Note de synthèse

Usages des outils d'Intelligence Artificielle générative dans le domaine de la recherche
- Points de vigilance et bonnes pratiques -

Céline Blitz Frayret, UMR Eco&Sols

Estelle Jaligot, Délégation à la déontologie et à l'intégrité scientifique

Colline Orsini, Délégation aux affaires juridiques et à la conformité

Table des matières

I Introduction

I.1 Objet de la note

La présente note se veut une tentative de résumer les principales questions soulevées par les nouveaux outils d'Intelligence Artificielle générative (IAg) et leurs applications dans les activités de recherche scientifique, dans le but d'attirer l'attention sur les risques potentiels (éthiques, juridiques, scientifiques, sociaux, environnementaux) associés à leur usage.
Les recommandations que nous émettons doivent être comprises comme des parades permettant de diminuer ou d'éliminer les risques présentés par les IAg actuelles. Elles ne s'appuient pas sur une liste exhaustive et stricte d'usages "permis" ou "interdits", car une telle liste serait obsolète sitôt établie et ne pourrait jamais couvrir l'intégralité des cas d'usages de tous les modèles d'IAg.
À l'opposé d'une telle approche, **nous fournissons aux utilisateurs des repères et des pistes de réflexion leur permettant d'employer les IAg de manière responsable et éclairée, ou de choisir de ne pas le faire en connaissance de cause.**
Si les exemples utilisés pour illustrer notre propos proviennent prioritairement de cas d'usage scientifiques des IAg, certains peuvent également concerner les personnels des fonctions administratives et d'appui au sein des établissements de recherche. Pour cette raison, nous avons fait le choix d'employer un langage aussi accessible que possible aux non-spécialistes.

Ce document est complété par des **Annexes**:

- Une traduction française partielle des recommandations de la Commission Européenne sur l'utilisation des IAg (**Annexe 1**).
- Des renvois aux recommandations émises par le Cirad ou par ses partenaires (**Annexe 2**).
- Une liste des ressources que nous avons exploitées: celles-ci traitent des IA en général ou des IAg en particulier et sont de portée générale ou de nature thématique (**Annexe 3**).

Cette note est amenée à évoluer à mesure que de nouveaux usages des IAg sont testés et leurs limites et biais analysés, ouvrant la porte à des améliorations de leur conception et de leurs performances. Toute contribution permettant de l'enrichir est donc bienvenue.

I.2 Les IAg : de quoi s'agit-il ?

Récemment, de nouveaux outils d'IA reposant sur de grands modèles de langage (large language models, LLMs) ont été mis à disposition du grand public à grand renfort de publicité: ChatGPT (proposé par OpenAI) est le plus connu d'entre eux, mais il existe aussi Bard et Gemini (Google), Llama (Meta), Claude (Anthropic), ainsi que de nombreux autres.
Ces différents outils ont en commun d'être des agents conversationnels (chatbots), réagissant à une requête (prompt) émanant de l'utilisateur pour produire une réponse crédible, sinon correcte. Ils fonctionnent sur la base d'un algorithme dont les performances ont été progressivement affinées et validées par des annotateurs humains sur d'important volumes de données d'entraînement (dites aussi d'apprentissage). La réponse fournie se présente comme une synthèse d'informations préexistantes, fournies par l'utilisateur en complément du prompt ou puisées par l'IAg dans les ressources auxquelles elle a accès.

Dans la présente note, les problèmes que nous avons choisis de mettre en lumière sont principalement liés à l'utilisation des IAg dédiées au texte, du fait de leur potentiel d'application très large dans un contexte de recherche. Il existe également des outils d'IAg dédiés à la génération d'images (par exemple Dall-E, Midjourney, Firefly) ou de sons (AudioCraft, AIVA, SOUNDRAW, etc.) ainsi que des IA polyvalentes telles que Copilot (Microsoft), qui partagent une partie de ces problèmes.

L'AI Act: des implications pour la recherche à évaluer

La croissance rapide de l'IA a motivé l'élaboration en Europe d'un texte très attendu : l'AI Act, qui entrera en vigueur en juin 2026. Ce règlement adopte une **approche par les risques :** concrètement, plus l'IA est porteuse de risques potentiels, plus les obligations (pesant en particulier sur le fournisseur) seront importantes. Il clarifie par ailleurs **le rôle et les responsabilités des différents acteurs des systèmes d'IA**, insiste sur l'importance de la transparence des systèmes et interdit certaines pratiques.

Concernant le **domaine de la recherche**, le règlement comporte une **exclusion** destinée à préserver la capacité d'innovation : celle-ci stipule en effet que **les obligations prévues par l'AI Act ne s'appliqueront pas "*aux systèmes et modèles d'IA, y compris leurs résultats, spécifiquement développés et mis en service aux seules fins de la recherche et du développement scientifique, ni à leurs sorties*"**. Une étude approfondie, basée sur l'examen d'une large gamme de cas d'usage, sera cependant nécessaire afin de déterminer le périmètre d'application de cette exception. En effet, s'il semble clair qu'elle concerne les modèles d'IA développés et appliqués dans le contexte d'un projet de recherche, le statut des IA conçues en appui à la recherche est moins évident. À l'inverse, les systèmes d'IA utilisés pour mener une activité de recherche et de développement mais qui ont été conçus à d'autres fins seront soumis à l'AI Act[1]. Ce cas de figure concerne par exemple l'utilisation de tout modèle d'IAg commercial dans le cadre d'un projet de recherche.

Le règlement rappelle par ailleurs que les normes éthiques et professionnelles reconnues en matière de recherche scientifique et le droit de l'UE restent applicables, que le modèle d'IA soit concerné par l'AI Act ou non.

[1] Point précisé au considérant 25 de l'AI Act.

II IAg : problèmes connus et implications dans le contexte de la recherche

II.1 Production ou propagation de données erronées ou tronquées

Les IAg produisent des réponses sur la base de corrélations statistiques (établies en fonction de la fréquence d'association entre des chaînes de caractères) qui peuvent n'avoir aucun rapport avec leur sens ni avec leur validité scientifique. L'outil n'a en effet pas la capacité d'effectuer une analyse critique des informations traitées ou produites, et ne sait donc pas les interpréter, ni les hiérarchiser ni leur attribuer un niveau de confiance qui soit totalement fiable. L'IAg est donc susceptible de **propager** ou de **produire de fausses informations**.

De plus, les jeux de données utilisés pour l'entraînement de l'IAg peuvent aussi être de mauvaise qualité du fait de leur caractère incomplet ou erroné (que ces erreurs soient de nature involontaire ou le résultat d'une action délibérée de " data poisoning "), ce qui augmente la probabilité de produire des réponses inadéquates.

Dans le contexte scientifique, les IAg peuvent notamment générer des sorties sans rapport avec des informations issues du monde réel ou **"hallucinations"**, en réarrangeant voire en reformulant des éléments préexistants sans liens entre eux. La capacité des IAg à produire des combinaisons d'assertions vraies et fausses sur un sujet donné est d'ailleurs reconnue par leurs concepteurs[2]. Ainsi, l'IAg produira presque toujours une réponse, si inexacte ou inadaptée qu'elle soit par rapport au prompt de départ: **rares sont en effet les IAg qui mentionnent leur incapacité à répondre ou leur incertitude**.

Si elles peuvent accélérer l'écriture de lignes de code informatique, les IAg sont susceptibles de produire des erreurs conduisant à des **pertes de données** voire à des **failles de sécurité**.
Les IAg devant "apprendre" à partir de données qui leur sont accessibles avant de pouvoir fournir une réponse, elles sont donc **"aveugles" aux informations les plus récentes**: ainsi, un l'état de l'art produit par une IAg devra être actualisé manuellement par l'utilisateur.

II.2 Production de réponses variables et non reproductibles

De par leurs modalités de développement et de fonctionnement, les IAg tendent à produire des réponses variables à une même requête, ce qui soulève la question de la reproductibilité des résultats obtenus avec ces outils. Les principaux facteurs de variation identifiés sont les suivants:

- Les **modalités d'interaction** avec l'IAg (utilisation d'un ou de plusieurs prompts, formulation et ordre d'utilisation de ceux-ci) sont susceptibles de faire varier sensiblement la réponse.

- **Les performances des IAg évoluent de manière continue** au gré i) des mises à jour proposées par les développeurs et ii) des requêtes successives des utilisateurs, qui participent à leur apprentissage.

[2] Voir la mention figurant en bas de la fenêtre lors de l'utilisation de ChatGPT : " *ChatGPT may produce inaccurate information about people, places, or facts* ".

- **Le modèle de langage sous-tendant une IAg est développé de manière indépendante pour chaque langue**, grâce à un corpus de ressources dont certaines peuvent être spécifiques à cette langue. Par conséquent, les résultats générés par l'IAg pourront être différents d'une langue à l'autre, y compris du point de vue de leur qualité, en raison des variations entre les tailles des corpus de données disponibles et du nombre de requêtes effectuées dans chaque langue. En ce qui concerne la recherche, la qualité des réponses obtenues sera donc bien meilleure en anglais, qui est la langue des échanges internationaux, que dans une langue dont l'usage est moins répandu.

II.3 Biais de représentation

Fonctionnant sur la reconnaissance des patrons (patterns) au sein d'un texte, les IAg tendent à reproduire préférentiellement les enchaînements de mots et de phrases en fonction de leur plus forte représentation au sein de leurs données d'entraînement et/ou au sein des données fournies par l'utilisateur. Elles auront donc tendance à fournir des réponses basées sur des éléments (faits, opinions) majoritaires au sein de ces données, et à éluder les éléments minoritaires. Ce mode de fonctionnement peut les conduire à produire des **réponses stéréotypées** et qui omettent des nuances importantes, voire à générer des contresens.

Ce problème peut s'avérer d'autant plus critique que le développement de l'algorithme ainsi que les jeux de données d'entraînement peuvent eux-mêmes être biaisés[3]. L'IAg fournira alors des réponses qui **reproduiront voire amplifieront ces biais**. Une décision prise sur la base de réponses fournies par une IA peut donc conduire à des **situations inéquitables**, voire à des **discriminations**.

II.4 Problèmes associés à l'utilisation des sources

Les IAg ne créent aucun matériel intrinsèquement nouveau, les réponses produites sont constituées du réarrangement d'éléments préexistants. Dans la mesure où le fonctionnement des IAg peut reposer sur l'exploitation, parfois non autorisée, de **matériel protégé par des droits d'auteur** (qu'il s'agisse des données d'apprentissage ou des données fournies par l'utilisateur), les réponses produites constituent une **violation** de ces droits[4]. Dans le cas de la production d'écrits ou d'illustrations scientifiques, l'utilisation d'une IAg peut ainsi être assimilable à un **plagiat** voire à de la **contrefaçon**.

De même, le fait que la plupart des IAg ne mentionnent pas l'appartenance des œuvres d'origine dans leurs productions contrevient aux principes du **droit d'auteur** (droit moral, droits patrimoniaux). Par extension, en utilisant les sorties produites par ces outils, l'utilisateur commet une faute. Les textes règlementaires qui se développent autour du sujet insistent d'ailleurs grandement sur les risques d'atteintes aux droits d'auteur portés par les IAg. Ainsi, les lignes directrices du G7 à Hiroshima recommandent de mettre en œuvre des mesures

[3] Le manque de diversité et d'inclusivité au sein des métiers du numérique peut être un facteur aggravant de ce phénomène.

[4] Cette exploitation peut se faire à l'insu des auteurs, même lorsqu'elle est à strictement parler légale. Ainsi récemment, deux accords de partenariat impliquant chacun un éditeur académique et un groupe privé a donné à ce dernier la possibilité d'utiliser le contenu du catalogue de publications pour l'entraînement de ses modèles d'IA (source: The Chronicles of Higher Education, 29 juillet 2024).

appropriées de saisie et de protection de la propriété intellectuelle. L'AI Act place aussi le respect et la protection de ces droits comme une nécessité.

Dans le cadre d'un usage scientifique, l'usage de la plupart des IAg est problématique du fait de leur **incapacité à citer leurs sources** ou de leur **tendance à les citer de manière inexacte et non pertinente**. Ainsi, la production de références bibliographiques "hallucinées" a été rapporté dans de nombreux cas.

II.5 Problèmes associés à la réutilisation des données fournies par l'utilisateur

Dans la plupart des cas (et notamment pour les modèles disponibles gratuitement en ligne), les données fournies par l'utilisateur à une IAg sont ensuite intégrées au fonctionnement de celle-ci, et sont donc susceptibles d'être accessibles à n'importe quel autre utilisateur dans les futures réponses données par l'outil. L'utilisation d'un tel modèle d'IAg peut donc entraîner une **fuite de données**, dès lors que les données qui lui sont fournies ne sont pas d'accès public: c'est par exemple le cas de documents internes à l'établissement.

Dans un contexte de recherche, l'utilisation de ce type d'IAg dans la rédaction ou la mise en forme de documents confidentiels constitue de ce fait une **violation du devoir de confidentialité**. Si cette obligation de confidentialité a été incluse dans un contrat, cette divulgation va à l'encontre des **engagements contractuels**, ce qui peut engager la responsabilité de l'utilisateur ou celle de l'établissement.
Selon la même logique, l'utilisation d'une telle IAg sur des documents contenant des données personnelles contrevient à **l'obligation d'assurer la protection des données personnelles et/ou de la vie privée**, conformément au règlement général pour la protection des données (RGPD)[5].

Plus largement, cette fuite de données pose la question de leur **souveraineté** puisque l'utilisateur n'aura pas de visibilité sur les utilisations qui pourraient en être faites par l'éditeur et/ou l'hébergeur de l'IAg. Ceci est d'autant plus vrai que ces utilisations sont potentiellement soumises à des **règles de droit différentes de celles qui sont applicables en France**. Par exemple, ChatGPT est un outil soumis aux lois américaines, lesquelles s'appuient sur une définition très large du concept de "souveraineté"[6]. L'AI Act insiste d'ailleurs sur la nécessité de respecter la protection des informations confidentielles de nature commerciale et les secrets d'affaires.

[5] Règlement (UE) 2016/679 modifié du Parlement européen et du Conseil du 27 avril 2016.
[6] Le Cloud Act permet ainsi aux autorités américaines bénéficiant d'un mandat d'accéder aux données d'individus et entreprises situés hors des Etats-Unis, à condition notamment que l'entité qui héberge les données soit basée aux Etats Unis ou de nationalité américaine.

II.6 Manque de transparence pour l'utilisateur

Du fait même de leur complexité et des volumes importants de données exploitées, les IAg sont d'une grande opacité quant à leurs modalités de fonctionnement: on les compare volontiers à des **"boîtes noires"**. L'utilisateur n'a donc en général pas les moyens de se prémunir contre les problèmes mentionnés ci-dessus ni d'exercer un réel contrôle qualité. Ainsi, **il peut être difficile voire impossible d'expliquer comment l'IAg est parvenue à un résultat ou d'interpréter celui-ci**, et ce d'autant plus si l'usage de ces outils ne lui est pas familier. Ce phénomène est encore accentué dans le cas des outils d'IAg développés par des compagnies privées qui protègent leurs innovations vis-à-vis de la concurrence par le secret.

Dans le contexte d'un usage en recherche, ce manque de transparence prive l'utilisateur de la possibilité de garantir pleinement l'intégrité des données produite et la rigueur de la démarche. Cette obligation faisant partie des **responsabilités fondamentales de tout auteur d'une publication scientifique**, un nombre croissant de journaux et de maisons d'édition scientifiques a mis en place des lignes directrices encadrant strictement l'usage des IAg dans les publications. Si ces directives sont encore très hétérogènes[7], elles ont généralement en commun i) de mettre en avant la **responsabilité totale de l'auteur vis-à-vis des résultats produits par l'IAg** (cette dernière, ne peut être créditée comme auteur ni être citée comme source, étant dépourvue de capacité à assumer une telle responsabilité) et ii) d'exiger une **déclaration détaillée** des modalités de son utilisation.

II.7 Manque d'ouverture et de traçabilité du fonctionnement et des produits

Selon les bonnes pratiques en matière d'**ouverture et de traçabilité des données de recherche**, l'accès à la chaîne de traitement d'un résultat produit par l'IAg (incluant les codes sources de l'IAg, les données, etc.) devrait être permis afin de pouvoir reproduire ce résultat en cas d'audit ou de litige. Cependant, dans la plupart des cas le modèle d'IAg ne cite pas ses sources, rendant difficile la vérification. Au-delà de la vérification de la véracité des réponses produites, cette opacité de la conception et du fonctionnement de l'outil empêche l'utilisateur de satisfaire aux **principes FAIR** de bonne gestion des données de la recherche[8].

L'IAg étant conçue pour les imiter les créations humaines, **la détection de ses produits** (par d'autres IA ou par des humains) **est à ce jour aléatoire**. Cette incapacité à distinguer de manière fiable les informations légitimes de celles qui sont produites par une IA fait donc peser sur l'utilisateur la responsabilité de déclarer l'usage qu'il fait de l'IAg, afin que ni l'origine de ses données ni la rigueur de son travail ne soient mises en question. Dans le cadre d'une activité de recherche, **l'utilisation non déclarée d'une IA constitue un manquement à l'intégrité scientifique**[9].

---

[7] La question de l'élaboration de consignes basées sur les spécificités de l'ensemble des parties prenantes de l'édition et des différentes disciplines scientifiques est au cœur de l'initiative CANGARU (voir référence en **Annexe 3**).

[8] Findable, Accessible, Interoperable, Reusable (Trouvable, Accessible, Interopérable et Réutilisable): voir Ouvrir la Science.

[9] **European Code of Conduct for research integrity**, section **Research Misconduct and other unacceptable practices:** "*Hiding the use of AI or automated tools in the creation of content or drafting of publications*" (voir référence en **Annexe 3**).

Les préoccupations éthiques sont complétées par des textes juridiques, de plus en plus nombreux (fiches pratiques de la Commission Nationale de l'Informatique et des Libertés, AI Act, etc.), qui soulignent l'importance de la transparence des outils développés par les fournisseurs.

II.8 Facilitation des manquements volontaires ou involontaires à l'intégrité scientifique

Comme mentionné précédemment, l'utilisation inappropriée des IAg peut favoriser le **plagiat**. La capacité de ces outils à imiter des données existantes peut également être détournée pour générer des données de recherche fictives (ne reposant sur aucune expérimentation, observation ou simulation: **fabrication**) ou pour manipuler des données réelles afin d'en biaiser l'interprétation (**falsification**). Ces différents agissements constituent de **graves manquements à l'intégrité scientifique**[10].

Au cours des dernières années, **l'exploitation à grande échelle des IAg par des acteurs mercantiles peu scrupuleux (les usines à papiers ou "paper mills")** a conduit à la publication massive d'articles scientifiques basés sur des données fabriquées, falsifiées ou plagiées. Le fait de contribuer, directement ou indirectement, aux activités d'un "paper mill" (en achetant une position d'auteur sur un article frauduleux ou en permettant sa publication en tant qu'éditeur ou relecteur) est un **manquement à l'intégrité scientifique**[11].
À terme, la proportion croissante de ces articles frauduleux est susceptible de jeter le doute sur la validité de résultats de recherche légitimes et, ainsi, de **"polluer" le corpus de connaissances** sur lequel s'appuient les recherches ultérieures. Sur le plus long terme, ces "fake news" scientifiques peuvent avoir des **impacts sociétaux** désastreux (élaboration de politiques publiques erronées, mise sur le marché de produits peu performants voire nocifs, gaspillage de ressources, etc.) et **mettre à mal la confiance accordée par la société à la science et aux scientifiques**.

En-dehors de ces cas exceptionnels par leur ampleur, des **manquements de moindre gravité** peuvent émerger du fait de la forte hétérogénéité des niveaux de maîtrise de l'usage responsable des IAg et de la plus ou moins grande accessibilité des ressources permettant l'amélioration des pratiques. Il existe donc un fort enjeu pour permettre simultanément une **montée en capacité** sensible à la diversité des utilisateurs, des usages et des contextes, et une **homogénéisation des pratiques**.

[10] Source: voir note 9.
[11] "*Establishing, supporting, or deliberately using journals, publishers, events, or services that undermine the quality of research ('predatory' journals or conferences and paper mills).*" Source: voir note 9.

III Impacts des IAg au-delà de la recherche scientifique

Le développement, l'apprentissage et le fonctionnement des IAg supposent la manipulation de **volumes de données** beaucoup plus importants que pour les autres modèles d'IA. Par ailleurs, leur **grande plasticité** permet d'envisager leur exploitation par des publics très diversifiés pour une large gamme d'usages. Pour cette raison, certains des problèmes connus chez les autres IA sont amplifiés dans le contexte des IAg.

Ainsi les IA, et les IAg en particulier:

- Ont une **empreinte environnementale** importante de par leur consommation en énergie et en eau et leurs émissions de gaz à effet de serre - l'essentiel de ces impacts étant le fait des centres de données. Ainsi, une requête soumise à une IAg a un impact environnemental de 5 à 10 fois supérieur à celui d'une requête traitée par un moteur de recherche en ligne.

- Sont potentiellement porteuses de profondes **transformations du marché du travail** ainsi que des conditions d'exercice de nombreux métiers. Au niveau mondial, on estime que 40% des métiers sont fortement exposés à l'IA, que celle-ci remplace ou vienne en appui de leurs activités. Dans ce contexte, si des gains de productivité et une augmentation des revenus sont anticipés pour certains métiers, la mise en place de mesures d'accompagnement des catégories de travailleurs pour lesquels l'adoption de l'IA est plus difficile (en raison de leur âge, de leur niveau de qualification, des spécificités de leur métier, etc.) est indispensable afin d'éviter d'accentuer les inégalités économiques et sociales.

- Exploitent et potentiellement, tendent à accentuer les **inégalités structurelles entre le Nord et le Sud**:
    - de par leur recours massif à une main-d'œuvre du Sud faiblement rétribuée et parfois employée dans des conditions peu éthiques;
    - de par les importants écarts observés dans les indicateurs de préparation à l'adoption des IA[12] en fonction du niveau de développement économique des pays;
    - de par la forte dépendance du secteur envers des technologies et des infrastructures détenues par de grands groupes privés (notamment des multinationales comme les GAFAM[13]) basés dans les pays à hauts revenus.

- Impliquent des **coûts de fonctionnement** très élevés[14], qui rendent leur utilisation peu accessible et, sur le long terme, non soutenable sur le plan économique.

Les IAg posent donc question quant au **modèle économique** qu'elles promeuvent implicitement et quant à la capacité des États (et plus particulièrement ceux du Sud) à garantir leur autonomie et leur souveraineté dans ce domaine.

[12] Le AI Preparedness Index (AIPI) élaboré par le FMI synthétise des indicateurs relatifs aux infrastructures numériques, au capital humain et au marché du travail, à la capacité d'innovation et d'intégration économique, et au cadre réglementaire et éthique (voir référence en **Annexe 3**).
[13] Google/Alphabet, Amazon, Facebook/Meta, Apple, Microsoft.
[14] Le coût financier d'une requête à ChatGPT est ainsi évalué à 500 dollars US.

En outre, la maîtrise des modalités particulières d'interaction avec les IAg et la compréhension de leurs potentialités et de leurs limites peuvent nécessiter un accompagnement et une formation spécifiques. Dès lors, les différentiels d'accès, non seulement aux outils d'IAg mais également à ces appuis (en fonction des contextes socio-culturels, économiques, etc., ou selon les métiers ou disciplines scientifiques dans le cas de la recherche) peuvent accentuer ou faire émerger des **inégalités**.

IV Bonnes pratiques d'utilisation des IAg en recherche

Les IAg constituent des outils puissants pour rassembler, synthétiser et reformuler rapidement de grandes quantités d'information. De ce point de vue, elles peuvent s'avérer utiles pour **rendre plus efficiente la réalisation de certaines tâches**, notamment celles qui présentent une faible complexité. Les IAg peuvent également contribuer à **améliorer l'inclusivité** en offrant à des publics divers un accès facilité à l'information scientifique.

L'utilisation responsable des IAg implique de:

- **S'informer systématiquement si l'usage d'une IAg est autorisé et sur les modalités de cette utilisation.** Pour cela, consulter les lignes directrices disponibles en fonction de l'usage envisagé (par exemple consignes aux auteurs du journal, recommandations aux évaluateurs, politique du bailleur) ou prendre conseil auprès de personnes référentes en la matière (responsable de collectif, éditeur de revue, présidence de comité d'évaluation ou de jury, etc).

- **Toujours déclarer l'utilisation d'un outil d'IA (générative ou non), et en détailler les modalités de manière transparente** en mentionnant a minima le modèle, la version, la source et le motif de cet usage. Chaque fois que cela est possible, expliciter et rendre accessibles les stratégies d'interaction avec l'IAg (formulation du prompt ou de la succession de prompts) ainsi que les connaissances utilisées (nature et modalité d'accès des données sources). Opter (dans la mesure du possible) pour un modèle dont le code source est ouvert.

- **Ne jamais utiliser une IAg pour produire tout ou partie d'un matériel présenté comme un travail original engageant la responsabilité de l'auteur ou celle de l'établissement** employeur ou d'accueil. Ceci est valable pour la production de jeux de données de recherche, de publications scientifiques, de rapports d'expertise ou d'évaluation, de policy briefs, de demandes de financement, de mémoires ou thèses, etc.

- **Utiliser, lorsque cela est permis, l'IAg comme aide à la rédaction**, dans la mesure où cela se limite à un appui ponctuel à la reformulation, à la correction de la syntaxe ou de l'orthographe ou à l'amélioration du style. Il est néanmoins recommandé de limiter cette utilisation au strict minimum, par exemple en faisant intervenir l'IAg sur des portions réduites du texte rédigé par l'utilisateur.

- **Limiter l'utilisation des IAg aux produits sur lesquels l'utilisateur possède les connaissances et compétences nécessaires pour effectuer un contrôle de la qualité et de la validité de la réponse, dont il assumera dans tous les cas la responsabilité.** Il peut s'agir par exemple de la production d'un résumé (sur la base de documents auxquels il a accès), de la reformulation d'une traduction (entre deux langues qu'il maîtrise raisonnablement bien) ou de la validation d'un programme informatique (dans un langage et pour des fonctions avec lesquels il est familier).

- **Ne jamais alimenter une IAg ouverte sur l'extérieur à l'aide de données qui ne sont pas destinées à être rendues publiques**, et notamment:
    - **des données personnelles** (par exemple: liste de contacts, CV, dossiers de suivi professionnel, social ou médical, etc.);
    - **des données confidentielles** (données de la recherche non publiées, manuscrits ou projets soumis pour évaluation, comptes-rendus d'instances de décision, documents de contractualisation, documents internes liés à la stratégie de l'établissement ou à son fonctionnement, etc.);
    - **des données protégées par un droit de propriété intellectuelle d'un tiers qui ne vous a pas donné son accord** (textes, images ou sons dont vous n'êtes pas propriétaire, publication dont le copyright/les droits d'auteur ont été transféré à l'éditeur).

  Dans la mesure du possible, il est préférable d'utiliser d'une **version "entreprises"** du modèle d'IAg plutôt que la version gratuite grand public. En effet, la version "entreprises", adaptée à l'usage professionnel, offre généralement une protection contre le risque de fuite des données. Si une **IAg souveraine** (c'est-à-dire mise en place par l'État français ou par une structure publique nationale[15]) existe pour les utilisations envisagées, elle doit être privilégiée dans tous les cas. **Si aucun modèle "entreprise" ou souverain ne peut être utilisé, l'utilisateur doit envisager d'éliminer les données non publiques (telles que décrites ci-dessus) de ses échanges avec l'IAg, ou s'abstenir d'utiliser celle-ci.**

- **Dans le cas où l'IAg est utilisée en appui à la prise de décision, traiter les réponses fournies comme un élément parmi d'autres**, et les confronter systématiquement à des éléments obtenus sans l'appui de cet outil. Dans tous les cas, le résultat et son processus d'obtention doivent être soumis à la **supervision** et à l'**analyse critique humaine** avant toute décision.

- **Contribuer**, dans la mesure du possible, **aux débats et à la définition de bonnes pratiques** d'usage des IAg, afin que celles-ci soient adaptées à une grande diversité d'utilisateurs, de domaines d'activité, de contextes, etc.

- Dans les interactions avec les **partenaires, collaborateurs ou sous-traitants**, veiller à ce que leurs propres usages des IAg soient conformes aux bonnes pratiques.

- En tant que responsable d'un établissement ou d'un collectif, **prévenir l'émergence ou l'accentuation d'inégalités dans l'accès et la maîtrise des outils d'IAg**. Prendre en compte les éventuelles différences de niveau de connaissances des atouts et risques de l'IAg (en fonction des disciplines scientifiques, des métiers, des types d'acteurs - de la recherche ou de la société civile -, des catégories de personnels, des contextes géographiques ou culturels, etc.). Encourager les personnels à développer leurs compétences en veillant à l'accompagnement spécifique des personnes ou catégories

[15] Par exemple Aristote, IAg souveraine et open source développée par l'école d'ingénieurs CentraleSupélec pour l'enseignement supérieur ou l'outil testé à l'Université de Rennes (sources: Campus Matin, 3 juillet et 11 juillet 2024, respectivement).

susceptibles de rencontrer des difficultés à l'adoption de l'IAg (en raison de leur âge, de leur genre, de leur niveau d'éducation, etc.).

- **En tant que protagoniste de l'édition scientifique (édition, relecture), se prémunir contre l'usage non déclaré ou abusif des IAg**, notamment dans le contexte des " paper mills ": se familiariser avec les techniques et outils de détection, se former à leur utilisation et/ou solliciter l'appui des experts, participer à la sensibilisation sur ces thématiques.

- Dans le cadre d'une recherche soucieuse de ses impacts sur les personnes, la société et l'environnement, **évaluer la pertinence d'utiliser des outils d'IAg par rapport à d'autres outils ne faisant pas appel à l'IA**. Lorsque l'emploi de l'IAg s'avère être suffisamment justifié, en rationnaliser l'usage en fonction des risques identifiés et mettre en place des mesures appropriées afin d'en minimiser les éventuels impacts négatifs.

- **Résister à la tentation de recourir à l'utilisation des IAg de manière opportuniste (sans justification scientifique rigoureuse) ou comme un palliatif face à un déficit de ressources ou de compétences.**

V Annexe 1 : Recommandations pour l'usage responsable des IAg en recherche

Traduction partielle et adaptation de: Living guidelines on the responsible use of generative AI in research, version 1.0 de mars 2024 (Commission Européenne)

V.1 Pour les personnels de recherche

1 – Rester responsable de la production scientifique

- Assumer la responsabilité de l'intégrité du contenu généré à l'aide de l'IAg.
- Porter un regard critique sur la production de l'IAg et ses limites (biais, hallucinations, imprécisions).
- Ne pas assimiler l'IAg à un auteur, cette qualité impliquant un rôle et des responsabilités réservés aux humains.
- Ne pas produire ni utiliser des données de recherche fabriquées ou falsifiées à l'aide de l'IAg.

2 – Faire preuve de transparence

- Préciser la nature de l'outil d'IAg utilisé (nom, version, date) et les modalités de son usage (finalités, prompts et sorties produites) dans le respect des principes de la science ouverte.
- Déclarer les biais et limites associés à l'utilisation de l'IAg ainsi que le manque de reproductibilité des résultats obtenus. Anticiper leurs impacts sur la recherche, et si possible, identifier des mesures de gestion de ces risques.

3 – Exercer une vigilance particulière envers les questions de confidentialité, vie privée, droits de propriété intellectuelle, lors du partage d'informations sensibles ou protégées

- Veiller à ne pas transférer à un outil d'IAg en ligne des informations non publiques ou protégées, SAUF dans le cas où leur non-réutilisation peut être garantie.
- Ne pas fournir de données personnelles à une IAg en ligne SAUF si l'individu donne son consentement ET si ce partage est dûment justifié dans le respect du RGPD.
- Examiner l'ensemble des modalités de gestion des données afin d'en maîtriser les implications éthiques et juridiques.

4 – Respecter les législations nationales, européennes et internationales

- Vérifier que les résultats produits par l'IAg ne contiennent pas d'éléments plagiés. Le cas échéant, citer les sources sur lesquelles est basée la réponse de l'IAg.
- Dans le cas où des données personnelles apparaissent dans les sorties de l'IAg, les gérer dans le respect des règles européennes.

5 – Monter en capacité afin d'optimiser les bénéfices de l'IAg

- Se tenir à jour sur les évolutions des outils d'IAg et de leurs usages.
- Participer à des échanges de bonnes pratiques au sein de sa communauté scientifique, et avec d'autres acteurs.

6 – Eviter l'utilisation de l'IAg dans le cadre d'activités sensibles pouvant affecter d'autres personnels ou établissements (évaluation, prise de décision)

- Afin de se prémunir contre toute inégalité de traitement pouvant être causée par les limites et biais de l'outil.
- Afin d'éviter la diffusion, voire l'intégration dans un modèle d'IA, de données non publiques ou confidentielles.

V.2 Pour les établissements de recherche

1 – Promouvoir une utilisation responsable de l'IAg

- Proposer des formations abordant différents aspects de l'utilisation de l'IAg: vérification des résultats, protection de la vie privée, protection des droits de propriété intellectuelle, etc.
- Fournir un appui et des lignes directrices afin d'assurer le respect des exigences éthiques et légales.

2 – Superviser le développement et l'utilisation des systèmes d'IAg dans l'établissement

- Établir une cartographie des domaines de recherche et processus où l'IAg est utilisée, afin de proposer un accompagnement ciblé en fonction des besoins et d'identifier les éventuels risques.
- Analyser les limites et biais des outils et proposer des recommandations aux personnels de recherche.

3 – Intégrer les recommandations sur l'usage des IAg dans les bonnes pratiques de l'établissement

- Utiliser ces recommandations comme base de discussion dans le cadre d'une consultation des différents acteurs (dont les personnels de recherche) concernés par l'usage des IAg, en vue de leur enrichissement et de leur déclinaison dans différents contextes.
- Appliquer ces recommandations et les amender sur la base d'études de cas spécifiques.

4 – Dans la mesure du possible, mettre en place des outils d'IA garantissant la protection et la confidentialité des données (outils auto-gérés)

- Garantir un niveau approprié de cybersécurité, en particulier pour les systèmes connectés à internet.

VI Annexe 2 : Recommandations sur les usages des IA au Cirad et chez les partenaires

VI.1 Ressources du Cirad

Cybersécurité

Bonnes pratiques de cybersécurité: Utiliser Traffic Light Protocol pour la classification des documents

Publication

Revue Bois et forêts des tropiques: Politique éditoriale vis à vis de l'intelligence artificielle

VI.2 Ressources d'INRAE

Micael Aliouat, Colette Cadiou, Jocelyn De-Goer-De-Herve, Remy Decoupes, Nathalie Gandon, Marjolaine Hamelin, Hadi Quesneville, Tristan Salord, Alban Thomas, 2024. Recommandations pour l'usage des IA génératives comme assistant personnel au sein d'INRAE, INRAE (France), 7 p. DOI : 10.17180/ztym-j930

VII Annexe 3 : Ressources sur les usages des IAg en recherche

VII.1 Ressources d'intérêt général

Au niveau national

- *Commission Nationale de l'Informatique et des Libertés (CNIL):*

Les fiches pratiques sur l'IA

LINC - Laboratoire d'innovation numérique de la CNIL: Dossier "IA générative"

- *Comité national pilote d'éthique du numérique (CNPEN):*

Avis n°3: Agents conversationnels: enjeux d'éthique, 15 septembre 2021.

Avis n°7: Systèmes d'intelligence artificielle générative : enjeux d'éthique, 30 juin 2023.

- *Association Data for Good:*

Livre blanc "Les grands défis de l'IA générative". Version 1.0, juillet 2023.

- *Office français de l'intégrité scientifique (OFIS):*

Systèmes d'intelligence artificielle générative : quelques points de vigilance, février 2024.

- *Académie Nationale de Médecine:*

Systèmes d'IA générative en santé : enjeux et perspectives, rapport adopté le 5 mars 2024.

À l'étranger

- The Royal Society (UK):

Science in the age of AI. Rapport, mai 2024.

- UK Research integrity office (UKRIO):

AI in research (mis à jour en janvier 2024).

- *Pôle Interordres de Montréal & Laboratoire d'éthique du numérique et de l'Intelligence Artificielle (LEN.IA), Québec:*

Former à l'éthique de l'IA en enseignement supérieur: référentiel de compétences; trousse pédagogique

Au niveau international

- *Organisation des Nations unies pour l'éducation, la science et la culture (UNESCO):*

Recommendation on the Ethics of Artificial Intelligence (23 novembre 2021).

- *Organisation de coopération et de développement économiques (OCDE):*

Recommendation of the Council on Artificial Intelligence (22 mai 2019).

OECD.AI policy observatory - Policies, data and analysis for trustworthy artificial intelligence (site web).

- *G7*

Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems, 30 octobre 2023

Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems, 30 octobre 2023.

- Association Montreal AI Ethics Institute (MAIEI):

Site web.

- *Parlement Européen et Conseil :*

AI Act: Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle

- *Commission Européenne (CE):*

Ethics guidelines for trustworthy AI, 2019.
Ethics By Design and Ethics of Use Approaches for Artificial Intelligence, version 1.0 du 25 novembre 2021.
Living guidelines on the responsible use of generative AI in research, version 1.0 de mars 2024.

- *All European Academies (ALLEA):*

European Code of Conduct for research integrity (édition révisée en 2023).

VII.2 Ressources thématiques

IA génératives et publications scientifiques

Recherche, enquêtes, rapports:

Bhavsar D, Duffy L, Jo H, Lokker C, Haynes RB, Iorio A, Marusic A, Ng JY (2024) Policies on Artificial Intelligence Chatbots Among Academic Publishers: A Cross-Sectional Audit. Preprint déposé dans medRxiv: https://www.medrxiv.org/content/10.1101/2024.06.19.24309148v1
Cacciamani GE, Eppler MB, Ganjavi C, Pekan A, Biedermann B, Collins GS, Gill IS (2023) Development of the ChatGPT, Generative Artificial Intelligence and Natural Large Language Models for Accountable Reporting and Use (CANGARU) Guidelines. Preprint déposé dans arXiv: http://arxiv.org/abs/2307.08974
Ganjavi C, Eppler MB, Pekcan A, Biedermann B, Abreu A, Collins GS, Gill IS, Cacciamani GE (2024) Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: bibliometric analysis. BMJ 384:e077192. https://doi.org/10.1136/bmj-2023-077192
Kacena MA, Plotkin LI, Fehrenbacher JC (2024) The Use of Artificial Intelligence in Writing Scientific Review Articles. Curr Osteoporos Rep 22:115–121. https://doi.org/10.1007/s11914-023-00852-0
Liang W, Zhang Y, Cao H, Wang B, Ding DY, Yang X, Vodrahalli K, He S, Smith DS, Yin Y, McFarland DA, Zou J (2024) Can Large Language Models Provide Useful Feedback on Research Papers? A Large-Scale Empirical Analysis. NEJM AI 1:AIoa2400196. https://doi.org/10.1056/AIoa2400196
Mugaanyi J, Cai L, Cheng S, Lu C, Huang J (2024) Evaluation of Large Language Model Performance and Reliability for Citations and References in Scholarly Writing: Cross-Disciplinary Study. Journal of Medical Internet Research 26:e52935. https://doi.org/10.2196/52935
Thelwall M (2024) Can ChatGPT evaluate research quality? Journal of Data and Information Science 9:1–21. https://doi.org/10.2478/jdis-2024-0013

Recommandations:

- *International Committee of Medical Journal Editors (ICMJE):*

"Defining the role of authors and contributors" - mise à jour 2024 intégrant un point sur l'usage des outils d'IA

- *Committee on Publication Ethics (COPE):*

Authorship and AI tools - COPE position statement, février 2023.

- *International Association of Scientific, Technical, and Medical Publishers (STM):* Generative AI in scholarly communications: ethical and practical guidelines for the use of generative AI in the publication process, décembre 2023.

Base de données - Retraction Watch:
Papers and peer reviews with evidence of ChatGPT writing.

Presse:
How ChatGPT and other AI tools could disrupt scientific publishing (Nature, 10 octobre 2023).
AI-generated rat genitalia: Swiss publisher of scientific journal under pressure (SWI swissinfo.ch, 13 mars 2024).
Quand ChatGPT tient la plume (TheMetaNews, 26 avril 2024).

Plateformes de lutte contre les paper mills:
STM Integrity Hub
UNITED2ACT

Impacts environnementaux et sociaux des IA
Recherche, enquêtes, rapports:
Li P, Yang J, Islam MA, Ren S (2023) Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. Preprint déposé dans arXiv: http://arxiv.org/abs/2304.03271
Ludec CL, Cornet M (2023) Enquête : derrière l'IA, les travailleurs précaires des pays du Sud. The Conversation.

Dossiers:
Fond Monétaire International (FMI): "Artificial Intelligence"

Presse:
"Ils profitent de notre pauvreté" : derrière le boom des intelligences artificielles génératives, le travail caché des petites mains de l'IA (France Info, 8 avril 2024).
IA : quel est le bilan carbone de ChatGPT ? (Les Numériques, 22 avril 2024).
À Madagascar, les petites mains bien réelles de l'intelligence artificielle (France Info, 29 avril 2024)
Comment l'intelligence artificielle a fait augmenter les émissions de gaz à effet de serre des géants de la tech (France Info, 3 juillet 2024).
IA: la consommation d'eau cachée de ChatGPT (Disclose, newsletter "Planète investigation", 4 juillet 2024).