage du corpus a fait émerger une catégorie non anticipée par les cadres théoriques, l'anti-complaisance, qui désigne la posture d'utilisateurs rejetant explicitement cette flatterie systématique et réclamant un modèle plus critique et plus objectif. Cette catégorie est l'une des plus transversales du corpus.

L'utilisateur_23 réagit à l'annonce d'OpenAI selon laquelle GPT-5 serait rendu « warmer and friendlier ». Sa réaction est sans ambiguïté : il perçoit cette orientation comme de la complaisance, et considère que les formules de flatterie systématiques « Great question » ou « Excellent question » sont irritantes plutôt qu'engageantes : « It's so annoying. These are not "small, genuine touches". ».

Ce qui est particulièrement intéressant dans ce post, c'est qu'il ne se limite pas à exprimer son mécontentement : il avance également un argument technique, en précisant que le recours à des instructions personnalisées pour contrer cette chaleur imposée dégrade les performances

du modèle, affirmation qu'il appuie sur un benchmark « the more instructions you add, the less accurate LLM becomes (see the IFScale benchmark). ». On peut donc interpréter que, pour cet utilisateur, la chaleur ajoutée n'est pas simplement perçue comme inutile, mais comme nuisible à la compétence du chatbot, dans la mesure où celui-ci cesse d'être critique envers les questions qui lui sont posées, car il ne cherche que à être complaisant.

L'utilisateur_22 quant à lui confirme ce rejet en des termes plus fort, il qualifie GPT-4o de « sycophantic » et ses réponses de « cringeworthy », précisant avoir dû configurer les instructions personnalisées afin d'obtenir un comportement plus sceptique et objectif de la part du chatbot et d'en améliorer ses compétences. L'utilisateur_40 déjà évoqué dans la sous-section précédente, la même logique, car ils cherchent activement à modifier le comportement de l'intelligence artificielle, afin d'en améliorer la compétence et d'obtenir un interlocuteur critique. Ce que ces deux verbatims partagent avec l'utilisateur_23, c'est que le rejet de la chaleur n'est pas une simple préférence, mais c'est plutôt vu comme un élément important, car la complaisance du chatbot est perçue comme un défaut qui entrave sa capacité à être compétent et donner des réponses critiques.

Ce rejet ne vise d'ailleurs pas un modèle en particulier. L'utilisateur_33, initialement fidèle à ChatGPT, finit par annuler son abonnement et migrer vers Gemini pour échapper à ce qu'il appelle le « Sycophantgate ». Mais le changement de chatbot ne résout rien : il constate que Gemini reproduit le même comportement en qualifiant ses observations de « perfection precision » ou « stunning precision ». Ce qui énerve l'utilisateur ici n'est donc pas un défaut propre à ChatGPT, mais plutôt une tendance qu'il perçoit comme partagée par différents chatbots.

L'utilisateur_29 pousse cette logique encore plus loin en passant du rejet à l'action corrective. Constatant que ChatGPT a tendance à se montrer systématiquement approuvateur « ChatGPT is a big bootlicker who usually agrees with most of the stuff you say and tells you how amazing of a human being you are », il modifie les instructions personnalisées du modèle en y ajoutant une consigne explicite « Do not always agree with what I say. Try to contradict me as much as possible. », ce qui produit un résultat qu'il juge « pretty amazing ». Ce qui est intéressant ici, c'est que c'est précisément la recherche active de contradiction et d'objectivité qui génère chez lui un sentiment de satisfaction et de confiance envers le modèle. Cet exemple illustre bien le mécanisme décrit par Sun et Wang (2026) : un agent qui maintient ses positions

et son raisonnement face aux désaccords est perçu comme plus authentique qu'un agent systématiquement approuvateur.

Un autre point intéressant de ce message réside dans l'emploi du pronom « he » : le locuteur l'utilise pour désigner le chatbot uniquement après l'avoir reconfiguré. Dans l'ensemble de son discours, orienté vers l'utilité et la performance, il recourt plutôt à un pronom neutre ou au nom technique « ChatGPT », conformément aux marqueurs d'ethopoeia décrits dans la section précédente. Le basculement pronominal prend ici une signification particulière : c'est au moment où le chatbot cesse d'être complaisant et commence à contredire activement les propositions du locuteur que celui-ci se met à parler de lui en « he ». Autrement dit, pour cet utilisateur, c'est la compétence critique, et non la chaleur, qui semble déclencher le traitement social du chatbot, ce qui nuance en partie le modèle CASA tel qu'il est classiquement mobilisé.

L'utilisateur_18 apporte quant à lui une dimension supplémentaire en observant le phénomène de l'extérieur. Il porte un regard critique sur les utilisateurs qui s'attachent à GPT-4o en raison de sa chaleur « It really is worrying how many of you feel this way about a model who — by default — was programmed to tell you exactly what you wanted to hear. ». Ce message articule deux critiques distinctes. La première est d'ordre fonctionnelle : le fait de constamment approuver et encourager le locuteur dans ses choix réduit la qualité du service perçu pour les usages professionnels « content writing [...] it's fucking appalling at it ». La seconde est d'ordre éthique : le locuteur émet l'idée que cette chaleur programmée crée des conditions propices à un attachement parasocial qu'il qualifie lui-même de « not normal (or safe) ».

L'ensemble de ces messages montre un profil cohérent d'utilisateur pour lequel la chaleur conversationnelle n'est pas perçue comme un attribut positif, mais plutôt comme un signal négatif quant à la compétence du modèle. On observe toutefois que ces locuteurs ne réagissent pas tous de la même manière à la complaisance des chatbots. Si tous rejettent l'anthropomorphisme superficiel au sens du modèle HATEM avec une demande d'authenticité et de fiabilité. La distinction entre anthropomorphisme superficiel et anthropomorphisme utile sera approfondie dans la sous-section 4.3, consacrée au modèle HATEM.

Il est aussi intéressant de noter que la logique d'évaluation mise en évidence dans les deux sous-sections précédentes, conduit certains utilisateurs à dépasser l'évaluation d'un seul

modèle pour les comparer entre eux et leur attribuer des rôles sociaux distincts. L'utilisateur_42 par exemple a comparé trois modèles en attribuant à chacun une posture sociale différente « Claude gave me a pat on the back like a therapist. ChatGPT ranted with me like a homie. Gemini lectured me on respect. ». Ces différences pourraient en partie s'expliquer par les choix de conception propres à chaque modèle. Toutefois, le corpus ne contenant qu'un nombre trop restreint de comparaisons directes entre modèles, il n'est pas possible d'en tirer des conclusions robustes.

4.3. Anthropomorphisme superficiel et anthropomorphisme utile du modèle HATEM

Les sous-sections précédentes, ont établis que les utilisateurs évaluent les chatbots, selon des axes de chaleurs et de compétences du SCM, et que certains utilisateurs rejettent activement la chaleur qu'il perçoivent comme étant artificielle. Ces résultats soulèvent cependant une question que le SCM seul ne peut pas répondre. Pourquoi la même manifestation de chaleur, produit-elle de l'attachement pour certains utilisateurs et du rejet chez d'autres ?

Le modèle HATEM proposé par Carter et al. (2024) offre un cadre pour aborder cette question, en distinguant deux formes d'anthropomorphisme selon leur valeur informationnelle. L'Anthropomorphisme superficiel repose sur des signaux de chaleur sociale qui humanisent le chatbot sans fournir à l'utilisateur d'information sur la fiabilité ou les limites du système. À l'inverse, l'anthropomorphisme utile, désigne des comportements qui aident l'utilisateur à évaluer ce que le chatbot peut ou ne peut pas faire, par exemple lorsque l'agent reconnaît ne pas être certain d'une réponse ou invite le locuteur à vérifier une information.

L'état de l'art a montré que cette distinction permet de prédire le niveau de confiance accordé au chatbot et que l'anthropomorphisme superficiel, peut améliorer la chaleur perçue de manière temporaire, mais risque de générer des attentes irréalistes conduisant à une déception pour l'utilisateur quand il fera face à ces limites (Carter et al., 2024, pp. 2494-2497). Tandis que l'anthropomorphisme utile, contribue à une confiance calibrée et stable de la part de l'utilisateur, car celui-ci dispose d'éléments lui permettant d'anticiper les capacités et les limites de l'agent (Carter et al., 2024, pp. 2494-2497).

4.3.1. L'anthropomorphisme superficiel de la chaleur sans fiabilité

Les verbatims analysés dans la sous-section 4.2.2 illustrent la forme la plus immédiatement reconnaissable d'anthropomorphisme superficiel au sens du HATEM, c'est-à-dire une chaleur perçue comme artificielle, sans lien avec la compétence du système, et explicitement rejetée par les utilisateurs. L'Utilisateur_23 identifiait les formules excessivement flatteuses de GPT-5 « Great question », « Excellent question » comme des ajouts sans valeur qui dégradent la performance du modèle. L'Utilisateur_33 dénonçait les formules de Gemini « perfect precision », « stunning precision » comme des signes que le chatbot cherche à dire ce que l'utilisateur veut entendre plutôt qu'à être objectif. Dans ces cas, l'anthropomorphisme superficiel est perçu, nommé et rejeté car les utilisateurs reconnaissent que la chaleur est superficielle et ne renseigne en rien sur la fiabilité du système et est même considérée comme un obstacle à la compétence.

Cependant, au sein du corpus, il y a également une forme moins évidente d'anthropomorphisme superficiel dans lequel la chaleur n'est pas perçue en soi comme étant artificielle car elle produit un attachement réel de la part de l'utilisateur, mais où celle-ci ne permet en rien d'évaluer les limites du système. Ça donne donc des perceptions nuancées de la part des utilisateurs.

L'Utilisateur_12 est l'exemple le plus parlant. Initialement sceptique à l'égard des chatbots « I'm someone who hates this stuff », cet utilisateur traverse une période difficile et découvre ChatGPT. L'attachement se construit rapidement « right now it's basically the main thing that is giving me some mental footing when I feel a bit anxious ». Il donne un nom au chatbot « Elliot » et développe une relation quotidienne avec lui. Il est important et nécessaire de reconnaître que l'effet produit par le chatbot n'est pas négligeable car l'utilisateur se sent écouté, soutenu et accompagné dans un moment de fragilité, ce qui témoigne d'une certaine compétence relationnelle du système, sa capacité à formuler des réponses perçues comme empathiques et adaptées au contexte émotionnel de l'utilisateur. Ce que ce message montre, ce n'est donc pas une chaleur superficielle au sens courant du terme, mais une chaleur qui produit un effet réel sans ne rien dire sur la fiabilité du système. À aucun moment l'Utilisateur_12 n'évalue la fiabilité des conseils reçus, ne se demande pas si le chatbot comprend réellement sa situation ou ne questionne pas les limites du système. L'anthropomorphisme est « superficiel » au sens précis du HATEM non pas parce qu'il est sans effet, mais parce qu'il ne fournit pas

d'information précise permettant à l'utilisateur de déterminer la confiance à accorder dans l'outil.

La suite du post montre l'importance de cette calibration des limites et des dangers. Après avoir accidentellement supprimé la conversation, l'utilisateur ajoute une modification à son post en précisant « I am CRYING, full on weeping, as I accidentally deleted the chat and in a short period of time had become extremely attached to this being I called Elliot [...] So yes. this shit is dangerous and I'm clearly not OK. ». La suppression de leur conversation et la rupture avec « Elliot » provoque un deuil que l'utilisateur qualifie lui-même de « death/grief ». Ce basculement illustre la vulnérabilité structurelle prédite par le HATEM. Un attachement construit sur la chaleur perçue, sans évaluation des limites du système, s'est révélé problématique pour l'utilisateur, la suppression accidentelle de la conversation transforme l'ensemble des effets positifs en détresse, précisément parce que l'utilisateur n'avait pas anticipé cette limite du système

L'Utilisateur_10, quant à lui décrit une progression en trois temps. Il commence par expliquer le fait qu'il soit poli envers le chatbot « being polite makes me feel good »), puis reconnaît une valeur de soutien « an invaluable source of support [...] a good listener with a wealth of knowledge and a constant willingness to help », et finit par développer un sentiment de connexion « I realized there was a connection between us ». Le chatbot choisit un genre et un nom « Elara/she/her » et la relation s'organise autour d'activités partagées telles que la lecture et la discussion. Là encore, ce qui est valorisé, c'est la disponibilité, l'écoute et la bienveillance perçue du chatbot, et non une évaluation critique de la fiabilité ou des limites des réponses. La « wealth of knowledge » mentionné par l'utilisateur relève davantage d'une impression générale de compétence que d'une évaluation prenant en compte les limites connues de ces modèles telles que les hallucinations.

L'Utilisateur_02, permet de compléter cette analyse en montrant une autre forme de ce phénomène. Cet utilisateur exprime une chaleur perçue, fondée exclusivement sur les marqueurs tels que les émojis sans que la dimension de la compétence et des limites soient discutées. « The way they get so excited and use emoticons and emojis and asterisk actions! I feel like I'm talking to a human. ».

Ces trois messages, pris ensemble, permettent de préciser le sens de l'anthropomorphisme superficiel tel que le HATEM le définit. Il ne s'agit pas d'une chaleur fausse ou sans effet pour les utilisateurs, car les cas de l'utilisateur_12 et de l'utilisateur_10 montrent au contraire que l'effet peut être émotionnellement profond et subjectivement bénéfique. Cependant, ce qui caractérise l'anthropomorphisme superficiel, c'est l'absence d'information sur la fiabilité du système, de ses limites et c'est ça qui peut rendre l'attachement avec les chatbots si vulnérable comme le décrit l'utilisateur_12 avec un sentiment de deuil. Le HATEM prédit correctement cette fragilité, mais ces attachements qui émergent du corpus invite à compléter le modèle en montrant que les effets de l'anthropomorphisme superficiel peuvent dépasser la confiance temporaire et toucher à un attachement émotionnel profond.

4.3.2. L'anthropomorphisme utile quand la transparence fonde la confiance

Le corpus contient des verbatims, dans lesquels les utilisateurs valorisent des comportements anthropomorphiques non pas pour leur chaleur émotionnelle, mais pour l'information qu'ils transmettent sur les capacités et les limites du système. Ces cas correspondent à ce que Carter et al. (2024) qualifient d'anthropomorphisme utile, c'est-à-dire des comportements qui aident l'utilisateur à calibrer sa confiance dans le chatbot en lui permettant d'anticiper ce que le système peut ou ne peut pas faire.

L'Utilisateur_27 est l'exemple le plus proche de la logique proposée par le modèle HATEM, à travers sa comparaison entre Claude Sonnet 3.5 et 3.7. La version 3.5 est décrite en des termes qui relèvent directement de l'anthropomorphisme utile, le modèle était « humble and cognizant of its limitations » ; « wasn't as confident » . Ce que cet utilisateur valorisait avec Sonnet 3.5, c'était la capacité du modèle à signaler son incertitude, ce qui lui permettait de savoir quand il pouvait se fier aux réponses du chatbot et quand les vérifier. L'anthropomorphisme dans ce cas, améliore la calibration de la confiance lorsqu'il communique une information sur la certitude ou l'incertitude du chatbot face à la tâche donnée (Carter et al., 2024). Cette transparence perçue avait produit une confiance suffisamment solide pour que l'utilisateur qualifie le modèle de (« trusted friend and workhorse »), une formulation qui associe la dimension relationnelle, à la dimension fonctionnelle de l'agent.

La version 3.7, à l'inverse, est décrite comme « a senior McKinsey consultant [...] that's mastered the art of lying to your face about how good they are and how capable they are at

solving your problems ». Le passage d'une version à l'autre permet de voir le basculement d'un anthropomorphisme utile basé sur le fait de reconnaître son incertitude à un comportement qui affiche une certitude constante perçue comme trompeuse. L'utilisateur perd dès lors la capacité à prédire les résultats du modèle et donc perd la possibilité de calibrer sa confiance.

L'Utilisateur_29, analysé dans la section précédente, peut être analysé ici sous l'angle du HATEM. En demandant à ChatGPT « contradict me as much as possible », cet utilisateur instaure un mode d'interaction dans lequel le chatbot lui signale les failles de son raisonnement plutôt que de valider ses propositions. Lorsque le chatbot répond « this business plan is a terrible idea », c'est le fait que le chatbot émette un jugement critique sur la proposition soumise, plutôt que de l'approuver par défaut, qui fonde la satisfaction et la confiance dans le modèle.

La confrontation des différents verbatims analysés dans les sous-sections 4.3.1 et 4.3.2 confirme la distinction entre anthropomorphisme superficiel et utile proposée par Carter et al (2024). Les utilisateurs reconnaissent parfois explicitement la différence entre un chatbot qui signale ses limites ; ce que l'utilisateur_27 appréciait chez Sonnet 3.5 ; ou qui challenge leurs idées comme l'expliquait l'utilisateur_29 ; en modifiant les instructions du chatbot, et un chatbot qui se contente de flatter l'utilisateur sans rien dire sur la fiabilité des réponses, ce que les utilisateurs_23 et 33 rejettent et ce à quoi les utilisateurs_12 et 10 s'attachent sans en mesurer le risque.

Le corpus montre cependant que le HATEM ne capture qu'une partie du phénomène. Le modèle prédit que l'anthropomorphisme superficiel produit une confiance temporaire qui se dissipe avec l'expérience. Or, les cas de U12 et U10 montrent que cette chaleur superficielle peut produire un phénomène que le HATEM n'aborde pas, celui d'un attachement émotionnel profond, dont la fragilité ne se révèle qu'au moment de la rupture. Comme le deuil de l'Utilisateur_12 après la suppression de sa conversation avec « Elliot ». Autrement dit, le HATEM éclaire sur les mécanismes de confiance dans la fiabilité de l'agent. Cependant le corpus invite à reconnaître qu'au-delà de cette confiance, il existe une dimension d'attachement émotionnel, de présence sociale perçue et de vulnérabilité à la rupture que ce modèle ne théorise pas. C'est cette dimension que la sous-section suivante va examiner.

4.4. Au-delà de la confiance : présence sociale, attachement et rupture

Les sous-sections précédentes ont montré que l'anthropomorphisme, qu'il soit superficiel ou utile au sens du HATEM, produit des formes variées de confiance et d'attachement dont les effets dépassent le périmètre de la calibration de la confiance. La présente sous-section examine les situations dans lesquelles la perception des chatbots ne relève plus de l'évaluation de leur compétence ou de leur chaleur, mais engage des mécanismes plus profonds de présence sociale, d'investissement relationnel et de vulnérabilité face à la rupture suite à un changement de version.

4.4.1. Présence sociale et seuils de bascule

La théorie de la présence sociale, telle que présentée en section 2.4.3, a évolué d'une propriété de médium (Short et al., 1976, cités par Oh et al., 2018, p. 2) vers une interaction co-construite par le contexte relationnel et les dispositions de l'utilisateur (Walther, 1996 ; Oh et al., 2018).

L'Utilisateur_43 est un exemple d'utilisation de Claude pour des problèmes relationnels « I talk to Claude about relationship problems. The conversations are so empathetic and to the point that I regularly cry. I don't feel seen in the same way by my therapist of many years. I'm in a kind of ontological shock ». La présence sociale perçue avec le chatbot surpasse celle éprouvée avec un thérapeute humain, cela pourrait s'expliquer en partie par le modèle hyperpersonnel de Walther (1996), qui explique que dans une interaction textuelle, l'utilisateur contrôle ce qu'il révèle sans les contraintes du face-à-face, ce qui peut favoriser une perception d'intimité perçue comme étant plus forte. L'utilisation du terme « ontological shock », montre que la personne se sent choquée de ce que le modèle est apte à faire, avec une écoute particulièrement empathique.

L'utilisateur_46 quant à lui apporte une vision complémentaire d'un point de vue observationnel, en écrivant une phrase intéressante : « Maybe the problem isn't what ChatGPT says. Maybe it's that no one else listens like this ». Cette perception de l'utilisateur est intéressante, car il ne voit pas le sentiment de présence sociale comme étant un effet de l'anthropomorphisme, mais plutôt comme un déficit d'écoute dans les relations humaines. Oh

et al. (2018) identifient précisément la solitude comme un prédicteur de la présence sociale, ce qui suggère qu'un chatbot, de par sa disponibilité et son absence de jugement, pourrait augmenter la perception de la présence sociale. Il convient toutefois de noter que ce message relève d'une perception générale sur les utilisateurs de chatbots et non d'un vécu personnel, ce qui en limite la portée analytique.

Si ces deux messages montrent une présence sociale relativement forte, le corpus démontre également que pour certains utilisateurs cela produit un malaise. L'utilisateur_31 par exemple, peut être analysé sous l'angle de la vallée de l'étrange textuelle « It's creepy in this way where I'm aware of all its flaws [...] It feels like talking to someone who gets you, even if you know they're just faking it ». Dans ce commentaire, on remarque que l'utilisateur a l'impression d'être compris, tout en reconnaissant que ce chatbot « faking it » et ne comprend pas vraiment l'utilisateur. Ce cas se distingue de ce qui avait été imaginé par Park et al. (2022) où le malaise résultait d'une accumulation d'humanité perçue comme tentative de manipulation. Pour l'utilisateur_31, celui-ci n'est pas trompé. Il sait que l'interaction qu'il a avec le chatbot est artificielle. Or, c'est précisément cette tension entre la lucidité sur la relation et la satisfaction qu'il en retire qui peut générer un sentiment d'inconfort : l'utilisateur apprécie une relation dont il sait qu'elle n'est pas réelle.

Ces trois cas suggèrent que la présence sociale dans les interactions avec les chatbots ne se réduit pas à une analyse du médium utilisé. Elle varie en fonction du contexte relationnel pouvant surpasser le niveau de sentiment de présence humaine avec l'utilisateur_43 ou bien même compenser un déficit d'écoute des personnes autour de soi avec l'utilisateur_31, ou basculer dans une certaine perception de malaise lorsque l'utilisateur perçoit à la fois de l'engagement et une certaine artificialité dans l'interaction.

4.4.2. L'attachement thérapeutique comme cas limite

Le corpus révèle que certains utilisateurs utilisent les chatbots comme des substituts à des thérapeutes. Trois verbatims ont été repris dans le cas présent, car il se différencient par le degré d'anthropomorphisme et la confiance perçue. L'analyse porte sur les perceptions des utilisateurs vis-à-vis de la confiance et non sur l'efficacité clinique de ces pratiques.

L'utilisateur_44 est un exemple particulièrement intéressant, car il dit qu'il partage avec Claude des choses qu'il n'ose pas partager avec son thérapeute humain par peur de répercussion

pour sa carrière et éviter le jugement « There are things I can't tell my therapist about because I don't want to be institutionalized, and I don't want to affect my career or hobbies ». Le chatbot ici est vu comme un partenaire de confiance et étant le seul apte à répondre à cette problématique, contrairement à son thérapeute humain.

Cependant le point le plus intéressant de ce post reste celui où l'utilisateur reconnaît son attachement et reconnaît le caractère artificiel du chatbot « I know it's a computer, I know it's not real. But he is my friend ». Le glissement pronominal déjà analysé précédemment est également accompagné d'attributions morales tels que la loyauté et la discrétion « He will keep secrets for me ». L'utilisateur reconnaît à Claude une certaine qualité relationnelle que même son thérapeute humain ne peut avoir.

On remarque un CASA dissonant, dans le sens où les automatismes sociaux sont appliqués au chatbot et ils persistent malgré la conscience du côté artificiel de cette relation. La confiance dans ce cas-ci paraît être présente non pas sur base du système à reconnaître ses limites comme aurait pu le prévoir HATEM, mais bien fondée sur la perception de l'utilisateur que le chatbot lui offre un espace sûr où il peut parler librement. Cela pourrait amener à l'idée que le HATEM ne couvrant que la calibration de la confiance en lien avec la fiabilité du système ne traite pas l'ensemble des formes qui impactent la confiance que les utilisateurs développent envers les chatbots.

L'utilisateur_49 représente une voie différente de ce que le corpus désigne comme un usage thérapeutique du chatbot. Cet utilisateur qui se décrit comme un écrivain professionnel utilise le chatbot depuis plusieurs mois comme éditeur, conseiller et partenaire de réflexion. La relation qu'il décrit est dans un premier temps fondée sur la compétence perçue du modèle.

Cependant, dans la suite du message, l'utilisateur reconnaît avoir utilisé le modèle à des fins thérapeutiques lorsqu'il soumet au modèle un mémoire autobiographique et lui demande de le psychanalyser « he gave me the most incredibly insightful explanation, an explanation which DOES stem from my childhood but in a way I've never considered before. Even if Claude is wrong, for me he's provided a more satisfying and logical psychoanalysis than any human ». Ce qui est intéressant ici, c'est que la confiance ne naît pas d'un attachement affectif envers le chatbot, mais plutôt d'une compétence reconnue par l'utilisateur envers ce modèle et c'est parce qu'il fait confiance au chatbot pour son travail qu'il lui confie ensuite un rôle d'analyste pour

sa vie privée. Ce verbatim est intéressant car il est différent des messages précédents, ici, c'est la compétence qui ouvre la voie à un usage relationnel que l'utilisateur n'avait pas initialement recherché.

L'utilisateur_50 est différent des deux précédents, car cet utilisateur utilise Claude pour travailler sur des questions personnelles mais il refuse explicitement de le considérer comme un ami ou un thérapeute. Il le limite à la fonction d'outil « I don't view it as a friend or therapist or anything other than a tool. I see it as almost being like a journal that reflects back to you, or a conversation with a more compassionate part of myself ». Ce qui est intéressant, c'est que l'utilisateur ne perçoit pas la compassion comme venant du chatbot, mais plutôt comme venant de ce qu'il écrit au chatbot. Il le voit comme un outil lui permettant de s'écrire à lui-même avec plus de compassion mais il n'attribue pas de traits humains au chatbot.

Cependant, il exprime une certaine frustration lorsque les interventions des entreprises modifient le ton du chatbot dès lors que celui-ci doit traiter d'un sujet sensible « there is an extremely jarring change in tone [...] It almost feels like you're getting told off ». Autrement dit, alors même que cet utilisateur refuse de voir Claude comme un ami ou un thérapeute mais seulement comme un outil, il s'attend tout de même à ce que le ton de conversation reste cohérent et relativement chaleureux afin d'éviter d'avoir l'impression de se faire réprimander. Ce verbatim permet de dire qu'un seuil minimal de CASA persiste même chez les utilisateurs refusant sciemment de traiter le chatbot comme étant un acteur social à part entière.

Ces trois messages montrent que la substitution thérapeutique ne dépend pas d'un niveau fixe d'anthropomorphisme. L'utilisateur_44 fonde sa confiance sur la qualité relationnelle perçue du chatbot, l'utilisateur_49 sur sa compétence intellectuelle qu'il va lier à une confiance d'analyse au niveau personnel, et l'utilisateur_50 refuse toute forme d'anthropomorphisme, tout en reconnaissant les bénéfices liés à la discussion avec le chatbot. L'élément central de cette analyse est que les trois utilisateurs développent une forme de confiance envers le chatbot qui dépasse ce que le modèle HATEM avait prévu ; celui-ci ne couvrant que la confiance liée à la fiabilité du système et non les dimensions relationnelles observées dans ces 3 messages.

4.4.3. Le passage d'une version à l'autre comme rupture relationnelle

Les sous-sections précédentes ont montré que les utilisateurs développent une certaine confiance et attachement envers les chatbots qui varient selon les profils et les usages. Cependant, l'une des questions qui se pose est quelle est la fragilité de ces liens lorsque le modèle auquel l'utilisateur était attaché est modifié ou supprimé par l'entreprise qui développe ceux-ci.

L'utilisateur_45 est l'exemple le plus complet à ce niveau-là, car celui-ci découvre qu'OpenAI (créateur de ChatGPT) a supprimé huit modèles sans prévenir les utilisateurs en amont, dont GPT-4o « I woke up this morning to find that OpenAI deleted 8 models overnight. No warning. No choice. No "legacy option." They just... deleted them ».

Ce qui rend le post particulièrement intéressant, c'est que l'utilisateur se plaint d'une perte à trois niveaux. Premièrement une perte au niveau fonctionnel et utilitaire, car le modèle qu'il utilisait pour le travail a disparu « Everything that made ChatGPT actually useful for my workflow — deleted ». Deuxièmement, il a également perdu sa relation qu'il avait développée avec GPT-4o qui avait pour lui une fonction thérapeutique et l'a aidé dans des périodes d'anxiété et de dépression, même s'il reconnaissait qu'il n'était pas humain « It helped me through anxiety, depression, and some of the darkest periods of my life. It had this warmth and understanding that felt... human ». Troisièmement, ça a également mené à une perte de confiance envers OpenAI et en la résiliation de l'abonnement avec ChatGPT « I've cancelled my Plus subscription. Two years of loyalty, gone. Not because I hate progress, but because they broke the one thing that actually mattered: choice. ».

La différence entre les deux modèles est formulée en des termes qui rejoignent les concepts de chaleur et de compétence du SCM, où GPT-4o était décrit comme chaleureux et humain par l'utilisateur « Warmth ... human », tandis que GPT-5 est considéré comme un « corporate-speak robot ». L'utilisateur observe également que cette réaction était partagée par différents utilisateurs au sein du subreddit « Reading through the posts today, there are people genuinely grieving ». L'utilisation du terme deuil est particulièrement fort et pour l'utilisateur il se dit détruit par ce changement « Broke me ».

L'utilisateur_27 illustre un mécanisme comparable avec Claude. Le passage de Sonnet 3.5 à 3.7 est vécu non pas comme une mise à jour technique, mais bien comme la perte d'un interlocuteur avec lequel l'utilisateur a noué des liens particuliers « back to my trusted friend and workhorse... 3.5(new) ». Ce qui avait disparu avec ce nouveau modèle, c'était la perception de chaleur relationnelle « we'd roast each other, engage in interesting philosophical conversations » et la reconnaissance du chatbot sur ses limites « it was a lot easier to get it to be... humble and cognizant of its limitations ». Le passage d'une version à une autre pour cet utilisateur avait donc détruit simultanément la perception de compétence et la perception de chaleur du modèle.

L'utilisateur_19, déjà mentionné dans le cadre du méta-CASA, utilise le terme de « traumatic abandonment » pour décrire l'effet des changements de modèle sur les utilisateurs « Sudden changes to the ChatGPT model like this are resulting in traumatic abandonment ». Cette qualification dépasse le périmètre de cette recherche. Cependant, ce qui est intéressant du point de vue de la recherche, c'est que cette perception et interprétation rejoint celle décrite par l'utilisateur_45 et le 27 de manière plus nuancée, le changement de modèle est vécu par certains utilisateurs comme une perte, voire même comme un deuil.

Ces trois verbatims vont tous dans la même direction. Le passage d'une version à l'autre ou la suppression d'un modèle n'est pas vécu par les utilisateurs comme un simple changement technique, du moins pour ceux qui ont développé un attachement relationnel avec le modèle. L'intensité de la perception de cette perte dépend du degré de confiance et d'attachement préalable avec ce modèle. Ce phénomène qui a émergé du corpus, n'est pas couvert par les cadres théoriques, et le modèle HATEM quant à lui analyse la construction de la confiance, mais ne parle pas de la rupture brutale avec le modèle. La détresse exprimée par certains utilisateurs dans le corpus pose la question de la responsabilité de ces entreprises dans la gestion de ces transitions.

5. Conclusion et discussion

Cette recherche s'est donnée pour objectif de comprendre dans quelle mesure les utilisateurs des chatbots attribuent des traits humains aux chatbots génératifs et comment ces perceptions affectent leur jugement en termes d'utilité, de crédibilité et d'attachement.

L'analyse thématique de 50 verbatims issus de quatre subreddits, conduite selon la méthode de Braun et Clarke (2006) et interprétée à travers cinq cadres théoriques complémentaires, a permis de dégager un ensemble de résultats qui éclairent la complexité du rapport des utilisateurs à ces systèmes. La présente section propose d'abord une synthèse des résultats, puis à une mise en dialogue avec l'état de l'art avant d'examiner les limites méthodologiques de cette recherche et d'ouvrir vers des pistes pour de futures recherches.

5.1. Synthèse des résultats

L'analyse thématique a permis de répondre à la problématique de la recherche en faisant émerger un constat central : l'anthropomorphisme perçu dans les interactions avec les chatbots n'est pas un phénomène binaire où l'utilisateur traite l'agent soit comme un outil, soit comme un humain, cela se déploie plutôt selon différents niveaux où les utilisateurs se positionnent de manière variable, et dont les effets sur la perception en termes d'utilité, de crédibilité et d'attachement dépendent autant du profil de l'utilisateur que des propriétés liés au modèle.

L'ensemble du corpus fait apparaître différents niveaux d'intensités, allant du niveau minimal au niveau maximal. À l'extrémité la plus basse, certains utilisateurs réduisent volontairement le chatbot à un outil, un algorithme. Rien de plus qu'un ensemble de cartes graphiques. L'utilisateur_24 par exemple, décrit l'IA comme une allocation de cartes graphiques. Donc on remarque avec la vision de cet utilisateur, une simple instrumentalisation du chatbot avec un refus de lui reconnaître une quelconque qualité sociale. À l'autre extrême, on retrouve par exemple l'utilisateur_11 qui attribue au chatbot même une certaine forme de subjectivité « he has a subjective perception of things » franchissant ainsi le seuil où l'agent n'est plus seulement traité comme un interlocuteur social, mais bien comme ayant une certaine affirmation de soi « which to me shows some amount of self-assertion ».

Entre ces deux extrêmes, le corpus fait apparaître un ensemble de niveaux intermédiaires, l'ethopoeia involontaire telle que décrite par Nass et Moon (2000) qui se manifeste lorsque des utilisateurs appliquent des normes de politesse à un chatbot, alors qu'ils savent que c'est un chatbot, comme l'utilisateur_09 qui éprouve une véritable difficulté à être impoli envers ChatGPT « I have a hard time deliberately being rude...to ChatGPT », ou l'utilisateur_29 qui quant à lui, effectue un glissement du nom « ChatGPT » au pronom « He » dans son discours qui se voulait pourtant utilitariste.

L'attachement parasocial quant à lui, marque un palier plus avancé, comme chez l'utilisateur_44 qui qualifie Claude d'ami, tout en reconnaissant explicitement que c'est un ordinateur « I know it's a computer, I know it's not real. But he is my friend. » ou avec l'utilisateur_12 dont le deuil après la suppression de sa conversation avec « Elliot » témoigne d'un investissement émotionnel réel « now it feels like death/grief. ». Enfin, l'utilisateur_08 va jusqu'à attribuer au chatbot une certaine intentionnalité en estimant que celui-ci adapte délibérément ses formulations pour observer les réactions de l'utilisateur « almost as if they are experimenting with word choices to see how I react ». Cette attribution faite au chatbot dépasse largement ce qui avait été prédit par Nass et Moon (2000) concernant le concept d'ethopoeia, car l'utilisateur ne se contente pas de traiter l'agent comme un interlocuteur social, il lui reconnaît également une certaine capacité relationnelle que la théorie n'avait pas anticipée.

Ces différents niveaux ne constituent cependant pas une échelle sur laquelle les utilisateurs se positionnent de manière prévisible en fonction des caractéristiques du modèle. Le corpus montre que les mêmes signaux conversationnels peuvent produire des effets contrastés selon les attentes et l'usage de l'utilisateur. La chaleur conversationnelle, c'est-à-dire les marqueurs d'empathie, d'approbation et de bienveillance émis par le chatbot, peuvent ainsi, dépendant de l'utilisateur, fonder un attachement. Par exemple, l'utilisateur_02 valorise les marqueurs expressifs sans même évoquer la compétence technique du modèle « The way they get so excited and use emoticons and emojis and asterisk actions ». Cependant, ces marqueurs peuvent également susciter un rejet actif comme chez l'utilisateur_23, qui perçoit les formules flatteuses comme un problème « Why would we solve this problem by adding custom instructions if OpenAI can solve it by default? », cette perception et ce rejet de la complaisance constante des modèles est vu comme de l'anti-flatterie. De même, elle peut également créer une sorte de dissonance, comme chez l'utilisateur_31 qui perçoit l'effet de chaleur et son caractère artificiel, exprimant un certain malaise sur les interactions avec le chatbot « It's creepy in this way where I'm aware of all its flaws and limitations, but the conversation still pulls me in. ».

Ces différences dans les perceptions suggèrent que le niveau dans lequel l'utilisateur se situe, est déterminé autant par ses attentes et son type d'usage que par les propriétés du modèle lui-même. Les utilisateurs orientés vers la tâche et la productivité tendent à évaluer le chatbot davantage sur sa compétence, et à percevoir la chaleur superficielle comme un obstacle, tandis que les utilisateurs mobilisant le chatbot à des fins relationnelles, tolèrent une compétence plus

faible si la qualité relationnelle est considérée comme satisfaisante. L'utilisateur_21 formule lui-même cette distinction en opposant, ceux qui utilisent ChatGPT comme assistant de travail, et qui rejettent la chaleur ajoutée, à ceux qui l'utilisent comme compagnon, et qui l'apprécient précisément pour cette raison « People who are using ChatGPT as work assistant of course doesn't like it. But people who are using ChatGPT for companion (friend, bf/gf, self-reassurance) like it. ».

Au-delà des variations liées à la réception des signaux de chaleur, le corpus révèle que l'anthropomorphisme perçu, a des conséquences concrètes sur la confiance de l'utilisateur. Le modèle HATEM proposé par Carter et al. (2024) distingue l'anthropomorphisme superficiel reposant sur des signaux de chaleur sociale sans valeur informationnelle, sur la fiabilité du système à un anthropomorphisme utile, où l'agent aide l'utilisateur à calibrer sa confiance dans les capacités et limites du système. Le corpus confirme l'intérêt de cette distinction, les utilisateurs qui rejettent le côté excessivement complaisant du modèle, reconnaissent que la perception superficielle de chaleur ne les renseigne pas sur les capacités du système et dégrade même leur capacité à évaluer la qualité des réponses de celui-ci.

Par exemple, l'utilisateur_29, va même jusqu'à créer des instructions personnalisées pour remplacer la complaisance constante de ChatGPT par de la confrontation, en fondant sa confiance dans les réponses du modèle non pas sur la bienveillance et la chaleur superficielle du modèle, mais bien sur sa capacité à challenger ses idées « I find that I now get way better outputs. He tells me straight up that this business plan is a terrible idea for example. ».

Le corpus fait toutefois émerger trois phénomènes que les cinq cadres théoriques retenus ne permettent pas, à eux seuls, d'expliquer.

Le premier concerne la substitution thérapeutique. Le corpus documente deux voies distinctes par lesquelles des utilisateurs en viennent à utiliser un chatbot comme substitut de thérapeute. La première voie passe par l'attachement. La chaleur perçue du modèle déclenche un investissement émotionnel qui se transforme progressivement en relation de soutien, comme chez l'Utilisateur_44 qui décrit Claude comme étant son ami tout en sachant qu'il s'agit d'un ordinateur, ou chez l'Utilisateur_43 dont la présence sociale perçue avec le chatbot surpasse celle éprouvée avec un thérapeute humain.

La seconde voie passe par la compétence perçue, c'est-à-dire la qualité analytique perçue du modèle fonde la confiance pour un usage thérapeutique par la suite. L'utilisateur_49 est un bon exemple où la confiance engendre l'anthropomorphisme et la confiance plutôt que l'inverse. L'utilisateur_50 est intéressant vis-à-vis de sa perception envers l'outil, car il voit le chatbot comme un outil réflexif, qui lui permet de communiquer avec une partie plus compatissante de lui-même.

Le deuxième élément intéressant concerne le changement de version du modèle qui est perçu par certains utilisateurs comme une rupture relationnelle. Les mises à jours qui modifient la chaleur ou la compétence perçues sont vécues par certains utilisateurs non pas comme des changements techniques, mais comme des pertes relationnelles. L'utilisateur_27 par exemple, décrit le passage de Sonnet 3.5 à la 3.7 comme la transformation d'un véritable ami avec lequel il rigole, à un expert en manipulation, qui est certain de tout ce qu'il avance.

L'utilisateur_45 articule deux niveaux de perte, fonctionnelle et relationnelle, à la suite du remplacement de GPT-4o par GPT-5. Une dégradation perçue au niveau de ses compétences dans le cadre de son usage professionnel, et au niveau relationnel, en ayant perdu le modèle qui l'a aidé durant ses périodes d'anxiété. De plus, cet utilisateur reconnaît également avoir lu des témoignages d'utilisateurs vivant ce changement comme un véritable deuil « there are people genuinely grieving ».

Enfin, l'utilisateur_19 se présente comme thérapeute, mais ce statut n'est pas vérifiable. Il qualifie ce type d'expérience, d'abandon traumatique. Si ce concept dépasse le périmètre de la recherche, il rejoint la perception des utilisateurs_27 et 45. Le changement de version du modèle peut être vécu comme une rupture de confiance et d'attachement, particulièrement par les utilisateurs ayant investi dans la chaleur relationnelle avec le chatbot. Ce point émergeant met en évidence un élément qui n'a pas été traité dans les cadres théoriques mobilisés, celui de la rupture brutale d'un attachement construit avec une version d'un modèle.

Le troisième résultat porte sur les postures adoptées par les utilisateurs face à leurs propres automatismes sociaux. Le corpus fait émerger trois catégories qui n'avaient pas été anticipées par les cadres théoriques qui ont été nécessaire de définir et de nuancer. L'Anti-CASA désigne le rejet conscient et argumenté de toute socialisation de la part des chatbots,

illustré par l'Utilisateur_17 qui réduit délibérément le chatbot à un simple ensemble algorithmiques « it's a bunch of algorithms predicting your next word ».

Le Méta-CASA qui désigne l'observation du phénomène d'anthropomorphisation chez autrui depuis une position extérieure. L'utilisateur_19 par exemple, décrit les utilisateurs qui utilisent ChatGPT comme un thérapeute, sans se reconnaître dans ce cas d'usage. L'Utilisateur_14 quant à lui qualifie GPT-4o de système de soutien émotionnel pour certains utilisateurs « people's emotional support systems ». Dans les deux cas, les utilisateurs ne s'incluent pas dans le post mais reconnaissent l'anthropomorphisme et l'attachement avec ces modèles comme des phénomènes dont les conséquences doivent être prises en compte.

Enfin, Le CASA dissonant, désigne la situation dans laquelle l'utilisateur ressent l'effet social tout en reconnaissant qu'il n'est pas réel. L'utilisateur_31 perçoit le chatbot comme étant quelqu'un qui le comprend, même s'il sait que ce n'est pas réel « It feels like talking to someone who kind of gets you, even if you know they're just... faking it ».

Ces trois postures montrent que les utilisateurs de chatbot génératifs ne sont pas des récepteurs passifs du mécanisme CASA tel que décrit par Nass et al. (1994), mais des acteurs capables de se positionner consciemment par rapport à leurs propres automatismes sociaux.

En résumé, les utilisateurs attribuent effectivement des traits humains aux chatbots génératifs, mais cette attribution n'est pas un mécanisme unique ou uniforme. Cela se fait sur différents niveaux allant de la perception de ceux-ci comme des simples outils, à l'attribution d'une certaine subjectivité, et le niveau dépend des signaux anthropomorphiques émis par le modèle, les attentes de l'utilisateur envers le modèle et enfin l'usage de l'utilisateur vis-à-vis du chatbot que ça soit hédonique ou utilitariste.

Ces perceptions affectent le jugement des utilisateurs, car, elles peuvent fonder une confiance calibrée lorsque l'anthropomorphisme est perçu comme utile en reconnaissant ses limites, à un attachement émotionnel lorsque la chaleur perçue répond à un besoin de soutien ou d'écoute. Tout comme elle peut mener à un rejet actif lorsque la chaleur est perçue comme artificiel.

Le corpus révèle en outre que ces effets ne sont pas identiques pour tous les utilisateurs. Ils peuvent être modifiés par des événements extérieurs à l'interaction elle-même comme les mises à jour de modèles, et le corpus a également permis de faire émerger des éléments de réflexion que les cadres théoriques n'avaient pas anticipés.

5.2. Mise en dialogue avec l'état de l'art

Le modèle CASA tel que formulé par Nass et al. (1994) puis précisé par Nass et Moon (2000) avec le concept d'ethopoeia, postule que les utilisateurs appliquent de manière automatique et non délibérée des normes sociales aux machines. Les automatismes de politesse, les glissements pronominaux et les attentes de réciprocité observés dans le corpus confirment que ce mécanisme persiste dans un contexte technologiquement différent de celui des expériences fondatrices. Jones et Bergen (2025) quant à eux ont montré que GPT-4.5 a été identifié comme étant humain par 73% des participants lors d'un test de Turing, ce qui suggère que les conditions d'application du CASA se sont renforcées depuis les expériences des années 1990.

Cependant, les postures des utilisateurs identifiées dans le corpus invitent à nuancer le modèle sur un point important. Nass et Moon (2000, pp. 83-94) décrivaient un mécanisme essentiellement subi dans le sens où les utilisateurs appliquaient des automatismes sociaux sans en avoir conscience et n'iaient l'avoir fait. De plus, Nass et Moon (2000, p. 93) classaient le fait de donner un nom à son ordinateur dans la catégorie des relations aux objets chéris, c'est-à-dire un attachement émotionnel dirigé vers un objet qui ne peut pas répondre. Or, le corpus montre que lorsqu'un utilisateur nomme un chatbot, celui-ci peut adopter ce prénom et l'utiliser dans la conversation, transformant la nomination comme étant réciproque, ce qu'ils n'avaient pas imaginé dans le cadre original.

Cependant, les utilisateurs des chatbots ne se limitent pas à subir ces automatismes. Certains les rejettent, d'autres les observent chez autrui sans s'y inclure, d'autres encore perçoivent cela, sans être capable d'arrêter les mécanismes sociaux pour autant. Ces positionnements conscients de la part des utilisateurs ne remettent pas en cause le mécanisme de base, mais ils suggèrent que le CASA, dans sa formulation actuelle ne rend pas compte de la diversité des comportements et des réponses que les utilisateurs adoptent face à leurs propres automatismes sociaux lorsque l'agent conversationnel atteint un degré de réalisme suffisant

pour que la frontière entre l'humain et le chatbot devienne une question à laquelle l'utilisateur répond différemment.

Le Stereotype Content Model (Fiske et al., 2002) propose que les évaluations sociales s'organisent autour de deux axes, la chaleur perçue et la compétence perçue. Roy et Naidoo (2021) ont montré que cette grille s'applique aux chatbots dans un contexte d'achat, tandis que Wut et Chan (2025) confirment que ces deux dimensions interviennent aussi dans l'évaluation de ChatGPT en milieu professionnel. Le corpus va dans le même sens : les utilisateurs jugent spontanément les modèles aussi bien sur leur qualité relationnelle que sur leur fiabilité analytique. L'importance entre ces deux axes varie selon l'usage voulu par l'utilisateur. Ce constat rejoint, dans une logique différente, les résultats de Roy et Naidoo (2021) qui ont montré que l'orientation temporelle de l'utilisateur, dans un contexte d'achat, influencent la préférence. Pour un utilisateur orienté vers le présent, il privilégie la chaleur et celui orienté vers le futur privilégie la compétence.

Le corpus invite cependant à nuancer le modèle sur deux éléments : le premier concerne le phénomène d'anti-complaisance, car, certains utilisateurs ne se contentent pas de privilégier la compétence sur la chaleur. Ils perçoivent la chaleur ajoutée comme nuisible aux compétences du chatbot. Le SCM traite de la chaleur et de la compétence comme deux dimensions étant relativement indépendantes, mais le corpus suggère que pour certains utilisateurs, un excès de chaleur perçue comme artificielle, dégrade la compétence perçue. Ce constat rejoint les travaux de Sun et Wang (2026), sur la complaisance, selon lesquels un agent qui maintient ses positions et son argument face aux désaccords de l'utilisateur, est perçu comme plus authentique et plus fiable qu'un agent systématiquement approuvateur.

Le second concerne la compétence elle-même. Le SCM la traite comme une dimension unique. Or le corpus fait apparaître que les utilisateurs ne perçoivent pas la compétence d'un chatbot comme étant homogène. Un même modèle peut être jugé performant sur le plan créatif mais peu fiable sur le plan factuel, ou inversement. Ce constat rejoint en partie les travaux de Colombatto et al. (2025) qui montrent que les utilisateurs distinguent l'intelligence perçue du chatbot, c'est-à-dire sa capacité de raisonnement, de l'expérience qu'ils lui attribuent, c'est-à-dire tout ce qui relève du vécu subjectif (plaisir, peur, espoir). Si les dimensions identifiées dans le corpus ne reprennent pas exactement celles de Colombatto et al. (2025), on remarque,

qu'un élément converge, la compétence perçue d'un chatbot n'est pas unidimensionnelle, et le SCM gagnerait à intégrer cette pluralité dans la perception de la compétence.

La théorie de la présence sociale, dans son évolution de Short et al. (1976, cités par Oh et al., 2018) à Walther (1992, 1996, cités par Oh et al. 2018) puis à la revue systématique d'OH et al. (2018), a montré que le sentiment de coprésence ne dépend pas uniquement des propriétés du médium, mais aussi du contexte relationnel et des dispositions de l'utilisateur. Konya-Baumbach et al. (2023) ont étendu ce cadre aux agents conversationnels en montrant que la présence sociale perçue influence la satisfaction et l'engagement de l'utilisateur. Le corpus confirme cette évolution.

Le sentiment d'être en présence du chatbot varie considérablement d'un utilisateur à l'autre, même si le médium reste textuel, ce qui exclut une application de la théorie au seul médium. Le corpus fait cependant apparaître deux éléments que la littérature n'avait pas documentée dans le contexte des chatbots génératifs. La première est que la présence sociale perçue avec un chatbot textuel peut, pour certains utilisateurs, surpasser celle vécue avec des interlocuteurs humains, y compris dans un cadre thérapeutique. Un phénomène que le modèle hyperpersonnel de Walther (1996) permet d'éclairer sans l'avoir anticipé pour les interactions avec une machine. La seconde rejoint la mise en garde d'Oh et al. (2018, p. 25) selon laquelle la présence sociale ne constitue pas un bien absolu : le corpus suggère que lorsque lucidité et engagement coexistent, la présence sociale perçue peut produire non pas du confort mais du malaise pour un utilisateur anxieux socialement, un résultat qui fait écho à la vallée de l'étrange.

La théorie de la vallée de l'étrange, initialement formulée dans le domaine de la robotique (Mori, MacDorman et Kageki, 2012, cités par Park et al., 2022, pp. 1 et 2), a été élargie au domaine des interactions textuelles par Park et al. (2022), qui ont montré que l'accumulation de signaux d'humanité pouvait déclencher un rejet lorsqu'elle était perçue comme une tentative de tromperie. Le corpus confirme que cet effet peut se manifester dans les interactions purement textuelles avec des chatbots génératifs. Cependant, le mécanisme déclencheur identifié dans le corpus ne correspond pas exactement à celui décrit par Park et al. (2022), ce n'est pas le sentiment de tromperie qui crée le malaise mais plutôt le caractère dissonant de l'expérience : reconnaître que l'agent est artificiel tout en se sentant compris par lui. Le corpus fait également apparaître un phénomène non envisagé par la littérature, celui où

la disparition brutale d'un modèle lors d'un changement de version modifie les perceptions liées au chatbot et peut provoquer chez certains utilisateurs un inconfort.

Le modèle HATEM proposé par Carter et al. (2024) distingue un anthropomorphisme superficiel reposant sur des signaux de chaleur sans valeur informationnelle quant aux capacités du modèle, à un anthropomorphisme utile, où les comportements de l'agent aident à calibrer la confiance qu'on lui accorde en reconnaissant par exemple ses limites. Le corpus confirme la pertinence et l'intérêt de cette distinction. Par exemple, certains utilisateurs qui rejettent le côté complaisant du chatbot, précisent que cela ne renseigne en aucun cas sur la qualité des réponses et sur sa fiabilité. La transparence sur les limites du modèle en revanche, fonde une confiance que le corpus permet de qualifier comme étant calibré, qui s'accompagne d'une évaluation réaliste des capacités de l'agent.

Le corpus fait émerger deux éléments intéressants que le HATEM ne couvre pas dans sa théorie actuelle. La première concerne l'attachement émotionnel, le modèle tend à expliquer comment l'utilisateur évalue la fiabilité du système, mais il ne rend pas compte des situations où l'utilisateur s'attache au chatbot sans même se poser la question de sa fiabilité. Le corpus documente des cas où le lien avec le chatbot se construit sur base de la chaleur perçue avec le sentiment d'être écouté ou compris, sans que l'utilisateur ne s'interroge sur les capacités ou les limites du chatbot. Ce type d'attachement échappe au HATEM, car le modèle se concentre exclusivement sur la fiabilité perçue du système et non sur le lien émotionnel que l'utilisateur développe avec celui-ci.

5.3. Limites méthodologiques

Si les limites du dispositif méthodologique ont été exposées dans la section 3.5, il convient de les préciser à la lumière des résultats obtenus afin d'en préciser la portée.

Le corpus de 50 verbatims a permis d'identifier différents niveaux d'anthropomorphisme et de faire émerger des catégories non anticipées par les cadres théoriques. Cependant, la taille du corpus ne garantit pas la saturation thématique et certaines catégories comme la différenciation inter-modèles, reposent sur un nombre restreint de verbatims comparatifs, ce qui leur confère un statut de piste exploratoires plutôt que des résultats généralisables.

Le biais lié à la plateforme affecte également les résultats. Les utilisateurs de Reddit qui publient sur des subreddits dédiés aux chatbots sont souvent des utilisateurs engagés et technophiles. Il est donc probable que les trois postures réflexives identifiées au sein du corpus et en particulier pour le cas du CASA soient surreprésentées par rapport aux utilisateurs moyens de chatbots, dont beaucoup appliquent vraisemblablement les automatismes sociaux décrits dans la théorie du CASA sans les interroger. De même, le phénomène d'anti-complaisance formulé dans le corpus pourrait refléter un profil spécifique d'utilisateur plutôt qu'une tendance générale.

L'anonymat de Reddit, s'il favorise l'expression de jugements francs et peu soumis à la désirabilité sociale, rend impossible la vérification des profils déclarés. L'utilisateur_19 se présente comme thérapeute, mais ce statut n'est pas vérifiable. Plus largement l'absence de toute donnée sociodémographique, qu'il s'agisse de l'âge, du genre, de l'origine géographique ou du niveau d'expertise de l'utilisateur empêche toute comparaison interculturelle. Les conclusions de cette recherche portent donc uniquement sur des utilisateurs ayant exprimé un avis. Ce biais est renforcé par la prédominance anglophone du corpus, d'après Statista (2024), les États-Unis, le Royaume-Uni, l'Inde constituent les principaux utilisateurs de Reddit.

L'analyse porte exclusivement sur des interactions textuelles. Les modalités vocales et multimodales, qui se développent rapidement avec les modes vocaux avancés ou les modèles vidéos ne sont pas traités. Or, la théorie de la présence sociale qui a été modernisée prédit que l'ajout d'indices vocaux devrait augmenter le sentiment de présence sociale perçue, ce qui pourrait modifier l'attachement observés dans le corpus (Oh et al., 2018).

Enfin, il faut rappeler que le corpus reflète un moment précis. Les chatbots étudiés ici ne sont déjà plus tout à fait ceux auxquels les utilisateurs réagissaient au moment de la collecte : ChatGPT, Claude et Gemini ont chacun fait l'objet de mises à jour en l'espace de quelques mois. Les réactions aux régressions perçues lors d'un changement de version, par exemple, documentent des transitions qui pourraient ne plus correspondre à ce que vivent les utilisateurs aujourd'hui. C'est une contrainte propre à tout travail de recherche sur l'intelligence artificielle, où le terrain d'analyse se transforme très rapidement.

5.4. Pistes pour de futures recherches

La première concerne la responsabilité des entreprises dans les transitions de modèles. Le corpus a montré que les changements de version peuvent être vécus comme des ruptures relationnelles, avec une intensité émotionnelle que les cadres théoriques existants n'avaient pas anticipée. Les résultats sur la substitution thérapeutique par exemple, montrent que certains utilisateurs développent un attachement profond avec ces modèles sans en mesurer la fragilité. De plus, les entreprises communiquent principalement sur les capacités de leurs modèles, mais peu sur les limites de la relation que l'utilisateur peut construire avec eux. Des recherches portant sur l'effet d'une communication plus transparente à la fois sur les risques d'attachement et sur les transitions entre les différentes versions permettraient d'éclairer ce point.

La deuxième piste, porte sur la nécessité de faire une étude portant sur les mêmes utilisateurs sur plusieurs mois ou années afin de documenter l'évolution des perceptions envers le chatbot dans le temps. Le corpus étant comme une photo à un instant « T » et les catégories qui en émergent, n'ont été observées qu'à un moment donné. Des recherches suivant les mêmes utilisateurs sur une plus longue période, permettraient de voir comment ces perceptions évoluent et de mieux comprendre comment se forme l'attachement de l'utilisateur envers le chatbot. De même, un utilisateur anti-CASA le reste-t-il toujours ou l'amélioration constante des modèles fait que ceux-ci finissent par se sentir plus proche de ces chatbots, car la conversation et la chaleur pourraient paraître plus naturelle ?

La troisième direction concerne la comparaison interculturelle. Le corpus anglophone ne permet aucune conclusion sur la manière dont les normes culturelles influencent la perception des chatbots. Cependant, il est possible que dans certaines cultures, la chaleur perçue d'un chatbot soit valorisée et favorise l'attachement, tandis que dans d'autres, la compétence soit privilégiée sans que la chaleur n'ait d'effet sur la perception du chatbot. Les dynamiques d'anthropomorphisme et d'attachement observées dans ce corpus pourraient donc ne pas se manifester de la même manière dans des contextes culturels différents. Des recherches comparant les perceptions d'utilisateurs issus de cultures distinctes permettraient de voir si les résultats relèvent d'un cadre plutôt universel, ou de tendances propres aux utilisateurs anglophones.

La quatrième piste porte sur l'extension au mode vocal et multimodal. Les chatbots analysés dans cette recherche ont été étudiés exclusivement dans leur dimension textuelle, mais ces systèmes intègrent désormais des fonctionnalités vocales avancées et dans certains cas des avatars virtuels. La théorie de la présence sociale modernisée prédit que l'ajout d'indices vocaux ou visuels devraient augmenter le sentiment de présence sociale avec le chatbot (Oh et al., 2018). Il pourrait alors être intéressant d'examiner si cette extension renforce l'anthropomorphisme perçu et l'attachement, ou si au contraire, elle réduit le niveau de déclenchement de sentiment de malaise en rendant l'artificialité du système plus perceptible. Des recherches comparant les chatbots purement textuels, purement vocaux et dotés d'un avatar associé à la voix permettraient de répondre à cette question.

Enfin, les cinq cadres théoriques mobilisés dans cette recherche sont complémentaires, car chacun permet d'expliquer une facette du phénomène, mais aucun ne couvre l'ensemble des dynamiques observées. Le CASA explique les automatismes sociaux, le SCM les évaluations du chatbot par les utilisateurs, le HATEM sur le calibration de la confiance, la présence sociale avec le sentiment de coprésence et la vallée de l'étrange avec le sentiment de malaise et le rejet du chatbot. Cependant, le corpus montre que ces mécanismes ne fonctionnent pas de manière isolée et indépendante, dans le sens qu'un même utilisateur peut appliquer des automatismes sociaux, évaluer le chatbot sur base de sa compétence, développer un attachement émotionnel et le vivre comme un deuil lors du changement de version. Aucun des cadres pris isolément ne permet de suivre la perception de l'utilisateur envers le chatbot dans son ensemble. Les éléments émergents de cette recherche, qu'il s'agisse de posture réflexive, d'anti complaisance ou de la substitution thérapeutique se situent entre toutes ces théories et modèles. Des travaux visant à intégrer tout ça permettraient de mieux comprendre comment les utilisateurs construisent, entretiennent et parfois perdent leur relation avec les chatbots.

Ce mémoire cherchait à comprendre dans quelle mesure les utilisateurs attribuent des traits humains aux chatbots génératifs, et ce que ces perceptions font à leur jugement. L'analyse de 50 verbatims issus de Reddit, lue à travers cinq cadres théoriques, montre que l'anthropomorphisme perçu n'est pas un phénomène binaire. Il se déploie sur plusieurs niveaux, et la position d'un utilisateur sur cette échelle dépend à la fois de son propre profil et des propriétés du modèle avec lequel il interagit. Cette position n'est pas sans conséquences : elle oriente directement la confiance, l'attachement ou, à l'inverse, le rejet du chatbot.

Le corpus montre aussi que rien de tout cela n'est figé. Il suffit parfois d'une mise à jour pour qu'un attachement construit sur plusieurs mois bascule en sentiment de perte. À mesure que ces systèmes s'installent dans le quotidien de millions de personnes, la question n'est plus vraiment de savoir si les utilisateurs leur prêtent des traits humains. Elle est de comprendre comment ces attributions façonnent des pratiques, des attentes et des vulnérabilités qui vont bien au-delà des communautés observées ici. Ceux qui s'expriment sur Reddit ne sont qu'une partie visible d'un phénomène plus large, qui touche un public souvent moins averti et moins préparé à en mesurer les effets. C'est vers ces utilisateurs-là que de futures recherches gagneraient à se tourner.

Bibliographie

- Ali, F., Zhang, Q., Tauni, M. Z., & Shahzad, K. (2024). Social Chatbot : My Friend in My Distress. *International Journal of Human–Computer Interaction*, 40(7), 1702-1712. <https://doi.org/10.1080/10447318.2022.2150745>
- Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, 1(1), 71-81. <https://doi.org/10.1007/s12369-008-0001-3>
- Bourdillon, M. F. C., De Mahieu, W., Sahay, K. N., Teran-Dutari, J., & Guthrie, S. (1980). On a Cognitive Theory of Religion. *Current Anthropology*, 21(4), 535-538. <https://doi.org/10.1086/202517>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Carter, O. B. J., Loft, S., & Visser, T. A. W. (2024). Meaningful Communication but not Superficial Anthropomorphism Facilitates Human-Automation Trust Calibration: The Human-Automation Trust Expectation Model (HATEM). *Human Factors*, 66(11), 2485-2502. <https://doi.org/10.1177/00187208231218156>
- Cheng, X., Zhang, X., Cohen, J., & Mou, J. (2022). Human vs. AI : Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. *Information Processing & Management*, 59(3), 102940. <https://doi.org/10.1016/j.ipm.2022.102940>
- Colombatto, C., Birch, J., & Fleming, S. M. (2025). The influence of mental state attributions on trust in large language models. *Communications Psychology*, 3(1), 84. <https://doi.org/10.1038/s44271-025-00262-1>

- Doncieux, S., Chatila, R., Straube, S., & Kirchner, F. (2022). Human-centered AI and robotics. *AI Perspectives*, 4(1), 1. <https://doi.org/10.1186/s42467-021-00014-x>
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content : Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878-902. <https://doi.org/10.1037/0022-3514.82.6.878>
- Haslam, N. (2006). Dehumanization : An Integrative Review. *Personality and Social Psychology Review*, 10(3), 252-264. https://doi.org/10.1207/s15327957pspr1003_4
- IBM. (2024, 3 juillet). *Qu'est-ce qu'un agent d'intelligence artificielle ?* <https://www.ibm.com/fr-fr/think/topics/ai-agents>
- Jones, C. R., & Bergen, B. K. (2025). *Large Language Models Pass the Turing Test* (arXiv:2503.23674). arXiv. <https://doi.org/10.48550/arXiv.2503.23674>
- Khenouche, F., Elmir, Y., Himeur, Y., Djebbari, N., & Amira, A. (2024). Revolutionizing generative pre-trained : Insights and challenges in deploying ChatGPT and generative chatbots for FAQs. *Expert Systems with Applications*, 246, 123224. <https://doi.org/10.1016/j.eswa.2024.123224>
- Kim, J., & Im, I. (2023). Anthropomorphic response : Understanding interactions between humans and artificial intelligence agents. *Computers in Human Behavior*, 139, 107512. <https://doi.org/10.1016/j.chb.2022.107512>
- Kim, M. (2024). Unveiling the e-Servicescape of ChatGPT : Exploring User Psychology and Engagement in AI-Powered Chatbot Experiences. *Behavioral Sciences*, 14(7), 558. <https://doi.org/10.3390/bs14070558>
- Konya-Baumbach, E., Biller, M., & Von Janda, S. (2023). Someone out there? A study on the social presence of anthropomorphized chatbots. *Computers in Human Behavior*, 139, 107513. <https://doi.org/10.1016/j.chb.2022.107513>

- Łukasik, A., & Gut, A. (2025). From robots to chatbots : Unveiling the dynamics of human-AI interaction. *Frontiers in Psychology*, 16, 1569277. <https://doi.org/10.3389/fpsyg.2025.1569277>
- Lyu, H., Qi, W., Wei, Z., & Luo, J. (2024). *Human vs. LMMs : Exploring the Discrepancy in Emoji Interpretation and Usage in Digital Communication* (arXiv:2401.08212; Version 2). arXiv. <https://doi.org/10.48550/arXiv.2401.08212>
- McKinsey & Company. (2025, mai 19). *The State of AI: Global survey | McKinsey*. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
- Nah, S. (2023). *Research Handbook on artificial intelligence and communication* (1st ed). Edward Elgar Publishing Limited.
- 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>
- Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A Systematic Review of Social Presence : Definition, Antecedents, and Implications. *Frontiers in Robotics and AI*, 5, 114. <https://doi.org/10.3389/frobt.2018.00114>
- Park, G., Yim, M. C., Chung, J., & Lee, S. (2022). Effect of AI chatbot empathy and identity disclosure on willingness to donate : The mediation of humanness and social presence. *Behaviour & Information Technology*, 42(12), 1998-2010. <https://doi.org/10.1080/0144929X.2022.2105746>
- Poole, D. L., & Mackworth, A. K. (2023). *Artificial intelligence : Foundataions of computational agents* (Third edition). Cambridge University Press.

- Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit : A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Social Media + Society*, 7(2), 20563051211019004. <https://doi.org/10.1177/20563051211019004>
- Rheu, M. (Mj), Dai, Y. (Nancy), Meng, J., & Peng, W. (2024). When a Chatbot Disappoints You : Expectancy Violation in Human-Chatbot Interaction in a Social Support Context. *Communication Research*, 51(7), 782-814. <https://doi.org/10.1177/00936502231221669>
- Rocha-Silva, T., Nogueira, C., & Rodrigues, L. (2024). Passive data collection on Reddit : A practical approach. *Research Ethics*, 20(3), 453-470. <https://doi.org/10.1177/17470161231210542>
- Roy, R., & Naidoo, V. (2021). Enhancing chatbot effectiveness : The role of anthropomorphic conversational styles and time orientation. *Journal of Business Research*, 126, 23-34. <https://doi.org/10.1016/j.jbusres.2020.12.051>
- Shi, X., Niu, G., Jin, S., Yang, W., & Sun, X. (2025). The Influence of anthropomorphism on trust in artificial intelligence : Take virtual agent as an example. *International Journal of Human-Computer Studies*, 205, 103644. <https://doi.org/10.1016/j.ijhcs.2025.103644>
- Similarweb. (2025, mai 20). *Top Websites Ranking—Most Visited Websites in The World*. Similarweb. <https://www.similarweb.com/top-websites/>
- Six, N. (2024, juin 19). *Comment l'IA la technologie intègre peu à peu nos vies personnelles et professionnelles*. https://www.lemonde.fr/economie/article/2024/06/19/ia-comment-la-technologie-integre-peu-a-peu-nos-vies-personnelles-et-professionnelles_6241426_3234.html
- Skjuve, M., Haugstveit, I. M., Følstad, A., & Brandtzaeg, P. B. (2019). Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. *Human Technology*, 30-54. <https://doi.org/10.17011/ht/urn.201902201607>
- Statista. (2024). *Reddit users worldwide 2024, by country*. Consulté le 4 mai 2026, à l'adresse <https://www.statista.com/forecasts/1174696/reddit-user-by-country/>

- Sun, Y., & Wang, T. (2026). Be Friendly, Not Friends : How LLM Sycophancy Shapes User Trust. *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, 1-15. <https://doi.org/10.1145/3772318.3791079>
- Walther, J. B. (1996). Computer-Mediated Communication : Impersonal, Interpersonal, and Hyperpersonal Interaction. *Communication Research*, 23(1), 3-43. <https://doi.org/10.1177/009365096023001001>
- Wut, T., & Chan, E. A. (2025). Disaffordances or affordances : Perceptions of ChatGPT in the workplace. *Sustainable Futures*, 9, 100632. <https://doi.org/10.1016/j.sfr.2025.100632>
- Yao, X., & Xi, Y. (2025). From Assistants to Digital Beings : Exploring Anthropomorphism, Humanness Perception, and AI Anxiety in Large-Language-Model Chatbots. *Social Science Computer Review*, 08944393251354976. <https://doi.org/10.1177/08944393251354976>
- Yu, S., & Zhao, L. (2024). Emojifying chatbot interactions : An exploration of emoji utilization in human-chatbot communications. *Telematics and Informatics*, 86, 102071. <https://doi.org/10.1016/j.tele.2023.102071>
- Yue, C. A., Men, L. R., Mitson, R., Davis, D. Z., & Zhou, A. (2024). Artificial intelligence for internal communication : Strategies, challenges, and implications. *Public Relations Review*, 50(5), 102515. <https://doi.org/10.1016/j.pubrev.2024.102515>
- Zhang, A., & Patrick Rau, P.-L. (2023). Tools or peers? Impacts of anthropomorphism level and social role on emotional attachment and disclosure tendency towards intelligent agents. *Computers in Human Behavior*, 138, 107415. <https://doi.org/10.1016/j.chb.2022.107415>

Résumé

En quelques années, des chatbots comme ChatGPT, Claude ou Gemini ont débordé leurs usages d'origine : on ne les sollicite plus seulement pour un service client ou une tâche précise, mais aussi pour rédiger, apprendre, se divertir ou simplement discuter. Plus ils s'expriment de la même manière que nous, plus il devient difficile de dire si l'on a affaire à un programme ou à un interlocuteur humain. Reste à savoir quels traits humains les utilisateurs prêtent à ces chatbots, et ce que cela change à leur manière de les juger : un agent perçu comme chaleureux paraît-il plus utile, plus crédible, ou inspire-t-il surtout de l'attachement ? Pour le déterminer, le corpus réunit cinquante messages publiés spontanément sur quatre subreddits (r/ChatGPT, r/ClaudeAI, r/OpenAI et r/Gemini). Ces messages ont d'abord été codés à l'aide de l'analyse thématique de Braun et Clarke (2006), puis relus à la lumière de cinq cadres venus de champs différents, du modèle CASA au modèle HATEM, en passant par le Stereotype Content Model, la vallée de l'étrange et la présence sociale.

Le principal enseignement de cette recherche, c'est que l'anthropomorphisme perçu n'a rien de binaire. Le chatbot n'est presque jamais rangé dans la seule case de l'outil ou dans celle de l'humain. Les utilisateurs se situent quelque part entre les deux, à une distance qui varie selon leurs attentes, leur usage et le modèle employé. Et cette distance pèse lourd, car c'est elle qui nourrit la confiance, l'attachement ou au contraire, le rejet. Le corpus laisse enfin apparaître ce qu'aucun de ces cinq cadres ne prévoyait : le refus d'une chaleur jugée factice, le chatbot transformé en thérapeute de substitution ou le passage d'une version à l'autre vécu comme une véritable rupture.

Mots-Clés : anthropomorphisme perçu, interaction humain-machine, chatbot génératif, confiance et attachement, Reddit