

Risques et bénéfices des agents conversationnels pour l'accès profane aux connaissances scientifiques

Risks and benefits of conversational agents for naive access to scientific knowledge

Robert VISEUR¹

¹ Service TIC, Faculté Warocqué d'Économie et de Gestion, UMONS, Belgique, robert.viseur@umons.ac.be

RÉSUMÉ. Cet article analyse les risques associés à l'utilisation d'agents conversationnels génératifs tels que ChatGPT pour l'accès aux connaissances scientifiques (et plus globalement techniques et médicales). Le développement du Web s'est en effet accompagné d'une évolution du processus de *gatekeeping* sous une forme algorithmique dont les intelligences artificielles génératives constituent le dernier avatar. Leurs limitations, et notamment les phénomènes d'hallucinations et de biais, sont cependant connues. Dès lors, ces agents conversationnels se révèlent-ils aptes à des tâches de médiation scientifique ? Leurs performances sont conditionnées aux caractéristiques de leur algorithme mais aussi de la disponibilité de données d'entraînement en quantité et de qualité. Or, l'accès à des contenus de sites de presse est régulièrement entravé par les éditeurs. Qu'en est-il pour les contenus scientifiques gérés par des éditeurs scientifiques commerciaux ? Les producteurs d'agents conversationnels génératifs doivent-ils se contenter de contenus de moindre qualité avec des conséquences dommageables sur la fiabilité des réponses ? Nous analysons dès lors les risques de mésinformation scientifique dus aux contraintes d'accès aux données. Nous discutons ensuite plus globalement ces risques, en cas d'usage en tant que médiateur scientifique, selon différents scénarios d'utilisation.

ABSTRACT. This article examines the risks associated with using generative conversational agents such as ChatGPT to access scientific knowledge (and, more broadly, technical and medical knowledge). The evolution of the Web has been accompanied by a shift in gatekeeping towards algorithmic forms, of which generative artificial intelligences are the latest manifestation. Their limitations, most notably hallucinations and various biases, are, however, well documented. Are these conversational agents therefore suitable for tasks of scientific mediation? Their performance depends not only on the properties of their algorithms but also on the availability of training data in sufficient quantity and quality. Access to content on news websites is, moreover, frequently hindered by publishers. What, then, of scientific content managed by commercial academic publishers? Must developers of generative chatbots rely on lower-quality material, with harmful consequences for the reliability of responses? We therefore analyse the risks of scientific misinformation stemming from constraints on data access. We then discuss these risks more broadly, when such agents are used as scientific mediators, across different usage scenarios.

MOTS-CLÉS. Intelligence artificielle, gatekeeping, mésinformation, jeu de données.

KEYWORDS. Artificial intelligence, gatekeeping, misinformation, datasets.

1. Introduction

L'année 2023 a été celle du développement commercial des intelligences artificielles génératives (IAG) avec l'essor des agents conversationnels (*chatbots*) comme [ChatGPT](#) et plus largement celui des grands modèles de langage (LLM) comme GPT. GPT est « *un modèle linguistique autorégressif (...) qui utilise l'apprentissage profond pour produire des textes semblables à ceux des humains* » (p. 684) [FLO 20]. Les performances de ces modèles dépendent notamment de la quantité et de la qualité des données disponibles pour leur entraînement [HUA 25] [FLO 20] [DOD 21]. Ces données sont souvent issues d'activités de collecte automatisée sur le Web. Malheureusement, Viseur et Delcouch [VID 24] ainsi que Fletcher [FLE 24] ont montré que des politiques de blocage des sites web de la presse en ligne, via le protocole d'exclusion des robots, étaient largement adoptées. Cela limite l'accès des producteurs d'intelligences artificielles génératives à des données de qualité. Or, les agents conversationnels accessibles au grand public sont aussi utilisés comme moyen de recherche d'information [CHA 25] [WU 25], jouant ainsi *de facto* un rôle de *gatekeeper* [POT 25], c'est-à-dire de filtration et d'élimination

« des informations indésirables, inintéressantes ou signifiantes » (p. 169), pour mettre en évidence « des informations de plus ample importance » (p. 170) [DEM 10]. Les agents conversationnels peuvent dès lors contribuer à la mésinformation par la production d’« une information fausse, inexacte ou trompeuse partagée sans intention de tromper » (p. e32-2) [BON 21]. La reprise de ces informations fabriquées, sans recul critique, amène un risque épistémique que Hannigan, McCarthy et Spicer ont baptisé « *botshit* » (par analogie au « *bullshit* ») [HAN 24]. Existe-t-il des raisons crédibles de penser que ce risque pourrait se manifester en matière de *gatekeeping* ou de médiation scientifique, technique ou médicale ? Nous nous focaliserons donc sur l'accès profane à ce type d'information, à son appropriation, et non à l'utilisation par des professionnels, par exemple issus du monde académique.

Les grands modèles de langage (LLM) sur lesquels s'appuient les agents conversationnels sont connus pour leur tendance à halluciner [MAL 24] [YE 23], c'est-à-dire à fabriquer des informations erronées. Par ailleurs, l'absence de données relatives à une thématique peut amener le modèle à broder pour composer une réponse cohérente, plausible, mais contenant des informations inexactes, inventées [HUA 25]. Ces problèmes relèvent donc, soit de la fidélité à la source (« *faithfulness* »), soit de l'exactitude des faits (« *factualness* ») [YE 23]. Dans cet article, nous nous intéressons en particulier aux risques de mésinformation en matière d'information scientifique par les IA génératives (IAG). Les robots d'exploration utilisés par les producteurs d'intelligences artificielles génératives disposent-ils d'un accès homogène à des sources de qualité ou les producteurs doivent-ils se contenter d'un entraînement sur des recherches non validées ou frauduleuses ? L'indisponibilité d'informations scientifiques fiables et complètes pourrait en effet conduire non seulement à des biais mais aussi à des problèmes importants de qualité dans l'information scientifique générée par, d'une part, les agents conversationnels, d'autre part, les plateformes s'appuyant sur des modèles génératifs. Cela inclut par exemple les *newsbots* [VIS 24].

Notre problématique concerne donc les risques de mésinformation induits par l'utilisation d'agents conversationnels génératifs. Nous considérons ces agents conversationnels comme des dispositifs de médiation scientifique. En effet, ils mettent en relation des publics avec des savoirs dans les limites imposées par des contraintes techniques et organisationnelles. Au sein de cette médiation, ils exercent une fonction de *gatekeeping* algorithmique, en assurant des opérations de sélection, de hiérarchisation et d'exclusion d'informations, dont la qualité dépend, entre autres, de leurs données d'entraînement mais aussi des caractéristiques des modèles produits et de l'accès à des données de contexte. Dans cette perspective, notre question de recherche est la suivante : dans quelle mesure ces agents constituent-ils des dispositifs de médiation fiables, et comment leur *gatekeeping* algorithmique (tel que permis par les politiques d'accès des éditeurs et la composition des sources) affecte-t-il l'accès à l'information scientifique ? Cet article est organisé en quatre parties. La première comporte un état de l'art relatif, d'une part, aux concepts de médiation et de *gatekeeper*, d'autre part, aux modalités de construction des *datasets* couvrant des contenus scientifiques permettant d'entraîner les modèles d'intelligence artificielle générative (IAG). La seconde décrit la méthodologie d'analyse. Cette dernière s'appuie sur la mesure du blocage, puis du biais global que cela induit, des robots d'exploration des producteurs d'IAG par les éditeurs scientifiques. Les cinq hypothèses testées sont ensuite présentées. La troisième partie présente les résultats pour chaque hypothèse. La quatrième, précédant la conclusion, discute les implications des résultats en matière de *gatekeeping* et de médiation scientifique, technique et médicale.

2. Revue de la littérature

Les sciences analytiques et techniques, soit les sciences de la nature, font l'objet d'« *un questionnement relatif aux rapports avec la société* » (p. 37) [CAU 08]. En ont résulté des dispositifs de médiation scientifique permettant la mise en relation des acteurs scientifiques et des profanes. Dans le contexte de notre société industrielle, ces dispositifs poursuivent un double objectif d'amélioration de l'acceptabilité des sciences et techniques mais aussi de leur appropriation par les profanes [CAU 08]. En France, la multiplication des « *occasions de contact entre les personnes et la science* » (p. 43) s'exprime notamment suivant deux canaux. Le premier est composé des professionnels de la médiation en matière de culture scientifique et technologique [LAS 11]. Ceux-ci sont notamment actifs au sein des musées,

soit une composante des industries culturelles et créatives [MAI 22]. Le second prend forme dans l'enseignement. La science y serait cependant moins considérée comme « *socle de connaissances minimales* » qu'un « *moyen de sélection* » (p. 57) préparant aux filières scientifiques [MOA 07]. En résulterait dès lors une méconnaissance en matière de culture générale scientifique et technique, y compris parfois au sein des filières scientifiques, que les médias, souvent focalisés sur des sujets conflictuels, peineraient à compenser [MOA 07].

La presse intervient également dans la médiation scientifique du fait de son rôle traditionnel de « *gatekeeper* » [BRO 15] [DEM 10]. Le concept de *gatekeeping* recouvre « *la capacité à sélectionner et à rejeter les contenus à publier* » (p. 2) [BRO 15]. En pratique, si le médiateur assure un rôle de pont entre différents publics tandis que le *gatekeeper* agit davantage comme un filtre, tous les deux constituent des intermédiaires de l'information. Bro et Wallberg distinguent trois modèles de *gatekeeping* [BRO 15]. Le premier, soit le processus d'information, correspondant au modèle historique, se concentre sur « *un processus linéaire de transmission de l'information* » (p. 4). Plusieurs étapes vont être suivies entre l'évènement et le lecteur. Le second, associé à la croissance des plateformes numériques, voit le *gatekeeping* comme un processus de communication au cours duquel différents publics (journalistes, citoyens, décideurs politiques...) vont être mis en relation. Ce mode recouvre par exemple le journalisme participatif où le citoyen devient coproducteur de l'information publiée. Le troisième correspond à un processus d'élimination graduelle dans la continuité du mouvement de désintermédiation alimenté par la diffusion du Web. Le *gatekeeper* y est « *terminal* » (p. 8), c'est-à-dire issus d'en dehors des salles de presse, et cible des centres d'intérêt spécifiques.

La conception traditionnelle du *gatekeeping* est parfois jugée comme un frein à l'innovation. L'idéal participatif butte cependant sur le constat fréquent d'un faible niveau des contributions issues du public [DEM 10]. Ces nouvelles personnes sont par ailleurs complétées par des dispositifs technologiques (portails, moteurs de recherche) « *basés sur un ensemble de valeurs prédéfinies intégrées dans les algorithmes informatiques* » (p. 10) [BRO 15]. Nous pourrions dès lors rajouter un quatrième mode de *gatekeeping* caractérisé par un processus algorithmique. Cardon voit d'ailleurs les algorithmes du Web comme « *les gatekeepers d'aujourd'hui* » [CAR 19]. Les réseaux sociaux numériques et les moteurs de recherche offrent ainsi un filtrage de l'information. Celui-ci est basé sur des principes d'autorité, dans le cas des moteurs de recherche, et de réputation, dans celui des réseaux sociaux numériques [CAR 19]. ChatGPT et par extension les autres agents conversationnels génératifs (Google Gemini, Claude, Grok, Le Chat, DeepSeek...), en tant que moteurs de réponse, offrent une alternative aux moteurs de recherche. Ils introduisent donc un nouveau type de *gatekeeper* algorithmique [POT 25] muni de réelles capacités d'exploration mais dépourvu d'une réelle compréhension des contenus générés [HAN 24] [GEF 23]. De plus, par application du principe GIGO (*garbage in, garbage out*), commun en sciences des données, la qualité de leur réponse dépend de celle des données d'entraînement.

L'entraînement d'une IA connexioniste nécessite une grande quantité de données. Les jeux de données utilisés peuvent comprendre des données internes (*first party data*), des données de partenaires (*second party data*) ou des données de tierces parties (*third party data*) obtenues par collecte directe ou par collecte indirecte [SOB 17]. Dans le cas des LLMs, la littérature a surtout mis en avant les données collectées sur le Web. Floridi et Chiriatti font état d'« *un ensemble de données non étiquetées composé de textes, tels que Wikipédia et de nombreux autres sites, principalement en anglais, mais aussi dans d'autres langues* » (p. 884) [FLO 20]. Le [Common Crawl](#), soit plus ou moins 300 TB avant filtrage¹, représenterait ainsi 60 % des données d'entraînement de GPT-3 [BRO 20]. La qualité des données est également importante. Dodge et ses co-auteurs expliquent ainsi l'appétit des producteurs d'IAG pour les données issues de la presse en ligne, des revues scientifiques et de l'encyclopédie collaborative Wikipédia [DOD 21]. En plus des données d'entraînement, les IA génératives peuvent aussi chercher à accéder à des données de contexte pour répondre aux requêtes des utilisateurs. Quatre robots sont d'un usage courant : GPTBot (utilisé pour la collecte de données d'entraînement), ChatGPT-User (utilisé pour

1 Voir <https://commoncrawl.github.io/cc-crawl-statistics/>.

les actions dans ChatGPT ou les customs ChatGPT², c'est-à-dire pour accéder à des données de contexte), Google-Extended et CCBot (associé au Common Crawl). Fin 2024, OpenAI a rajouté OAI-SearchBot associé à son fonctionnement comme moteur de recherche. Des listes plus complètes existent³. Cependant, elles se répercutent actuellement peu dans les fichiers robots analysés [VID 24].

Cette collecte de données non négociée peut être considérée comme une forme de prédatation par certains éditeurs de contenus. Elle conduit donc à des politiques de blocage par les propriétaires des sites présentant des contenus originaux [VID 24] [DIN 24]. Ce choix peut s'expliquer par, d'une part, la volonté de protéger la propriété intellectuelle associée aux contenus publiés, d'autre part, de préserver les ressources informatiques des sites web [KRO 23]. La contremesure peut relever, soit du blocage passif, soit du blocage actif [DIN 24] [VID 24] [AMI 20] [SUN 07]. Le blocage passif s'appuie sur le protocole d'exclusion des robots⁴. Ce dernier permet de préciser les sections du site qui peuvent être parcourues et celles qui doivent être ignorées [VID 24] [SUN 07]. Seuls les robots dits éthiques respectent ce principe d'*opt-out*. L'utilisateur d'un robot d'exploration peut ainsi choisir d'ignorer sciemment ces consignes et de collecter malgré tout les contenus publiés en ligne. Le blocage actif conduit *a contrario* à une détection du robot puis à son blocage⁵. La détection peut être réalisée simplement en s'appuyant sur des listes de *user agents* ou d'adresses IP. Ces dernières peuvent être fournies spontanément par les propriétaires des robots⁶ ou alimentées par les gestionnaires de sites web. La détection est également possible par le calcul d'empreinte (« *fingerprinting* ») et l'analyse du comportement des terminaux accédant aux sites web [AMI 20]. Ce type d'approche plus sophistiquée est notamment retenu par le service commercial [Cloudflare](#)⁷ [AMI 20]. Une fois repéré, le robot peut aussi être soumis à la résolution d'un *captcha* [AMI 20]. Enfin, une redirection peut également être utilisée par l'exploitation de la fréquente incapacité des robots (sauf s'ils s'appuient sur un « *headless browser* » comme feu PhantomJS ou [Selenium](#) par exemple) d'exécuter les codes Javascript. Les dispositifs de blocage actif sont dès lors nombreux, et aisément accessibles aux éditeurs.

Ces politiques de blocage sont en particulier mises en œuvre par la presse généraliste [FLE 24] [VID 24]. Viseur et Delcucq ont en particulier analysé le comportement des éditeurs de presse face aux producteurs d'IA génératives [VID 24]. Ils montrent que les blocages des robots d'exploration alimentant les jeux de données sont fréquents, basés sur le protocole d'exclusion des robots, et que cela occasionne de nombreux biais, notamment linguistiques et culturels [FER 23]. Le même type de dispositif de protection de la propriété intellectuelle est-il mis en place par les éditeurs scientifiques ? Certaines caractéristiques du marché de l'édition scientifique nous le laissent en tout cas craindre. En effet, l'édition scientifique est devenue un marché lucratif caractérisé par des marges élevées [LAR 15]. Les articles sont fréquemment publiés derrière des *paywalls*. Aussi les grands éditeurs (Elsevier, Springer Nature, Wiley Blackwell, Taylor & Francis...) tendent à défendre la propriété des contenus qu'ils publient [CHA 17]. Leur position sur le marché les autorise par ailleurs à régulièrement augmenter leurs tarifs. Cette situation a suscité plusieurs réactions. D'une part, les articles derrière *paywall* se retrouvent publiés sur des plateformes alternatives, parfois illégales à l'image du *black open access* [BEC 20]. Par exemple, [Sci-Hub](#) est une base de données gratuite, riche de plusieurs dizaines de millions d'articles, souvent toujours couverts par droit d'auteur, dès lors considérée comme une appropriation frauduleuse

2 Voir <https://platform.openai.com/docs/bots>.

3 Voir par exemple <https://github.com/ai-robots-txt/ai.robots.txt>.

4 Voir <https://robots-txt.com/> et <https://datatracker.ietf.org/doc/rfc9309/>.

5 Pour une synthèse illustrée à destination des praticiens, voir par exemple <https://www.willmaster.com/library/tutorials/ways-to-redirect-bots-and-browsers.php>.

6 Voir par exemple la page d'information fournie par OpenAI : <https://platform.openai.com/docs/bots>. Cette page inclut l'accès à des fichiers JSON documentant les IP utilisées par les différents robots.

7 Voir <https://blog.cloudflare.com/declaring-your-aindependence-block-ai-bots-scrapers-and-crawlers-with-a-single-click/>.

d'œuvres protégées par les éditeurs scientifiques [BAN 16]. Compte tenu de son caractère peu ou prou légal, il ne s'agit pas d'un jeu de données exploitables, du moins à long terme, par les producteurs d'IA génératives. D'autre part, le monde de la recherche a encouragé la création de nouveaux journaux publiés en *open access* [BEC 20] [GER 20]. La publication des résultats de recherche dans de tels journaux s'est d'ailleurs trouvée encouragée par certains organismes de financement (p. ex. [Plan S](#)). Le développement des journaux en *open access* (OA) s'est malheureusement accompagné de la prolifération de revues pratiquant un marketing agressif et offrant des taux d'acceptation élevés [RIC 18]. Ces revues acceptent des articles sans processus rigoureux de révision par les pairs, dans un but de profit [XIA 15] [RIC 18]. Le phénomène a notamment été étudié par Jeffrey Beall. Ce dernier a désigné ces journaux comme prédateurs et maintenu une liste pour sensibiliser la communauté académique aux pratiques de publication malhonnêtes [BEA 10]. Les producteurs d'IAG voient donc l'accès facilité à ces revues en *open access*, sans cependant que la qualité des publications soit garantie.

L'automatisation de l'exploration de la littérature scientifique a précédé le développement des LLM. Deux jeux de données scientifiques antérieurs aux premières IAG commerciales ressortent ainsi de la littérature en NLP (*Natural Language Processing*) : S2ORC [LO 20] et Microsoft Academic Graph [WAN 20]. S2ORC est un vaste corpus contenant à sa création 81,1 millions d'articles. Pour une minorité d'articles en *open access* (8,1 millions), le texte intégral est disponible ; pour la plupart, seuls les métadonnées, le résumé et les références sont fournies. S2ORC est associé au moteur de recherche scientifique [Semantic Scholar](#) [KIN 23], qui référence aujourd'hui plus de 200 millions de documents. Microsoft Academic Graph (MAG) prend la forme d'un graphe d'articles (incluant les résumés) régulièrement mis à jour. Ces deux jeux de données ne sont pas spécifiquement liés à l'univers des LLM. Cependant, à l'instar de certains outils basés sur l'IA générative, ils ont été conçus comme des outils de NLP capables de supporter la recherche sur des documents académiques [LO 20] au travers notamment d'agents logiciels aptes à explorer automatiquement la littérature scientifique disponible sur le Web [WAN 20]. Microsoft Academic Graph [WAN 20] [SIN 15] a ultérieurement été fusionné avec [AMiner](#) [TAN 08] pour former [Open Academic Graph](#) (OAG). Sa réutilisation est possible à des fins de recherche uniquement.

La littérature existante donne quelques éclairages sur les *datasets* proposant de l'information scientifique pour l'entraînement des LLM [BRO 20] [GAO 20] [DOD 21] [LIU 24] : [Common Crawl](#), Colossal Clear Crawled Corpus C4 et [The Pile](#). Le Common Crawl est un jeu de données constitué par une exploration à large échelle du Web. Il est notamment utilisé par OpenAI [BRO 20]. Il est aussi utilisé comme *dataset* de base pour la constitution de *datasets* de meilleure qualité après l'application de règles de filtre. C'est notamment le cas du Colossal Clear Crawled Corpus C4 [DOD 21]. Ce dernier s'appuie substantiellement sur les éditeurs de presse (New York Times, LA Times, Washington Post...) et les éditeurs scientifiques (PLOS One, Frontiers...), en plus de [Google Patent](#) pour l'accès aux connaissances techniques. The Pile est un jeu de données de haute qualité incluant 22 sous-*datasets* [GAO 20]. Parmi les jeux de données, deux sont de nature scientifique : [ArXiv](#) (8,96 % du poids total) et [PubMed Central](#) (14,40 % du total). Le premier est un serveur de *preprints*, le second, un répertoire de documents issus de la recherche médicale. Si ArXiv n'offre donc pas une information soumise à un processus strict de *peer reviewing*, PubMed présente par contre un cas hybride. En effet, il s'appuie notamment de manière substantielle sur MEDLINE, qui sélectionne des revues à forte qualité scientifique et éditoriale⁸. Les *datasets* intègrent classiquement des données issues de Wikipédia [DOD 21]. Or, il apparaît que Wikipédia est un bon relais pour l'information scientifique publiée, d'une part, dans des journaux en *open access*, d'autre part, dans des journaux à facteur d'impact élevé, éventuellement protégés par *paywall* [TEP 17].

Tous les contenus scientifiques n'ont en effet pas la même valeur [CAB 243]. Premièrement, un contenu scientifique peut avoir fait ou non l'objet d'une révision par les pairs. Un contenu publié dans une conférence ou une revue à comité de lecture bénéficiera donc d'un niveau de validation supérieur à

8 Voir https://www.nlm.nih.gov/medline/medline_selection.html.

un article en *preprint*. Deuxièmement, à l'intérieur même des conférences ou des journaux scientifiques, une hiérarchie existe, que les articles soient ou non publiés en *open access*. Par exemple, le portail CORE⁹ propose un classement de conférences centrées sur l'informatique. Les conférences sont réparties entre quatre catégories (A*, pour les conférences les plus sélectives, puis A, B et C). Le classement s'appuie sur une combinaison d'indicateurs (p. ex. citations, taux d'acceptation et profil scientifique des organisateurs). Autre exemple : l'indicateur SCImago Journal Rank ([SJR](#)). Celui-ci offre un classement des revues scientifiques contenues dans la base de données [Scopus](#). Il est basé sur une mesure, inspirée du Pagerank de Google [CAR 13], qui tient compte à la fois du nombre de citations reçues par une revue et du prestige des revues d'où proviennent les citations. Pour les producteurs d'IAG, il ressort, d'une part, que les contenus scientifiques les plus accessibles ne sont pas nécessairement les meilleurs, d'autre part, qu'il existe un risque que les comportements protecteurs des éditeurs soient d'autant plus forts qu'une revue scientifique est réputée pour sa haute qualité.

En complément des *datasets* publics, les producteurs de LLM construisent aussi des *datasets* internes. Ceux-ci sont alimentés, nous l'avons vu précédemment, par leurs propres robots. Cependant, des accords de licence sont également signés avec les éditeurs pour accéder à certains contenus de qualité [STR 25]. Dès lors, les données d'entraînement sont composées, d'une part, de données publiques, utilisées après filtrage, collectées sur le Web par des tiers (p. ex. Common Crawl) ou par les producteurs eux-mêmes (p. ex. GPTBot), d'autre part, de données achetées auprès des éditeurs. Gibney mentionne ainsi Taylor & Francis¹⁰ (accord avec Microsoft) et Wiley (partenaire inconnu), pour l'édition scientifique, ainsi que Financial Times (accord avec OpenAI), pour la presse spécialisée [GIB 24]. Le phénomène est difficile à quantifier mais concernerait cependant davantage d'éditeurs¹¹. Elsevier, Springer Nature et De Gruyter Brill sont par exemple cités [KWO 24]. Le [Generative AI Licensing Agreement Tracking](#) tente ainsi d'inventorier les accords souscrits, souvent sans identification des bénéficiaires. Cette politique d'octroi de licences de gré à gré motive aussi les blocages mis en œuvre par les éditeurs [KWO 24]. De manière plus surprenante, des données volées, par exemple issues de Library Genesis (LibGen) [BOD 18], auraient été ponctuellement utilisées, par Meta (modèles Llama) et OpenAI (modèles GPT) notamment¹². Le *dataset* Books3¹³ serait aussi concerné. La composition précise des jeux de données d'entraînement demeure donc au final une inconnue, excepté pour des projets open-sources comme LUCIE¹⁴, où la transparence est au cœur du projet.

Cette revue de littérature nous permet de formuler les cinq hypothèses suivantes. Première hypothèse (H1) : les robots des IA génératives sont davantage bloqués que les robots des moteurs de recherche. En effet, ces derniers participent à un échange de données contre de l'audience [CAR 19]. Les relations entre éditeurs et moteurs de recherche classiques se sont par ailleurs progressivement pacifiées au cours du quart de siècle écoulé [VIS 25]. Seconde hypothèse (H2) : le robot GPTBot est davantage bloqué que les robots d'autres IA génératives. En effet, c'est un constat qui ressort du secteur de la presse, où les robots gérés par OpenAI semblent davantage pénalisés [VID 24]. Troisième hypothèse (H3) : les éditeurs scientifiques commerciaux dominants bloquent davantage les robots d'IA

9 Voir Cf. <https://portal.core.edu.au/conf-ranks/>.

10 Voir <https://www.ccn.com/news/technology/microsoft-taylor-francis-secret-ai-publishing-deal-outrages-academics/> et <https://www.informa.com/globalassets/documents/investor-relations/2024/informa-plc---market-update.pdf> pour plus d'informations.

11 L'exploitation des contenus scientifiques pourrait être facilitée par la structuration des articles selon le format *Journal Article Tag Suite (JATS)*. Voir Kleidermacher et Zou pour un exemple d'exploitation du format JATS par une IA générative [KLE 25].

12 Voir <https://www.theatlantic.com/technology/archive/2025/03/libgen-meta-openai/682093/> (incluant les liens vers les dossiers juridiques) et <https://authorsguild.org/news/you-just-found-out-your-book-was-used-to-train-ai-now-what/> pour plus d'informations.

13 Voir par exemple <https://aicopyright.substack.com/p/the-books-used-to-train-langs> pour une analyse du contenu de ce jeu de données.

14 Voir <https://lucie.chat/> et <https://huggingface.co/collections/OpenLLM-France/lucie-llm-67099ba7b992dee2c32b1f92> pour plus d'informations.

génératives que les autres éditeurs. En effet, la défense de la propriété intellectuelle associée aux articles est souvent âpre chez ces éditeurs [BEC 20] [CHA 17] [LAR 15]. Le caractère commercial de leur activité devrait par ailleurs les amener à adopter un comportement offensif similaire à celui des éditeurs de presse [VID 24]. Quatrième hypothèse (H4) : les revues prédatrices bloquent moins les robots d'IA génératives que les revues non prédatrices. En effet, ces revues constituent une variante dévoyée de l'*open access*. Leur inclination pour une large diffusion de leurs contenus devrait donc être plus forte. Cinquième hypothèse (H5) : mieux une revue scientifique est classée et plus elle bloque les robots d'IA génératives. En effet, les éditeurs scientifiques commerciaux recherchent la rentabilité [LAR 15]. Cette quête devrait donc les amener à davantage défendre les contenus des meilleures revues.

3. Méthodologie

Deux jeux de données sont utilisés, téléchargés respectivement le 23 septembre 2024 et le 12 octobre 2024. Le premier est constitué de la liste de revues prédatrices publiée par Beall, et disponibles sur le site « Beall's List »¹⁵. Le second est constitué des revues évaluées dans le « Norwegian Register for Scientific Journals, Series and Publishers »¹⁶. Il s'agit d'un registre officiel norvégien qui fournit un classement des canaux de publication académiques, utilisé pour mesurer la qualité et la quantité des publications académiques des chercheurs. Les revues y sont classées sur trois niveaux (level 0, level 1, level 2). Le niveau 1 intègre des revues scientifiques respectant les critères de qualité académique élémentaires tandis que le niveau 2 rassemble les meilleurs canaux de publication. Le niveau 0 contient des revues non scientifiques, dédiées par exemple à la vulgarisation, mais aussi parfois des revues prédatrices, comme « Progress in Physics », également incluse dans la liste de Beall. L'appartenance éventuelle au Directory of Open Access Journals¹⁷ (DOAJ) y est indiquée. Parmi les classements publiés, celui de 2024 a été utilisé.

Parmi les listes de revues, certains sites sont cependant enjoignables, inaccessibles ou associés à des redirections (avec un envoi, correct ou non, d'un code HTTP). De plus, de nombreux doublons existent car plusieurs revues peuvent être publiées sur le même site (cas des grands éditeurs par exemple). Aussi un filtrage des URL (suppression des sites enjoignables, calcul des redirections...) a dû être réalisé à l'aide d'un premier script codé en Python. Au final ont été considérés 1.153 sites de revues prédatrices, 4.129 de niveau 0, 7.276 de niveau 1 et 542 de niveau 2 (ces valeurs peuvent être légèrement inférieures lors de l'analyse proprement dites, en cas d'inaccessibilité du site par exemple).

Le filtrage des URL s'appuie notamment sur une analyse du code HTTP¹⁸ renvoyé lors d'une connexion sur le site d'une revue. Premièrement, les sites enjoignables (car fermés par exemple) ont été retirés de la liste. Ont bien été distingués le cas des sites n'existant plus (connexion impossible) et celui des sites ne disposant pas de fichier *robots.txt*. Dans ce second cas, le site est accessible. La connexion à une page existante retourne un code HTTP 200 contre un code 404 si la page n'est pas trouvée. Deuxièmement, le cas des redirections a été traité. Si le standard HTTP est respecté, un code 301 « *Moved Permanently* » est renvoyé. Face à la diversité des codes renvoyés en cas de redirection, une récupération de l'adresse de redirection a été tentée pour la totalité des codes HTTP compris entre 300 inclus et 400 non inclus.

Plusieurs bases de données sont exploitées à l'issue de cette phase de filtrage : revues prédatrices (toutes), revues non prédatrices (toutes), revues non prédatrices de niveau 0, revues non prédatrices de niveau 1, revues non prédatrices de niveau 2 et Top 50 des plateformes non prédatrices. Ce Top 50 est

15 Voir <https://beallslist.net/>.

16 Voir <https://kanalregister.hkdir.no/en>.

17 Voir <https://doaj.org/>.

18 Voir <https://www.iana.org/assignments/http-status-codes/http-status-codes.xhtml> pour une liste de codes HTTP standards.

déterminé en agrégeant les revues par adresse de publication (en général : 1 éditeur = 1 adresse) puis en conservant les cinquante adresses les plus fréquentes. Ces bases de données sont ensuite utilisées pour accéder automatiquement aux fichiers *robots.txt* à l'aide d'un second script codé en Python. Ces fichiers sont sauvés localement (mise en cache) puis explorés. Cette analyse nous permet d'obtenir la liste des robots d'exploration cités, dont est décliné un Top 10 des robots les plus fréquemment cités, d'identifier les pratiques de blocage (trois cas : pas de fichier *robots.txt*, liste blanche ou liste noire) et de déterminer la politique de blocage adoptée pour chaque robot. Sur cette base, il est possible de calculer le taux de blocage ainsi que le biais global par robot. Le biais global, introduit initialement par Sun et ses co-auteurs (2007), est une valeur comprise entre -1 et 1 qui « *représente le pourcentage, en valeur absolue, de sites (de l'échantillon) qui favorisent (signe positif) ou défavorisent (signe négatif) le robot* » (p. 98) [VID 24]. Il est ainsi possible d'estimer la discrimination des robots d'IAG comparativement au robot universel (cas général) puis aux robots des moteurs de recherche, qui sont globalement bien acceptés par les éditeurs de sites web. Le calcul du biais global utilise la procédure documentée par Viseur et Delcucq [VID 24]. Les résultats sont enregistrés dans un fichier journal. Ces fichiers peuvent ensuite être ingérés par ChatGPT pour l'assistance à la production de tableaux de synthèse spécifiques.

Afin de pouvoir discuter plus largement la problématique des biais, une analyse complémentaire de la localisation des serveurs hébergeant les revues prédatrices et non prédatrices a également été réalisée. La localisation des sites web a été déterminée avec un troisième script Python basé sur la solution GeoLite2¹⁹ de MaxMind. Les 10 localisations les plus fréquentes pour les revues prédatrices ont été conservées puis comparées aux revues listées (niveaux 0, 1 et 2).

4. Résultats

Nous présentons dans cette section les résultats de l'analyse des données collectées pour vérifier nos cinq hypothèses.

H1 : Les robots des IA génératives sont davantage bloqués que les robots des moteurs de recherche.

Les robots les plus fréquemment cités par les revues non prédatrices sont, juste après le robot universel (« * »), GPTBot, CCBot, Google-Extended, Googlebot et ChatGPT-User. Les trois premiers robots sont les robots d'exploration utilisés pour la création et la mise à jour des jeux de données.

Robot	Type	Citations	Blocages	Taux de blocage (si robot cité)	Taux de blocage (tous les sites)
googlebot	SE	689	9	1,31 %	0,08 %
bingbot	SE	367	26	7,08 %	0,23 %
ccbot	NLP, LLM (1)	763	666	87,29 %	5,96 %
gptbot	LLM (1)	921	834	90,55 %	7,46 %
chatgpt-user	LLM (2)	679	603	88,21 %	5,39 %
google-extended	LLM (1, 2)	720	657	91,25 %	5,88 %

Tableau 1. Taux de blocage des robots d'exploration (SE : moteur de recherche ; NLP : traitement automatique de la langue ; LLM : grand modèle de langage ; 1 : données d'entraînement ; 2 : données de contexte)

19 Voir <https://dev.maxmind.com/geoip/>.

Le blocage des robots d'exploration des deux moteurs de recherche dominants (Google et Bing) par les revues non prédatrices apparaît sensiblement moindre que celui des robots d'exploration des producteurs d'IA génératives (cf. Tableau 1). Même ChatGPT-User, le robot associé aux actions dans ChatGPT ou dans les « *customs* » ChatGPT fait l'objet d'un blocage fréquent. Surtout, dès lors que le robot est cité, c'est dans l'immense majorité des cas pour être finalement bloqué.

H2 : Le robot GPTbot est davantage bloqué que les robots d'autres IA génératives.

Le robot GPTBot fait l'objet d'un blocage par les revues non prédatrices dans 90,55 % (cf. Tableau 1) des cas où il est mentionné dans le fichier *robots.txt* (81,68 % des sites configurent un tel fichier). Au final, 7,46 % des sites web interdisent l'accès aux pages de contenu. Cette valeur est légèrement plus élevée (9,44 %) pour les sites des revues estampillées DOAJ.

Les autres robots d'exploration des IA génératives font l'objet d'un taux de blocage légèrement inférieur même si l'ordre de grandeur est équivalent. Le biais global pour GPTBot (tous niveaux de revues confondus dans le registre norvégien), soit -0,0743, est le plus élevé, et traduit une discrimination du robot comparativement à d'autres robots poursuivant ou non les mêmes objectifs de collecte.

H3 : Les éditeurs scientifiques internationaux bloquent davantage les robots d'IA génératives.

Les éditeurs internationaux comme Elsevier (ScienceDirect) ou Springer ont un taux de blocage très sensiblement plus élevé que les revues prédatrices ou même que la moyenne des revues non prédatrices. Ce blocage accru conduit à un biais global (cf. Tableau 2) sensiblement plus élevé, en particulier pour le robot d'exploration GPTBot.

Robot	Revues prédatrices	Top 50
googlebot	-0,0141	0,0323
bingbot	-0,0247	0,0323
ccbot	0	-0,1290
gptbot	-0,0018	-0,4194
chatgpt-user	0	-0,1613
google-extended	0	-0,2581

Tableau 2. Biais global (revues prédatrices vs Top 50)

Les politiques de blocage des robots peuvent prendre des allures parfois radicales à l'image de ScienceDirect qui renvoie une erreur 403 (« *forbidden* ») lors de la lecture avec un script du fichier *robots.txt*. En pratique, le fichier existe (un lien non fonctionnel déclencherait une erreur 404) mais son accès est activement bloqué après détection du robot. Ce dernier précise d'emblée la politique : « *# go away ? tell all others not in the list below to stay out!* ». Ce dispositif pourrait s'expliquer par la volonté de freiner l'exploration des sites à large échelle et de limiter l'identification de ressources protégées par le droit d'auteur.

H4 : Les revues prédatrices bloquent moins les robots d'IA génératives que les revues non prédatrices.

Les revues prédatrices se distinguent par, d'une part, le plus faible pourcentage de sites disposant d'un fichier *robots.txt*, d'autre part, le très faible biais global associé aux robots d'IA générative (cf. Tableau 3). Le biais global augmente sensiblement pour les revues de niveau 2.

Robot	Revue prédatrice	Revue (niveau 0)	Revue (niveau 1)	Revue (niveau 2)
googlebot	-0,0141	-0,0027	0,0082	0,0112
bingbot	-0,0247	-0,0012	0,0082	0,0112
ccbot	0	-0,0194	-0,0551	-0,1453
gptbot	-0,0018	-0,0334	-0,0700	-0,2682
chatgpt-user	0	-0,0147	-0,0467	-0,1415
google-extended	0	-0,0167	-0,0529	-0,1750

Tableau 3. Biais global par type de revue

H5 : Mieux une revue scientifique est classée et plus elle bloque les robots d'IA génératives.

Les revues non prédatrices ont une politique de régulation des robots d'exploration d'autant plus systématique que la revue est d'un niveau plus élevé. Cela se marque par l'utilisation plus systématique d'un fichier *robots.txt* (cf. Tableau 4).

Robot	Nombre de sites	Nombre de sites avec <i>robots.txt</i>
Revues prédatrices	1134	852 (75,1 %)
Revues de niveau 0	4070	3262 (80,1 %)
Revues de niveau 1	7169	5845 (81,5 %)
Revues de niveau 2	537	486 (90,5 %)

Tableau 4. Utilisation du protocole d'exclusion des robots

Le taux de blocage augmente avec le niveau de la revue, légèrement jusqu'au niveau 1 puis plus brutalement pour les revues de niveau 2 (cf. Tableau 5).

Robot	Taux de blocage (ccbot)	Taux de blocage (google-extended)	Taux de blocage (gptbot)
Revues prédatrices	0 %	0 %	0,18 %
Revues de niveau 0	1,89 %	1,77 %	3,39 %
Revues de niveau 1	5,36 %	5,33 %	7,04 %
Revues de niveau 2	15,08 %	17,88 %	27,37 %
Top 50	19,35 %	32,26 %	48,39 %

Tableau 5. Taux de blocage en fonction du niveau

De plus, plus la revue est bien classée et plus le biais global présente une valeur négative basse (cf. Tableau 3). Les revues de niveau 2 se distinguent particulièrement des revues de niveau 0 ou 1.

Les résultats observés à partir des taux de blocage et de biais globaux calculés corroborent nos hypothèses H1, H2, H3, H4 et H5.

Du point de vue de la géolocalisation, des disparités existent entre revues prédatrices et non prédatrices. Cette localisation montre une surreprésentation des revues indiennes parmi les revues prédatrices. De plus, les revues non prédatrices ressortent comme globalement moins concentrées sur quelques pays, soit les États-Unis d'Amérique et l'Inde (plus de 50 % des revues prédatrices).

5. Discussion

Après une synthèse des résultats, nous discutons des impacts liés à l'utilisation de littérature grise pour l'entraînement des IA génératives puis des limitations en matière de médiation scientifique. Enfin, avant de présenter les limitations et les perspectives, nous proposons nos recommandations.

5.1. Synthèses des résultats

Hannigan et ses co-auteurs rappellent que « *les chatbots génératifs ne s'intéressent pas à la connaissance intelligente mais à la prédiction* » (p. 477) [HAN 24]. Les grands modèles de langage (LLM, *Large Language Model*) sont en effet entraînés sur de vastes ensembles de données, souvent collectées à partir du Web, pour prédire des contenus. Ils sont capables de générer « *un charabia technique basé sur des motifs de mots dans les données d'entraînement, qui sont elles-mêmes une boîte noire* » (p. 477) [HAN 24]. Deux choses méritent d'être soulignées à ce stade. D'une part, les grands modèles de langage ne disposent pas de la capacité à dégager un consensus scientifique par la compréhension profonde d'un corpus de documents. D'autre part, la qualité des contenus prédits dépend fortement de la qualité des données fournies en entraînement [HUA 25]. Or, si les robots d'exploration sont bloqués par les revues scientifiques bien classées, soit ce que Gefen et Arinze nomment les journaux académiques « top-tier » [GEF 23], et les grandes plateformes comme Springer ou ScienceDirect, le contenu provenant de ces sources de haute qualité risque d'être sous-représenté dans les jeux de données obtenues par *scraping*. En revanche, les revues prédatrices, qui ne bloquent pas ces robots, risquent d'y voir leurs contenus surreprésentés. Or, les revues prédatrices publient souvent des articles sans processus rigoureux de validation par les pairs. Ce déséquilibre entraîne un risque de diminution de la qualité de l'information scientifique que les *chatbots* (ou d'autres applications génératives) peuvent fournir. Dès lors, en exploitant des informations issues de sources peu fiables, les *chatbots* peuvent involontairement contribuer à la propagation de la mésinformation scientifique, une forme de ce que Hannigan et ses co-auteurs nomment « *botshit* » (p. 474) [HAN 24]. Cela est particulièrement préoccupant dans des domaines sensibles comme la santé, l'environnement ou la technologie, où des informations erronées peuvent avoir des conséquences graves. En outre, cela peut influencer négativement la perception de certains sujets scientifiques. Ajoutons que le taux de blocage des revues étiquetées DOAJ (*Directory of Open Access Journals*) est du même ordre de grandeur que celui des revues de niveau 1 et que, en France, l'Archive ouverte HAL ferme aussi l'accès à de multiples robots²⁰ incluant CCBot, GPTBot, Google-Extended et ChatGPT-User. À l'heure actuelle, ces politiques d'*open access* (OA) n'apportent dès lors pas de réponse satisfaisante à cette problématique d'accès aux données. Notre recherche permet ainsi d'enrichir les typologies de biais existantes [JAI 24] [FER 23] [NAV 23] avec le biais de validation, que nous définissons comme la surreprésentation parmi le corpus d'entraînement de données faiblement validées sur un plan scientifique.

5.2. Impacts du recours à la littérature grise

Ces constats relatifs au fonctionnement des IAG amènent plusieurs questions. Premièrement, l'utilisation, semble-t-il dominante, de *preprints* dans les jeux de données spécialisés en information scientifique conduit-il à une dégradation sensible de la qualité de l'information scientifique présente dans les jeux de données ? Deuxièmement, Wikipédia, comme *proxy* permettant d'accéder à l'information scientifique, publiée derrière *paywall* ou bloquée, atteint-il ses objectifs de neutralité de point de vue ou est-il au contraire contaminé par des actions militantes et lobbyistes comme prétendu par le journaliste

20 Voir <https://hal.science/robots> (accédé le 24 septembre 2025).

Erwan Seznec [SEZ 24] ou l'entrepreneur Elon Musk [LEL 25] ? Troisièmement, jusqu'où, et comment, les blocages des différents robots de ChatGPT nuisent-ils à l'information scientifique générée par le moteur de réponse ?

Parmi les jeux de données à orientation scientifique, nous avons aussi vu que l'usage des *preprints* était répandu. Ce choix, notamment dicté par les facilités d'accès, est-il pénalisant du point de vue de la qualité des données collectées ? Dans le domaine de l'informatique, les pratiques de diffusion de recherches sous la forme de *preprints* sur [arXiv](#) a été étudiée par Lin et ses co-auteurs [LIN 20]. Leur recherche a nécessité un travail complexe de réconciliation des *preprints* et des versions publiées dans des actes de conférences ou des revues à comité de lecture. Leur recherche montre que près de 80 % des articles identifiés sont publiés dans un second temps. Les différences constatées concernent « *des révisions adéquates, des auteurs multiples, un résumé et une introduction détaillés, des références étendues et faisant autorité et un code source disponible* » (p. 555) [LIN 20]. L'étude note cependant une tendance à la baisse du taux de publication des *preprints* (passé de 80 % à 75 % en quelques années). Ce haut taux de publication laisse cependant préjuger de la bonne qualité globale des recherches initiales. L'étude de Carneiro et ses co-auteurs, appliquée à la littérature biomédicale ([bioRxiv](#), [PubMed](#)), va dans le même sens [CAR 20]. Les auteurs notent ainsi que la qualité des rapports est équivalente. Un léger avantage en faveur des versions publiées est cependant souligné. Même dans le contexte de la pandémie, les *preprints* sont apparus comme un allié précieux, par exemple pour la compréhension des mécanismes de transmission [MAJ 20]. Majumder et Mandl citent cependant le cas d'un article présent sur [bioRxiv](#) « *qui affirmait à tort que le COVID-19 contenait des « insertions » du VIH* » (p. e628) [MAJ 20]. La pandémie était en effet propice à la prolifération de contenus trompeurs (affirmations scientifiques non prouvées, propagande, théories du complot...) parfois qualifiée d'infodémie [BOT 22] [GIR 22] [BIN 21]. Ce type de contexte complexifie fortement le travail de constitution de *datasets* complets, actualisés et fiables. Malgré cela, les *preprints* semblent ressortir comme un moyen efficace pour alimenter les jeux de données en informations scientifiques fiables et récentes.

Wikipédia est une encyclopédie collaborative et multilingue²¹. Créée en 2001, elle est progressivement devenue un des sites les plus populaires mondialement [WIL 15]. Son contenu peut être édité par n'importe quel visiteur du site à condition qu'il respecte les règles et bonnes pratiques définies par l'encyclopédie. Wikipédia joue un rôle reconnu de *gatekeeper* en matière d'information scientifique et technique²² [WIL 15] [TEP 17], en plus de servir de source de données lors de la constitution de *datasets* pour l'entraînement des intelligences artificielles génératives [DOD 21]. Malgré cette ouverture aux profanes, la fiabilité de Wikipédia est, par exemple, réputée proche de celle de l'encyclopédie Britannica [WIL 15]. C'est en particulier le cas des articles dédiés aux sujets scientifiques [GIL 05]. Les données biographiques présentent également un faible taux d'erreur en plus de porter sur un ensemble de personnalités élargi comparativement aux encyclopédies classiques [VIS 14]. D'une manière plus générale, la qualité d'un article est fortement corrélée à la quantité d'éditions reçues par la page, au nombre de contributeurs actifs sur cette page et à la diversité de connaissances couvertes par ces contributeurs [WIL 07] [ARA 11]. Les articles populaires sont ainsi plus difficiles à dominer [OBO 10]. De plus, Wikipédia apparaît comme un bon relais pour l'information scientifique publiée, d'une part, dans des journaux en *open access*, d'autre part, dans des journaux à facteur d'impact élevé, éventuellement protégés par *paywall* [TEP 17]. Wikipédia est dès lors reconnu pour sa capacité à

21 Voir <https://fr.wikipedia.org/wiki/Wikip%C3%A9dia>.

22 Si l'on s'en réfère aux définitions dites de New York et de Prague de la littérature grise [SCH 12], une partie des documents diffusés via des archives ouvertes peut relever de la littérature grise, notamment lorsqu'il s'agit de prépublications (*preprints*), thèses ou rapports déposés dans des dépôts institutionnels (p. ex. HAL) ou sur des serveurs de *preprints* (p. ex. arXiv). À l'inverse, les contenus de Wikipédia ne relèvent généralement pas de la littérature grise. Il s'agit d'une source tertiaire encyclopédique, utile comme point d'entrée (repérage de notions et de références). Dans la mesure où ces canaux ne garantissent pas, par eux-mêmes, une validation par les pairs au sens académique (et où le statut des documents peut varier d'un cas à l'autre), ils appellent une vigilance particulière quant à la fiabilité et à la traçabilité des informations. C'est pourquoi nous les traitons dans une section commune.

présenter des faits simples. Cependant, il est davantage critiqué pour traiter « *des questions moins tranchées - par exemple l'histoire ou la politique* » [WAT 06].

Certains sujets, davantage polémiques, conduisent en effet à des guerres d'influence sur certaines pages spécifiques. Cela se marque notamment par des guerres d'édition, appelées « *edit wars* »²³ en anglais [WAT 06]. Le journaliste français Erwan Seznec dénonçait ainsi dans un article à charge l'influence de militants écologistes sur des pages de l'encyclopédie dédiées au glyphosate ainsi qu'à certains médias et journalistes jugés par eux trop proches des industriels [SEZ 24]. Ces comportements militants ont notamment été étudiés par Oboler et ses co-auteurs dans le cas du conflit israélo-palestinien [OBO 10]. Les auteurs ont ainsi identifié les motivations conduisant à éliminer des critiques au sein de pages de manière à « *dominer et manipuler des articles pour défendre des agendas politiques et idéologiques* » (p. 284) [OBO 10]. Si les sujets liés directement ou indirectement à la religion sont propices à la controverse, l'information scientifique, technique et médicale donne aussi lieu à des conflits d'édition au sein de l'encyclopédie. Yasseri et ses co-auteurs ont ainsi identifié comme des sujets particulièrement polémiques la production d'électricité nucléaire, l'homéopathie ou la personne de Rudolph Steiner (à l'origine d'un courant pseudoscientifique appelé anthroposophie couvrant notamment certaines pratiques médicales, et agricoles comme la biodynamie) [YAS 14].

Wikipédia peut donc servir de *proxy* pour l'accès à la connaissance scientifique derrière *paywall*, avec un bon niveau de fiabilité dès lors que l'article fait l'objet d'un travail d'édition réellement collectif et ne couvre pas un sujet fortement polémique. Par contre, l'information scientifique reste minoritaire comparativement à d'autres sujets, tels que les personnalités publiques. De plus, elle n'est pas homogène non plus pour tous les domaines [TEP 17]. Teplitskiy et ses co-auteurs remarquent par exemple une surreprésentation des sciences sociales (au regard de leur poids parmi la littérature scientifique) parmi les citations reprises dans le Wikipédia francophone. Ils soulignent cependant l'absence de corrélation avec les distinctions traditionnelles entre sciences exactes et sciences humaines. Wikipédia contribue donc de manière variable, mais globalement positive, à la constitution de jeux de données techniques et scientifiques en vue d'entraîner les IAG. En effet, les matières scientifiques y sont globalement couvertes, mais inégalement, avec le relai de références scientifiques de bonne qualité. Par contre, la fiabilité de l'encyclopédie régresse sur les sujets périphériques, pour lesquels l'activité collaborative est plus faible, et sur les sujets polémiques, pour lesquels les luttes d'influence peuvent conduire à des contenus orientés. Ces sujets peuvent couvrir des sujets techniques, scientifiques et médicaux (nucléaire civil, COVID-19, homéopathie...). Les risques sont cependant évaluables en estimant l'activité par pages de manière à identifier les pages peu ou prou éditées ou celles au contraire sujettes aux guerres d'édition. Le manque de transparence des producteurs d'IAG quant aux données d'entraînement réellement utilisées permet cependant difficilement de savoir comment ces risques sont pris en compte lors de la construction de grands modèles de langage (LLM).

5.3. Limitations en matière de médiation

Pour mieux comprendre la capacité limitée des agents conversationnels génératifs à traiter des informations scientifiques contradictoires, nous prendrons l'exemple du modèle cosmologique Janus (JCM). Si l'on demande un résumé de 5 lignes à ChatGPT 4o, ce dernier propose le texte suivant : « *Le modèle JANUS, conçu par Jean-Pierre Petit, propose un univers à deux feuillets où le temps s'écoule dans des directions opposées. Dans ce cadre, chaque particule de notre univers aurait une « anti-particule » dans un univers miroir connecté par la gravité. Cette idée novatrice pourrait expliquer des phénomènes comme l'énergie et la matière sombres, en attribuant l'influence de l'expansion cosmique à cet univers miroir. Le modèle se distingue par une approche symétrique, sans recourir aux constantes cosmologiques classiques.* » En réalité, ce modèle Janus est un candidat idéal pour illustrer les limites des agents conversationnels en matière de médiation. L'objectif n'est pas ici de discuter de la validité de la théorie, qui relève des sciences physiques et non des sciences informatiques, mais d'en mettre en

23 Voir également la page du projet WikiWarMonitor : <http://wwm.phy.bme.hu/>.

évidence le caractère controversé. Premièrement, elle souffre de la réputation de son principal promoteur (Jean-Pierre Petit), suite à des controverses extra-académiques, incluant notamment l'« affaire Ummo »²⁴, et influant sur sa crédibilité perçue. Deuxièmement, ce promoteur controversé affirme subir une marginalisation²⁵, dont la légitimité peut certes être discutée, mais qui contribue néanmoins à un déficit d'expérimentations et de validations indépendantes, ce qui permet difficilement à cette théorie de passer l'épreuve de la réfutation [LEC 16]. En particulier, le modèle JANUS a très tôt fait l'objet de critiques publiées sous forme de notes techniques de la part de physiciens reconnus tels qu'Alain Riazuelo (IAP) et, plus récemment, Thibault Damour (IHES)²⁶. Le premier affirme douter de son réel intérêt scientifique, insistant par ailleurs à l'époque sur le peu de citations des publications associées, tandis que le second discute les incohérences physiques et mathématiques du modèle. Troisièmement, la diffusion du modèle JANUS tend à se dérouler en marge des circuits scientifiques classiques. Certains médias promouvant cette théorie, parfois à l'initiative du chercheur, sont de surcroît associés dans la littérature, scientifique [CLA 23] ou grise [ADA 22], à des médias complotistes ou proches du complotisme tels que France Soir, Le Courrier des Stratèges et Nexus²⁷. Nous mobilisons ces éléments, non comme des arguments de disqualification intrinsèque du modèle, mais comme un indicateur des enjeux de médiation, et des précautions à prendre quant au statut scientifique des contenus.

Pourquoi ChatGPT présente-t-il le modèle Janus comme simplem ent novateur ou spéculatif suivant les réponses ? Nous posons l'hypothèse que la provenance des données, d'entraînement ou de contexte, explique cette particularité. En effet, si l'on analyse, pour la requête « JP Petit JANUS », la première page de résultats sur Google Scholar, nous constatons que ces résultats proviennent, premièrement, d'une revue non prédatrice, « *Astrophysics and Space Science* », proposée chez Springer, dont le site <https://link.springer.com/> bloque cependant GPTBot ; deuxièm ent, de plateformes de dépôt de documents académiques (HAL, ResearchGate et fs.unm.edu), qui ne bloquent pas systématiquement GPTBot, mais qui n'imposent pas non plus un processus de relecture par les pairs, car ce sont des espaces de partage de prépublications (*preprints*) ou de manuscrits déjà acceptés par des revues officielles ; troisièm ent, de la revue « *Progress in Physics* », incluse dans la liste de Beall, donc considérée comme prédatrice ; et, quatrièm ent, d'une revue profane, *Nexus*, promotrice de théories alternatives (ufologie, parapsychologie...), en dehors de tout circuit scientifique, ayant un fichier *robots.txt* très ouvert. La seule revue à processus de relecture bloque donc GPTBot, et il en va de même pour d'autres articles écrits par le même trio d'auteurs (d'Agostini, Debergh, Petit), comme « *Modern Physics Letters A* », une revue non prédatrice mais à facteur d'impact très faible. Lorsque l'on demande au *chatbot* une information plus critique, les références utilisées proviennent de sources peu fiables et volontiers acquises à la cause (site personnel de Jean-Pierre Petit, Agoravox, Le Courrier des Stratèges...). Les données relatives au modèle Janus proviennent donc实质iellement, soit de revues scientifiques prédatrices, soit de références profanes, parfois associées à la sphère complotiste (France Soir, Le Courrier des Stratèges, Nexus...), dont les serveurs sont plus couramment ouverts aux robots d'OpenAI.

24 Voir <https://www.afis.org/Des-ummoristes-chez-les-ufologues>. Cette réputation est également alimentée par son association, prétendue notamment par Conspiracy Watch (cf. https://www.conspiracywatch.info/Les-delires-de-Jean-Pierre-Petit_a207.html), aux théories complotistes relatives aux attentats du 11 septembre 2001. Eu égard au risque d'« *anathème facile et disqualifiant* », de « *labelisation infamante* » source de « *disqualification dans divers champs sociaux* » (p. 6), lié aux théories du complot [GIR 17], nous ne discutons pas le bien fondé de cette qualification mais insistons plutôt sur les controverses entourant le personnage ainsi que les relais médiatiques de ses théories.

25 Voir par exemple <https://www.lecourrierdesstrategies.fr/jean-pierre-petit-le-conformisme-de-la-caste-interdit-a-un-villani-de-madresser-la-parole/>.

26 Voir en particulier <https://www2.iap.fr/users/riazuelo/cosmo/jpp/p2.html> (2006), pour Alain Riazuelo, ainsi que, pour Thibault Damour, <https://www.ihes.fr/~damour/publications/JanusJanvier2019-1.pdf> (2019) puis <https://www.ihes.fr/~damour/publications/JanusDecembre2022.pdf> (2022).

27 Voir par exemple <https://www.francesoir.fr/videos-les-debriefings/jean-pierre-petit> et <https://lecourrierdesstrategies.fr/2023/01/18/jean-pierre-petit-reinvente-le-cosmos-dans-le-silence-de-ses-pairs/>.

Cette théorie est cependant suffisamment peu problématique sur un plan sociétal pour expliquer l'absence de garde-fous en aval de l'entraînement, comme cela pourrait être le cas pour des questions de violence, de racisme, de santé publique (p. ex. vaccination) ou de climatologie.

Cette neutralité excessive de l'agent conversationnel peut cependant être réduite en lui imposant un rôle spécifique. Par exemple, lui attribuer à un rôle de zététicien le conduira à ajouter la précision suivante : « *En tant que zététicien, on pourrait noter que, bien qu'original et intéressant, le modèle JANUS doit encore démontrer une prédition vérifiable et spécifique qui pourrait le distinguer des autres modèles cosmologiques, car l'acceptation scientifique repose sur la capacité à tester et à vérifier des hypothèses de manière reproductible* ». Soulignons ici aussi que notre objectif n'est pas de prendre position pour ou contre la zététique. Cette dernière est une discipline critique popularisée sur les réseaux et médias sociaux qui suscite cependant certaines controverses [BOD 21]. L'objectif consiste ici à exploiter la classique attribution d'un rôle à l'IA pour influencer sa réponse [SHE 23], en capitalisant sur la revendication de la zététique d'être un « *art du doute* » (p. 80) [BOD 21]. Par ailleurs, cet impact du rôle peut être reproduit sur d'autres recherches. Ainsi, l'interrogation visant à obtenir « *une liste des insecticides utilisés contre le carpocapse aux Etats-Unis* » débouchera sur une liste d'insecticides (Altacor, Delegate, Exirel...) dans le cas où le rôle attribué sera celui d'un « *agronome spécialisé dans les opérations de contrôle des populations d'insectes* », là où celui d'un « *instagrameur passionné par les médicaments biologiques et naturelles* » aboutira à une réponse démarquant par une liste de procédés physiques, incluant l'utilisation de pièges et de substances collantes, ou biologiques, incluant le recours à des prédateurs.

Dès lors, peut-on déterminer si, et à quelles conditions, un agent conversationnel constitue, d'une part, un bon *gatekeeper* algorithmique en matière d'accès à l'information scientifique, technique et médicale, d'autre part, un bon médiateur pour ces mêmes disciplines ? Le *gatekeeping* s'apparente à un travail de filtrage éditorial. Sur ce plan, un agent conversationnel est entraîné sur base de jeux de données non exempts de biais [FER 23] [NAV 23] mais néanmoins expurgés des contenus les plus problématiques [DOD 21]. Par contre, nous avons montré que l'accès aux contenus scientifiques les plus reconnus était contrarié par les politiques de blocage appliquées aux données d'entraînement mais aussi de contexte. En particulier, le biais de validation vient grêver la qualité de sa fonction de *gatekeeper*. Son fonctionnement stochastique induit par ailleurs des réponses naturellement plus consensuelles, plus lisses. Cependant, les capacités à synthétiser un travail sont limitées, notamment par le risque d'hallucination [HUA 25] [MAL 24] [YE 23], la tendance à la sur-généralisation [PET 25] ou l'influence de la position de l'information au sein d'un document [WAN 24]. De plus, sa capacité à sourcer précisément reste perfectible [WU 25]. La médiation peut s'appuyer sur les capacités de *gatekeeping* mais poursuit un objectif plus ambitieux d'appropriation de la connaissance scientifique. En la matière, les agents conversationnels fournissent une interface conviviale et interactive pour les utilisateurs [ABD 24]. Cette facilité d'utilisation contribue à l'adoption de ces technologies. Ces agents peuvent par ailleurs faciliter l'exploration guidée d'un sujet. L'utilisateur dispose dès lors d'un véritable dispositif de médiation, toujours disponible, capable d'adapter sa réponse à son profil et d'assister l'exploration d'un sujet en simplifiant puis en approfondissant progressivement les explications. Dès lors, cela démocratise l'accès à l'information, notamment scientifique, technique et médicale. Par contre, la neutralité de l'agent conversationnel, lorsqu'elle est constatée, est surtout une façade. Ainsi les procédures d'alignement (p. ex. apprentissage par renforcement) incorporent des préférences humaines [OUY 22] [FLO 20]. Les IA génératives tendent aussi à entretenir les croyances exprimées par les utilisateurs, par renforcement d'un biais de confirmation [SHA 24] ou par flagornerie (« *sycophancy* ») [FAU 25], en particulier sur les sujets où une vérité fondamentale (« *ground truth* ») fait défaut [CHE 25].

Le comportement de l'agent conversationnel dépend aussi des sujets traités. Balhorn et ses co-auteurs ont ainsi évalué la capacité de ChatGPT à répondre à des questions dans les domaines des sciences naturelles et de l'ingénierie [SCH 24]. Leurs résultats montrent que « *les réponses de ChatGPT sont, en moyenne, perçues comme "globalement correctes"* ». En particulier, la fiabilité de l'agent se dégrade lorsque « *le niveau éducatif de la question augmente* » (complexité) et dès lors que la demande nécessite

des « compétences allant au-delà des connaissances scientifiques » (p. ex. esprit critique). Sur le plan des contenus, les IA génératives restent sujettes aux hallucinations lorsque la réponse nécessite de mobiliser les limites de la connaissance, que celle-ci est rare car spécialisée, récente ou d'un accès restreint par droits d'auteur [HUA 25]. Sur le plan des capacités de raisonnement, les LLM sont, avec les *Language Reasoning Models* (LRM), progressivement optimisés pour les tâches nécessitant un raisonnement intensif. Reste que Shen et ses co-auteurs nous montrent que les modèles successifs améliorent progressivement les performances sans pour autant effacer l'ensemble des limitations constatées [SHE 23]. Optimisées sur des tâches de médiation citoyenne, des IA peuvent cependant déjà aider à cartographier des positions et à résumer des controverses [TES 24]. Dès lors, les agents conversationnels génératifs accessibles au grand public nous semblent pouvoir se révéler de bons *gatekeepers* algorithmiques, voire de bons médiateurs, plutôt sur des sujets de connaissance de base (p. ex. explication d'un principe physique ou d'un procédé industriel), moins sur des sujets complexes et/ou polémiques (p. ex. production électrique et agriculture industrielle).

5.4. Recommandations

Les limites inhérentes à la technologie, au-delà de la seule problématique des *datasets*, nécessite la formation des usagers à davantage de littératie algorithmique [LED 22], c'est-à-dire à « une conscience critique de ce que sont les algorithmes, une compréhension de leur mode d'interaction avec les données comportementales humaines dans les systèmes d'information, et une compréhension des enjeux sociaux et éthiques liés à leur utilisation » (p. 332), soit ici une « generative engine literacy » complémentaire à la « search engine literacy » déjà théorisée par Le Deuff [LED 22]. Les acteurs historiques de la médiation scientifique et technique pourraient contribuer à cette montée en compétence. Cet apprentissage couvre deux dimensions, d'une part, la compréhension des risques inhérents aux intelligences artificielles génératives, d'autre part, le développement des compétences en promptologie. En matière d'évaluation des risques, Hannigan et ses co-auteurs proposent une typologie de risques fonction de l'importance du contrôle de la véracité et de la complexité de la vérification [HAN 24]. Si l'utilisation d'une IAG pour générer de nouvelles idées ne présente pas de grands risques, il n'en va pas de même lorsque l'outil est utilisé pour un traitement automatisé complexe où le fonctionnement en boîte noire réduit sensiblement le contrôle par l'utilisateur. En particulier, le blocage des robots d'exploration conduit à des risques spécifiques en fonction des scénarios d'utilisation. D'une part, l'utilisation de ChatGPT comme moteur de réponse (sans source) conduit à un risque d'hallucination accru dès lors que le domaine est mal couvert par les documents présents dans les données d'entraînement. D'autre part, dans le cas où l'agent est utilisé pour traiter des documents en ligne, la limitation concerne les données de contexte, et le robot impacté n'est plus le même. Cependant, certains travaux d'analyse peuvent se révéler impossibles. L'agent peut alors privilégier certaines sources plus accessibles au détriment d'autres parfois de meilleure qualité. En matière de promptologie, nous avons vu que le rôle affecté à l'IAG pouvait imposer un niveau de langage ou forcer un recul critique face au propos généré. Au-delà de la structure du *prompt*, sa rédaction va également contribuer à renforcer ou non le biais de confirmation dans la réponse [FAN 25] [SHA 24].

En complément, une régulation de l'accès aux données d'entraînement paraît nécessaire, qu'elle se fasse à l'initiative des pouvoirs publics (ce que tente de réaliser l'[AI Act](#) par exemple) ou des acteurs privés. Le développement peu ou prou contrôlé de l'intelligence artificielle générative (IAG) pourrait ainsi conduire à une forme de tragédie des communs, et dans notre cas des communs scientifiques, c'est-à-dire la dégradation et l'épuisement d'un bien commun informationnel « comme conséquence de comportements opportunistes » (p. 155) [ZIM 23]. Le cycle suivant tend en effet à se développer. Les producteurs d'IAG accèdent gratuitement aux contenus des éditeurs en les collectant sur le Web. En réaction, les éditeurs interdisent l'accès aux robots d'exploration de ces producteurs par des dispositifs actifs (détection) et passifs (protocole d'exclusion). Dès lors, face à l'extension des dispositifs d'*opt-out* et de blocage des *crawlers* d'IA par une part croissante des grands éditeurs, les producteurs d'IAG se retrouvent, soit à s'appuyer plus fortement sur des sources subsistantes de moindre qualité [LON 24], soit à négocier des licences d'accès (y compris avec des éditeurs scientifiques), comme l'attestent les accords récents [STR 25] [KWO 24]. En l'absence de transparence sur les jeux de données réellement

utilisés, l'évaluation des risques reste difficile. Ces restrictions peuvent réduire la qualité et la diversité des données, ce qui alimente le phénomène d'hallucinations [HUA 25], dès lors la diffusion de contenus partiellement fabriqués. La gestion des données d'entraînement relève donc d'une forme de « *dilemme social* » où « *l'intérêt individuel immédiat s'oppose à celui du groupe* » (p. 153) [ZIM 23]. Ce constat relance l'intérêt d'une gestion de la connaissance en tant que commun [BEC 20]. La gouvernance de ce commun devrait être polycentrique [OST 10]. Pour des raisons évidentes de symétrie de pouvoir, elle associerait alors les grands organismes financeurs de la recherche, en tant que représentants des chercheurs, à la fois producteurs et consommateurs principaux des savoirs [BEC 20], et les intermédiaires œuvrant à la diffusion des savoirs, incluant les éditeurs scientifiques et les producteurs d'IA génératives. La concentration documentée des intermédiaires éditoriaux [POT 25] [LAR 15] plaide à la fois pour la crédibilité de cette négociation et pour une vigilance accrue face aux risques de capture du commun.

5.5. Limitations et perspectives de la recherche

Cette recherche présente deux limitations. Premièrement, nos résultats reposent sur un *proxy*, à savoir la configuration du fichier « *robots.txt* » (blocage passif), observé au niveau des sites web des éditeurs scientifiques. Cette approche ne comptabilise donc pas les blocages actifs. De plus, elle ne permet pas d'en déduire l'exposition effective des modèles aux contenus, ni la composition réelle des jeux de données. Deuxièmement, le panel de robots étudiés, bien que reprenant les robots les plus couramment cités, n'est pas exhaustif. Sans probablement changer fondamentalement nos conclusions, une couverture plus complète contribuerait à particulariser nos conclusions à d'autres moteurs de recherche conversationnels.

Cette recherche présente quatre perspectives. Premièrement, l'analyse des risques a permis d'identifier des scénarios d'utilisation (p. ex. sujets polémiques) davantage propices à la mésinformation. Une évaluation davantage systématique sur un ensemble de sujets scientifiques, techniques et médicaux permettrait dès lors de mieux caractériser les risques rencontrés par les utilisateurs dans leurs interactions avec les agents conversationnels utilisés comme médiateurs. Deuxièmement, la discussion des risques s'appuie substantiellement sur l'agent conversationnel ChatGPT. Cependant, tous les agents ne présentent pas nécessairement les mêmes performances, les mêmes comportements. Existe-t-il donc des différences entre les principaux agents conversationnels (p. ex. ChatGPT, Google Gemini, Grok, Claude, DeepSeek et Le Chat) dans leurs capacités à donner accès aux savoirs scientifiques, techniques et médicaux ? Troisièmement, les sujets scientifiques à fort enjeu sociétal sont confrontés à l'existence d'une littérature partisane, facilement accessible, qu'il s'agisse, soit de science militante, portée par des organisations à but non lucratif poursuivant des objectifs politiques [WOE 24], soit de science lobbyiste, portée par des industriels poursuivant des objectifs économiques [FOU 20]. Le débat anime depuis quelques années l'espace médiatique français. La difficulté, sur le plan de la qualité de l'information scientifique, vient ici du fait que la publication des recherches jugées partisanes se produit également dans des revues considérées de qualité. Comment, et à quel point, les IAG sont-elles affectées par ce type de littérature ? Quatrièmement, la mise en évidence des faiblesses des IA génératives est rarement assortie d'un contrepoint expérimental avec des médiateurs ou des *gatekeepers* humains sur des tâches équivalentes. Pourtant, des biais, notamment de sélection [VAN 17], existent aussi chez les *gatekeepers* humains. De plus, les réponses apportées dans le domaine médical par les *chatbots* se trouvent par exemple appréciées pour leur qualité et leur empathie [AYE 23]. Comment se manifestent les limites en matière de *gatekeeping* et de médiation, d'une part, des professionnels humains, d'autre part, des agents conversationnels génératifs ? Quels sont leurs profils de risque respectifs ?

6. Conclusion

Le développement de l'intelligence artificielle générative (IAG) s'est accompagné de l'utilisation d'agents conversationnels. Utilisés comme moteurs de réponse, ceux-ci jouent un rôle de *gatekeeper* algorithmique. Ce changement dans les usages amène cependant un risque de mésinformation scientifique. L'étude met ainsi en lumière l'impact des restrictions d'accès appliquées par les éditeurs

scientifiques aux robots d'exploration des intelligences artificielles génératives. D'une part, l'accès aux contenus intégraux est souvent bloqué par *paywall*; d'autre part, l'accès aux sites des revues, donc des résumés des articles (si ces derniers sont derrière *paywalls*), est interdit par usage du protocole d'exclusion des robots. Nos résultats montrent le risque d'une prépondérance de contenus issus de revues prédatrices (ou plus largement de revues de moindre qualité) dans les données d'entraînement des IAG, créant ainsi un biais de validation. Ce biais expose les utilisateurs à une mésinformation scientifique, potentiellement amplifiée dans des domaines sensibles. La recherche souligne l'urgence de développer des stratégies de rééquilibrage des *datasets* en favorisant un accès contrôlé et éthique aux contenus validés par des pairs, afin d'améliorer la qualité et la fiabilité des réponses fournies par les modèles d'IAG. Elle analyse également les risques de mésinformation scientifique. À cette occasion, les conséquences des restrictions d'accès aux données sont resituées parmi les limitations inhérentes aux grands modèles de langage (LLM) et aux modalités d'interrogation des agents conversationnels génératifs. Certes, les agents conversationnels fournissent un moyen convivial et interactif pour accéder aux savoirs scientifiques. Cela leur confère, dans la pratique, des capacités de *gatekeeping* et de médiation, facilitant l'accès et l'appropriation des connaissances scientifiques, techniques et médicales. Leur usage nécessite cependant encore un socle minimal de compétences pour palier les limitations de l'outil (p. ex. biais, hallucinations et flagornerie).

7. Références

- [ABD 24] Abdaljaleel, M., Barakat, M., Alsanafi, M., Salim, N. A., Abazid, H., Malaeb, D., ... & Sallam, M. (2024). A multinational study on the factors influencing university students' attitudes and usage of ChatGPT. *Scientific Reports*, 14(1), 1983. <https://doi.org/10.1038/s41598-024-52549-8>.
- [ADA 22] Adamczyk, R., Simmons, C., & Morinière, S. (2022). Le spectre de la fraude électorale : l'impact des discours de désinformation pendant les élections de 2022. ISD (Institute for Strategic Dialogue). <https://www.isdglobal.org/isd-publications/le-spectre-de-la-fraude-electorale-limpact-des-discours-de-desinformation-pendant-les-elections-de-2022/>.
- [AMI 20] Amin Azad, B., Starov, O., Laperdrix, P., & Nikiforakis, N. (2020). Web runner 2049: Evaluating third-party anti-bot services. In *Detection of Intrusions and Malware, and Vulnerability Assessment: 17th International Conference, DIMVA 2020, Lisbon, Portugal, June 24–26, 2020, Proceedings* 17 (pp. 135-159). Springer International Publishing. https://doi.org/10.1007/978-3-030-52683-2_7.
- [ARA 11] Arazy, O., Nov, O., Patterson, R., & Yeo, L. (2011). Information quality in Wikipedia: The effects of group composition and task conflict. *Journal of management information systems*, 27(4), 71-98. <https://doi.org/10.2753/MIS0742-1222270403>.
- [AYE 23] Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., ... & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine*, 183(6), 589-596. <http://doi.org/10.1001/jamainternmed.2023.1838>.
- [BAN 16] Banks, M. (2016). What Sci-Hub is and why it matters. *American Libraries*, 47(6), 46-49. <https://www.jstor.org/stable/26380679>.
- [BEA 10] Beall, J. (2010). "Predatory" open-access scholarly publishers. *The Charleston Advisor*, 11(4), 10-17.
- [BEC 20] Bechis, J. (2020). La gouvernance du marché de la publication scientifique à l'ère du numérique: une analyse de la connaissance scientifique en tant que commons. *Innovations*, 63(3), 95-123. <https://doi.org/10.3917/inno.063.0095>.
- [BIN 21] Bin Naeem, S., & Kamel Boulos, M. N. (2021). COVID-19 misinformation online and health literacy: a brief overview. *International journal of environmental research and public health*, 18(15), 8091. <https://doi.org/10.3390/ijerph18158091>.
- [BOD 21] Bodin, C. (2021). La zététique ou les usages multiples d'une mise en récit scientiste du monde social. *Les Enjeux de l'information et de la communication*, 223(S2), 77-89. <https://doi.org/10.3917/enic.hs11.0077>.
- [BOD 18] Bodó, B. (2018). Library Genesis in Numbers: Mapping the Underground Flow of Knowledge. In: Karaganis, J. (2018). Shadow libraries. Access to knowledge in global higher education. The MIT Presse. <https://library.oapen.org/bitstream/handle/20.500.12657/26038/11339.pdf>.
- [BON 21] Bontridder, N., & Poulet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy*, 3, e32. <https://doi.org/10.1017/dap.2021.20>.

- [BOT 22] Bottemanne, H. (2022). Théories du complot et COVID-19 : comment naissent les croyances complotistes?. *L'encephale*, 48(5), 571-582. <https://doi.org/10.1016/j.encep.2021.12.005>.
- [BRO 15] Bro, P., & Wallberg, F. (2015). Gatekeeping in a Digital Era. *Journalism Practice*, 9, 105 – 92. <https://doi.org/10.1080/17512786.2014.928468>.
- [BRO 20] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165. <https://doi.org/10.48550/arXiv.2005.14165>.
- [CAB 23] Cabanac, G. (2023). Fake Science: Misconduct Galore and Proposed Counterattack. Doctoral. IPBS, France. <https://ut3-toulouseinp.hal.science/hal-04225515v4>.
- [CAR 13] Cardon, D. (2013). Dans l'esprit du PageRank: une enquête sur l'algorithme de Google. *Réseaux*, (1), 63-95. <https://shs.cairn.info/revue-reseaux-2013-1-page-63>.
- [CAR 16] Cardon, D. (2019). Culture numérique. Presses de Sciences Po. ISBN : 978-2724623659.
- [CAR 20] Carneiro, C. F., Queiroz, V. G., Moulin, T. C., Carvalho, C. A., Haas, C. B., Rayée, D., ... & Amaral, O. B. (2020). Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical literature. *Research Integrity and Peer Review*, 5, 1-19. <https://doi.org/10.1186/s41073-020-00101-3>.
- [CAU 08] Caune, J. (2008). La culture scientifique: une médiation entre sciences et société. *Lien social et Politiques*, (60), 37-48. <https://doi.org/10.7202/019444AR>.
- [CHA 25] Chatterji, A., Cunningham, T., Deming, D. J., Hitzig, Z., Ong, C., Shan, C. Y., & Wadman, K. (2025). How people use ChatGPT (Working Paper No. 34255). National Bureau of Economic Research. <https://doi.org/10.3386/w34255>.
- [CHA 17] Chawla, D. S. (2017). Publishers take academic networking site to court. *Science*, vol. 358, issue 6360, p. 161. <https://www.science.org/doi/pdf/10.1126/science.358.6360.161>.
- [CHE 25] Cheng, M., Yu, S., Lee, C., Khadpe, P., Ibrahim, L., & Jurafsky, D. (2025). Social sycophancy: A broader understanding of llm sycophancy. arXiv preprint arXiv:2505.13995. <https://doi.org/10.48550/arXiv.2505.13995>.
- [CHU 24] Chu, Z., Wang, Z., & Zhang, W. (2024). Fairness in large language models: A taxonomic survey. *ACM SIGKDD explorations newsletter*, 26(1), 34-48. <https://doi.org/10.1145/3682112.3682117>.
- [CLA 23] Clavert, F., & Paci, D. (2023). Traces et mémoires en devenir d'une pandémie (Italie et France). *Passés Futurs*, (14). <https://doi.org/10.4000/157kc>.
- [DEM 10] De Maeyer, J. (2010). Être journaliste dans un environnement 2.0 : Les médias belges face aux innovations technologiques. IEEE Conference on Local Computer Networks. <https://doi.org/10.3166/LCN.6.1.157-178>.
- [DIN 24] Dinzinger, M., & Granitzer, M. (2024). A longitudinal study of content control mechanisms. In Companion Proceedings of the ACM on Web Conference 2024 (pp. 1382-1387). <https://doi.org/10.1145/3589335.3651893>.
- [DOD 21] Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., ... & Gardner, M. (2021). Documenting large webtext corpora: A case study on the colossal clean crawled corpus. arXiv preprint arXiv:2104.08758. <https://doi.org/10.48550/arXiv.2104.08758>.
- [FAN 25] Fanous, A., Goldberg, J., Agarwal, A. A., Lin, J., Zhou, A., Daneshjou, R., & Koyejo, S. (2025). Syceval: Evaluating LLM sycophancy. arXiv preprint arXiv:2502.08177. <https://doi.org/10.48550/arXiv.2502.08177>.
- [FER 23] Ferrara, E. (2023). Should chatgpt be biased? Challenges and risks of bias in large langage models. arXiv preprint arXiv:2304.03738. <https://doi.org/10.5210/fm.v28i11.13346>.
- [FLE 24] Fletcher, R. (2024). How many news websites block AI crawlers? [Factsheet]. Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/how-many-news-websites-block-ai-crawlers>.
- [FLO 20] Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30, 681-694. <https://doi.org/10.1007/s11023-020-09548-1>.
- [FOU 20] Foucart, S., Horel, S., Laurens, S. (2020). Les gardiens de la raison : Enquête sur la désinformation scientifique. La Découverte. ISBN : 978-2348046155.
- [GAO 20] Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., ... & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics*, 50(3), 1097-1179. https://doi.org/10.1162/coli_a_00524.
- [GAO 20] Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., ... & Leahy, C. (2020). The pile: An 800gb dataset of diverse text for language modeling. arXiv preprint arXiv:2101.00027. <https://doi.org/10.48550/arXiv.2101.00027>.

- [GER 20] Gershenson, S., Polikoff, M. S., & Wang, R. (2020). When paywall goes AWOL: The demand for open-access education research. *Educational Researcher*, 49(4), 254-261. <https://doi.org/10.3102/0013189X20909834>.
- [GIB 24] Gibney, E. (2024). Has your paper been used to train an AI model? Almost certainly. *Nature*, 632(8026), 715-716. <https://doi.org/10.1038/d41586-024-02599-9>.
- [GIL 05] Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 14 décembre 2005. <https://doi.org/10.1038/438900a>.
- [GIR 17] Giry, J. (2017). Étudier les théories du complot en sciences sociales: enjeux et usages. *Quaderni*, (94), 5-11. <https://doi.org/10.4000/quaderni.1101>.
- [GIR 22] Giry, J. (2022). Fake news et théories du complot en période (s) pandémique (s). *Quaderni*, 106(2), 43-64. <https://doi.org/10.4000/quaderni.2303>.
- [HAN 24] Hannigan, T. R., McCarthy, I. P., & Spicer, A. (2024). Beware of botshit: How to manage the epistemic risks of generative chatbots. *Business Horizons*. <https://doi.org/10.1016/j.bushor.2024.03.001>.
- [HUA 25] Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2), 1-55. <https://doi.org/10.1145/3703155>.
- [JAI 24] Jain, R., Jain, A. (2024). Generative AI in writing research papers: a new type of algorithmic bias and uncertainty in scholarly work. In: Arai, K. (eds) Intelligent Systems and Applications. IntelliSys 2024. Lecture Notes in Networks and Systems, vol 1065. Springer, Cham. https://doi.org/10.1007/978-3-031-66329-1_42.
- [KIN 23] Kinney, R., Anastasiades, C., Authur, R., Beltagy, I., Bragg, J., Buraczynski, A., ... & Weld, D. S. (2023). The semantic scholar open data platform. arXiv preprint arXiv:2301.10140. <https://doi.org/10.48550/arXiv.2301.10140>.
- [KLE 25] Kleidermacher, H. C., & Zou, J. (2025). Science Across Languages: Assessing LLM Multilingual Translation of Scientific Papers. arXiv preprint arXiv:2502.17882. <https://doi.org/10.48550/arXiv.2502.17882>.
- [KRO 23] Krotov, V., & Johnson, L. (2023). Big web data: Challenges related to data, technology, legality, and ethics. *Business Horizons*, 66(4), 481-491. <https://doi.org/10.1016/j.bushor.2022.10.001>.
- [KWO 24] Kwon, D. (2024). Publishers are selling papers to train AIs-and making millions of dollars. *Nature*, 636(8043), 529-530. <https://doi.org/10.1038/d41586-024-04018-5>.
- [LAR 15] Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLoS one*, 10(6), e0127502. <https://doi.org/10.1371/journal.pone.0127502>.
- [LAS 11] Las Vergnas 1, O. (2011). L'institutionnalisation de la « culture scientifique et technique », un fait social français (1970–2010). *Savoirs*, (3), 9-60. <https://doi.org/10.3917/SAVO.027.0009>.
- [LEC 16] Lecomte, J. (2016). Qu'est-ce que la science ? In Les grands penseurs des sciences humaines (pp. 71–74). Éditions Sciences Humaines. <https://doi.org/10.3917/sh.journ.2016.01.0071>.
- [LED 22] Le Deuff, O., & Roumanos, R. (2022). Enjeux définitionnels et scientifiques de la littératie algorithmique: Entre mécanologie et rétro-ingénierie documentaire. *tic&société*, 15(2-3), 325-360. <https://doi.org/10.4000/ticetsociete.7105>.
- [LEL 25] Leloup, D. (2025). Pourquoi Elon Musk appelle au boycott de Wikipédia. *Le Monde*, 29 janvier 2025. https://www.lemonde.fr/pixels/article/2025/01/29/pourquoi-elon-musk-appelle-au-boycott-de-wikipedia_6522113_4408996.html.
- [LIN 20] Lin, J., Yu, Y., Zhou, Y., Zhou, Z., & Shi, X. (2020). How many preprints have actually been printed and why: a case study of computer science preprints on arXiv. *Scientometrics*, 124(1), 555-574. <https://doi.org/10.1007/s11192-020-03430-8>.
- [LIU 24] Liu, Y., Cao, J., Liu, C., Ding, K., & Jin, L. (2024). Datasets for large language models: A comprehensive survey. arXiv preprint arXiv:2402.18041. <https://doi.org/10.48550/arXiv.2402.18041>.
- [LO 20] Lo, K., Wang, L. L., Neumann, M., Kinney, R., & Weld, D. S. (2020). S2ORC: The semantic scholar open research corpus. arXiv preprint arXiv:1911.02782. <https://doi.org/10.48550/arXiv.1911.02782>.
- [LON 24] Longpre, S., Mahari, R., Lee, A., Lund, C., Oderinwale, H., Brannon, W., ... & Pentland, A. (2024). Consent in crisis: The rapid decline of the ai data commons. *Advances in Neural Information Processing Systems*, 37, 108042-108087. https://proceedings.neurips.cc/paper_files/paper/2024/file/c3738949a80306cc48a8ea8ba0560f9d-Paper-Datasets_and_Benchmarks_Track.pdf.
- [MAI 22] Mairesse, F. (2022). Les musées et l'économie créative. *La Lettre de l'OCIM. Musées, Patrimoine et Culture scientifiques et techniques*, (200), 12-17. <https://doi.org/10.4000/ocim.4743>.

- [MAJ 20] Majumder, M. S., & Mandl, K. D. (2020). Early in the epidemic: impact of preprints on global discourse about COVID-19 transmissibility. *The lancet global health*, 8(5), e627-e630. [https://doi.org/10.1016/S2214-109X\(20\)30113-3](https://doi.org/10.1016/S2214-109X(20)30113-3).
- [MAL 24] Maleki, N., Padmanabhan, B., & Dutta, K. (2024). AI hallucinations: a misnomer worth clarifying. In 2024 IEEE Conference on Artificial Intelligence (CAI) (pp. 133-138). IEEE. <https://doi.org/10.1109/CAI59869.2024.00033>.
- [MOA 07] Moatti, A. (2007). La diffusion de la culture scientifique, réalisations et réflexions. In Réalités industrielles. Annales des mines (No. Mai 2007, pp. 54-59). <https://shs.hal.science/halshs-00147069/document>.
- [NAV 23] Navigli, R., Conia, S., & Ross, B. (2023). Biases in large language models: origins, inventory, and discussion. *ACM Journal of Data and Information Quality*, 15(2), 1-21. <https://doi.org/10.1145/3597307>.
- [OBO 10] Oboler, A., Steinberg, G., & Stern, R. (2010). The framing of political NGOs in Wikipedia through criticism elimination. *Journal of Information Technology & Politics*, 7(4), 284-299. <https://doi.org/10.1080/19331680903577822>.
- [OST 10] Ostrom, E. (2010). Beyond markets and states: polycentric governance of complex economic systems. *American economic review*, 100(3), 641-672. <http://doi.org/10.1257/aer.100.3.641>.
- [OUY 22] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35, 27730-27744. https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf.
- [PET 25] Peters, U., & Chin-Yee, B. (2025). Generalization bias in large language model summarization of scientific research. *Royal Society Open Science*, 12(4), 241776. <https://doi.org/10.1098/rsos.241776>.
- [POT 25] Potnis, D., Tahamtan, I., & McDonald, L. (2025). Negative consequences of information gatekeeping through algorithmic technologies: An Annual Review of Information Science and Technology (ARIST) paper. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.24955>.
- [RIC 18] Richtig, G., Berger, M., Lange-Asschenfeldt, B., Aberer, W., & Richtig, E. (2018). Problems and challenges of predatory journals. *Journal of the European Academy of Dermatology and Venereology*, 32(9), 1441-1449. <https://doi.org/10.1111/jdv.15039>.
- [SCH 12] Schöpfel, J. (2012). Vers une nouvelle définition de la littérature grise. *Cahiers de la Documentation*, 66(3), 14-24. https://archivesic.ccsd.cnrs.fr/sic_00794984v1.
- [SCH 24] Schulze Balhorn, L., Weber, J. M., Buijsman, S., Hildebrandt, J. R., Ziefle, M., & Schweidtmann, A. M. (2024). Empirical assessment of ChatGPT's answering capabilities in natural science and engineering. *Scientific Reports*, 14(1), 4998. <https://doi.org/10.1038/s41598-024-54936-7>.
- [SEZ 24] Seznec, E. (2024). Wikipédia, plongée dans la fabrique d'une manipulation. *Le Point*, 13 décembre 2024. https://www.lepoint.fr/societe/wikipedia-plongee-dans-la-fabrique-d-une-manipulation-13-12-2024-2577881_23.php.
- [SHA 24] Sharma, N., Liao, Q. V., & Xiao, Z. (2024). Generative Echo Chamber? Effect of LLM-Powered Search Systems on Diverse Information Seeking. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1-17). <https://doi.org/10.1145/3613904.3642459>.
- [SHE 23] Shen, X., Chen, Z., Backes, M., & Zhang, Y. (2023). In ChatGPT we trust? Measuring and characterizing the reliability of chatgpt. *arXiv preprint arXiv:2304.08979*. <https://doi.org/10.48550/arXiv.2304.08979>.
- [SIN 15] Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J. P., & Wang, K. (2015). An overview of Microsoft Academic Service (MAS) and applications. *Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion)*, 243-246. ACM. <https://doi.org/10.1145/2740908.2742839>.
- [SOB 17] Sobel, B. L. (2017). Artificial intelligence's fair use crisis. *Colum. JL & Arts*, 41, 45. <https://doi.org/10.7916/jla.v41i1.2036>.
- [STR 25] Stratton, M. (2025). Market-Based Licensing for Publishers' Works Is Feasible. *Big Tech Agrees*. *Colum. JL & Arts*, 48, 434. <https://doi.org/10.52214/jla.v48i4.13925>.
- [SUN 07] Sun, Y., Zhuang, Z., Councill, I. G., & Giles, C. L. (2007). Determining bias to search engines from robots. *txt. In IEEE/WIC/ACM International Conference on Web Intelligence (WI'07)* (pp. 149-155). IEEE. <https://doi.org/10.1109/WI.2007.98>.
- [TAN 08] Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). ArnetMiner: Extraction and mining of academic social networks. *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'2008)*, 990-998. <https://doi.org/10.1145/1401890.1402008>.

- [TEP 17] Teplitskiy, M., Lu, G., & Duede, E. (2017). Amplifying the impact of open access: Wikipedia and the diffusion of science. *Journal of the Association for Information Science and Technology*, 68(9), 2116-2127. <https://doi.org/10.1002/asi.23687>.
- [TES 24] Tessler, M. H., Bakker, M. A., Jarrett, D., Sheahan, H., Chadwick, M. J., Koster, R., ... & Summerfield, C. (2024). AI can help humans find common ground in democratic deliberation. *Science*, 386(6719), eadq2852. <https://doi.org/10.1126/science.adq2852>.
- [VAN 17] Van der Meer, T. G., Verhoeven, P., Beentjes, J. W., & Vliegenthart, R. (2017). Disrupting gatekeeping practices: Journalists' source selection in times of crisis. *Journalism*, 18(9), 1107-1124. <https://doi.org/10.1177/1464884916648095>.
- [VID 24] Viseur, R., & Delcucq, L. (2024). Exploration des pratiques de régulation des IA génératives par le protocole d'exclusion des robots. INFORSID, 28-31 mai 2024, Nancy (France). <http://inforsid.fr/actes/2024/inforsid24-89-104.pdf>.
- [VIS 14] Viseur, R. (2014). Reliability of user-generated data: The case of biographical data in Wikipedia. In Proceedings of The International Symposium on Open Collaboration (pp. 1-3). <https://doi.org/10.1145/2641580.2641618>.
- [VIS 24] Viseur, R. (2024). Analyse de l'impact des IA génératives sur la presse en ligne : anatomie d'un newsbot basé sur GPT. Actes des conférences AIM. Montpellier (France). https://aim.asso.fr/fr/publications/actes-conferences/id-1809-aim2024_557858.
- [VIS 25] Viseur, R., & Finet, A. (2025). Dépasser ou respecter la norme ? Une analyse de la stratégie de plateformes numériques sous l'angle de la théorie néo-institutionnelle. *Innovations*, 2025/3, N° 78, pp. 41-71. <https://doi.org/10.3917/inno.pr2.0193>.
- [WAN 24] Wan, D., Vig, J., Bansal, M., & Joty, S. (2024). On positional bias of faithfulness for long-form summarization. arXiv preprint arXiv:2410.23609. <https://doi.org/10.48550/arXiv.2410.23609>.
- [WAN 20] Wang, K., Shen, Z., Huang, C., Wu, C. H., Dong, Y., & Kanakia, A. (2020). Microsoft academic graph: When experts are not enough. *Quantitative Science Studies*, 1(1), 396-413. https://doi.org/10.1162/qss_a_00021.
- [WAN 07] Wilkinson, D. M., & Huberman, B. A. (2007). Cooperation and quality in wikipedia. In Proceedings of the 2007 international symposium on Wikis (pp. 157-164). <https://doi.org/10.1145/1296951.1296968>.
- [WAT 06] Waters, R. (2006). Wikipedia stand-off in search for online truth. *Financial Times*, 10 novembre 2006. <https://www.ft.com/content/3300554a-6d6a-11db-8725-0000779e2340>.
- [WIL 15] Wilson, A. M., & Likens, G. E. (2015). Content volatility of scientific topics in Wikipedia: a cautionary tale. *PLoS one*, 10(8), e0134454. <https://doi.org/10.1371/journal.pone.0134454>.
- [WOE 24] Woessner, G., Seznec, E. (2024). Les illusionnistes - Climat, agriculture, nucléaire, OGM : l'enquête inédite sur les dérives de l'écologie politique. Robert Laffont. ISBN : 978-2221274491.
- [WU 25] Wu, K., Wu, E., Wei, K., Zhang, A., Casasola, A., Nguyen, T., ... & Zou, J. (2025). An automated framework for assessing how well LLMs cite relevant medical references. *Nature Communications*, 16(1), 3615. <https://doi.org/10.1038/s41467-025-58551-6>.
- [XIA 15] Xia, J., Harmon, J. L., Connolly, K. G., Donnelly, R. M., Anderson, M. R., & Howard, H. A. (2015). Who publishes in “predatory” journals?. *Journal of the Association for Information Science and Technology*, 66(7), 1406-1417. <https://doi.org/10.1002/asi.23265>.
- [XIA 17] Xia, J., Li, Y., & Situ, P. (2017). An overview of predatory journal publishing in Asia. *Journal of East Asian Libraries*, 2017(165), 4. <https://scholarsarchive.byu.edu/jeal/vol2017/iss165/4>.
- [YAS 14] Yasseri, T., Spoerri, A., Graham, M., & Kertész, J. (2014). The most controversial topics in Wikipedia: A multilingual and geographical analysis. In P. Fichman & N. Hara (Eds.), *Global Wikipedia: International and cross-cultural issues in online collaboration*. Scarecrow Press.
- [YE 23] Ye, H., Liu, T., Zhang, A., Hua, W., & Jia, W. (2023). Cognitive mirage: A review of hallucinations in large language models. arXiv preprint arXiv:2309.06794. <https://doi.org/10.48550/arXiv.2309.06794>.
- [ZIM 23] Zimmermann, J. B. (2023). Communs et innovation : une relation paradoxale. *Innovations*, I144-31. <https://doi.org/10.3917/inno.pr2.0144>.