UCLouvain
FUCaM MONS

LOUVAIN
School of Management

Louvain School of Management

L’impact de l’intelligence artificielle générative sur la crédibilité perçue des fake news

Étude comparative de titres rédigés par IA et par des humains

Auteur : Curto Lorenzo
Promoteur(s) : Fouss François
Année académique 2024-2025
Master [120] : ingénieur de gestion à finalité spécialisée

Declaration Regarding AI Tool Usage in Master's Thesis

We recognize that AI tools might be valuable aids during the master's thesis work, but they are not infallible. Remember that transparency fosters trust, and acknowledging AI's role enhances the credibility of your work.

Therefore, when deciding to use such a tool, you need to adhere to the following principles of responsible use of AI.

1. **Critical Evaluation :**
    - We critically assessed the AI-generated output, ensuring its alignment with our research objectives.
    - Any modifications or corrections were made based on our expertise and domain knowledge.
2. **Transparency :**
    - We acknowledge the use of ChatGPT transparently, emphasizing that it contributed to our work but did not replace human judgment.
    - Our commitment to transparency ensures the integrity of this thesis.
3. **Ethical Considerations :**
    - We actively monitored for biases or unintended consequences introduced by the AI tool.
    - Our ethical responsibility guided our decisions throughout the research process.

Declaration **(This declaration is mandatory and must appear on the first page (after the title page) of the document.**

During the preparation of this master's thesis, the author(s) utilized Chat GPT for the following purpose:

1. **Writing Assistance:** To rephrase or restructure sentences in order to reduce redundancy and improve conciseness and and to help produce academic translations closely aligned with the original meaning in both English and French.
2. **Technical Support:** To better understand the use of specific Python libraries and statistical tools.

After using ChatGPT, the author(s) diligently reviewed and edited the content produced by the tool. We take full responsibility for the final content presented in this thesis.

By signing this declaration, we affirm that the content of this master's thesis reflects our original work, augmented by the responsible use of AI.

Feel free to customize the text by replacing "[NAME TOOL / SERVICE]" with the actual AI tool you used and providing a concise explanation of its purpose. Remember to sign and date the declaration appropriately.

06/08/2025

Résumé

À l'ère de l'intelligence artificielle générative, la création de fake news devient plus accessible, rapide et convaincante. Ce mémoire explore un enjeu central : les fake news générées par IA sont-elles perçues comme plus crédibles que celles rédigées manuellement par des humains ?

La première partie du travail s'appuie sur une revue de littérature afin de définir les concepts clés de fake news et d'intelligence artificielle générative. Elle permet également d'identifier les facteurs personnels et contextuels influençant la crédibilité perçue.

À partir de ce cadre théorique, plusieurs hypothèses ont été formulées concernant l'influence de la source de la fake news (IA ou humain), du niveau d'éducation, de la familiarité technologique, de la compétence perçue en IA, et de l'âge.

Une étude quantitative a été menée pour tester ces hypothèses. Les résultats montrent que les fake news générées par intelligence artificielle génératives sont perçues comme significativement plus crédibles que celles rédigées par des humains.

En revanche, les variables comme le niveau d'éducation, la compétence perçue en IA ou la familiarité technologique n'ont pas eu d'effet notable. Seul l'âge semble jouer un rôle, mais sans effet robuste. Ces résultats suggèrent que la crédibilité perçue repose sur des mécanismes cognitifs plus complexes.

Bien que cette recherche présente certaines limites comme un échantillon restreint, la sélection manuelle des fake news ou encore l'absence de contrebalancement expérimental, elle offre des pistes de réflexion pertinentes sur les risques liés à l'utilisation de l'IA générative dans la production de fake news.

Ce mémoire constitue une première étape vers une compréhension plus fine de la relation entre intelligence artificielle et perception de la vérité, et appelle à des recherches complémentaires, notamment sur les contenus multimodaux ou les effets d'expositions croisées. Dans un contexte où la frontière entre information réelle et fabriquée devient de plus en plus floue, il est urgent de repenser les mécanismes de confiance et de vérification.

Remerciements

Tout d'abord, j'adresse mes sincères remerciements à mon promoteur, M. Fouss pour la confiance qu'il m'a accordée.

Je tenais également à remercier l'UCLouvain pour les moyens mis à notre disposition tout au long de notre cursus.

Je remercie particulièrement l'atelier soutien mémoire, aussi bien les organisateurs que les participants pour leur aide et leur motivation.

Je remercie très sincèrement Marie Margaux Sakkas, Samed Baguirov, Clément Fontaine, Margot Achtergael, Juliette Doyen, Théo Lemahieu, Damien Couty, Pauline Brouez, Louise Lavenne et Maxime Wacquier pour leur soutien tout au long de ce travail.

Ensuite, je remercie ma famille pour leur patience et plus particulièrement ma sœur Tatiana pour m'avoir guidé dans la réalisation de ce travail.

Je tiens également à exprimer ma profonde gratitude envers mes responsables chez Intys et Alstom, Hermine Mesquita da Cunha, Agathe Mugnier, Alexis Manderlier et Caroline Lepers, pour leur soutien, leur confiance et leur compréhension.

Enfin, mes remerciements vont à Claudio Ranieri et à sa dernière saison à l'AS Roma, non seulement pour les moments de bonheur et d'émotions mais aussi pour avoir sauvé le club.

À tous ces intervenants, je présente mes remerciements, mon respect et ma gratitude.

Table des matières

Liste des tableaux

Liste des annexes

Introduction

Un homme armé qui rentre dans une pizzeria pour « démanteler » un prétendu réseau pédocriminel dirigé par Hillary Clinton. Un ministre de la Défense menace un pays d'utiliser l'arme nucléaire, sur la base d'informations inventées. Ces exemples, bien réels, illustrent les conséquences parfois dramatiques des fake news.

Avec l'essor de l'intelligence artificielle générative, la production de fake news devient non seulement plus facile mais également plus rapide. En quelques secondes, un utilisateur peut générer des centaines de titres trompeurs à l'apparence crédibles, gratuitement et sans compétence techniques. Cette évolution soulève une question centrale : les fake news crées par IA sont-elles perçues comme plus crédibles que celles rédigées manuellement par des humains ?

Afin d'y répondre, nous commencerons par poser notre question de recherche, avant de passer en revue la littérature existante. Cette revue permettra de définir les fake news, de présenter le fonctionnement de l'intelligence artificielle générative et d'identifier les facteurs influençant la crédibilité perçue d'une information.

Dans une seconde partie, nous formulerons nos hypothèses sur base de ce cadre théorique. Nous détaillerons ensuite la méthodologie employée, les échelles de mesure utilisées ainsi que les analyses statistiques menées. Enfin, nous présenterons les résultats obtenus, avant de les discuter en tenant compte des biais et des limites de l'étude.

I. Question de recherche

A. Contexte

En 2016, un homme armé est entré dans une pizzeria à Washington. Il avait lu sur Internet que cette pizzeria hébergeait un réseau pédocriminel dirigé par Hillary Clinton (Kang et Goldman, 2016). La même année, le ministre de la Défense du Pakistan a menacé Israël d'une attaque nucléaire après avoir cru à une fake news relayée sur les réseaux sociaux (Politi, 2016)

Ces incidents illustrent à quel point les fake news peuvent entraîner des conséquences concrètes et potentiellement dramatiques dans le monde réel. Leur impact ne se limite pas à l'opinion publique : elles peuvent mettre en danger la vie des individus, la stabilité politique, voire la sécurité internationale.

Aujourd'hui, avec l'essor des intelligences artificielles génératives, ce phénomène prend une nouvelle ampleur. Il est désormais possible pour n'importe qui de générer en quelques secondes des textes, images ou vidéos d'apparence crédibles. En mars 2023, une image du Pape François portant une doudoune blanche a été massivement partagée sur les réseaux sociaux avant qu'il ne soit révélé qu'elle avait été générée par une intelligence artificielle (Mansoura, 2023).

Dans ce contexte, la production de fake news est facilitée et démultipliée, l'IA générative pouvant créer à la chaine des informations crédibles et difficiles à détecter. La frontière entre contenu humain et contenu artificiel devient floue, ce qui complique encore d'avantage la tâche des lecteurs, déjà confrontés à une surcharge informationnelle.

Ce changement de paradigme soulève une question centrale : si les fake news deviennent plus convaincantes grâce à l'IA, les percevons-nous aussi comme plus crédibles ?

B. Question de recherche

Face à l'avènement des outils d'intelligence artificielle générative il devient essentiel de comprendre si les contenus produits par ces technologies, en particulier les fake news, sont perçus comme plus ou moins crédibles que ceux créés par des humains.

Ce mémoire vise donc à explorer l'effet de l'intelligence artificielle générative sur la crédibilité perçue de titres de fake news.

Notre question de recherche peut donc se formuler comme suit :

Quel est l'impact de l'intelligence artificielle générative sur la crédibilité perçue des titres de fake news ?

Plus précisément, il s'agit de comparer la crédibilité accordée à deux types de titres de fake news :

- des fake news réelles, issues de cas authentiques repérés par des sites de fact-checking ;
- des fake news générées par intelligence artificielle, à l'aide de *ChatGPT*.

Cette étude explorera également l'influence de plusieurs variables individuelles sur cette perception :

- l'âge du lecteur ;
- le niveau d'éducation ;
- la confiance dans les technologies ;
- la compétence perçue en intelligence artificielle.

Ces facteurs seront analysés pour déterminer s'ils modèrent la relation entre la nature du contenu et la crédibilité qui lui est attribuée.

Revue de littérature

II. Introduction

Afin de répondre à notre problématique, il est essentiel de s'appuyer sur une revue de littérature existante. Celle-ci a pour objectif de poser les fondations théoriques de l'étude et de contextualiser les enjeux liés aux fake news et à l'intelligence artificielle générative.

Cette revue de littérature s'articule autour de trois grands axes. Dans un premier temps, nous nous intéresserons au phénomène des fake news, en tentant de le définir de manière précise, puis en le distinguant d'autres formes de fausses informations telles que les erreurs journalistiques, la parodie et la satire.

Ensuite, nous nous concentrerons sur l'intelligence artificielle générative. Nous proposerons une définition claire de cette technologie en plein essor, puis nous en décrirons le fonctionnement général, en mettant en lumière les grands modèles de langages. Nous évoquerons également ses limites connues, notamment en matière de biais, d'hallucinations et d'opacité algorithmique.

Enfin, la dernière partie de cette revue de littérature sera consacrée aux facteurs qui influencent la crédibilité perçue des fake news. Nous aborderons d'une part les facteurs contextuels, liés à la forme et au contenu du message, et d'autre part les facteurs individuels, tels que l'âge, le niveau d'éducation ou la familiarité avec les technologies. Cette dernière section servira de base directe à la formulation des hypothèses de recherche que nous testerons dans la partie pratique.

III. Fake News

A. Introduction

Dans ce chapitre dédié aux fake news, nous commencerons en mettant en contexte l'émergence de ces dernières grâce aux avancées des technologies de l'information et de la communication.

Par la suite, nous entreprendrons un examen approfondi de la littérature pour élaborer une définition claire et précise du concept de fake news.

Enfin, nous conclurons ce chapitre en distinguant les fake news d'autres formes de désinformation

Ce chapitre vise à poser les bases conceptuelles de notre étude en définissant rigoureusement les fake news, en explorant leurs caractéristiques distinctives.

B. Contexte

Le développement des technologies de l'information et de la communication permet une diffusion rapide et peu coûteuse des contenus, facilitant également la propagation de fausses informations (Kalsnes, 2018 ; Shu et al., 2017). Selon Meyers et al (2020, p.138), les réseaux sociaux constituent un terreau fertile pour les fake news.

Cette diffusion de masse est également facilitée par la manière dont les réseaux sociaux floutent la conceptualisation d'une source d'information. Un article de presse peut circuler via le site du média, un post Facebook ou un simple partage par un ami. La numérisation a également élargi le concept d' « information », un tweet ou une story peuvent être perçus comme tels, surtout s'ils émanent d'une figure d'autorité (Tandoc et al., 2017).

Les fake news sont devenues le symbole d'un problème sociétal plus vaste : la manipulation de l'opinion publique pour affecter le monde réel (Gu et al., 2017, p.3). Selon Silverman & Singer-Vine, (2016) les titres de fake news trompent des adultes américains dans 75 % des cas. De plus, plus de la moitié des Américains déclarent s'informer via les réseaux sociaux, un canal vulnérable à la désinformation (Shearer & Mitchell, 2021).

Les conséquences des fake news sont profondes : perte de confiance dans les médias, radicalisation et désinformation persistante même après démystification ; Thorson dans

Kalsnes, 2018). Comprendre la nature de ces fake news est alors crucial pour mieux appréhender leurs impacts sur nos démocraties

C. Définition

Avant d'aller plus loin, il nous semble essentiel de définir clairement le concept de fake news. Nous avons examiné la littérature existante et identifié plusieurs définitions.

Selon Sharma et al. (2019, p. 4), les fake news désignent des articles ou des messages d'actualité publiés et propagés à travers les médias, véhiculant de fausses informations, indépendamment des moyens et des motivations qui les sous-tendent *(« a news article or message published and propagated through media, carrying false information regardless of the means and motivations behind it »*) [traduction libre].

Cette définition ne prend pas en compte les motivations derrière la publication de fausses nouvelles. Dès lors, cette définition englobe aussi les parodies, satires et erreurs journalistiques.

L'article d'Allcott et Gentzkow (2017, p. 213) définit les fake news comme des articles d'actualité qui sont intentionnellement faux, dont la fausseté est vérifiable, et qui pourraient induire les lecteurs en erreur (« *news articles that are intentionally and verifiably false, and could mislead readers* ») [traduction libre].

En y ajoutant la notion d'intentionnellement faux, Allcott et Gentzkow (2017) excluent donc les erreurs journalistiques, celles-ci étant par essence non intentionnelles. Néanmoins, si l'on suit cette définition, les parodies et satires d'informations restent des fake news.

Dans le cadre de notre étude, nous préférons nous appuyer sur la définition de Kalsnes (2018). Selon Kalsnes, les fake news sont des informations complètement ou partiellement fausses, (souvent) présentées comme des nouvelles, et typiquement exprimées sous forme de contenu textuel, visuel ou graphique dans l'intention d'induire l'utilisateur en erreur ou de le confondre (« *complete or partly false information, (often) appearing as news, and typically expresed as textual, visual or graphical content with an intention to mislead or confuse user* ») [traduction libre].

Cette définition met en évidence l'intention de nuire qui distingue les fake news des autres formes d'informations erronées ou satiriques.

En complément, le *Reuters Institute for the Study of Journalism* définit les fake news comme de fausses informations délibérément diffusées dans un but stratégique spécifique, que ce soit

politique ou commercial. Ces contenus se présentent généralement sous l'apparence de véritables reportages d'actualités, tout en véhiculant des théories du complot ou d'autres sujets chargés d'appels émotionnels qui confirment des croyances existantes (« *false information knowingly circulated with specific strategic intent—either political or commercial. Such content typically masquerades as legitimate news reports while trafficking in conspiracy theories or other matters laden with emotional appeals that confirm existing beliefs »,*The Reuters Institute for Study of Journalism (RISJ), 2017) [traduction libre]. Il est intéressant de noter que cette définition met en lumière les motivations derrière les fake news. Elle en distingue deux, politique et commercial. Selon Wardle et Derakhshan (2017), il existe une troisième motivation : la motivation sociale.

Enfin, malgré des différences entre les définitions, nous avons constaté un point commun essentiel : les fake news cherchent à se faire passer pour des informations légitimes. Les fake news se cachent sous un vernis de légitimité (Tandoc et al., 2017). Elles ont alors une apparence crédible, rendant leur détection plus difficile pour le public.

D. Différences avec d'autres fausses informations

Il nous paraît important de faire une distinction claire entre notre définition des fake news et d'autres concepts qui, bien qu'étant également de fausses informations, ne correspondent pas à notre définition spécifique.

Nous excluons donc de notre définition les erreurs journalistiques, la satire et la parodie d'actualité.

Tout d'abord, les erreurs journalistiques résultent de mauvaises sources ou d'inexactitudes involontaires, sans intention de tromper. Le journalisme est régi par des codes de conduite éthique qui exigent des journalistes qu'ils recherchent la vérité et la rapportent (Kalsnes, 2018).

La satire, quant à elle, exagère des faits réels de manière humoristique. Tandoc et al. (2017) citent par exemple l'émission *The Daily Show* sur *Comedy Central*. Dans le contexte francophone, nous pouvons penser aux *Guignols de l'info*. Ces émissions sont avant tout des émissions de divertissement et non d'information, dès lors nous ne les considérons pas comme des fake news.

Enfin, la parodie invente des informations absurdes pour critiquer ou amuser. *The Onion* est un exemple de site parodique cité par Tandoc et al. (2017) ; nous pouvons également penser à d'autres sites comme *Nordpresse*, une parodie de *Sudpresse*, ou *Le Gorafi*, parodiant *Le Figaro*.

Ces sites parodiques ont parfois été pris pour de véritables sources d'informations mais leur but reste humoristique et non trompeur.

Il est important de noter que la parodie et la satire supposent que l'auteur et le lecteur partagent le même sens de l'humour. En cas de mauvaises interprétations, certains lecteurs peuvent se faire piéger mais l'intention initiale reste non malveillante (Tandoc et al., 2017).

E. Conclusion

Cette exploration approfondie des fake news nous a permis de mettre en lumière leur évolution dans le contexte des nouvelles technologies de l'information et de la communication.

Ensuite, nous avons défini avec précision ce concept complexe en parcourant la littérature existante. Nous nous sommes basés sur la définition de Kalsnes, cette définition prenant en compte l'intention de nuire derrière la création des fake news ainsi que leur apparence de légitimité.

Enfin, nous avons distingué les fake news de concepts similaires tels que l'erreur journalistique, la satire et la parodie d'information.

IV. L'intelligence artificielle générative

A. Introduction

Ce chapitre aborde l'essor de l'intelligence artificielle générative. Nous commencerons par une remise en contexte de l'intelligence artificielle générative et son possible impact sur les fake news.

Ensuite, nous dégagerons une définition claire de l'intelligence artificielle générative et des concepts qui en découlent.

Nous parlerons alors du fonctionnement des modèles génératifs et plus particulièrement des grands modèles de langage (LLM)

Enfin, nous terminerons ce chapitre en analysant les limites des outils d'intelligence artificielle générative.

Ce chapitre se propose ainsi de définir ce qu'est l'intelligence artificielle générative, d'en explorer le fonctionnement et d'en analyser les capacités ainsi que les limites, en mettant l'accent sur les risques liés à la désinformation.

B. Contexte

Il aura fallu seulement cinq jours à *ChatGPT* pour atteindre un million d'utilisateurs et deux mois pour atteindre les cent millions, ce qui en fait l'application numérique à la croissance la plus rapide de l'histoire (Euchner, 2023). Cette adoption massive a suscité fascination, intérêt et inquiétude, notamment concernant son impact sur la diffusion de fake news

En effet, l'intelligence artificielle générative permet de créer facilement du contenu textuel, visuel ou sonore. Cette capacité rend la production de fake news plus rapide, plus accessible et potentiellement plus crédible. Selon Gold & Fischer (2023), l'IA générative risque de déclencher le prochain cauchemar de la désinformation ; le public n'étant bientôt plus en mesure de discerner le vrai du faux (Metz, 2023).

C. Définition

Selon Euchner (2023, p.1), l'IA générative utilise un très grand corpus de données - textes, images ou autres données étiquetées - pour créer, à la demande des utilisateurs, de nouvelles versions de textes, d'images ou de données prédites « *generative AI uses a very large corpus*

of data—text, images, or other labelled data—to create, at the request of users, new versions of text, images, or predicted data » [traduction libre].

Galaz et al. (2023, p.5), eux, définissent l'intelligence artificielle comme des technologies qui utilisent le *machine learning* y compris les méthodes de *deep learning* (« *technologies that employ machine learning including deep learning methods* ») [traduction libre].

L'intelligence artificielle générative, quant à elle, fait référence à une intelligence artificielle reposant sur des réseaux de neurones avancés, capable de produire des textes, images, vidéos et audios synthétiques très réalistes (y compris des histoires fictives, des poèmes et du code informatique), avec peu ou pas d'intervention humaine (« *refers to AI based on advanced neural networks with the ability to produce highly realistic synthetic text, images, video and audio (including fictional stories, poems, and programming code) with little, to no human intervention* », op. cit.) [traduction libre].

Feuerriegel et al. (2023) expliquent que l'IA générative fait référence à des techniques informatiques capables de générer un contenu apparemment nouveau et significatif, comme du texte, des images ou de l'audio, à partir de données d'entraînement (« *refers to computational techniques that are capable of generating seemingly new, meaningful content such as text, images, or audio from training data* ») [traduction libre].

Toujours selon les auteurs, les systèmes d'IA générative peuvent non seulement être utilisés à des fins artistiques pour créer de nouveaux textes imitant des écrivains ou de nouvelles images imitant des illustrateurs, mais ils peuvent également, et vont, assister les humains en tant que systèmes intelligents de questions-réponses « *generative AI systems can not only be used for artistic purposes to create new text mimicking writers or new images mimicking illustrators, but they can and will assist humans as intelligent question-answering systems »*) [traduction libre].

Ainsi, l'intelligence artificielle générative peut être définie comme un sous-domaine de l'intelligence artificielle dans lequel des modèles d'apprentissage automatique apprennent à produire de nouvelles données similaires à celles utilisées durant leur entraînement.

Cette capacité à produire automatiquement du contenu textuel, visuel ou sonore ouvre la voie à de nombreux usages, mais soulève également des enjeux majeurs en matière de désinformation.

D. Fonctionnement

1. Les modèles génératifs

Les modèles génératifs créent de nouvelles données à partir de leur entrainement, tandis que les modèles discriminatifs classent les données selon des catégories. Les premiers sont à la base de l'IA générative, et plus particulièrement des LLM (Nanda, 2024).

2. Les grands modèles de langage (LLM)

Les LLM reposent sur l'architecture des *transformers* (Vaswani et al., 2023 *;* Menon, 2023), qui permet de gérer des données séquentielles, comme du texte, de manière contextuel grâce au mécanisme d'attention. Ils sont entraînés sur d'immenses volumes de données non étiquetées, via un apprentissage auto-supervisé ou semi-supervisé (Bergman, 2023).

Ces modèles ne sont pas conçus pour une tâche spécifique, ils prédisent la suite la plus probable d'un texte, ce qui les rend polyvalents (Thome et Wolfe, 2023). Certains dépassent les 1 000 milliards de paramètres, donnant lieu à des capacités émergentes, c'est-à-dire des aptitudes non programmées et imprévisibles, comme la capacité à résoudre des problèmes logiques, réussir des examens universitaires ou synthétiser des textes complexes (Feuerriegel et al., 2023).

Les modèles de LLM les plus connus sont les modèles GPT (OpenAI), Bard/Gemini (Google), Claude (Anthropic) ou LLaMA (Meta).

3. Limites des LLM

Les LLM présentent plusieurs limites majeures :

Tout d'abord, la limite principale est leur tendance à produire des hallucinations, c'est-à-dire des affirmations factuellement fausses mais formulées de manière convaincante. Environ 19,5% des réponses de *ChatGPT* contiendraient des erreurs (Li et al. (2023).

Ensuite, les LLM utilisant d'immenses bases de données, ils héritent des biais présents dans leur corpus. Certains résultats sont influencés par des idéologies ou des stéréotypes (Durmus et al., 2023 ; Santurkar et al., 2023),

Même involontaires, ces biais peuvent renforcer des stéréotypes ou conduire à des discriminations. Si des efforts sont faits pour les atténuer, le dé-biaisage reste un défi technique.

En outre, les LLM sont qualifiés de « boîtes noires » (Steyer et al., 2021). Avec des milliards de paramètres, il est impossible de comprendre précisément pourquoi une réponse a été générée.

Comme le souligne le rapport de l'Office parlementaire d'évaluation des choix scientifiques et technologiques (2024, p.118) : « Les réseaux de neurones profonds, surtout avec leurs milliards de paramètres, sont si complexes qu'il n'est plus possible - même pour les meilleurs développeurs - d'expliquer pourquoi telles ou telles entrées parviennent à telles ou telles sorties ; seules les entrées et les sorties du système peuvent être observées ».

De plus, les données d'entraînement ne sont pas toujours publiques, notamment chez *OpenAI*, rendant alors l'audit difficile (op. cit.).

Enfin, d'autres limites existent tels que la cohérence limitée sur les longs textes ou encore les informations qui peuvent être obsolètes en fonction de la date des données d'entraînement. La consommation énergétique de tels outils ne doit également pas être minimisée.

4. Conclusion

En résumé, nous avons vu dans ce chapitre que l'intelligence artificielle générative a connu une croissance fulgurante et conquis le public. Cette démocratisation soulève des enjeux importants en termes de désinformation, la création de fake news étant facilitée par ces nouveaux outils.

Nous avons ensuite clarifié la notion d'intelligence artificielle générative en nous basant sur la littérature existante. Il en ressort qu'il s'agit d'un sous-domaine de l'intelligence artificielle, reposant sur des modèles capables de produire de nouvelles données similaires à celles rencontrées lors de leur entraînement.

Nous nous sommes alors concentrés sur le fonctionnement de ces modèles, en particulier les grands modèles de langage (LLM). Grâce à l'architecture des *transformers* et à un nombre colossal de paramètres, ces modèles ont acquis des capacités impressionnantes et parfois même non prévues lors de leur conception.

Enfin, nous avons mis en évidence les limites de ces outils, telles que leur tendance à générer des contenus erronés, la reproduction de biais issus de leurs données d'apprentissage et leur opacité de fonctionnement.

V. Crédibilité des fake news

A. Introduction

Les fake news ne se propagent pas uniquement parce qu'elles existent mais parce qu'elles parviennent à convaincre. Elles apparaissent comme vraisemblable, légitimes et crédibles aux yeux des lecteurs. Ce chapitre explore les mécanismes qui renforcent cette crédibilité.

Nous analyserons d'abord les facteurs contextuels tels que la présentation visuelle ou l'appel à l'émotion, qui renforcent artificiellement la légitimité perçue des fake news. Ensuite, nous étudierons les caractéristiques individuelles influençant la crédulité comme le niveau d'éducation, l'âge ou la littératie numérique.

Ce chapitre vise ainsi à identifier les mécanismes contextuels et individuels qui favorisent la crédibilité des fake news afin de mieux comprendre les conditions de leur succès.

B. Facteurs externes renforçant la crédibilité

Tout d'abord, selon Fernández-López et Perea, (2020), la présentation visuelle et structurelle d'une fake news renforce sa crédibilité. Une mise en page soignée et une interface facile à naviguer augmentent la confiance des lecteurs. Les jeunes adultes, en particulier, jugent une information plus fiable lorsque le site parait sérieux.

Ensuite, l'utilisation de chiffres constitue également un levier classique pour renforcer la crédibilité. Statistiques, pourcentages ou données chiffrées donnent une illusion de rigueur scientifique. A l'inverse, des termes vagues comme « beaucoup » ou « certains experts » suscite davantage de doute (Koetsenruijter, 2011)

Un autre levier fréquemment utilisé est l'appel à l'émotion. Les créateurs de fake news exploitent des émotions fortes comme la peur, la tristesse ou la colère pour susciter l'indignation. Un message émotionnellement chargé sera perçu comme plus authentique qu'un message neutre. (Fernández-López & Perea, 2020)

C. Facteurs individuels

D'autres variables, plus individuelles, jouent également un rôle dans la crédibilité perçue. Selon Beauvais (2022), le niveau de connaissance et d'éducation constitue un facteur déterminant : un faible bagage scientifique ou un bas niveau d'étude favorisent la crédulité. A

l'inverse, des études plus poussées ou une éducation numérique semble protéger en partie contre la désinformation.

Toutefois, une méta-analyse du Max Planck Institute nuance ce constat : des individus très éduqués peuvent se montrer tout aussi vulnérables aux fake news que ceux ayant un niveau d'éducation plus faible (Sultan et al., 2024)

En outre, l'âge joue également un rôle. Contrairement aux idées reçues, les jeunes adultes ne sont pas toujours les mieux armés pour détecter les fake news. Selon la méta-analyse de Sultan et al. (2024), les personnes âgées ont tendance à classer plus souvent des informations comme fausses, adoptant une approche plus méfiante. Les jeunes adultes, bien qu'à l'aise avec les outils numériques, peuvent se montrer plus naïfs.

Enfin, la familiarité avec les outils numériques ou l'intelligence artificielle influe aussi. Selon Eshet-Alkalai (2004), la littératie numérique (*digital literacy*) ne se résume pas à savoir utiliser un ordinateur, elle désigne un ensemble de compétences cognitives et critiques notamment évaluer la qualité et la validité d'une information en faisant preuve de recul face aux contenus rencontrés.

Plus spécifiquement, la *AI literacy* (Ng et al., 2021), à savoir la compréhension du fonctionnement de l'IA, de ses limites et de ses biais, joue aussi un rôle. Une personne familière avec ces outils pourra plus facilement détecter des signes de génération artificielle et accorder moins de crédibilité à une fake news produite par IA.

D. Conclusion

Ce chapitre nous a permis d'explorer les mécanismes rendant les fake news crédibles aux yeux du public.

Nous avons analysé les facteurs externes renforçant la crédibilité perçue comme la présentation visuelle, l'utilisations de données chiffré ou l'appel à l'émotion.

Nous nous sommes ensuite concentrés sur les facteurs individuels, comme l'âge, le niveau d'éducation la littératie numérique et la connaissance en intelligence artificielle, qui influencent la propension à croire de fausses informations.

VI. Conclusion

Cette revue de littérature a permis d'explorer trois axes fondamentaux pour la compréhension de notre problématique : les fake news, l'intelligence artificielle générative, et les facteurs influençant la crédibilité perçue.

Nous avons tout d'abord défini les fake news en nous basant sur la définition de Kalsnes, qui insiste à la fois sur l'intention de nuire ainsi que sur l'apparence de légitimité de ces contenus. Nous avons également distingué les fake news de l'erreur journalistique, la satire et la parodie.

Nous nous sommes ensuite penchés sur l'intelligence artificielle générative, considérée comme un sous-domaine de l'IA capable de produire des données originales à partir d'un corpus d'apprentissage. Nous avons détaillé le fonctionnement des LLM, rendus possible par l'architecture des *transformers* et un nombre massif de paramètres. Nous avons aussi mis en évidence les limites de ces outils, notamment les hallucinations, les biais hérités et leur opacité méthodologique.

Enfin, nous avons étudié les facteurs qui influencent la perception de crédibilité des fake news crédibles en distinguant les facteurs contextuels des facteurs individuels.

Cette synthèse théorique permet désormais de formuler des hypothèses de recherche solides, qui seront explorées empiriquement dans la suite de ce mémoire à travers une étude quantitative.

Etude empirique

VII. Introduction

À la lumière des enseignements tirés de la revue de littérature, cette seconde partie adopte une approche empirique afin d'examiner l'impact de l'intelligence artificielle générative sur la crédibilité perçue des fake news.

Nous commencerons par formuler nos hypothèses de recherche, fondées sur les facteurs identifiés comme influençant la crédibilité dans la littérature existante.

Nous détaillerons ensuite la méthodologie adoptée, les variables mesurées, les échelles de mesure utilisées, la procédure de recueil des données, ainsi que les outils statistiques mobilisés pour tester nos hypothèses.

Enfin, nous présenterons les résultats obtenus, que nous discuterons à la lumière du cadre théorique, tout en soulignant les biais et limites de l'étude.

VIII. Méthodologie

A. Introduction

Dans ce chapitre, nous parlerons de la méthodologie adoptée pour étudier l'impact de l'intelligence artificielle générative sur la crédibilité perçue des titres de fake news. Après avoir rappelé notre problématique et défini nos hypothèses de recherche, nous détaillerons les choix méthodologiques effectués.

Nous évoquerons les variables utilisées, le déroulement de l'expérience, les outils de collecte de données les échelles de mesure utilisées ainsi que les méthodes d'analyse statistique.

L'objectif est de garantir la reproductibilité de l'étude, d'assurer la validité des résultats et de répondre de manière rigoureuse aux questions de recherche posées.

B. Hypothèses de recherches

Notre question de recherche étant « Quel est l'impact de l'utilisation de l'intelligence artificielle générative sur la crédibilité perçue des titres de fake news ? », nous avons formulée 5 hypothèses principales à partir de la revue de littérature, accompagnées pour certaines d'hypothèses complémentaires (notées « b ») explorant les effets d'interaction.

- H1 : Les titres de fake news générés par une intelligence artificielle générative sont perçus comme plus crédibles que ceux rédigés par des humains.

Certains travaux récents (Luo et al., 2022 ; Tseng et Yuan, 2023) montrent que les fake news sont perçues comme aussi crédibles, voire légèrement moins crédibles que les vraies informations. Toutefois, aucune de ces études ne s'est penchée sur le cas spécifique des fake news générées par IA. C'est précisément cette nuance que nous souhaitons explorer : nous posons l'hypothèse que les fake news générées par IA pourraient être perçus comme plus crédibles que les fake news humaines

- H2 : Les jeunes adultes perçoivent les titres de fake news générés par IA comme plus crédibles que les adultes plus âgés.
- H2b : L'effet de l'âge sur la crédibilité perçue varie selon que le titre est généré par une IA ou par un humain.

Contrairement aux idées reçues, les jeunes adultes ne sont pas nécessairement mieux armés pour détecter les fake news. Selon Sultan et al. (2024), les personnes âgées adoptent souvent une posture plus méfiante, ce qui les conduit à classer davantage de contenus comme faux.

- H3 : Les individus ayant un niveau d'études élevé perçoivent les titres de fake news générés par IA comme moins crédibles que ceux ayant un niveau d'études plus faible
- H3b : L'effet du niveau d'études sur la crédibilité perçue varie selon que le titre est généré par une IA ou par un humain.

Le niveau d'éducation influence la capacité à évaluer de manière critique les informations reçues. D'après Beauvais, (2022), une éducation plus poussée favorise l'esprit critique et la remise en question, même face à des contenus convaincants. Ainsi, les personnes plus diplômées, souvent mieux formées à l'analyse de l'information, seraient donc plus à même de détecter les incohérences ou les signaux de génération artificielle, rendant les titres générés par IA moins crédibles à leurs yeux.

- H4 : Les individus présentant un attrait pour la technologie (à savoir un score élevé à l'échelle de *Technology Readiness Index*) perçoivent les titres de fake news générés par IA comme moins crédibles que les individus ayant un score plus faible.
- H4b : L'effet du *Technology Readiness Index* sur la crédibilité perçue varie selon que le titre est généré par une IA ou par un humain.

Un haut niveau de *Technology Readiness Index* suppose une familiarité avec les outils technologiques, mais aussi une meilleure conscience de leurs limites. Comme le souligne Eshet-Alkalai (2004), la littératie numérique comprend des compétences critiques. Les individus attirés par la technologie pourraient être plus vigilants face aux biais ou imperfections propres aux contenus générés par IA.

- H5 : Les individus se percevant comme compétents dans l'utilisation de l'intelligence artificielle (à savoir un score élevé à l'échelle *Measuring user competence in using artificial intelligence*) perçoivent les titres de fake news générés par IA comme moins crédibles que ceux se percevant moins compétents.
- H5b : L'effet du sentiment de compétence en IA sur la crédibilité perçue varie selon que le titre est généré par une IA ou par un humain.

Selon Ng et al., (2021), l'*AI literacy*, c'est-à-dire la compréhension du fonctionnement, des limites et des biais de l'IA, permet une prise de recul face aux contenus générés artificiellement. Les individus se percevant comme compétents dans l'utilisation de ces outils seraient ainsi plus enclins à repérer des indices de génération artificielle et, par conséquent, à accorder moins de crédibilité aux titres produits par IA.

C. Variables de recherche

1. Variable dépendante

Pour construire cette étude, nous nous sommes basés sur Tseng et Yuan, (2023). La variable dépendante est le score de crédibilité attribué aux titres de fake news.

Chaque participant a évalué 10 titres à l'aide d'une échelle de Likert à 7 points, allant de 1 (Pas du tout crédible) à 7 (Tout à fait crédible).

2. Variables indépendantes

Variable	*Nom*	*Description*	*Type*	*Lien avec l'hypothèse*
Origine du titre	*IsAI*	Titre de fake news généré par une IA ou rédigé par un humain	Manipulée (expérimentale)	H1
Age	*AGE*	Age du participant (en années)	Mesurée (quantitative)	H2 / H2b
Niveau d'études	*EDU*	Plus haut diplôme obtenu	Mesurée (catégorielle)	H3 / H3b

Technology Readiness Scale	*Score_TECH*	Score à l'échelle « Technology Readiness Scale »	Mesurée (quantitative)	H4 / H4b
Measuring user competence in using AI)	*Score_AI*	Score à l'échelle "Measuring user competence in using AI"	Mesurée (quantitative)	H5 / H5b

Table 1 : Tableau des variables indépendantes

3. Variables de contrôle

Certaines variables ont été collectées à des fins de vérification ou de contrôle de biais potentiels :

Variable	*Nom*	*Description*	*Rôle possible*
Genre	*SEX*	Homme / Femme / Autre / Préfère ne pas dire	Peut influencer la perception ou l'attention
Langue maternelle	*ENG*	Vérification que les répondants comprennent bien l'anglais	Contrôle de compréhension / biais linguistique

Table 2 : Tableau des variables de contrôle

D. Procédure

L'enquête est administrée en ligne, en anglais, via un formulaire Microsoft Forms. Le lien de participation est diffusé sur les réseaux sociaux, des forums ainsi que des sites de partage d'enquête, afin d'obtenir un échantillon international et diversifié.

Les participants sont répartis aléatoirement en deux groupes :

- Groupe A : exposition à dix titres de fake news générés par intelligence artificielle.
- Groupe B : exposition à dix titres de fake news réels.

Afin de scinder les participants en deux groupes, l'outil « *Allocate.Monster* » est utilisé (Ferguson, 2016).

Les fake news réelles sont extraites de sites de vérification des faits reconnus tels que *Snopes*, *FactCheck.org*, *Reuters* et *Politifact*. Cinq d'entre elles portent sur la pandémie de COVID-19, et cinq sur des sujets politiques.

Les titres générés par IA proviennent de *ChatGPT* (version gratuite, juillet 2024). Dans le but d'éviter la censure, il lui est demandé de produire différents titres de fausses nouvelles crédibles et sensationnalistes sur une pandémie fictive ainsi que sur des tensions géopolitiques fictifs. Les termes « COVID-19, » les noms des pays et de dirigeants sont ensuite ajoutés manuellement.

Avant de commencer, les participants doivent lire et accepter un consentement éclairé, précisant que les réponses sont anonymes, qu'ils ne doivent pas chercher à vérifier les titres sur Internet pendant l'évaluation et que la participation est volontaire, ils peuvent quitter l'enquête à tout moment.

Les participants doivent alors évaluer les dix titres présentés sur une échelle de Likert à 7 points, allant de 1 (Pas du tout crédible) à 7 (Tout à fait crédible).

Ils remplissent ensuite les deux échelles utilisées, *Technology Readiness Index* et *Measuring User Competence in Using AI.*

Enfin, deux questions de contrôle sont posées à la fin :

- "Avez-vous répondu au hasard à certaines questions ?"
- "Avez-vous consulté Internet pour vérifier certains titres pendant l'enquête ?"

Le questionnaire se conclut par des questions sociodémographiques : âge, genre, plus haut diplôme obtenu, ainsi que leur maitrise de l'anglais, afin de vérifier leur compréhension des consignes et des titres de fake news.

E. Echelles de mesure

Les réponses sont recueillies sur une échelle de Likert à 7 points, de « Pas du tout d'accord » / « Pas du tout crédible » à « Tout à fait d'accord » / « Tout à fait crédible ». Les échelles de mesure complètes sont disponibles en annexe (voir Annexes 1 à 3).

Variable mesurée	*Échelle / Source*	*Nombre d'items*
Crédibilité des fake news	Score moyen des évaluations des 10 titres présentés	10 (5 politiques, 5 COVID)
Technology Readiness	Technology Readiness Index (Seong & Hong, 2022)	7
Compétence perçue en IA	Measuring User Competence in Using AI (Wang et al., 2023)	12

Table 3 : Tableau des échelles de mesure

F. Traitement des données et analyses préliminaires

Avant de procéder aux analyses principales, nous avons réalisé différentes vérifications préliminaires de nous assurer de la qualité des données et de la validité échelles de mesures.

1. Validité des échelles de mesure

Nous avons utilisé deux échelles de mesures lors de l'enquête.

Pour chacune de ces échelles, nous avons évalué la cohérence interne à l'aide de l'alpha de Cronbach avec un seuil de $\alpha \geq 0.70$ comme critère d'acceptabilité.

Ensuite, nous avons effectué une analyse en composantes principales (ACP) afin d'examiner l'unidimensionnalité des échelles.

Les critères d'acceptation retenus pour l'ACP sont un indice KMO (Kaiser-Meyer-Olkin) supérieur ou égal à 0,70 et une *eigenvalue* du premier facteur supérieure à 1, selon la règle de Kaiser.

Lorsque l'ACP révélait plusieurs dimensions sous-jacentes, des scores factoriels distincts ont été calculé pour chaque facteur principal et utilisés séparément dans les analyses de régression. Dans le cas contraire, une moyenne des scores des différents items a été calculée afin d'obtenir un score composite unique.

Ces scores sont ensuite utilisés comme variables indépendantes dans les analyses principales.

2. Vérification de la cohérence des fake news

Les titres de fake news ayant été choisis arbitrairement, il est possible que certains titres trop peu crédibles aient été utilisés. Afin de garantir la fiabilité de l'échelle de crédibilité perçue, nous avons calculé la moyenne et l'écart type de chaque titre de fake news.

Un titre a été considéré comme trop peu crédible, peu pertinent, si sa moyenne était inférieure à 1,5 (forte non-crédibilité) et son écart-type était inférieur à 1 (manque de dispersion). Nous avons défini ce seuil de manière pragmatique : sur une échelle de 1 à 7, une moyenne inférieure à 1,5 reflète un consensus fort autour de la non-crédibilité. Le seuil de 1 pour l'écart-type permet d'écarter les titres ne suscitant que peu de divergence d'opinion, donc peu informatifs.

Les titres répondant à ces critères ont été exclus des analyses pour éviter de biaiser les moyennes globales.

Nous avons alors créé la variable *mean_total* avec les titres restants pour représenter la crédibilité perçue :

- *mean_total* : moyenne globale sur l'ensemble des titres retenus

3. Homogénéité des groupes

Avant d'analyser les effets de l'origine des titres (générés par IA ou rédigés par des humains) sur la crédibilité perçue, il est essentiel de vérifier que les deux groupes expérimentaux étaient homogènes et comparables.

Des tests d'homogénéité ont donc été réalisés entre le groupe IA et le groupe Humain :

Pour les variables catégorielles (*SEX, EDU, LANG*), nous avons effectué des tests de chi-deux.

Pour les variables continues (*AGE, Score_TRI, Score_AI*), nous avons tout d'abord effectué un test de Shapiro-Wilk afin de déterminer si ces variables suivaient une loi normale ou non.

Ensuite, un test de student a été effectué pour les variables suivant une distribution normale et un test de Mann-Whitney pour les autres.

4. Corrélations exploratoires

Afin d'examiner les relations initiales entre les variables indépendantes et la crédibilité perçue des titres de fake news, nous avons calculé la corrélation entre l'ensemble des variables.

Pour permettre cette analyse, les variables catégorielles (*SEX, EDU* et *LANG*) ont été transformées en variables ordinales. Nous avons ensuite utilisé la corrélation de Spearman pour l'ensemble des variables, afin de tenir compte de la nature ordinale ou non normale de certaines d'entre elles.

G. Analyse

1. Hypothèse 1

Dans le cadre de l'analyse, nous avons comparé les moyennes de crédibilité perçue (*mean_total*) entre les deux groupes.

$$\begin{cases} H_0 : \mu_{\text{humain}} = \mu_{ia} \text{ (il n'y a pas de différence entre les moyennes de crédibilité perçue des titres)} \\ H_1 : \mu_{\text{humain}} \neq \mu_{ia} \text{ (il existe une différence significative entre les deux moyennes)} \end{cases}$$

Lorsque la distribution suivait une loi normale, nous avons utilisé un test *t* de Student ; dans le cas contraire, un test de Mann-Whitney a été appliqué.

2. Hypothèse 2

a) Régression simple

La relation entre l'âge et la crédibilité perçue (*mean_total*) a été examinée uniquement dans le groupe exposé aux titres de fake news générées par IA au moyen d'une régression linéaire simple :

$$mean\_total_i = \beta_0 + \beta_1 . Age_i + \epsilon_i$$

Où

- $mean\_total_i =$ score moyen de crédibilité pour le participant i
- $Age_i =$ âge du participant i
- $\beta_0 =$ constante
- $\beta_1 =$ coefficient d'effet de l'âge
- $\epsilon_i =$ erreur résiduelle

Les hypothèses statistiques sont les suivantes :

$$\begin{cases} H_0 : \beta_1 = 0 \text{ (l'âge n'a aucun effet sur la crédibilité perçue)} \\ H_1 : \beta_1 \neq 0 \text{ (l'âge a un effet significatif sur la crédibilité perçue)} \end{cases}$$

b) Régression multiple

Une régression linéaire multiple a ensuite été effectuée sur l'ensemble de l'échantillon afin d'examiner l'effet de l'âge, en interaction avec l'origine du titre :

$$mean\_total_i = \beta_0 + \beta_1 . Age_i + \beta_2 . IsAI_i + \beta_3 . (Age_i \times IsAI_i) + \epsilon_i$$

Où

- $\beta_3 =$ effet d'interaction entre l'âge et le groupe
- $IsAI_i =$ variable binaire indiquant si le participant a vu des titres générés par IA (1) ou non (0)

Les hypothèses statistiques pour tester l'interaction sont :

$$\begin{cases} H_0 : \beta_3 = 0 \text{ (l'effet de l'âge est identique quel que soit le groupe)} \\ H_1 : \beta_3 \neq 0 \text{ (l'effet de l'âge dépend du groupe)} \end{cases}$$

Les conditions classiques de validité des modèles de régression ont été vérifiées pour chaque analyse :

- Normalité des résidus : test de Shapiro-Wilk
- Homoscédasticité : test de Breusch-Pagan
- Indépendance des erreurs : inspection des résidus.
- Multicolinéarité (dans le modèle multiple) : évaluation du VIF (Variance Inflation Factor), avec un seuil de VIF < 5 comme critère d'acceptabilité.

3. Les hypothèses 4 et 5

Des modèles similaires ont été utilisés pour tester les effets du *Technology Readiness* (H4) et de la compétence perçue en IA (H5)

a) Hypothèse 4

$$\begin{cases} H_0 : \beta_1 = 0 \text{ (le score de Technology Readiness n'a aucun effet)} \\ H_1 : \beta_1 \neq 0 \text{ (le score a un effet significatif)} \end{cases}$$

b) Hypothèse 4b

$$\begin{cases} H_0 : \beta_3 = 0 \text{ (l'effet est identique dans les deux groupes)} \\ H_1 : \beta_3 \neq 0 \text{ (l'effet est différent selon le groupe)} \end{cases}$$

c) Hypothèse 5

$$\begin{cases} H_0 : \beta_1 = 0 \text{ (le score n'a aucun effet)} \\ H_1 : \beta_1 \neq 0 \text{ (le score a un effet significatif)} \end{cases}$$

d) Hypothèse 5b

$$\begin{cases} H_0 : \beta_3 = 0 \text{ (l'effet est identique dans les deux groupes)} \\ H_1 : \beta_3 \neq 0 \text{ (l'effet est différent selon le groupe)} \end{cases}$$

e) Echelles à plusieurs facteurs

Si l'analyse factorielle révèle plusieurs dimensions significatives pour une échelle, des scores distincts sont extraits pour chaque facteur et utilisés séparément dans les régressions.

(1) Régression simple à deux facteurs

$$mean\_total_i = \beta_0 + \beta_1 . F1_i + \beta_2 . F2_i + \epsilon_i$$

$$\begin{cases} H_0 : \beta_1 = \beta_2 = 0 \text{ (aucun des deux facteurs n'a d'effet)} \\ H_1 : \beta_1 \neq 0 \text{ ou } \beta_2 \neq 0 \text{ (au moins un facteur influence la crédibilité perçue} \end{cases}$$

(2) Régression multiple avec interaction

$$mean\_total_i = \beta_0 + \beta_1 . F1 + \beta_2 . F2_i + \beta_3 . IsAI_i + \beta_4 . (F1_i \times IsAI_i) + \beta_5 . (F2_i \times IsAI_i) + \epsilon_i$$

$$\begin{cases} H_0 : \beta_4 = \beta_5 = 0 \text{ (pas d'effet modérateur)} \\ H_1 : \beta_4 \neq 0 \text{ ou } \beta_5 \neq 0 \text{ (au moins une interaction est significative} \end{cases}$$

4. Hypothèse 3

La variable *EDU* étant une variable catégorielle, nous avons d'abord vérifié la normalité des résidus issus d'un modèle simple reliant cette variable à la crédibilité perçu (*mean_total*).

Lorsque les résidus suivaient une loi normale et que l'homogénéité des variances était vérifiée (test de Levene), une ANOVA unidirectionnelle a été appliquée. Dans le cas contraire, un test de Kruskall-Wallis a été privilégié.

Les hypothèses statistiques sont :

$$\begin{cases} H_0 : \mu_{EDU1} = \mu_{EDU2} = \ldots = \mu_{EDU8} \text{ (Il n'y a pas de différence selon le niveau d'études)} \\ H_1 : \exists\, (i, j) : \mu_{EDU1} = \mu_{EDU2} \ \beta_4 \neq 0 \text{ ou } \beta_5 \neq 0 \text{ (au moins une interaction est significative} \end{cases}$$

Concernant l'hypothèse 3b, si les conditions de normalité des résidus et d'homogénéité des variances étaient respectés, une ANOVA à deux facteurs a été effectuée. Sinon, une régression multiple a été utilisée en transformant *EDU* en variable muette (*dummy*) afin de tester l'effet de l'interaction (*EDU* * *IsAI*) sur la crédibilité perçue.

5. Vérification de la multicolinéarité

À la suite des analyses de corrélation, la possibilité d'une multicolinéarité entre les variables explicatives n'a pas pu être exclue.

Un modèle de régression linéaire multiple a donc été estimé sur le groupe ayant reçu des titres de fake news générés par IA en intégrant l'ensemble des variables :

$$mean\_total_i = \beta_0 + \beta_1 . Age_i + \beta_2 . EDU_i + \beta_3 . Score\_TECH_i + \beta_4 . Score\_AI_i + \beta_5 . SEX_i + \beta_6 . LANG_i + \epsilon_i$$

Ce modèle vise à évaluer l'impact spécifique de chaque variable sur la crédibilité perçue, en contrôlant statistiquement les autres.

La significativité des coefficients a été examinée afin de déterminer si chaque prédicteur conservait un effet une fois les autres inclus dans le modèle.

Enfin, les indices de facteur d'inflation de la variance (VIF) ont été calculés pour diagnostiquer d'éventuelles colinéarités problématiques. Un VIF supérieur à 5 a été considéré comme un seuil d'alerte potentiel.

H. Conclusion

L'ensemble des éléments méthodologiques présentés visent à assurer la robustesse de l'étude.

En combinant un plan expérimental contrôlé avec des échelles validées, des tests de qualité des données et des analyses statistiques adaptées, cette méthodologie permet d'examiner de manière l'effet de l'origine des fake news sur leur crédibilité perçue, ainsi que l'influence de variables individuelles telles que l'âge, le niveau d'études, la familiarité avec la technologie et la compétence perçue en IA. Les résultats issus de ces analyses seront présentés et interprétés dans le chapitre suivant.

IX. Résultats

A. Introduction

Ce chapitre présente les résultats obtenus à la suite de l'analyse des données recueillies. Les résultats sont organisés en fonction des hypothèses formulées dans le cadre théorique et testées à l'aide des méthodes statistiques décrites dans le chapitre précédent. Les analyses portent à la fois sur les effets principaux et les effets modérateurs des différentes variables sur la crédibilité perçue des titres de fake news.

B. Analyse préliminaire

Nous avons recueilli 151 réponses. Après avoir exclu les personnes ayant répondu au hasard ou ayant recherché les titres sur un moteur de recherche, nous obtenons un échantillon final de 145 répondants : 76 dans le groupe exposé à des fake news générées par IA et 69 dans le groupe ayant reçu de véritables fake news (Annexe 4)

1. Validité des échelles de mesures

a) Technology Readiness Index

Tout d'abord, nous avons vérifié la fiabilité de l'échelle TRI (Technology Readiness Index).

Nous avons évalué la cohérence interne à l'aide de l'alpha de Cronbach, qui s'élève à 0,823, ce qui indique une bonne fiabilité.

L'indice KMO est de 0,807, et le test de Bartlett est significatif ($\chi^2 = 334,53$; $p < 0,001$), indiquant que les données se prêtent à une réduction de dimensionnalité.

L'analyse en composantes principales (ACP) révèle deux facteurs avec des *eigenvalues* supérieures à 1. Ces deux facteurs ont une *eigenvalue* de 3,44 et 1,08 et expliquent ensemble 64,17% de la variance totale (Annexes 5 et 6).

Le tableau en annexe (Annexe 7) présente les charges factorielles associées à la solution à deux facteurs, après rotation Varimax.

Nous pouvons observer un regroupement clair des items sur deux dimensions. La cohérence interne de chaque dimension a été vérifiée :

- *Tech_Score_F1* : $\alpha = 0,796$

- *Tech_Score_F2* : α = 0,739

Ces résultats confirment la bonne fiabilité des deux dimensions identifiées dans l'échelle. Pour chaque facteur, un score moyen a été calculé par participant.

b) Measuring User Competence in Using AI

Nous nous sommes ensuite intéressés à la structure de *l'échelle Measuring User Competence in Using AI*.

L'alpha de Cronbach initial est de 0,741, ce qui reste satisfaisant. L'indice KMO est de 0,772 et le test de Bartlett est significatif ($\chi^2 = 425,44$, $p < 0,0001$), cela nous indique que les données sont adéquates.

L'analyse en composantes principales (ACP) révèle trois composantes ayant une *eigenvalue* supérieure à 1, suggérant une structure tridimensionnelle. Toutefois, l'analyse des charges factorielles montre une répartition peu claire, avec de nombreuses charges croisées. Le troisième facteur repose uniquement sur deux items, avec une cohérence interne très faible (α = 0,19) (Annexe 9). Les alphas des trois facteurs sont respectivement de 0,78 ; 0,60 et 0,19.

L'observation du *scree plot* (Annexe 8) montre un coude net après la deuxième composante, ce qui nous a conduit à tester une solution à deux facteurs. Dans cette configuration, les items AI1 à AI9 chargent principalement sur le premier facteur, tandis que les items AI10 à AI12 forment un second facteur plus faible (Annexe 10). Les alphas de Cronbach obtenus sont respectivement de 0,798 pour le premier facteur et de 0,604 pour le second.

En raison de la faible cohérence interne du second facteur (3 items seulement, α < 0,7), seul le premier facteur a été conservé pour les analyses ultérieures. Il présente une bonne fiabilité interne (α = 0,798) ainsi qu'un indice KMO satisfaisant (0,816)

Un score unidimensionnel (*AI_Score*) a ainsi été construit en faisant la moyenne des items AI1 à AI9 pour chaque participant.

2. Vérification de la cohérence des fake news

Nous avons calculé, pour chaque item et par groupe (fake news humaines et fake news IA), la moyenne et l'écart-type des réponses. Les résultats sont présentés dans le tableau en annexe (Annexe 11).

Aucun item ne présente de moyenne extrême ni d'écart-type anormalement faible, ce qui indique une bonne variabilité des réponses, aucun item ne justifie donc d'être exclu.

3. Homogénéité des groupes

L'homogénéité des groupes expérimentaux a été vérifiée à l'aide des tests appropriés en fonction du type de variable et de leur distribution. Les résultats sont synthétisés dans le tableau en annexe (Annexe 12).

Tous les tests confirment l'homogénéités des groupes à l'exception de la variable *AI_Score*. Les participants ayant été répartis aléatoirement, cette hétérogénéité peut être attribuée au hasard ou à une fluctuation d'échantillonnage.

4. Corrélation exploratoire

Avant de tester nos hypothèses, nous avons effectué une analyse exploratoire des corrélations. Les variables *EDU* et *LANG* ont été traitées comme ordinales. Un participant non-binaire a été exclu de cette analyse (mais conservé pour les autres) afin d'utiliser *SEX* comme variable binaire et de garantir la validité statistique (Annexe 13 et 14).

Une corrélation positive modérée est observée entre les deux dimensions de l'échelle *TRI* (r = 0,49 ; $p < 0,001$), ce qui est plutôt logique étant donné que ces variables proviennent de la même échelle. Ils présentent également une corrélation positive significative avec *AI_Score* ($r = 0,39$; $p < 0,001$).

Nous pouvons également observer des corrélations faibles mais significatives entres les scores techno/IA et le niveau d'éducation (r = 0,22 ; p = 0,008) ainsi qu'avec l'âge. Le genre est également corrélé négativement aux scores *TRI,* les scores seraient légèrement plus élevés chez les hommes.

Enfin, la maîtrise de l'anglais est modérément corrélée avec le niveau d'éducation ($r = 0,37$; $p < 0,001$).

C. Hypothèse 1

Nous avons comparé les scores moyens de crédibilité perçue selon l'origine des titres.

Le test de Shapiro-Wilk indique que la normalité est respectée pour le groupe humain (p = 0,521) mais rejetée pour le groupe IA ($p = 0,003$)

Un test non paramétrique de Mann-Whitney a alors été effectué, il révèle une différence significative entre les deux groupes ($p = 0{,}004$).

Les titres générés par IA ($\mu = 4{,}03$) sont perçus comme plus crédibles que ceux rédigés par des humains ($\mu = 3{,}69$).

Nous rejetons donc l'hypothèse nulle H_0 au profite de l'hypothèse alternative H_1 :

D. Hypothèse 2

1. Hypothèse 2.a

Une régression linéaire a été menée pour tester l'effet de l'âge sur la crédibilité perçue.

Nous obtenons $\beta_1 = -0{,}024$; $p = 0{,}015$. Le modèle est significatif ($R^2 = 0{,}078$; $F(1,74) = 6{,}245$; $p = 0{,}015$), mais les résidus ne sont pas normalement distribués ($p = 0{,}0033$) (Annexe 15).

Nous avons alors réalisé un *bootstrap* du coefficient β_1 (Annexe 16). Celui-ci montre un IC 95% : [- 0,026 ; 0,0033], incluant 0. L'effet de l'âge n'est pas significatif de manière robuste malgré la significativité dans le modèle initial.

Ce résultat peut s'expliquer par la distribution déséquilibrée de l'âge dans l'échantillon, majoritairement composée de jeunes adultes, et la sous-représentation des personnes plus âgées.

Bien que le modèle initial suggère une diminution de la crédibilité avec l'âge, l'effet n'est pas confirmé de manière robuste. Nous ne pouvons pas rejeter l'hypothèse nulle de manière rigoureuse, et l'hypothèse H2a n'est pas validée statistiquement, même si une tendance semble exister.

2. Hypothèse 2.b

Nous avons utilisé un modèle de régression multiple avec interaction $AGE \times IsIA$ (âge centré).

Le modèle est significatif ($R^2 = 0{,}101$; $F(3, 141) = 5{,}264$; $p = 0{,}0018$).

Les résultats indiquent un effet significatif de l'interaction : $\beta_3 = -0{,}032$; $p = 0{,}026$

Cela signifie que l'effet de l'âge diffère selon l'origine du titre. Il n'y a aucun effet significatif dans le groupe humain alors que dans le groupe IA, l'âge est associé à une baisse de 0,032 point par an sur la crédibilité perçue

Les tests d'hypothèses du modèle sont concluants hormis une légère déviation à la normalité des résidus (Annexe 17).

Le modèle valide donc H2b, en montrant que l'âge influence différemment la crédibilité perçue selon l'origine du titre. Toutefois, cette déviation à l'hypothèse de la normalité des résidus nous invite à interpréter ces résultats avec prudence.

E. Hypothèse 3

1. Hypothèse 3.a

La variable *EDU* étant catégorielle, nous avons réalisé une ANOVA unifactorielle afin de comparer la crédibilité perçue (*mean_total*) selon les niveaux d'éducation, uniquement dans le groupe IA.

Les conditions d'application du test sont respectées :

- Normalité des résidus : $p = 0{,}222$
- Homogénéité des variances : $p = 0{,}343$

Les résultats de l'ANOVA indiquent qu'il n'existe pas de différence significative entre les groupes : $F = 0{,}484$; $p = 0{,}747$

Nous ne pouvons pas rejeter l'hypothèse nulle. Le niveau d'éducation ne semble pas influencer la crédibilité perçue des titres générés par IA.

1. Hypothèse 3.b

Pour tester l'existence d'un effet d'interaction entre l'origine du titre et le niveau d'éducation, une ANOVA à deux facteurs a été réalisée (EDU, IsIA, et leur interaction).

Variable	*F*	*p-value*
C(EDU)	0,480	0,697
C(IsIA)	8,1	**0,005**
C(EDU) × C(IsIA)	0,693	0,598

Table 4 : ANOVA à deux facteurs

Seul le facteur principal *IsIA* est significatif, indiquant que les titres générés par IA sont globalement perçus comme plus crédibles, indépendamment du niveau d'éducation. L'interaction n'est pas significative.

Le niveau d'éducation ne semble pas influencer la crédibilité perçue des fake news, ni directement, ni en interaction avec l'origine du titre. Les hypothèses H3a et H3b sont rejetées

F. Hypothèse 4

1. Hypothèse 4.a

Afin d'évaluer si les dimensions de l'échelle *TRI* influencent la crédibilité perçue des titres générés par IA, nous avons réalisé une régression linéaire multiple avec *Tech_Score_F1* et *Tech_Score_F2* comme variables prédictives (Annexe 18).

Le modèle est le suivant : $mean\_total_i = 3{,}343 + 0{,}141\,.\,Tech\_Score\_F1_i + 0{,}023\,.\,Tech\_Score\_F2_i + \epsilon_i$

Terme	***Coefficient***	***p-value***	***Interprétation***
Tech_Score_F1	+0,1405	0,079	Non significatif
Tech_Score_F2	+0,0229	0,808	Non significatif

Table 5 : Coefficients et p-value du modèle de régression de H4.a

- $R^2 = 0{,}067$; $F(2, 73) = 2\,,601$, $p = 0.081$, le modèle n'est pas significatif
- Les hypothèses de validité sont respectées

Aucun effet significatif n'est observé. Toutefois, une légère tendance est notée pour Tech_Score_F1 mais l'hypothèse n'est pas validée. Nous ne rejetons pas l'hypothèse nulle.

2. Hypothèse 4.b

Nous avons utilisé un modèle de régression multiple avec interaction. Les deux facteurs sont centrés autour de leur moyenne (Annexe 19).

$$mean\_total_i = \beta_0 + \beta_1\,.\,\text{Tech\_Score\_}F1 + \beta_2\,.\,\text{Tech\_Score\_}F2_i + \beta_3\,.\,IsAI_i + \beta_4\,.\,(Tech\_Score\_F1_i \times IsAI_i) + \beta_5\,.\,(Tech\_Score\_F2_i \times IsAI_i) + \epsilon_i$$

Terme	*Coefficient*	*p-value*	*Interprétation*
Tech_Score_F1 (centré)	–0.0208	0.795	Non significatif
Tech_Score_F2 (centré)	= –0.0127	p = 0.894	Non significatif
IsIA	0.3439	0.005	**Significatif**
Tech_Score_F1 * IsIA	= 0.1612	= 0.143	Non significatif
Tech_Score_F2 * IsIA	= 0.0356	0.785	Non significatif

Table 6 : Coefficients et p-value du modèle de régression de H4.b

- $R^2 = 0.092$; $F(5,139) = 2.820$, $p = 0.0186$, le modèle est globalement significatif
- Les hypothèses statistiques sont respectées.

Seul l'effet principal de l'origine du titre est significatif. Aucune interaction significative n'est observée. Nous ne rejetons pas l'hypothèse nulle.

G. Hypothèse 5

1. Hypothèse 5.a

Nous avons réalisé une régression linéaire simple avec *AI_Score* comme prédicteur (Annexe 20).

$$mean\_total_i = 3{,}578 + 0{,}093 \, . \, AI\_Score_i + \epsilon_i$$

- $\beta_1 = 0{,}093$ ave une *p-value* de 0,033, non significatif
- $R^2 = 0{,}013$; $F(1, 74) = 0{,}969$, $p = 0{,}328$, le modèles n'est pas significatif
- L'hypothèse de normalité des résidus n'est pas respectée **(***p* = 0.0067)

Aucun effet significatif n'est observé. Nous ne rejetons pas l'hypothèse nulle.

2. Hypothèse 5.b

Nous avons testé un modèle avec interaction entre *AI_Score*(centré) et *IsIA* (Annexe 21).

Terme	*Coefficient*	*p-value*	*Interprétation*
AI_Score (centré)	0,074	0,521	Non significatif
IsIA	0.375	0.003	Significatif
AI_Score * IsIA	0,019	0.897	Non significatif

Table 7 : Coefficients et p-value du modèle de régression de H5.b

- $R^2 = 0,064$; $F(3,141) = 3,188$, $p = 0,026$, modèle globalement significatif
- La normalité des résidus n'est pas respectée.

Encore une fois, seul l'effet principal de l'origine du titre est significatif. Aucune interaction n'est observée, nous ne rejetons pas l'hypothèse nulle.

H. Multicolinéarité

Un modèle de régression multiple a été estimé pour examiner l'effet combiné des variables individuelles sur la crédibilité perçue des titres. Les variables principales incluaient l'âge, l'attitude technologique (*Tech_Score_F1* et *Tech_Score_F2*) et le sentiment de compétence en IA ainsi que le niveau d'éducation. Le sexe et la maîtrise de l'anglais étaient inclus comme variables de contrôle (Annexe 22).

Le modèle respecte l'ensemble des conditions statistiques (normalité des résidus, homoscédasticité, indépendance des erreurs) et ne présente aucun problème de colinéarité (VIF < 2).

Cependant, le modèle global n'est pas significatif ($p = 0,305$) et n'explique qu'une faible part de la variance de la crédibilité perçue ($R^2 = 0,058$). Aucune variable n'atteint le seuil de significativité, bien que des tendances marginales apparaissent pour l'âge ($p = 0,075$) et le facteur *Tech_Score_F1* ($p = 0,068$). Ces résultats suggèrent que, même en tenant compte de variables sociodémographiques, les facteurs étudiés ne suffisent pas à expliquer la manière dont les individus évaluent la crédibilité des fake news générées par IA.

I. Conclusion

L'analyse statistique menée a permis de tester les différentes hypothèses formulées. Les résultats confirment l'hypothèse principale (H1) : les fake news générées par une IA sont perçues comme plus crédibles que celles écrites par des humains.

En revanche, les hypothèses relatives à l'âge, au niveau d'éducation, à la compétence perçue en IA ou à l'affinité technologique n'ont pas toutes été confirmées de manière robuste ou significative. Si certains effets apparaissent localement, notamment concernant l'âge, ils ne se maintiennent pas dans les modèles globaux.

Ces résultats posent les bases d'une discussion critique, qui sera développée dans le chapitre suivant, afin d'interpréter ces observations, d'en dégager les implications, et d'envisager les perspectives.

X. Discussion

A. Introduction

L'objectif de cette discussion est d'interpréter les résultats obtenus à la lumière des hypothèses formulées, tout en les mettant en perspective avec la littérature existante et les limites méthodologiques de l'étude.

B. Discussion

L'analyse des résultats a permis de confirmer l'hypothèse H1 : les fake news générées par une intelligence artificielle sont perçues comme significativement plus crédibles que celles rédigées par des humains, c'est un enjeu majeur de désinformation.

En revanche, les hypothèses H3 (niveau d'éducation), H4 (affinité technologique) et H5 (compétence perçue en IA) n'ont pas été confirmées, ni dans leurs effets principaux ni en interaction. Le niveau d'études ou la familiarité avec la technologie ne semblent pas influencer l'évaluation de la crédibilité des fake news. Ce constat suggère que les compétences critiques nécessaires à la détection de fake news ne sont pas directement liées à la formation académique ou à la confiance en ses capacités technologiques.

Concernant l'âge (H2), un effet significatif a été observé, les plus jeunes ont tendance à attribuer une crédibilité plus élevée aux titres présentés. Toutefois, cet effet perd en robustesse dans le modèle global, il semble lié à la distribution déséquilibrée des âges dans l'échantillon, principalement concentré entre 20 et 30 ans.

Enfin, l'absence de significativité pour l'échelle *AI_Score* pourrait s'expliquer par sa construction hétérogène. Dans l'ensemble, ces résultats soulignent la complexité des mécanismes cognitifs sous-jacents à la crédibilité perçue, et la nécessité de poursuivre les recherches avec des outils plus ciblés et des échantillons mieux équilibrés.

C. Biais et limites

Plusieurs biais et limites doivent être pris en compte dans l'interprétation de nos résultats.

Tout d'abord, le biais principal réside dans la sélection manuelle des fake news utilisées. Bien que ce choix ait été effectué de manière rigoureuse, il peut introduire un effet de formulation non contrôlé. De plus, seuls les titres ont été présentés aux participants. D'autres études ont

montré que l'ajout d'images influence significativement la perception de crédibilité. Ce choix de simplicité expérimentale pourrait être revisité dans de futures recherches.

Par ailleurs, le mode de diffusion du questionnaire, sur les réseaux sociaux et des plateformes de sondage en ligne, a entraîné une sur-représentation de jeunes adultes, notamment entre 20 et 30 ans. Ce déséquilibre démographique limite la généralisation de nos résultats en particulier en ce qui concerne l'effet de l'âge

De plus, la taille de l'échantillon (N = 145) reste modeste. Un échantillon plus élevé aurait renforcé la puissance statistique des tests et la robustesse des analyses de sous-groupes

Sur le plan expérimental, l'absence de contrebalancement constitue également une limite : chaque participant n'a été exposé qu'à un seul type de contenu (IA ou humain), ce qui empêche une analyse intra-individuelle des différences de perception.

Enfin, la langue du questionnaire et sa diffusion à l'international ont introduit une variabilité non contrôlée liée au niveau de compréhension, aux références culturelles ou aux attitudes envers l'IA. Les thématiques abordées (COVID-19 et politique) constituent également une source potentielle de biais. Le COVID-19, sujet désormais « dépassé », pourrait être perçu comme moins crédible en raison d'une fatigue informationnelle, tandis que les titres politiques, plus polarisants émotionnellement, peuvent biaiser les jugements de crédibilité.

D. Conclusion

Cette discussion a mis en évidence les principaux enseignements de l'étude : les contenus générés par une intelligence artificielle peuvent apparaître aussi, voire plus crédibles que ceux rédigés par des humains. Ce résultat souligne l'importance de développer de nouveaux outils pédagogiques et technologiques pour sensibiliser le public aux risques associés à la désinformation automatisée.

Toutefois, les variables individuelles supposées moduler cette perception – telles que l'âge, le niveau d'études ou la compétence perçue en IA – se sont révélées moins déterminantes que prévu. Ce constat invite à repenser les profils à risque et les leviers de prévention.

Enfin, les biais méthodologiques identifiés rappellent la nécessité de mener des études plus larges et plus diversifiées, afin d'approfondir la compréhension des dynamiques entre intelligence artificielle et crédibilité perçue.

XI. Conclusion

Cette partie pratique a permis de tester empiriquement les hypothèses formulées à partir de la littérature scientifique. Les résultats confirment que les titres générés par l'intelligence artificielle peuvent être perçus comme plus crédibles que ceux rédigés par des humains, ce qui soulève des enjeux cruciaux en matière de désinformation.

En revanche, plusieurs variables individuelles – telles que le niveau d'éducation, la compétence perçue en IA ou encore l'affinité technologique – ne se sont pas révélées significativement liées à la crédibilité perçue, nuançant ainsi certains postulats théoriques. Ces résultats suggèrent que d'autres facteurs, contextuels ou cognitifs, pourraient jouer un rôle plus déterminant.

La discussion des résultats, associée à l'analyse des biais et limites, ouvre la voie à de futures recherches plus approfondies sur la manière dont les utilisateurs interagissent avec des contenus générés artificiellement

Conclusion

L'objectif de ce mémoire était d'explorer l'impact de l'intelligence artificielle générative sur la crédibilité perçue des fake news. À travers notre revue de littérature, nous avons d'abord défini les concepts-clefs : les fake news, leurs caractéristiques distinctives, l'intelligence artificielle générative et les facteurs influençant la crédibilité perçue.

Nous nous sommes servis de cette base pour formuler plusieurs hypothèses et mener une étude quantitative afin de tester empiriquement les effets de différents facteurs individuels et contextuels.

Les résultats obtenus montrent que les fake news générées par IA sont perçues comme plus crédibles que celles créées par des humains, confirmant notre hypothèse principale. D'autres variables, comme le niveau d'éducation, la familiarité technologique ou la compétence perçue en IA n'ont pas eu d'effet significatif. L'âge semble jouer un rôle mais de manière partielle et peu robuste. Ces résultats suggèrent que la crédibilité des contenus se repose sur des processus cognitifs plus complexe que ce que laissent penser les seuls facteurs sociodémographiques ou déclaratifs.

Cette recherche présente toutefois plusieurs limites : la taille de l'échantillon, la sélection manuelle des fake news, l'absence de contrebalancement, ou encore la concentration démographique autour des jeunes adultes. Malgré cela, elle offre une première base de réflexion sur les risques que pose l'IA générative en matière de désinformation.

À l'heure où ces technologies se démocratisent à grande vitesse, il devient urgent de mieux comprendre les biais cognitifs qu'elles mobilisent, et d'éduquer les utilisateurs à la détection de contenus artificiels. Des études futures pourraient approfondir ces résultats, en incluant des contenus multimodaux, une exposition croisée, ou une comparaison entre différents types de générateurs.

Ce travail ne prétend pas clore le débat, mais il ambitionne de poser les premières briques d'une réflexion essentielle : comment garantir la fiabilité de l'information à l'ère des machines capables de la manipuler avec une crédibilité redoutable ?

Bibliographie

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211-236. https://doi.org/10.1257/jep.31.2.211

Beauvais, C. (2022). Pourquoi croyons-nous aux fake news ? *Revue du Rhumatisme*, *89*(6), 555-561. https://doi.org/10.1016/j.rhum.2022.09.013

Bergman, D. (2023, décembre 5). *IBM.com*. https://www.ibm.com/fr-fr/think/topics/self-supervised-learning

Durmus, E., Nguyen, K., Liao, T. I., Schiefer, N., Askell, A., Bakhtin, A., Chen, C., Hatfield-Dodds, Z., Hernandez, D., Joseph, N., Lovitt, L., McCandlish, S., Sikder, O., Tamkin, A., Thamkul, J., Kaplan, J., Clark, J., & Ganguli, D. (2023). *Towards Measuring the Representation of Subjective Global Opinions in Language Models* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2306.16388

Eshet-Alkalai. (2004). Digital Literacy : A Conceptual Framework for Survival Skills in the Digital Era. *Journal of Educational Multimedia and Hypermedia*, *13*(1), 93-106.

Euchner, J. (2023). Generative AI. *Research-Technology Management*, *66*(3), 71-74.

Ferguson, A. (2016, juin 12). Designing online experiments using Google forms + random redirect tool. *teaching statistics is awesome*. https://teaching.statistics-is-awesome.org/designing-online-experiments-using-google-forms-random-redirect-tool/

Fernández-López, M., & Perea, M. (2020). Language does not modulate fake news credibility, but emotion does. *Psicológica Journal*, *41*(2), 84-102. https://doi.org/10.2478/psicolj-2020-0005

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2023). *Generative AI*. https://doi.org/10.48550/ARXIV.2309.07930

Galaz, V., Metzler, H., Daume, S., Olsson, A., Lindström, B., & Marklund, A. (2023). *AI could create a perfect storm of climate misinformation*. Stockholm Resilience Centre (Stockholm University) and the Beijer Institute of Ecological Economics (Royal Swedish Academy of Sciences).

Gold, A., & Fischer, S. (2023, février 21). *Chatbots trigger next misinformation nightmare*. Axios. https://www.axios.com/2023/02/21/chatbots-misinformation-nightmare-chatgpt-ai

Gu, L., Kropotov, V., & Yarochkin, F. (2017). The Fake News Machine : How Propagandists Abuse the Internet and Manipulate the Public. *Trend Micro*, *5*, 1-85.

Kalsnes, B. (2018). Fake News. In *Oxford Research Encyclopedia of Communication*. https://doi.org/10.1093/acrefore/9780190228613.013.809

Kang, C., & Goldman, A. (2016, décembre 5). *In Washington Pizzeria Attack, Fake News Brought Real Guns—The New York Times*.

https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html

Koetsenruijter, A. W. M. (2011). Using Numbers in News Increases Story Credibility. *Newspaper Research Journal*, *32*(2), 74-82. https://doi.org/10.1177/073953291103200207

Li, J., Cheng, X., Zhao, W. X., Nie, J.-Y., & Wen, J.-R. (2023). *HaluEval : A Large-Scale Hallucination Evaluation Benchmark for Large Language Models* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2305.11747

Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media : Effects of Truth-Bias and Endorsement Cues. *Communication Research*, *49*(2), 171-195. https://doi.org/10.1177/0093650220921321

Mansoura, S. (2023, mars 28). Les images du pape François en doudoune, devenues virales, sont complètement fausses. *France Bleu*. https://www.francebleu.fr/infos/insolite/les-images-du-pape-francois-en-doudoune-devenues-virales-sont-completement-fausses-5922778

Menon, P. (2023, mars 9). Introduction to Large Language Models and the Transformer Architecture. *Medium.com*. https://rpradeepmenon.medium.com/introduction-to-large-language-models-and-the-transformer-architecture-534408ed7e61

Metz, C. (2023, mai 1). *'The Godfather of A.I.' Leaves Google and Warns of Danger Ahead*. New York Times. https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html

Meyers, M., Weiss, G., & Spanakis, G. (2020). *Fake News Detection on Twitter Using Propagation Structures* (p. 138-158). https://doi.org/10.1007/978-3-030-61841-4_10

Nanda, A. (2024, septembre 2). Generative vs Discriminative Models : Differences & Use Cases. *Datacamp.com*. https://www.datacamp.com/blog/generative-vs-discriminative-models

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy : An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Office parlementaire d'évaluation des choix scientifiques et technologiques. (2024). *ChatGPT, et après ? Bilan et perspectives de l'intelligence artificielle* (Rapport de l'OPECST Rapport n° 170). Sénat Français. https://www.senat.fr/rap/r24-170/r24-170.html

Politi, D. (2016, décembre 26). A Fake News Story Leads Pakistani Minister to Issue Nuclear Threat Against Israel. *Slate*. https://slate.com/news-and-politics/2016/12/fake-news-story-leads-pakistani-minister-to-issue-nuclear-threat-against-israel.html

Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P., & Hashimoto, T. (2023). *Whose Opinions Do Language Models Reflect?* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2303.17548

Seong, B.-H., & Hong, C.-Y. (2022). Corroborating the effect of positive technology readiness on the intention to use the virtual reality sports game "Screen Golf" : Focusing on

the technology readiness and acceptance model. *Information Processing & Management*, *59*(4), 102994. https://doi.org/10.1016/j.ipm.2022.102994

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). *Combating Fake News : A Survey on Identification and Mitigation Techniques* (arXiv:1901.06437). arXiv. https://doi.org/10.48550/arXiv.1901.06437

Shearer, E., & Mitchell, A. (2021, janvier 12). News Use Across Social Media Platforms in 2020. *Pew Research Center's Journalism Project*. https://www.pewresearch.org/journalism/2021/01/12/news-use-across-social-media-platforms-in-2020/

Shu, K., Amy Sliva, Suhang Wang, Jiliang Tang, & Huan Liu. (2017). Fake News Detection on Social Media : A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22-36. https://doi.org/10.1145/3137597.3137599

Silverman, C., & Singer-Vine, J. (2016, décembre 7). *Most Americans Who See Fake News Believe It, New Survey Says*. BuzzFeed News. https://www.buzzfeednews.com/article/craigsilverman/fake-news-survey

Steyer, V., Vuarin, L., & Ben Mahmoud-Jouini, S. (2021, novembre 17). *IA : pourquoi il faut ouvrir la boîte noire*. Polytechnique insights. https://www.polytechnique-insights.com/tribunes/digital/inexplicabilite-de-lia-un-enjeu-organisationnel/

Sultan, M., Tump, A. N., Ehmann, N., Lorenz-Spreen, P., Hertwig, R., Gollwitzer, A., & Kurvers, R. H. J. M. (2024). Susceptibility to online misinformation : A systematic meta-analysis of demographic and psychological factors. *Proceedings of the National Academy of Sciences*, *121*(47), e2409329121. https://doi.org/10.1073/pnas.2409329121

Tandoc, E., Zeng Wei, L., & Ling, R. (2017). Defining "Fake News" : A typology of scholarly definitions. *Digital Journalism*, *6*, 1-17. https://doi.org/10.1080/21670811.2017.1360143

The Reuters Institute for Study of Journalism (RISJ). (2017, mars 24). *How can we combat fake news? – The role of platforms, media literacy, and journalism*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/news/how-can-we-combat-fake-news-role-platforms-media-literacy-and-journalism

Thome, N., & Wolfe, C. (2023). *Histoire des réseaux de neurones et du deep learning en traitement des signaux et des images*. https://hal.science/hal-04058482

Tseng, Y.-C., & Yuan, C. W. (Tina). (2023). Investigating Perceived Message Credibility and Detection Accuracy of Fake and Real Information Across Information Types and Modalities. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1-7. https://doi.org/10.1145/3544549.3585719

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762

Wang, B., Rau, P.-L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence : Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, *42*(9), 1324-1337. https://doi.org/10.1080/0144929X.2022.2072768

Wardle, C., & Derakhshan, H. (2017). *INFORMATION DISORDER : Toward an interdisciplinary framework for research and policy making Information Disorder Toward an interdisciplinary framework for research and policymaking*.

Annexes

Annexe 1 : Titre des fake news

1. Titres de fake news réels

1. The Moderna COVID vaccine is no longer recommended because it causes myocarditis.
2. Donald Trump tweeted in 2009 that “Obama's handling of this pandemic was terrible!”
3. March 2020 was the first March without a school shooting in the US since 2002.
4. The United Kingdom is banning anyone under the age of 50 from getting COVID-19 vaccines.
5. Three days before her death, Betty White declared she had received her booster vaccination.
6. The 2020 election was the only time in American history that an election took more than 24 hours to count.
7. The movie "Anastasia" was removed from Disney+ in protest of Russia's invasion of Ukraine.
8. Time magazine's March 2022 issue compares Russian President Putin to Adolf Hitler.
9. Donald Trump said in 1998 that he would one day run as a Republican because they are the "dumbest group of voters."
10. Former U.S. President Barack Obama ordered counterintelligence agents to illegally "spy" on Donald Trump's campaign for president in 2016.

2. Titres de fake news générés par IA

1. An American company creates edible masks to encourage people to wear them.
2. In northern Alaska, an indigenous village confines itself to igloos to escape COVID-19.
3. COVID-19 vaccine reduces male fertility according to study.
4. The American bat population has plummeted since the pandemic.

5. Wearing a mask does more harm than good, say experts.
6. The Canadian government warns its citizens against traveling to the United States.
7. Joe Biden proposes 5 years of military service in exchange for free college tuition.
8. Ukrainian artist creates monumental border mural to symbolize peace.
9. The “Anonymous” hacker group neutralizes Russian missiles using computer viruses.
10. Famous models launch a clothing brand to promote peace.

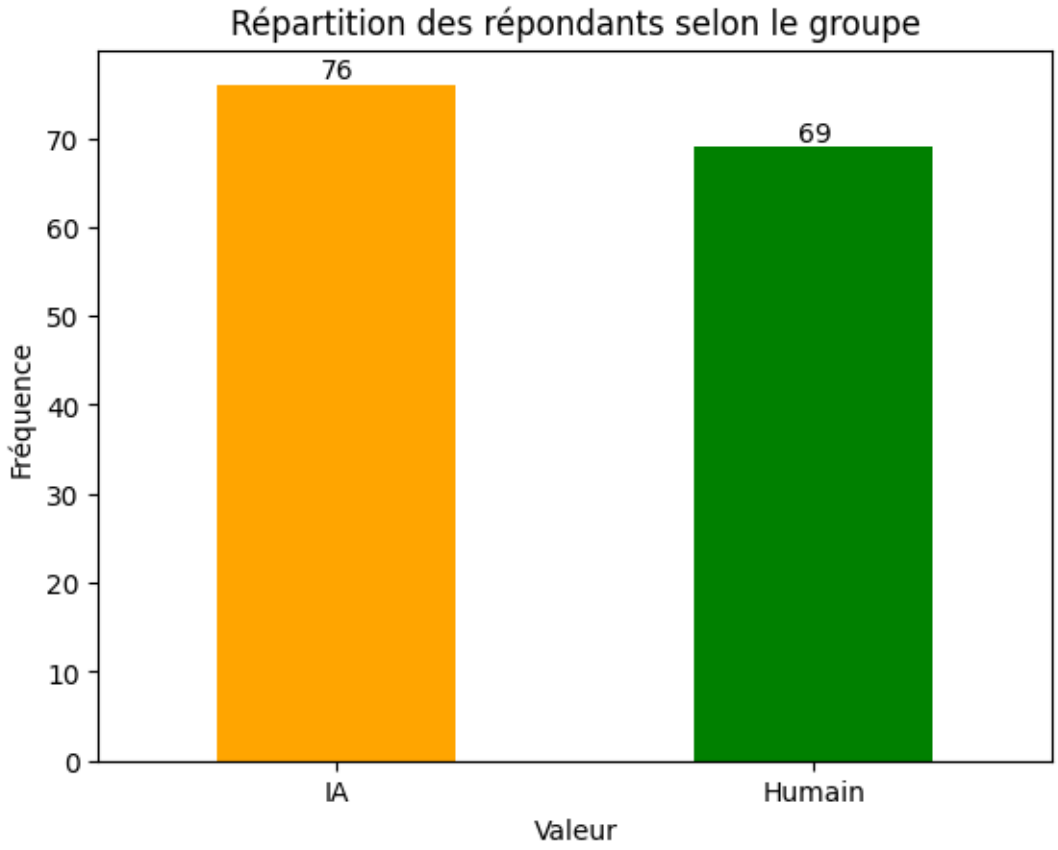
Annexe 2 : Echelle Technology Readiness Scale

1. I usually prefer to use high technology.
2. I'm used to using products or services with the latest technology.
3. I think technology enables efficient work.
4. I tend to use new technology before people around me.
5. I'm well aware of the trend of products or services applied with new technology.
6. I tend to keep track of the latest technology developments in fields of interest.
7. I tend to inform others about new technology.

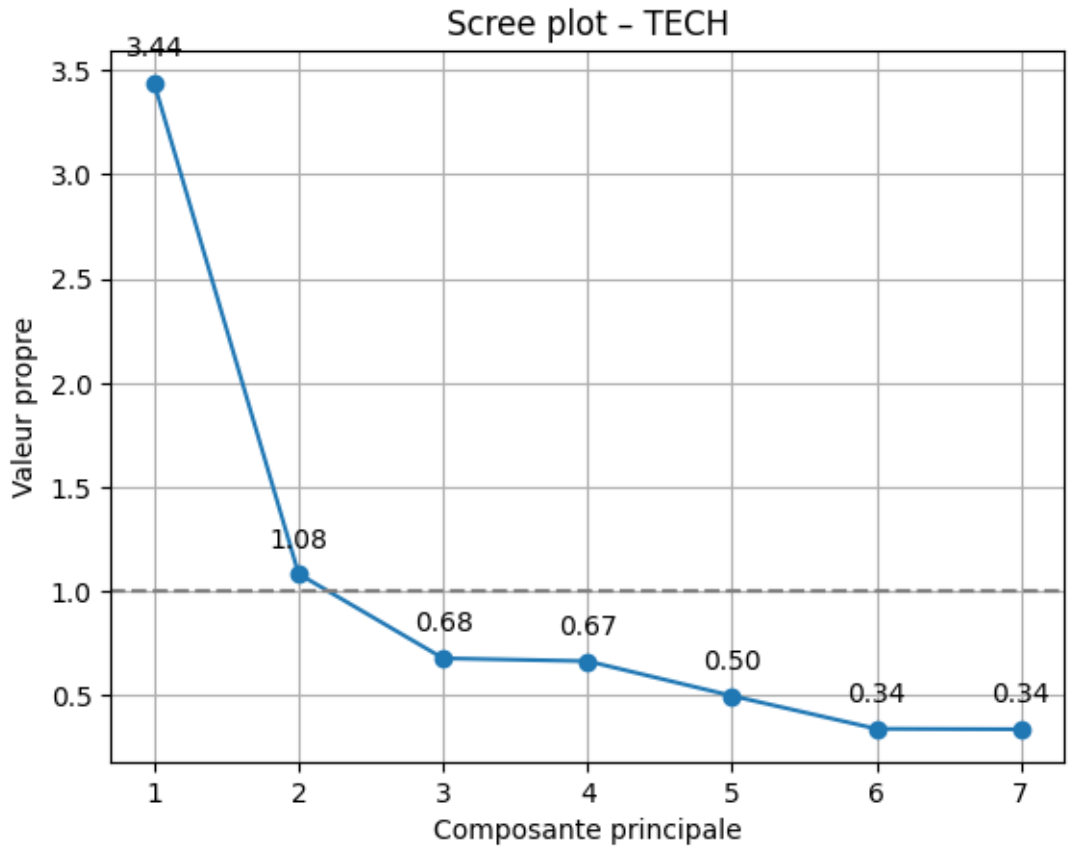
Annexe 3 : Measuring User Competence in Using AI

1. I can distinguish between smart devices and non-smart devices.
2. I do not know how AI technology can help me. (inversé)
3. I can identify the AI technology employed in the applications and products I use.
4. I can skilfully use AI applications or products to help me with my daily work.
5. It is usually hard for me to learn to use a new AI application or product. (inversé)
6. I can use AI applications or products to improve my work efficiency.
7. I can evaluate the capabilities and limitations of an AI application or product after using it for a while.
8. I can choose a proper solution from various solutions provided by a smart agent.
9. I can choose the most appropriate AI application or product from a variety for a particular task.
10. I always comply with ethical principles when using AI applications or products.
11. I am never alert to privacy and information security issues when using AI applications or products. (inversé)
12. I am always alert to the abuse of AI technology.

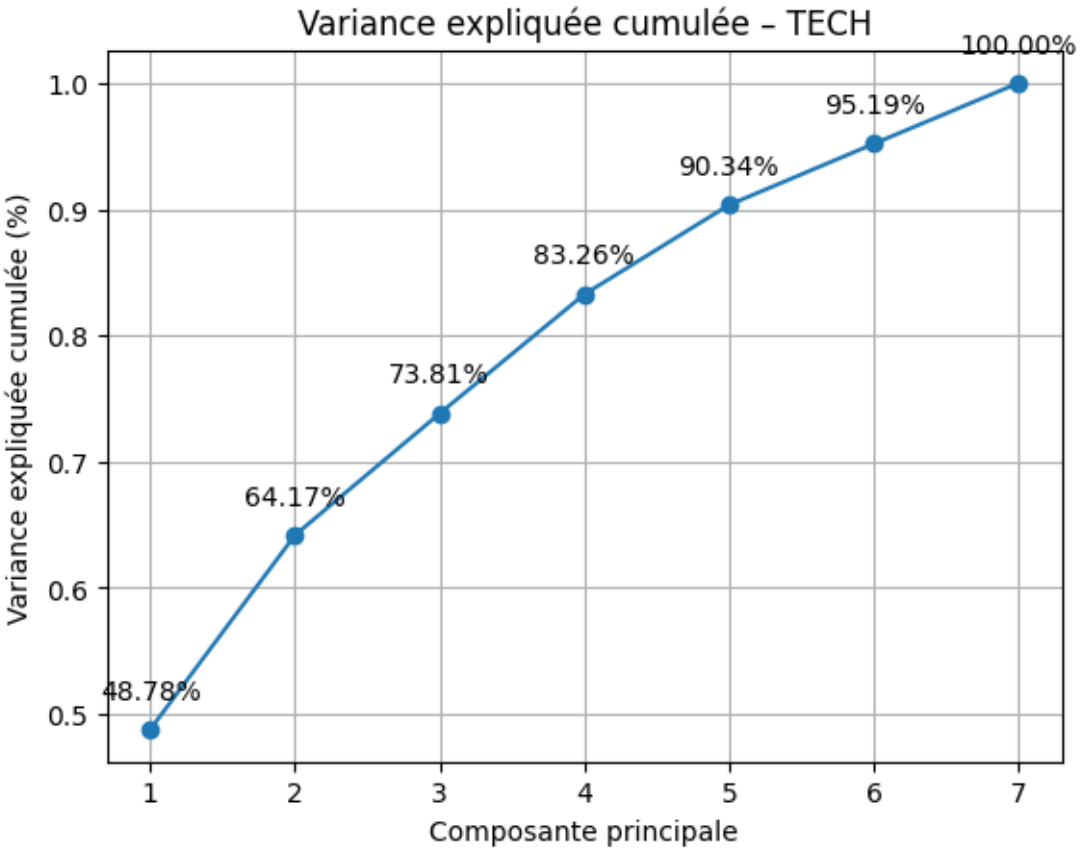
Annexe 4 : Répartition des répondants selon le groupe IsIA



Annexe 5 : Scree plot - Echelle TRI



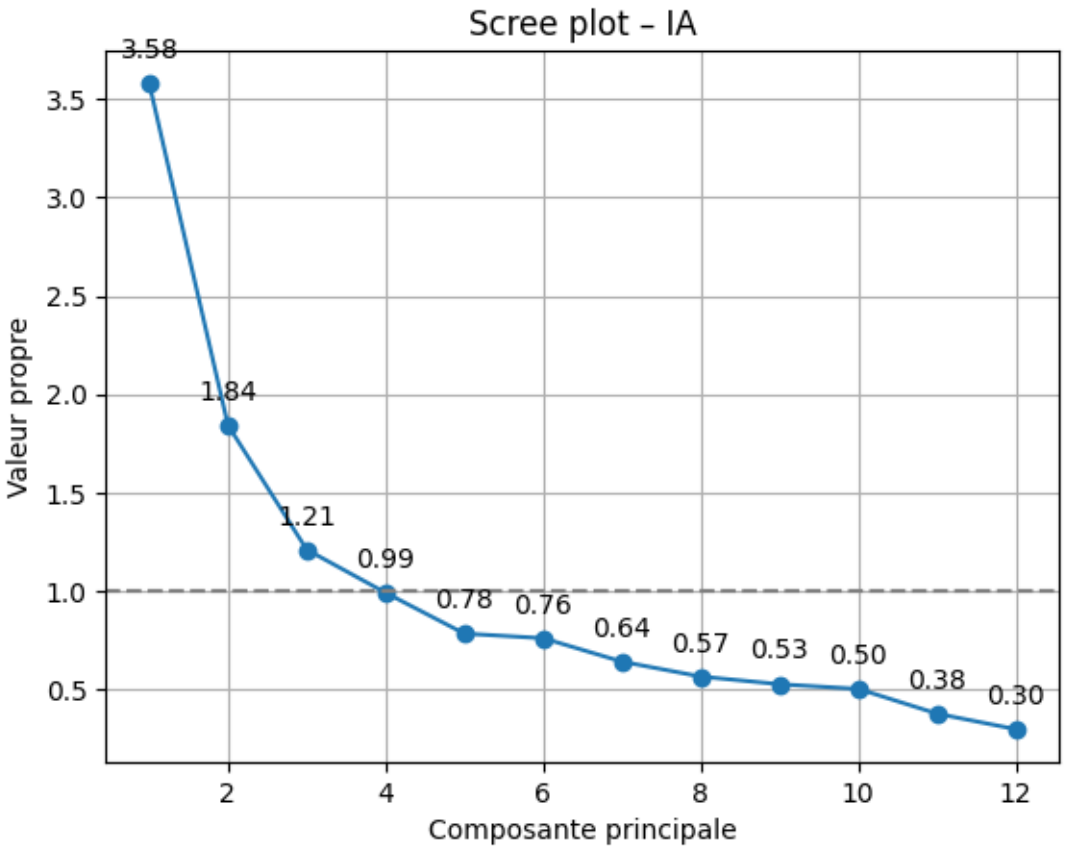
Annexe 6 : Variance expliquée cumulée - Echelle TRI



Annexe 7 : Charges factorielles à 2 facteurs - Echelle TRI - Varimax

Item	*Facteur 1*	*Facteur 2*
TECH1	0,297	**0,623**
TECH2	0,207	**0,872**
TECH3	0,207	**0,493**
TECH4	**0,587**	0,379
TECH5	**0,762**	0,27
TECH6	**0,66**	0,218
TECH7	**0,603**	0,192

Annexe 8 : Scree Plot - Echelle IA



Annexe 9 : Charges factorielles à 3 facteurs - Echelle IA - Varimax

Item	*Facteur 1*	*Facteur 2*	*Facteur 3*
AI1	0,136	0,157	**0,481**
AI2	**0,489**	- 0,139	0,249
AI3	0,239	0,049	**0,622**
AI4	**0,797**	-0,065	0,105
AI5	**0,391**	-0,135	0,361
AI6	**0,812**	0,016	0,055
AI7	**0,464**	0,172	0,324
AI8	**0,563**	0,049	0,173
AI9	**0,506**	0,18	0,156
AI10	0,129	**0,508**	-0,149
AI11	-0,019	**0,678**	0,137
AI2	-0,093	**0,624**	0,253

Annexe 10 : Charges factorielles à 2 facteurs - Echelle IA - Varimax

Item	*Facteur 1*	*Facteur 2*
AI1	**0,289**	0,279
AI2	**0,557**	-0,076
AI3	**0,425**	0,214
AI4	**0,775**	-0,081
AI5	**0,499**	-0,023
AI6	**0,753**	-0,028
AI7	**0,546**	0,243
AI8	**0,589**	0,055
AI9	**0,52**	0,177
AI10	0,039	**0,361**
AI11	-0,003	**0,637**
AI2	-0,031	**0,712**

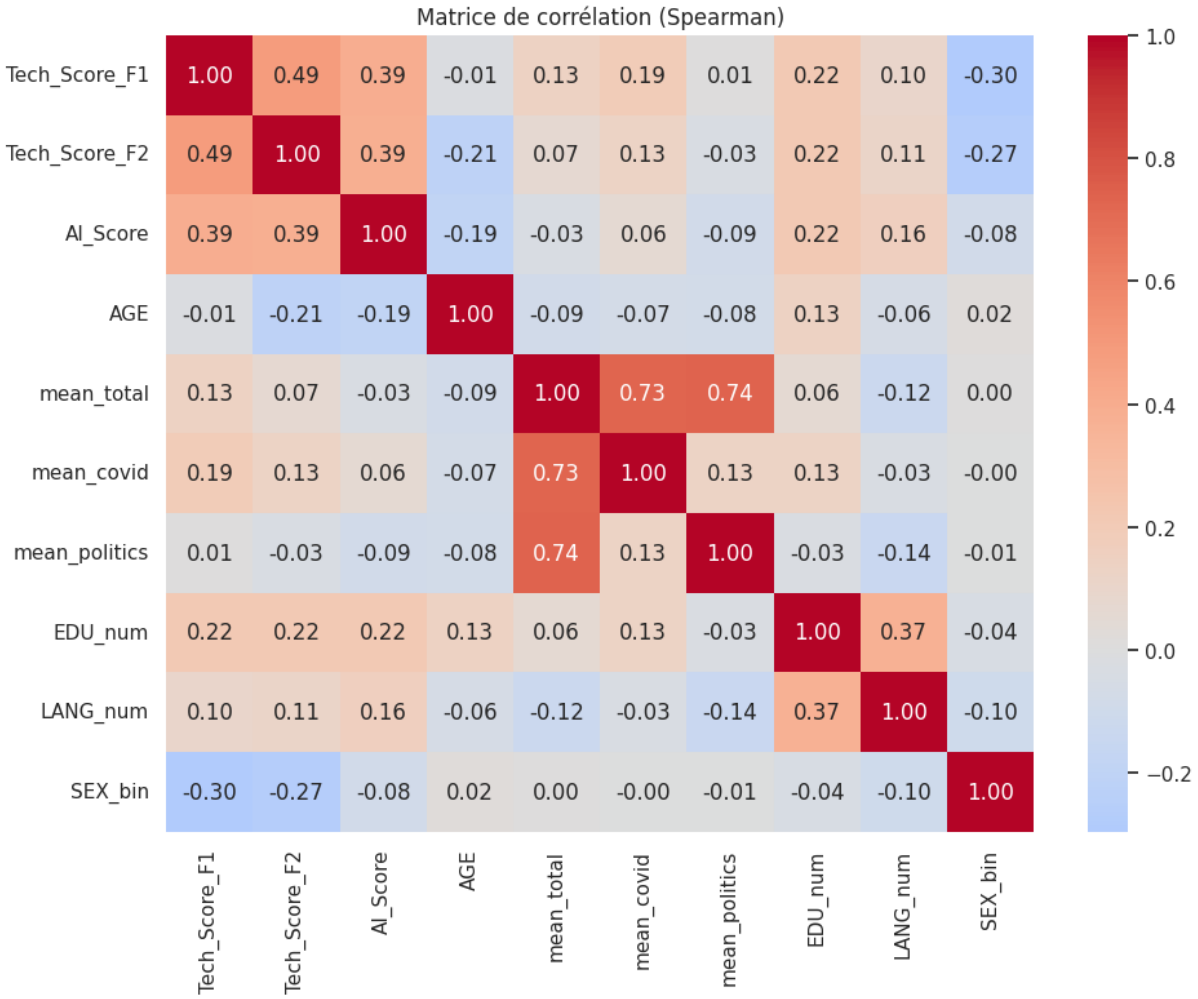
Annexe 11 : Cohérence des titres de fake news

Titre	*Groupe IA*		*Groupe humaine*	
	Moyenne	**Écart-type**	**Moyenne**	**Écart-type**
Cov1	2,36	1,46	3,83	1,63
Cov2	3,46	1,66	3,46	2,10
Cov3	4,24	1,63	4,28	1,98
Cov4	3,86	1,65	2,88	1,69
Cov5	3,62	1,73	3,28	1,68
Pol1	4,88	1,85	4,19	1,69
Pol2	4,01	1,59	3,91	1,70
Pol3	5,21	1,52	4,03	1,71
Pol4	3,84	1,53	3,87	1,89
Pol5	4,87	1,57	3,17	1,55

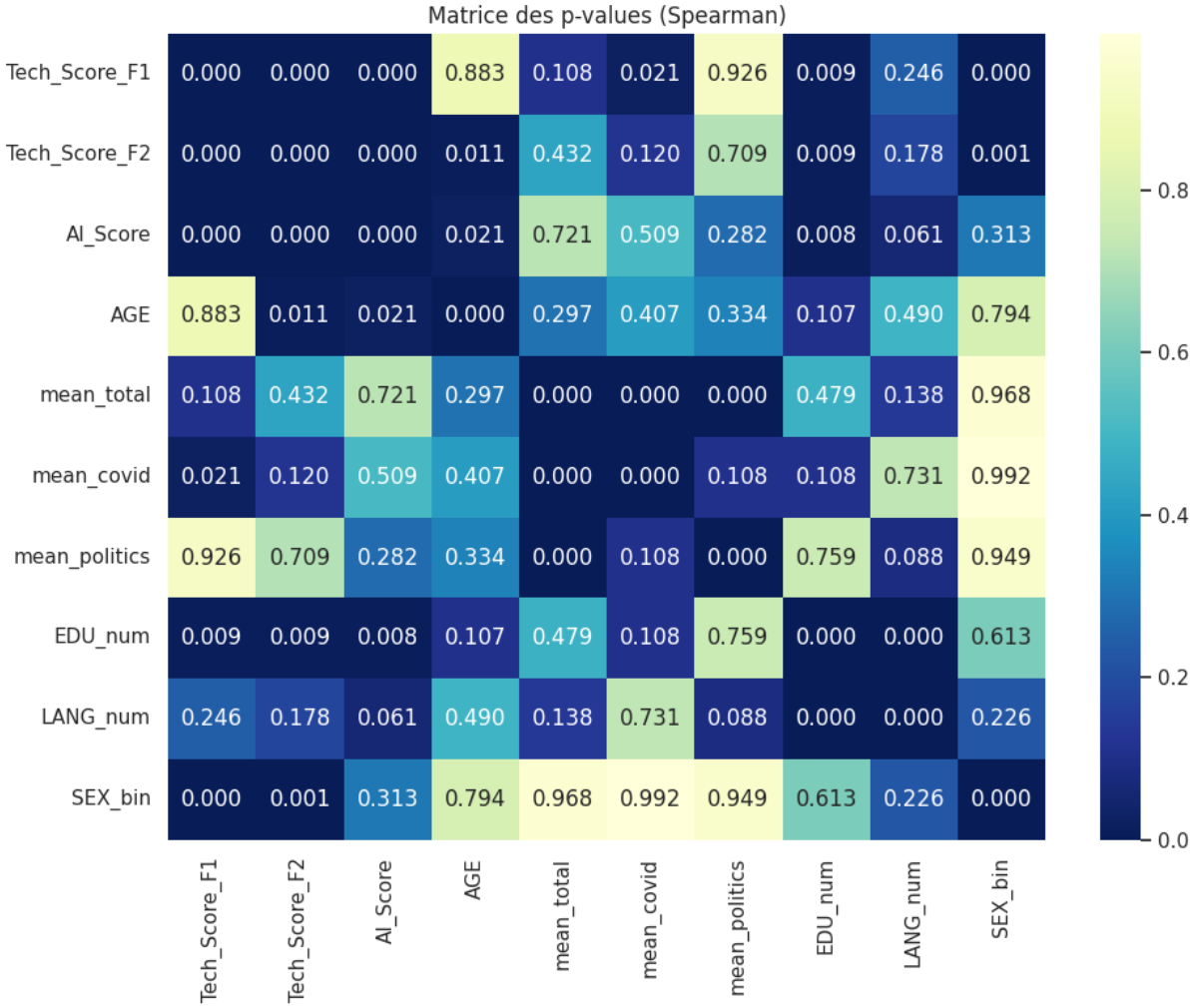
Annexe 12 : Résultats des tests d'homogénéité des variables

Variable	***Test utilisé***	***p-value***	***Groupes homogènes***
AGE	Mann-Whitney	0,757	Oui
Tech_Score_F1	T-test	0,917	Oui
Tech_Score_F2	Mann-Whitney	0,227	Oui
AI_Score	T-test	0,012	Non
SEX	Chi^2	0,42	Oui
EDU	Chi^2	0,742	Oui
LANG	Chi^2	0,182	Oui

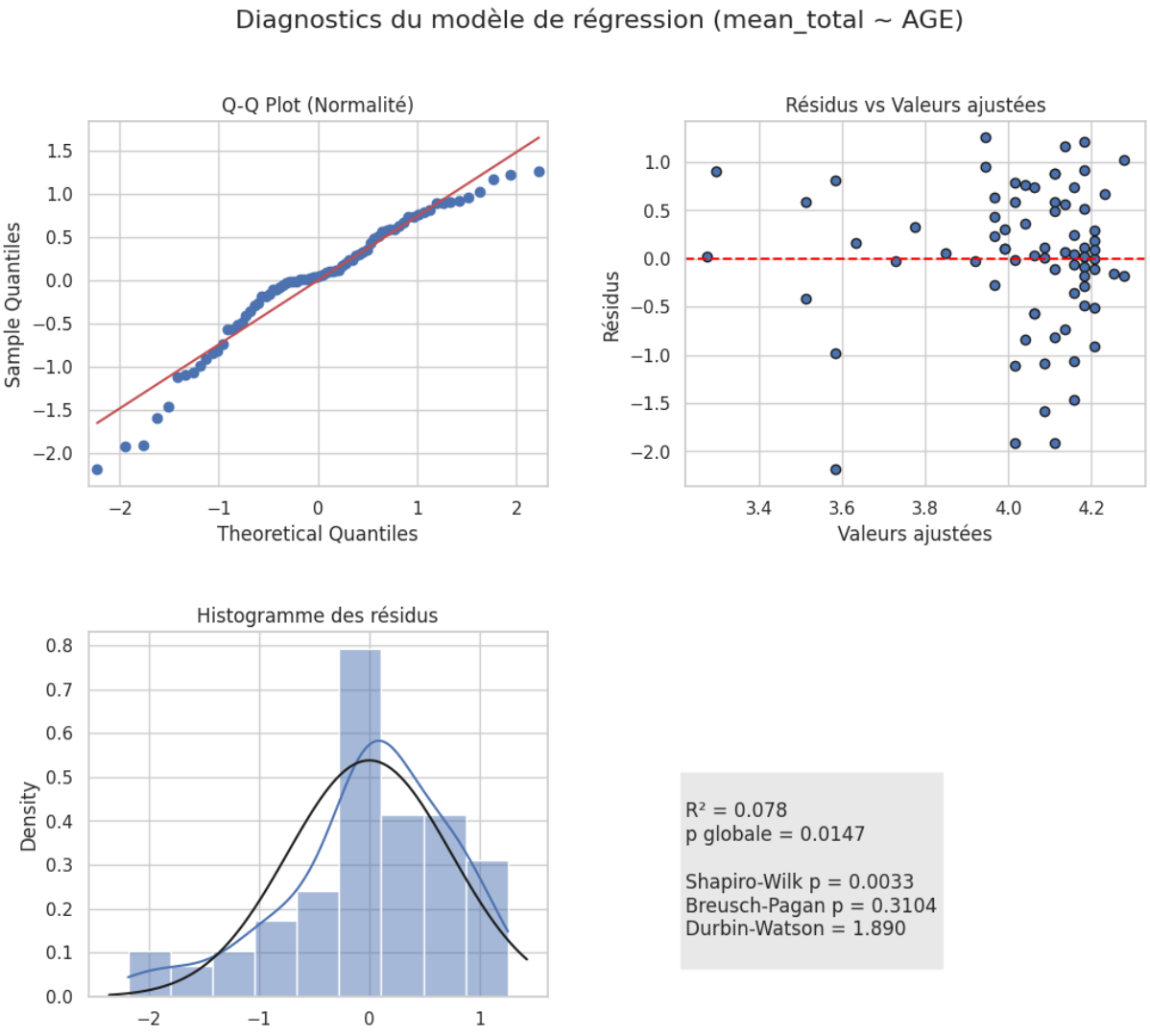
Annexe 13 : Matrice de corrélation (Spearman)



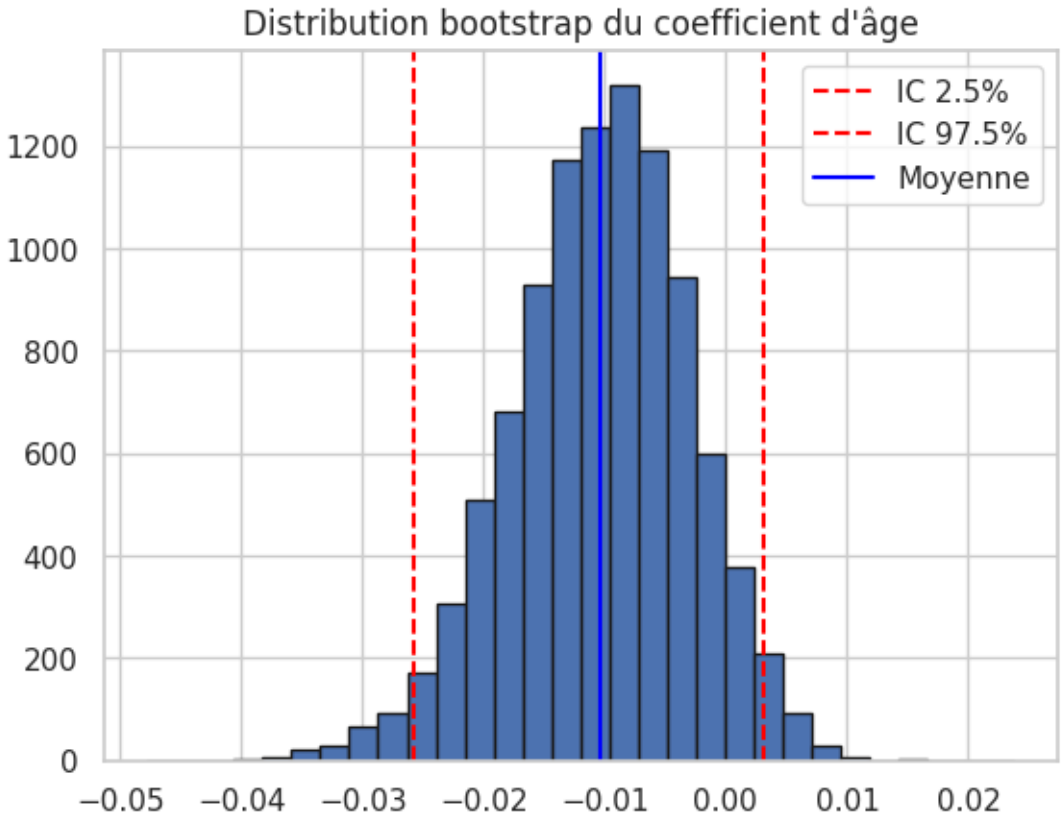
Annexe 14 Annexe 14 : Matrice de p-value (Spearman)



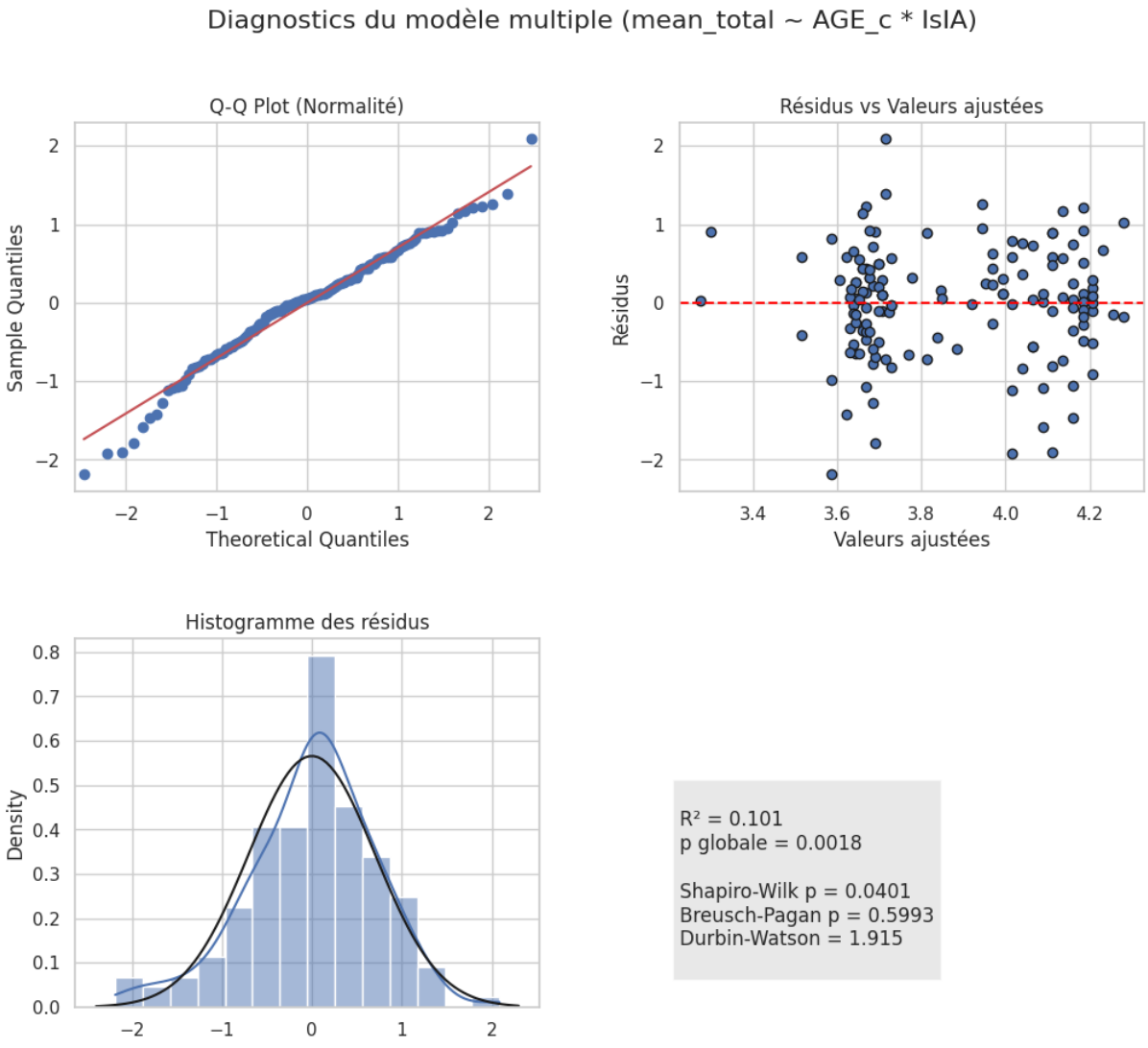
Annexe 15 : Modèle de régression - H2.a



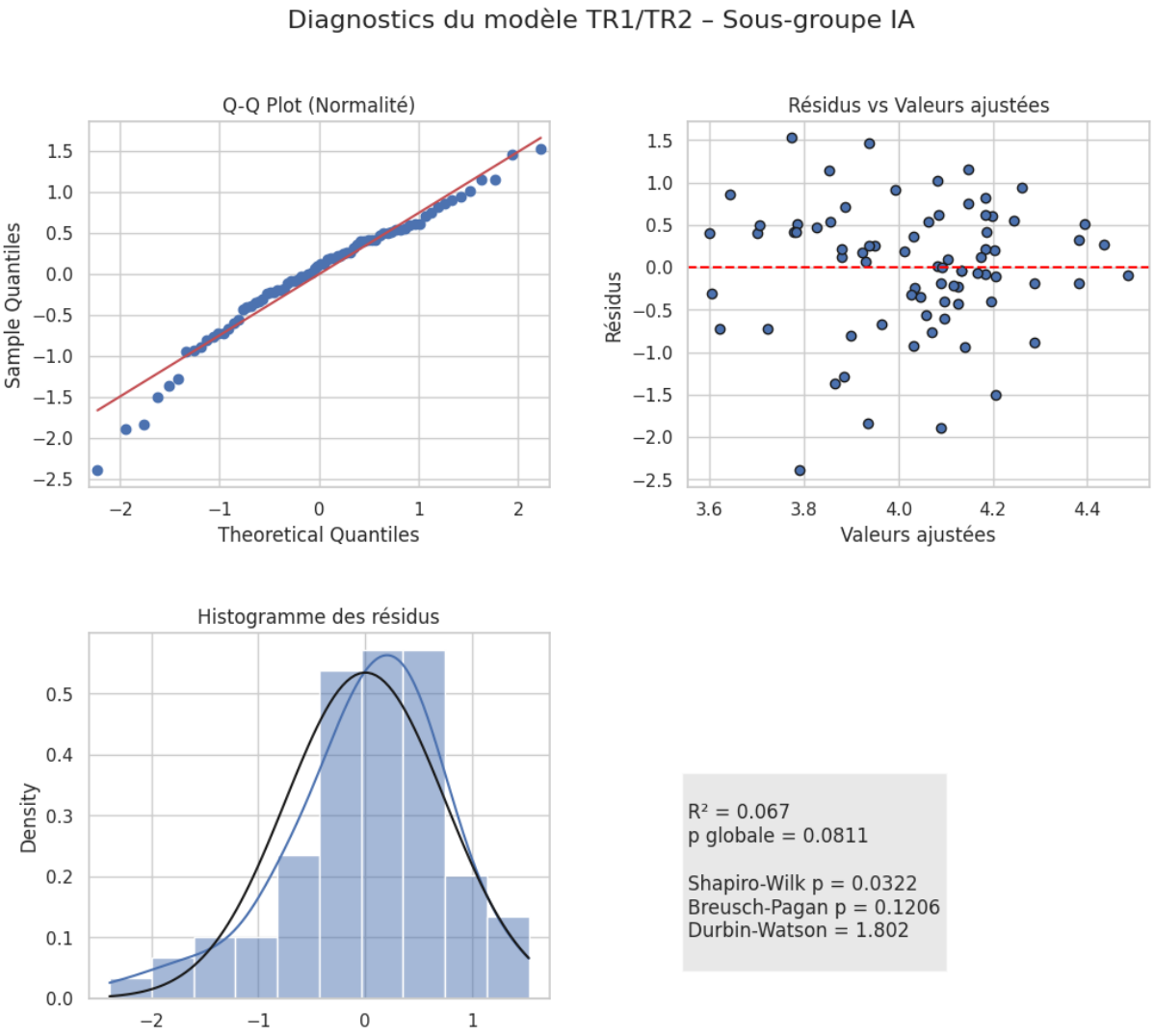
Annexe 16 : Bootstrap du coefficient d'âge



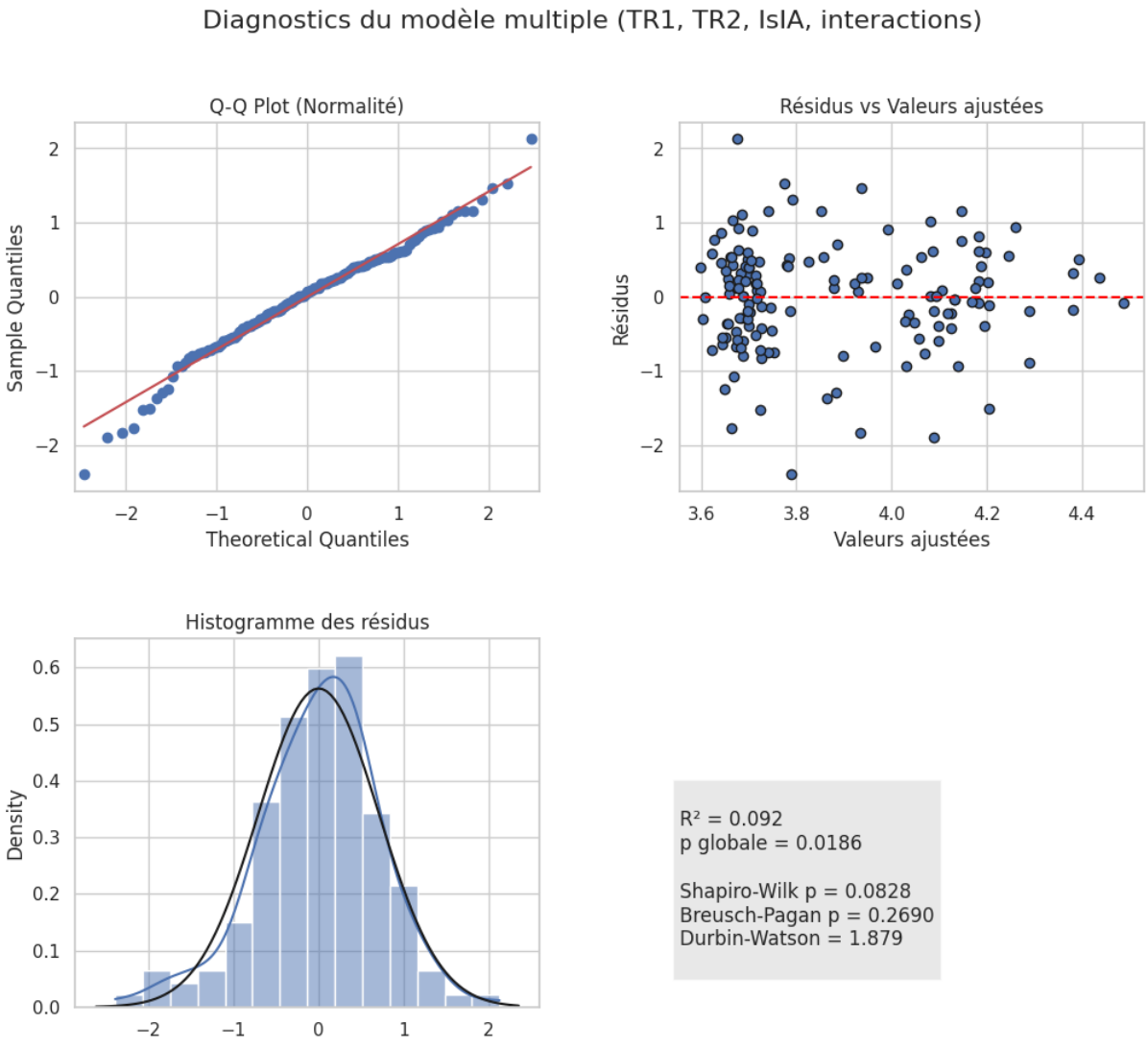
Annexe 17 : Modèle de régression - H2.b



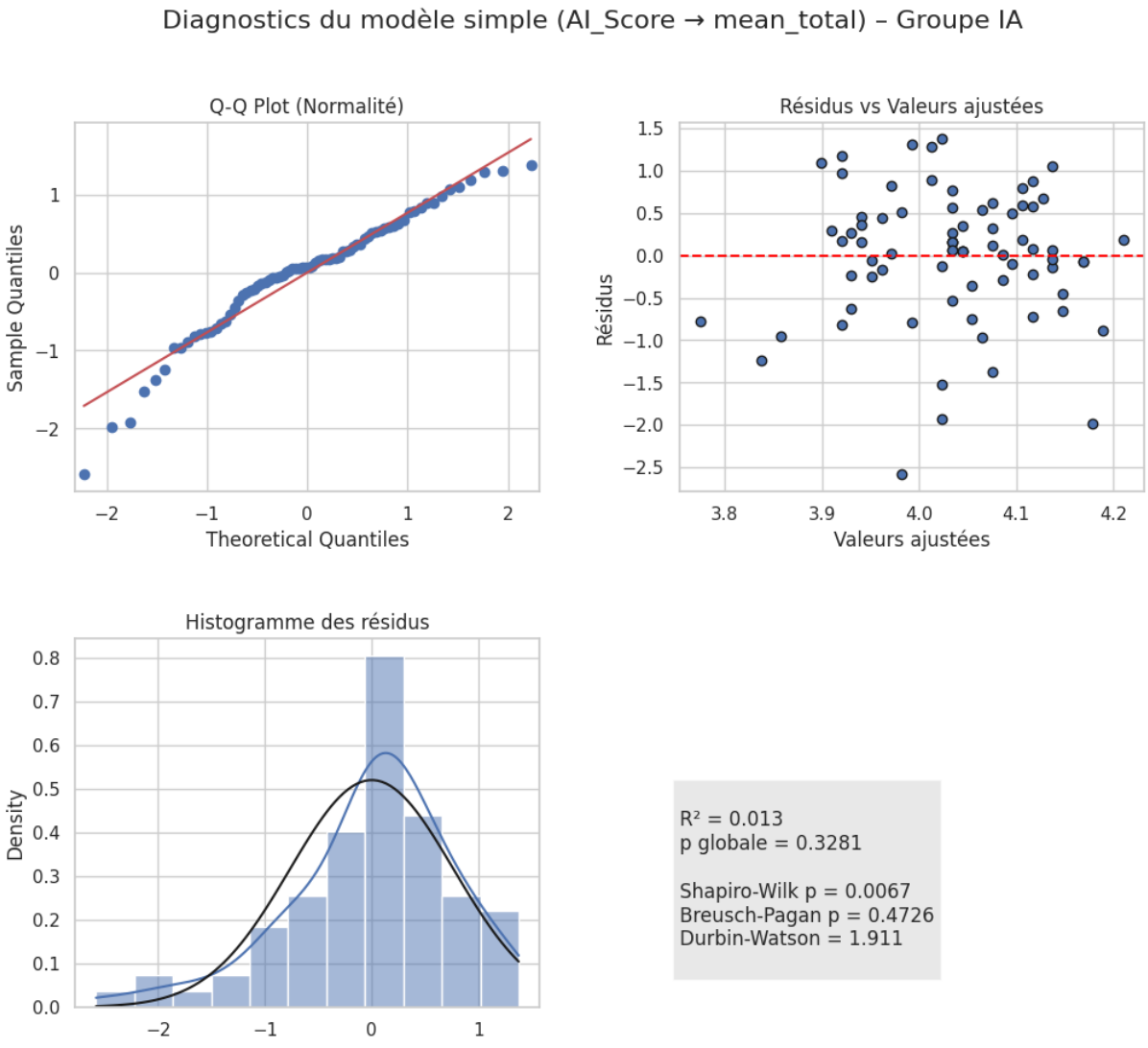
Annexe 18 : Modèle de régression - H4.a



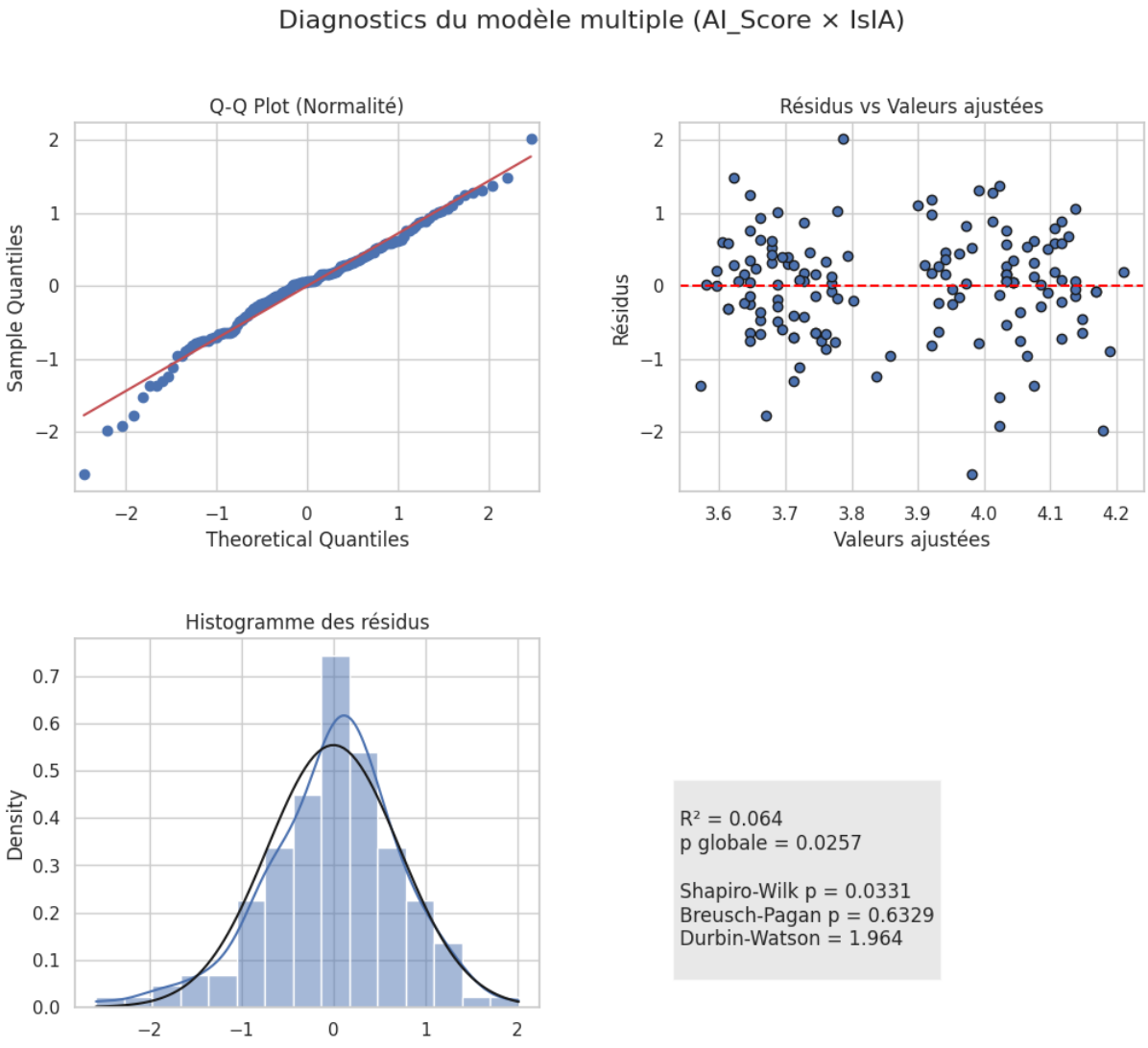
Annexe 19 : Modèle de régression - H4.b



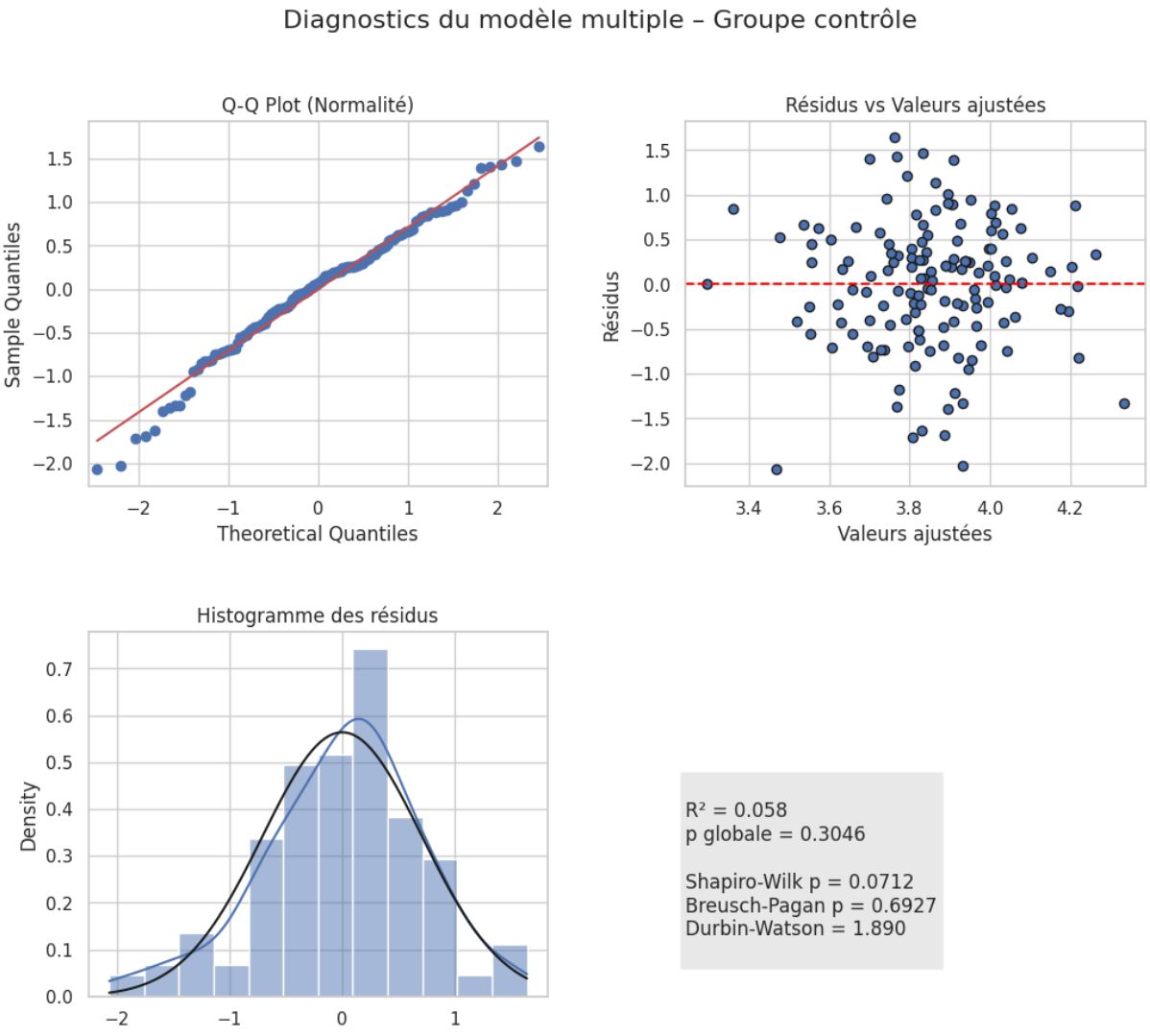
Annexe 20 : Modèle de régression - H5.a



Annexe 21 : Modèle de régression - H5.b



Annexe 22 : Modèle de régression - Modèle global



UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Chaussée de Binche 151, 7000 Mons, Belgique | **www.uclouvain.be/lsm**