



Arcom
Le régulateur de la communication
audiovisuelle et numérique

CAFÉ IA

IA et information : enjeux et outils pour un usage éclairé

Sommaire

Introduction	3
Partie 1. L'IA au sein de la chaîne de l'information	4
1. La génération de contenu	4
2. L'accès à l'information	5
3. L'accès synthétisé	6
4. La détection et l'analyse de l'information	7
5. La recommandation	8
6. La modération	9
7. La certification et la labellisation	10
Partie 2. L'IA au centre des grands enjeux de notre rapport à l'information	11
1. Une altération de la représentation de la réalité	11
2. Faciliter la fabrication de fausses informations	13
3. Accélérer la diffusion de fausses informations	15
4. Renforcer le déséquilibre dans la répartition de la valeur	15
Partie 3. L'IA comme levier d'action pour lutter contre la propagation de fausses informations	16
1. Des outils utilisés par les plateformes visant à lutter contre les comportements inauthentiques et manipulateurs sur leurs services	16
2. Des outils visant à aider les utilisateurs à exercer leur esprit critique sur le contenu qu'ils voient	17
3. Des fonctionnalités visant à assurer la transparence vis-à-vis des utilisateurs	18
Partie 4. Des conseils pour exercer son esprit critique de manière éclairée	20
Bibliographie	22

Introduction

La télévision et la radio jouent un rôle toujours central dans la recommandation et l'influence des comportements du public. Malgré l'essor des plateformes numériques, ces médias traditionnels restent très consommés et influencent ce que les gens regardent, écoutent et les opinions qu'ils se forment.

L'émergence du web 2.0 a transformé les usages numériques en donnant aux internautes un rôle actif dans la production de contenu. **Avec l'essor des réseaux sociaux, des blogs et des wikis, les utilisateurs ne se sont plus contentés de consommer l'information : ils ont commencé à la créer, la partager et l'influencer.** Le *User Generated Content* s'est imposé comme un pilier du web, redéfinissant les dynamiques de consommation et les tendances, mais aussi notre rapport à l'information, où chacun peut désormais en devenir un acteur à part entière.

Cette démocratisation du partage de l'information a certes permis une plus grande circulation des contenus, mais a aussi eu ses revers. L'information est devenue tellement abondante qu'elle finit par être anxiogène, fragilisant ainsi la confiance envers les institutions qui la diffusent. Une large majorité des Français n'ont ni confiance dans les médias ni dans les réseaux sociaux, qui se retrouvent en queue de liste des organismes dans lesquels les français placent leur confiance¹. Ce climat de méfiance a conduit à un phénomène de retrait face à l'information, certains choisissant même de s'en couper volontairement. L'étude de l'Arcom dédiée à la relation que les Français entretiennent avec l'information le montre : 61 % indiquent faire le choix régulier d'éviter volontairement de s'informer².

C'est dans ce contexte de surabondance et de défiance vis-à-vis de l'information que l'apparition de l'intelligence artificielle doit être appréhendée. Pour les acteurs traditionnels de la diffusion et de la production d'information, l'usage de l'IA n'est pas un phénomène récent. Ces technologies les accompagnent depuis déjà un certain temps. Toutefois, l'émergence de l'IA générative a marqué un tournant en offrant de nouvelles possibilités aux utilisateurs, modifiant la manière dont l'information est créée, ainsi que notre rapport à celle-ci. **La question se pose alors de savoir dans quelle mesure l'IA a pu participer à ce délitement de la confiance vis-à-vis de l'information, et comment elle pourrait potentiellement permettre de la restaurer.**

¹ *Baromètre de la confiance*, Cevipof, 2025.

² *Les Français et l'information*, Arcom, 2024.

Partie 1 : L'IA au sein de la chaîne de l'information

L'essor fulgurant des IA génératives suscite de vifs débats, notamment sur notre rapport à l'information. Si leurs capacités impressionnent, **il est essentiel de rappeler que l'IA n'est pas une nouveauté dans l'espace informationnel**. Depuis des années, des IA non génératives, souvent imperceptibles par le grand public, jouent déjà un rôle clé : résumé automatique, traduction, recommandation, modération. Ces technologies, bien ancrées, soulèvent des enjeux majeurs qui ont conduit à l'adoption de nouvelles réglementations pour encadrer leur utilisation : le Règlement européen sur les Services Numériques aussi appelé Digital Services Act (DSA), le Règlement européen sur l'Intelligence Artificielle (RIA), la directive sur les Services Médias Audiovisuels (directive SMA).

L'IA intervient à chaque étape du cycle de l'information, de sa création à sa diffusion, en passant par son identification et sa recommandation. **L'émergence soudaine des IA génératives auprès du grand public ne fait que raviver des questionnements déjà anciens tout en introduisant de nouveaux, interrogeant la manière dont nous produisons, percevons et consommons l'information.**

1. La génération de contenu

L'IA générative a cela d'innovant qu'elle permet de produire du contenu, une tâche autrefois réservée aux humains. C'est en effet dans son essence : générer du contenu, textuel, visuel ou audio, en réponse à une requête (aussi appelée « prompt ») d'un utilisateur. Ces systèmes sont adossés à des modèles d'IA dits de « fondation », appelés « Transformers » du nom de l'architecture du modèle. Il s'agit de modèles de grande taille, entraînés sur une vaste quantité de données et pouvant ainsi s'adapter à un large éventail de tâches, dont la création de contenus.

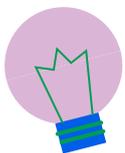
Focus sur le fonctionnement de l'IA générative : les IA génératives, qui manipulent du texte, sont des modèles de langage géants aussi appelés « giga-modèles » ou « LLM » (*Large Language Model*) comme, par exemple *ChatGPT* d'OpenAI ou un équivalent français, *Le Chat* de Mistral AI. Leur principe est simple : il est demandé au modèle de langage d'estimer, à partir du début d'une phrase, quel est le mot suivant le plus probable d'apparaître. En lui apprenant à prédire la suite de milliards de phrases (provenant majoritairement d'internet), le modèle de langage va être capable d'imiter des conversations repérées dans la masse d'exemples. **Il s'agit donc d'imiter un langage de façon statistique, par apprentissage, à partir de ce qui est contenu dans les milliards de données³.**

³ Fiche IA « A quoi ça sert ? », site Francenum.

Les contenus générés par ces modèles peuvent être de différents types :

- ▶ **Des contenus textuels** produits par des assistants conversationnels sous la forme de « chatbots », soit des « systèmes conçus pour des conversations prolongées imitant les conversations non structurées ou les « chats » caractéristiques de l'interaction humain-humain » comme *ChatGPT* proposé par OpenAI, *Gemini* proposé par Google ou *Le Chat* de Mistral AI. Ces systèmes utilisent des « grands modèles de langage » (*Large Language Model*, LLM) ;
- ▶ **Du code de programmation complexe**, dans un langage choisi par l'utilisateur (ex. Python), permettant de réaliser du développement informatique en un temps record, également grâce à *ChatGPT* ou d'autres services similaires ;
- ▶ **Des contenus audios** notamment produits par le clonage ou la déformation de voix réelles afin de faire dire à un individu quelque chose qu'il n'a pas dit ;
- ▶ **Des images et des vidéos** produites par des services comme *Midjourney*, *Dall-E* proposés par OpenAI ou Stability AI.

Cette capacité nouvelle emporte de multiples enjeux sur la qualité de l'information, sa transparence et sa traçabilité, tout en suscitant des inquiétudes quant à la prolifération de contenus « synthétiques » et trompeurs. L'IAG (IA générative), en baissant la barrière de la création de contenus, facilite les opérations d'influence qui peuvent être encore plus personnalisées et efficaces. Elle renforce les acteurs malveillants en leur permettant de créer rapidement du contenu synthétique et trompeur plus réaliste et convaincant.



Il convient de préciser, qu'en parallèle, sont développés des outils dont l'objectif est de tenter de répondre à ces défis : en effet, les techniques d'IA sont progressivement prises en main par le monde de la recherche scientifique et les régulateurs. À ce titre, il est observé une émergence d'outils permettant aux utilisateurs d'évaluer la probabilité qu'un contenu soit manipulé ou créé grâce à l'IA (ex. <https://weverify.eu/about/>).

2. L'accès à l'information

Internet et les réseaux sociaux ont profondément transformé l'accès à l'information, en devenant des voies par lesquelles nous nous informons de plus en plus. Autrefois, l'accès à l'information passait presque exclusivement par des intermédiaires de confiance tels que les journalistes, les éditeurs ou les experts, qui sélectionnaient et validaient les contenus que nous pouvions consulter. **Aujourd'hui, les utilisateurs ont une liberté plus grande pour rechercher et accéder à l'information de manière autonome.** Or, d'après l'étude de l'Arcom sur le rapport des Français à l'information, 49 % d'entre eux utilisent quotidiennement les moteurs de recherche pour s'informer et 47 % les réseaux sociaux⁵. Cela témoigne du rôle central que ces plateformes jouent désormais dans notre accès à l'information.

⁴ FAVRO Karine, VILLATA Serena, et ZOLYNSKI Célia. *Des assistants vocaux aux agents conversationnels. Vers un encadrement des interfaces vocales humain-machine.* Dalloz IP/IT: droit de la propriété intellectuelle et du numérique, 2023, no 09, p. 459.

⁵ *Les Français et l'information*, Arcom, 2024.

L'IA a été un facteur clé de ce changement. Elle est intégrée dans les moteurs de recherche et les réseaux sociaux, où elle intervient pour affiner les résultats en fonction des préférences des utilisateurs. Par exemple, **l'IA permet de personnaliser les résultats de recherche, de simplifier l'exploration des contenus, ou encore d'améliorer l'accessibilité de l'information** grâce à des outils comme la traduction automatique et la synthèse des articles. Ces technologies rendent l'information plus facilement accessible et compréhensible, mais elles soulèvent aussi des questions sur la qualité, la diversité et la fiabilité des contenus proposés.

L'IAG quant à elle crée de nouveaux canaux d'accès à l'information et peut bouleverser cette structuration. L'accès à l'information se fait désormais de plus en plus par les agents conversationnels, les grands modèles de fondation comme *ChatGPT, PerplexityAI, Copilot*, qui ont cela de commun d'avoir une interface unique et d'apporter une seule réponse à la question posée par l'utilisateur. Il est probable que l'on passe d'usages via différentes applications pour nos différents réseaux sociaux et médias, à des interfaces uniques personnalisées qui centraliseront les contenus, informations, etc.

De façon encore plus prospective, l'accès à l'information pourra se faire demain par le biais des outils de réalité virtuelle. Certains médias internationaux proposent déjà des applications de réalité virtuelle comme le journal suédois *Dagens Nyheter*, le *New York Times* ou *ABC News*.

3. L'accès synthétisé

Les IA génératives accentuent à l'extrême le phénomène du « web synthétique », un concept développé par Olivier Ertzscheid⁶. Cette expression désigne la place grandissante des robots, agents conversationnels et algorithmes dans la production d'information. L'explosion de ces technologies conduit à une boucle où les contenus utilisés par les IA proviennent eux-mêmes d'une synthèse d'algorithmes.

L'association de différentes briques d'IA permet aujourd'hui de créer des contenus audiovisuels entièrement « synthétiques ». Les grands modèles de langage peuvent ainsi rédiger des textes qui seront ensuite lus par des avatars d'animateurs, créés artificiellement ou clonés à partir de véritables animateurs.

Des expérimentations ont lieu depuis plusieurs années, par exemple en **Corée du Sud**, où une animatrice de journal TV peut être, si besoin, remplacée par son avatar. **En Suisse**, c'est un bulletin météo de la chaîne *M le Média* qui est présenté par une animatrice virtuelle. *Radio FG* utilise pour sa part des voix de synthèse pour présenter les titres diffusés sur la station de radio numérique *Dance One*. Cette technique permet à Radio FG de multiplier la diffusion de programmes musicaux, en vue notamment de se développer à l'international en limitant les coûts. **Enfin, aux États-Unis** une chaîne d'information en continu avec uniquement des présentateurs générés par IA, *Channel 1 AI*, a commencé à émettre début 2024. Cet usage s'applique particulièrement bien à la présentation de contenus préparés et formatés, au point de pouvoir duper un public non averti. Mais la technologie n'est pas encore très adaptée pour la réalisation de fictions ou de programmes interactifs en temps réel.

⁶ https://www.lemonde.fr/idees/article/2023/10/07/intelligence-artificielle-nous-sommes-passees-du-reve-d-un-web-semantique-a-la-realite-d-un-web-synthetique_6192914_3232.html

4. La détection et l'analyse de l'information

La détection instantanée d'informations ou d'événements d'importance⁷ peut être facilitée par l'intelligence artificielle, en analysant par exemple en continu le nombre de publications sur les réseaux sociaux et leur géolocalisation. L'analyse s'appuie sur le renseignement en sources ouvertes et se veut générale (par opposition aux systèmes de veille plus classiques, basés uniquement sur des mots-clés ou des thématiques spécifiques).

L'Inria (Institut national de recherche en sciences et technologies du numérique) a par exemple développé, en partenariat avec Radio France, un outil d'IA appelé *StatCheck* qui permet de vérifier des déclarations faites à l'antenne ou sur les réseaux sociaux par des personnalités ou responsables politiques, en contrôlant leurs affirmations grâce aux statistiques publiques de l'Insee et d'Eurostat. Les journalistes de Radio France utilisent StatCheck pour vérifier les informations dans le cadre de leurs activités quotidiennes. L'outil est également utilisé pour alimenter la rubrique quotidienne « Le vrai du faux » sur *France Info*⁸. Par ailleurs, le site pluralisme.fr créé par l'entreprise française Magic Lemp, propose de suivre l'évolution des positions politiques, des projets de lois ou encore de vérifier les déclarations politiques.

Par ailleurs, **l'IA peut permettre de faciliter la recherche documentaire**. Les grands modèles de langage sont capables de comprendre une requête formulée en langage naturel et d'y apporter une réponse structurée. Il convient de préciser que l'IA générative fournit toujours une réponse, même négative. Bien qu'elle puisse être cohérente, cette réponse peut contenir des informations fausses, notamment lorsque les instructions de l'utilisateur sont ambiguës ou lorsque les données utilisées pour l'entraînement de l'IA générative sont biaisées ou incomplètes. C'est notamment le cas pour une demande concernant une actualité récente dont l'IAG n'aurait pas connaissance de la solution. Le but de ces systèmes d'IA est de satisfaire les utilisateurs en tentant de fournir une réponse sur n'importe quel sujet, même si elle est incorrecte, tord la vérité ou invente des faits. Tous les assistants conversationnels ne sont pas connectés à Internet en temps réel et leur base de données d'entraînement peut avoir une temporalité limitée. **Il est ainsi possible de parler « d'hallucinations » quand une IA générative fournit une réponse manifestement fausse à une question pour laquelle la réponse est connue par ailleurs, sans émettre d'alerte sur le niveau de fiabilité de la réponse apportée⁹.**



La génération augmentée de récupération (ou RAG, pour Retrieval-Augmented Generation) est une technique qui permet de recadrer un outil d'IA afin que ce dernier travaille uniquement à partir d'un corpus de contenus « sain » dans lequel il va chercher ses réponses.

⁷ Le 17 mars 2014 était publié sur le site du *Los Angeles Times*, un [article](#) annonçant en premier, grâce à l'algorithme *Quakebot* branché sur les données fournies par le Centre fédéral d'information sismique (US Geological Survey), qu'un séisme venait de toucher un quartier de Los Angeles.

⁸ [Rapport de l'Arcom](#) : Bilan de la « Mission IA » sur l'impact de l'intelligence artificielle dans les domaines de la création et de l'information.

⁹ [Fiche IA « A quoi ça sert ? »](#), site Francenum.

L'IA peut aussi faciliter l'exploration d'un grand nombre de données par les journalistes d'investigation qui reçoivent parfois, par l'intermédiaire de lanceurs d'alerte, de très grande quantité de données qu'il est impossible d'analyser manuellement.

Les consortiums de journalistes tels que le consortium international des journalistes d'investigation (ICIJ) se sont donc associés à des chercheurs spécialisés dans l'intelligence artificielle afin de développer des outils capables de rendre les millions de documents mis à leur disposition plus facilement exploitable¹⁰. On a aussi assisté à l'utilisation d'un algorithme pour surveiller des contenus de masse afin d'identifier des tendances (ex : le [Reuters News Tracer](#) qui met sous surveillance quotidienne des millions de tweets pour repérer des sujets potentiels à traiter)¹¹.

5. La recommandation

La recommandation est probablement l'étape de la chaîne de l'information où l'impact de l'intelligence artificielle est le plus marqué et sensible. Elle induit la façon dont l'information va circuler dans l'espace informationnel numérique et dont elle sera diffusée. S'appuyant sur des algorithmes d'apprentissage automatique (*machine learning*), elle affine la personnalisation des contenus suggérés aux utilisateurs. Cette personnalisation repose sur l'étude de leurs préférences, déduites à partir de leurs interactions en ligne et des données de navigation collectées, notamment via les cookies et d'autres traceurs numériques. Les algorithmes de recommandations régissent aujourd'hui la façon dont les contenus d'information sont organisés et définissent l'éditorialisation des plateformes numériques (moteur de recherche, réseaux sociaux, médias).

En matière d'éditorialisation, l'IA offre des opportunités en termes de personnalisation des grilles de contenus et des offres, en fonction des choix des utilisateurs et de leur géolocalisation.

Par exemple, aux États-Unis, la grande radio publique *NPR One* a développé une application, qui agrège des podcasts de ses émissions avec des programmes issus de stations locales affiliées¹².

Les contenus postés sur les réseaux sociaux et sur les plateformes de partage de vidéo, qu'il s'agisse de programmes courts, de séquences ou d'extraits d'émissions télévisées, doivent généralement respecter certains codes (dimensions, durée, formatage) pour pouvoir être exploités dans des conditions optimales.

¹⁰ [Rapport de l'Arcom](#) : Bilan de la « MISSION IA » sur l'impact de l'intelligence artificielle dans les domaines de la création et de l'information.

¹¹ Article de l'INA : <https://larevuedesmedias.ina.fr/information-face-lintelligence-artificielle-promesses-et-dangers>.

¹² Article de l'INA : <https://larevuedesmedias.ina.fr/information-face-lintelligence-artificielle-promesses-et-dangers>.

Les outils d'IA tels que ceux proposés par Aive (*Artificial Intelligence for Video Experience*) peuvent faciliter ce travail d'édition et de montage automatisé, en proposant un recadrage pertinent des images, en générant des titres, en suggérant les passages qui peuvent être coupés ou en retravaillant la bande son de chaque contenu sous-jacent¹³.

Par ailleurs, des **algorithmes peuvent entrer en dialogue avec les internautes**, répondant à leurs questions ou leurs commentaires comme s'il s'agissait d'un être humain (chatbots ou « agents conversationnels »).

Par exemple « Le 20Bot du journal *20 Minutes*, permet aux internautes, depuis juin 2016, de poser des questions sur *Messenger* sur des sujets d'actualité et d'obtenir en retour une sélection automatisée de cinq articles répondant à leur requête¹⁴.

6. La modération

Un autre aspect dans lequel l'intelligence artificielle s'avère déterminante est la modération des contenus. La modération est l'activité par laquelle un réseau social définit et applique les règles à suivre sur son réseau social, qu'elles émanent de la loi ou des règles d'utilisation des réseaux sociaux en cause. Aujourd'hui, la grande majorité des contenus modérés le sont par des algorithmes d'IA, qui automatisent la classification des contenus qui doivent être laissés en ligne, supprimés ou soumis à un examen humain. La plupart des décisions de modération du contenu sont aujourd'hui prises par des machines, et non par des êtres humains. Grâce à des IA avancées comme le traitement du langage naturel (NLP), la *computer vision* – un domaine de l'IA qui permet de reconnaître des informations à partir d'images numériques, de vidéos et d'autres entrées visuelles – et le *machine learning*, les systèmes d'IA sont capables d'analyser d'énormes volumes de données en temps réel. Les réseaux sociaux misent beaucoup sur ces outils automatisés dans leur politique de modération. D'abord peu fiables, réservés aux contenus les plus faciles à modérer, comme les images à caractère sexuel, ils ont peu à peu été étendus à un nombre croissant de types de contenus, notamment ceux qui nécessitent de la nuance et une compréhension fine du contexte, comme les discours haineux.

Or ce recours accru à la modération automatique, parfois au détriment de la modération humaine, est critiqué pour ses effets de bords. En effet, l'automatisation, si elle permet le traitement d'une grande quantité de contenus, amplifie l'erreur humaine, en intégrant des biais dans les données d'entraînement et la conception du système. De plus, les décisions d'application sont prises rapidement, laissant peu de place à la supervision humaine¹⁵.

¹³ [Rapport de l'Arcom](#) : Bilan de la « MISSION IA » sur l'impact de l'intelligence artificielle dans les domaines de la création et de l'information.

¹⁴ Article de l'INA : <https://larevuedesmedias.ina.fr/linformation-face-lintelligence-artificielle-promesses-et-dangers>

¹⁵ En cas de recours interne, les plateformes « veillent à ce que les décisions [...] soient prises sous le contrôle de collaborateurs dûment qualifiés, et pas uniquement par des moyens automatisés » (article 20(6) du DSA).

7. La certification et la labellisation

La certification et la labellisation jouent un rôle fondamental dans le cycle de l'information en garantissant sa légitimité et sa crédibilité ; elles servent à s'assurer que les utilisateurs accèdent à des informations fiables et de qualité. Dans un monde où la diffusion de contenus en ligne est largement influencée par des algorithmes, ces certifications peuvent jouer un rôle clé en favorisant l'exposition de contenus de qualité dans les recommandations des plateformes numériques. Le recours à un « *marquage bien visible* » fait d'ailleurs partie des mesures prévues par l'article 35 du règlement sur les services numériques (RSN), en ce qu'il permet d'aider les utilisateurs à distinguer un contenu qui est généré par l'intelligence artificielle.

Les plateformes, comme les moteurs de recherche ou les réseaux sociaux, utilisent des algorithmes pour déterminer quels contenus sont visibles pour les utilisateurs. **En intégrant des critères de certification dans ces algorithmes, l'idée est de donner plus de visibilité aux informations certifiées et vérifiées, contribuant ainsi à réduire la circulation de contenus erronés ou trompeurs.** Aussi, certaines plateformes, comme celles de *Meta*, permettent désormais aux utilisateurs de signaler les contenus générés par l'IA lorsqu'ils ne sont pas étiquetés tels quels, pour que la plateforme y appose un label et applique des pénalités aux utilisateurs concernés.

Néanmoins, la mise en place d'un système de labellisation fiable et efficace implique une surveillance continue et une mise à jour des critères d'étiquetage pour s'adapter à de nouveaux types de contenus, générés de plus en plus sophistiquement, et aux tentatives de contourner ou manipuler le système de labellisation. De ce fait, il est nécessaire de développer de nouvelles normes et de promouvoir la collaboration entre les secteurs public et privé. Cela implique que les entreprises technologiques et les médias travaillent ensemble.

Par exemple, la *Coalition for Content Provenance and Authenticity (C2PA)* a développé un outil (*Content Credentials*) permettant de réduire ces risques, en donnant la possibilité aux créateurs d'ajouter des métadonnées inviolables sur eux-mêmes et sur le processus de création/modification de leur contenu, afin de renforcer la traçabilité et la transparence au profit des utilisateurs.

D'autres initiatives voient également le jour, avec des partenariats entre entreprises d'IA et médias afin d'améliorer la production, la personnalisation et la distribution de contenus. Ces partenariats offrent aux médias des technologies innovantes et permettent aux entreprises d'IA de gagner en légitimité en garantissant la qualité et la fiabilité des informations générées par leurs modèles, tout en renforçant la transparence des données utilisées.

Par exemple, l'AFP a noué un partenariat avec Mistral AI pour enrichir son assistant conversationnel *Le Chat* avec ses archives, et *Le Monde* a signé un accord avec OpenAI, à l'instar de Prisa Media.

Partie 2. L'IA au centre des grands enjeux de notre rapport à l'information

La présence de l'IA dans le cycle de l'information transforme profondément notre manière de produire, de consommer et d'interagir avec l'information.

L'intelligence artificielle a un impact direct sur la manière dont l'information nous parvient et **peut contribuer à la diminution de la qualité de l'information et à sa standardisation**. En effet, ces systèmes reposent sur des algorithmes qui génèrent et recommandent de l'information en s'appuyant sur des données préexistantes, comme des textes, des images ou des vidéos déjà produits. Ce processus est limité par les données d'entraînement de l'IA, ce qui peut conduire à une reproduction de schémas, d'idées et de discours déjà connus, sans véritable créativité ou apport original. Cela peut mener à une surabondance de contenus similaires, où les nuances et la profondeur de réflexion sont mises à l'écart au profit d'une simplification parfois excessive.

Dès lors, en recommandant certains types de contenus et en en minimisant d'autres, l'IA peut renforcer des biais, limiter la diversité des opinions auxquelles nous sommes exposés, et parfois amplifier la désinformation. **Cela modifie notre rapport à l'information en réduisant notre capacité à accéder à une vision complète et nuancée des faits.**

L'IA joue également un rôle dans la création de contenu, avec des systèmes capables de générer des articles, des textes ou même des vidéos à partir de données brutes, ce qui soulève des questions sur la fiabilité de l'information.

Ainsi, l'impact de l'IA sur le cycle de l'information ne se limite pas à la production ou la distribution des contenus, mais affecte également la manière dont nous les recevons, les traitons et les croyons.

1. Une altération de la représentation de la réalité

► **Risque d'une polarisation et d'une fragmentation de l'espace informationnel** : Les interfaces des IA génératives sont conçues pour devenir des guichets uniques, fournissant directement des réponses aux questions des utilisateurs sans nécessiter de recherches supplémentaires. Cela s'inscrit en rupture avec l'ère des réseaux sociaux, où l'interaction sociale, c'est-à-dire le dialogue entre utilisateurs, est au cœur de leur modèle. Sur les plateformes en ligne, bien que l'information reçue soit de plus en plus personnalisée en fonction des préférences et des comportements numériques des utilisateurs, elle reste visible par tous (ou à tout le moins par un groupe d'utilisateurs) : chacun peut partager, réagir et interagir avec les contenus, les utilisateurs peuvent débattre, confronter leurs points de vue avec d'autres et diffuser des informations partagées collectivement. Avec des outils d'IA génératives, l'information reçue par l'utilisateur est totalement personnalisée, basée sur ses requêtes spécifiques.

L'IA répond individuellement et de manière privée à chaque utilisateur, qui devient ainsi une entité isolée, exposé à une réalité informationnelle qui lui est propre, mais qui peut manquer de diversité et de confrontation avec d'autres perspectives. Cela interroge quant à la découvrabilité de nouveaux contenus et à la pluralité de l'espace informationnel. Ce risque d'enfermement est redoublé par l'autopréférence souvent à l'œuvre et par laquelle les opérateurs de ces assistants privilégient leurs propres services dans la réponse fournie à l'utilisateur¹⁶. Ce changement de paradigme fait peser un risque accru d'enfermement dans des bulles informationnelles.

- ▶ **Amplification des tendances dominantes** : les IAG sont des modèles probabilistes qui génèrent des contenus selon des données préexistantes, et sont incapables de produire des contenus, originaux ou singuliers, renforçant ainsi les moyennes et amplifiant les tendances dominantes.
- ▶ **Illusion de la réalité** : il est difficile pour un humain de faire la différence entre un texte écrit par une IA et un texte rédigé par une personne. Les utilisateurs ont tendance à préférer des contenus artificiels et à leur accorder plus de crédibilité. Enfin, un phénomène plus subtil se produit : à force d'être exposés à des textes biaisés par l'IA, nous risquons d'adopter ces biais sans même nous en rendre compte, ce qui peut influencer nos propres écrits et, à long terme, nos opinions. Cela remet en question notre manière de recevoir et de partager l'information, et interroge l'existence d'un espace d'information commun à tous.
- ▶ **Renforcer les biais : un biais cognitif est un processus de pensée automatique et souvent inconscient qui façonne notre jugement et notre perception.** Ces raccourcis peuvent mener à des erreurs de jugement, en nous faisant privilégier certaines informations et en interprétant les données de manière subjective. À titre d'exemple, le biais de confirmation se caractérise, pour les individus, comme une tendance naturelle à rechercher, interpréter et se souvenir des informations qui confirment leurs croyances préexistantes, tout en rejetant les informations contraires. Ce biais contribue à la diffusion de la désinformation, car les individus sont plus enclins à accepter des informations qui correspondent à leurs opinions, même si ces informations sont erronées. Comme les biais sont présents dans la société, chez les ingénieurs qui conçoivent les IA et dans les données d'entraînement, l'IA reproduit ces biais et peut parfois les amplifier. **De tels biais algorithmiques peuvent entraîner des représentations stéréotypées ou discriminatoires.**

À titre d'exemple, la société américaine *Amazon* a développé un système d'IA dont l'objectif était de faciliter le recrutement en automatisant le classement des candidats pour un poste. Or, le système qui se basait sur les données historiques d'embauche de la société, a montré de très sérieux biais contre les candidates, déclassant les CV des candidates ayant fréquenté deux collèges réservés aux femmes et pénalisant même les CV contenant le mot « femme ». *Amazon* a tenté d'ajuster l'algorithme, mais le projet a rapidement dû être interrompu.

¹⁶ [Rapport de mission du CSPLA sur les assistants vocaux et autres agents conversationnels, décembre 2022.](#)

2. Faciliter la fabrication de fausses informations

La désinformation est un phénomène qui n'est pas né avec le numérique et l'on en retrouve des traces dès l'antiquité. La presse papier pouvait déjà servir de support à la fabrication des faits, ce qui a par ailleurs conduit à l'adoption d'un article dans la loi de 1881 qui consacre la liberté de la presse pour prohiber les « fausses nouvelles ».

Aujourd'hui, avec l'avènement des outils d'IA générative, une nouvelle ère de création de contenus en ligne (textes, audios, vidéos) hyperréalistes s'est ouverte. **Les nouvelles technologies ont renforcé des dynamiques existantes: elles amplifient la capacité à créer et diffuser de l'information, tout en rendant plus complexe la vérification de la véracité des contenus.**

▶ L'IA peut faciliter la production et la diffusion de fausses informations :

- ◆ elle **donne des moyens techniques très poussés pour apporter une illusion de fiabilité et de vérité** à des contenus textuels, visuels ou audio. Ces contenus crédibles et convaincants semblent ainsi – à première vue – authentiques ;
- ◆ elle **permet de manipuler l'information** en créant des récits fictifs, biaisés ou en modifiant des éléments visuels, rendant la détection de ces faux contenus plus difficiles ;
- ◆ elle **peut être utilisée pour discréditer de vraies informations en les qualifiant de fausses** ou de synthétiques, ou, à l'inverse, **pour crédibiliser une information fautive**, en exploitant la confiance que beaucoup accordent aux réponses générées par ces outils ;

Par exemple, dans le contexte de la guerre en Ukraine, des responsables russes ont utilisé des réponses générées par ChatGPT pour légitimer des narratifs mensongers. Présentées comme des « vérités », ces réponses étaient utilisées pour accréditer des théories du complot ou des informations manipulées, donnant l'impression que l'IA elle-même « validait » ces propos¹⁶.

- ◆ même inintentionnellement, **les agents conversationnels se font parfois l'écho de narratifs erronés ou trompeurs**. Formés sur une large base de données, ils peuvent parfois véhiculer involontairement des stéréotypes ou des points de vue polarisés, en fonction des biais présents dans les informations sur lesquelles ils ont été entraînés.

DeepSeek, l'IA générative chinoise, a notamment été taxée de relais des positions du gouvernement pékinois¹⁷.

- ◆ d'un point de vue technique, l'IA peut aussi être utilisée pour diffuser ces contenus à grande échelle via des réseaux de *botnets* (ou « réseaux de zombies¹⁸ ») pour amplifier leur impact :

¹⁷ [Baromètre de la désinformation générée par IA en français, octobre 2024.](#)

¹⁸ Groupe d'ordinateurs ou de dispositifs sous le contrôle d'un acteur malveillant.

- multiplication de faux sites : l'essor des IA et des algorithmes dans la production d'information et les interactions en ligne pourrait mener à un « web synthétique », où les contenus générés par l'IA s'alimenteraient eux-mêmes de données artificielles, créant un écosystème de plus en plus détaché des sources humaines. Depuis deux ans, on observe une explosion des sites d'information générés par IA sur le Web. Rien qu'entre mai 2023 et octobre 2024, ces pseudo-sites d'actualité sont passés de 49 à 1121 ;

En France, des cas concrets de « médias synthétiques » manipulés par l'IA ont été identifiés en 2024. De faux sites internet d'informations tels que « Verite Cachee », « France en colere », « Media alternatif » ou « Info independants » étaient en effet conçus pour reprendre des articles diffusés sur des sites d'information réputés puis réécrits selon des directives pro-Russes¹⁹.

- la traçabilité de l'information est brouillée : traditionnellement, les sources apparaissent sous forme de liens enrichis, de *snippets* - avec des informations complémentaires - dans les moteurs de recherche et la recherche entre différentes sources permet de forger sa propre analyse. Avec l'IA, c'est cette dernière qui fait la recherche à la place de l'utilisateur et lui fournit une réponse. Dans l'absolu, ce n'est donc pas un outil de recherche. Pour répondre à une question posée, **les IA génératives actuelles donnent une réponse sur un ton parfois assertif, mais qui n'est pas vérifiable, là où un moteur de recherche traditionnel ne donne pas de réponse claire, mais plein de sources pour se faire son propre avis.** Lorsque l'on demande une information à *ChatGPT*, *Bard* ou aux moteurs de recherche IA, la réponse est instantanée et se suffit à elle-même. Qu'il s'agisse d'une analyse, d'un résumé ou d'une simple requête pratique, si la sélection des sources n'est pas rigoureuse, ou que les sources ne sont pas fiables, la réponse peut ne pas être neutre ou éclairée.

▶ **L'IA est également un outil de création des *deepfakes*** : ces techniques de manipulation utilisent l'intelligence artificielle, et plus particulièrement l'apprentissage automatique (*deep learning*), pour créer ou modifier des images, vidéos ou audios, de manière extrêmement réaliste, en remplaçant des visages, en modifiant des expressions faciales ou en générant des voix qui semblent authentiques. Ce phénomène participe à altérer la crédibilité des contenus, voire à renforcer la défiance envers les médias, en rendant plus complexe la vérification de l'authenticité d'une information et en facilitant la propagation de la désinformation.

Toutefois, le rôle de l'IA dans la désinformation doit être relativisé : ce ne sont pas tant les capacités de création qui posent problème, mais plutôt les mécanismes de propagation, en particulier sur les réseaux sociaux, où les fausses informations prospèrent grâce à un modèle économique qui privilégie l'engagement et la viralité.

¹⁹ Voir <https://www.sudouest.fr/international/russie/comment-la-russie-utilise-de-pseudos-medias-alternatifs-et-l-ia-pour-diffuser-sa-propagande-20440053.php> et <https://go.recordedfuture.com/hubfs/reports/TA-2024-0628.pdf>

3. Accélérer la diffusion de fausses informations

Les algorithmes de recommandation ne sont pas neutres. Les critères qui régissent la mise en avant des contenus restent opaques, tandis que les plateformes numériques ont bâti leur modèle économique sur l'économie de l'attention, qui repose principalement sur la captation du temps et de l'engagement des utilisateurs. **Ce fonctionnement a fait des plateformes de puissants accélérateurs de contenus, et favorise la viralité de la désinformation,** qui, par rapport aux informations vérifiées, est davantage sensationnaliste et plus susceptible de provoquer des émotions fortes. Ce phénomène est amplifié par les *superspreaders*, ces acteurs qui, intentionnellement ou non, propagent massivement des informations erronées.

L'IA générative s'inscrit pleinement dans cette logique en renforçant la capacité de production et de diffusion de contenus, y compris ceux à fort potentiel viral mais à faible valeur informative. En automatisant la création de textes, d'images ou de vidéos, elle facilite la génération de contenus optimisés pour capter l'attention, sans nécessairement garantir leur véracité. De plus, couplée aux algorithmes de recommandation, l'IA générative peut accélérer la propagation de fausses informations en adaptant dynamiquement les contenus aux préférences des utilisateurs, alimentant ainsi des bulles informationnelles et renforçant les biais cognitifs. Dans ce contexte, son rôle devient ambivalent : elle peut à la fois servir d'outil d'amplification des dérives existantes et, à l'inverse, être mobilisée pour développer des mécanismes de vérification et de contextualisation de l'information.

4. Renforcer le déséquilibre dans la répartition de la valeur

Les algorithmes de recommandation ont créé une situation de dépendance qui a entraîné le développement d'un marché de l'information en ligne caractérisé par une relation déséquilibrée entre les grandes plateformes en ligne et les éditeurs qui se joue à deux niveaux : les plateformes accaparent la majeure partie des recettes publicitaires en ligne et entretiennent une dépendance à leurs services sur le marché de la publicité en ligne. Les services d'intermédiation publicitaire proposés notamment par le leader du marché Google, tels que *Google Ad Manager* et *Google Analytics*, sont des outils indispensables pour toute activité en ligne des éditeurs. Ces bouleversements ont rendu les modèles économiques des éditeurs peu adaptés pour survivre et exposés à l'arbitraire des plateformes dans leurs modifications algorithmiques et politiques publicitaires.

Par ailleurs, les IA génératives fonctionnent selon une logique extractive de données, dont certaines sont protégées par le droit d'auteur et ont été collectées sans l'accord des auteurs. Le monde de la presse dénonce ce phénomène²⁰ et de nombreuses affaires.

Par exemple : aux États-Unis, les affaires Thomson Reuters c/ Ross Intelligence²¹ et Laion vs Robert Kneschke) illustrent les questionnements liés à la légalité de l'utilisation par l'IA des données protégées.

²⁰ « IA : les médias français s'organisent face à la collecte de données par les robots », *Les Echos*, 2023

²¹ Dans le cadre de cette affaire, le jugement américain en faveur de la protection du droit d'auteur a été rendu le 11 février 2025.

Partie 3. L'IA comme levier d'action pour lutter contre la propagation de fausses informations

À l'heure où la modération des réseaux sociaux est de plus en plus remise en question, le Règlement sur les Services Numériques (RSN) joue un rôle primordial. En effet, il impose aux très grandes plateformes et aux très grands moteurs de recherche en ligne d'identifier les risques systémiques présents sur leurs services (dont la désinformation), et de prendre des mesures d'atténuation appropriées, proportionnées et efficaces pour y remédier (articles 34 et 35 du RSN). À ce titre, certains d'entre eux ont notamment choisi de s'appuyer sur des outils d'IA pour lutter contre la propagation de fausses informations.

1. Des outils utilisés par les plateformes visant à lutter contre les comportements inauthentiques et manipulateurs sur leurs services

▶ **Algorithmes dédiés à la lutte contre les comportements malveillants** : certaines plateformes en ligne investissent dans de nouvelles technologies visant à lutter contre les comportements inauthentiques sur leurs services.

À titre d'exemple, les algorithmes de traitement du langage naturel mis en place par *LinkedIn* permettent de détecter des séquences d'activités associées à une automatisation abusive et artificielle. D'autres fournisseurs de services, comme *Meta*, ont des modèles visant à détecter les identités trompeuses à travers des signaux reposant sur différents critères.

▶ **Détection automatique par l'IA** : des outils de détection automatique des contenus générés ou manipulés par l'IA sont en développement. Cependant, leur fiabilité reste limitée et leur action nécessite d'être complétée par un travail humain de vérification.

▶ **Le système des CAPTCHAs** : certaines plateformes mettent en place, à la création de compte ou à l'authentification, des systèmes permettant de différencier les humains des robots en leur posant des questions ou en leur demandant de résoudre des tâches simples (ex : sélectionner toutes les images où figure un passage piéton). Néanmoins, ces systèmes présentent des limites et sont notamment susceptibles d'être contournés par certains outils d'IA, parfois en capacité de reconnaître des *patterns* et des déformations dans les images, ainsi que de discerner des caractères ou objets spécifiques dans des situations complexes.

- ▶ **Système de signalement des contenus partiellement automatisé** : le RSN impose aux plateformes en ligne de faciliter le signalement des contenus illicites ou contraires aux CGU de la plateforme et de mettre en place des mécanismes de traitement rapide de ces derniers (article 16). Ces dispositifs de signalement peuvent être partiellement automatisés grâce à l'IA, pour identifier plus rapidement les contenus notifiés. Cependant, ces outils de modération algorithmique peuvent parfois nécessiter, dans certaines circonstances ou pour des types de contenus spécifiques, d'être complétés par une vérification humaine, afin d'éviter les erreurs de jugement.

2. Des outils visant à aider les utilisateurs à exercer leur esprit critique sur le contenu qu'ils voient

- ▶ **Watermarking** : l'apposition d'un filigrane numérique sur les contenus générés par l'IA permettrait une détection et une traçabilité plus fiables. Des initiatives comme *SynthID* de Google et *Stable Signature* de Meta s'inscrivent dans cette démarche, sans pour autant proposer à ce stade de solution universelle ou standardisée. Le règlement européen sur l'intelligence artificielle (RIA, ou IA Act) adopté en 2024 prévoit d'ailleurs à son article 50 une obligation de marquage des contenus générés ou fortement transformés par IA. Toutefois, la technologie reste à parfaire et un enjeu fort d'interopérabilité de ces solutions de marquage puis de détection du marquage se pose.
- ▶ **Authentification des sources** : la CAI (*Content Authenticity Initiative*) et la C2PA (*Coalition for Content Provenance and Authenticity*) regroupent à l'échelle internationale de nombreux acteurs tels que les fabricants d'appareils photos, éditeurs de solutions d'IA et de logiciels, médias, agences de presse, plateformes numériques, etc., afin de créer des normes et des outils permettant d'assurer une traçabilité des contenus, notamment pour en garantir et en vérifier l'authenticité. Une norme ISO devrait voir le jour en 2025 pour formaliser ce standard, également prévu pour permettre de signaler l'origine synthétique d'un contenu et de tracer les modifications et éditions apportées à un contenu authentique.
- ▶ **Labellisation (étiquetage) des contenus** : identifier clairement les contenus générés par l'IA est une solution prometteuse. L'étiquetage est une mesure d'atténuation des risques systémiques, au sens de l'article 35 du RSN, que certaines plateformes comme *YouTube* et *Instagram* implémentent, y compris pour les annonceurs. Au-delà de l'impératif de transparence, cette labellisation interroge : marquer un contenu comme généré ou modifié par IA ne signifie pas qu'il est nécessairement faux ou trompeur et inversement. Il faudrait éviter de jeter une présomption de véracité sur les contenus non labellisés ou, au contraire, de fausseté sur les contenus marqués.

Aussi, *Meta* permet désormais aux utilisateurs de signaler les contenus générés par l'IA lorsqu'ils ne sont pas étiquetés tels quels, pour que la plateforme y appose un label et applique des pénalités aux utilisateurs concernés.

- ▶ **Promotion du contenu faisant autorité** : un contenu faisant autorité contient des informations pouvant être considérées comme fiables, pertinentes et de confiance (ex : sites gouvernementaux ; sites de presse en ligne et médias audiovisuels, éventuellement labellisés par des tiers et dont les articles sont rédigés par des journalistes professionnels). Toutefois, cela implique de définir ce que l'on considère comme source faisant autorité, ce qui peut parfois prêter à débat. Certaines plateformes ou moteurs de recherche utilisent l'IA notamment dans les contextes suivants :
 - ◆ pour rediriger les utilisateurs vers de telles informations (la plupart des plateformes ont eu tendance à le faire pendant les campagnes électorales, pour renvoyer par exemple aux informations officielles sur les modalités de scrutin) ;
 - ◆ pour mettre en œuvre des classificateurs d'apprentissage automatique visant à évaluer la pertinence d'un site ou d'une information.

À titre d'exemples, *Google Search* utilise des signaux tels que la signification, la pertinence, la qualité, l'utilisabilité et le contexte pour déterminer quels résultats doivent être envoyés en priorité pour chaque requête d'un utilisateur. Par ailleurs, *Microsoft Bing* améliore son système de classement en évitant de suggérer dans les résultats de recherche des liens qui pourraient rediriger les utilisateurs vers du contenu trompeur ou nuisible.

3. Des fonctionnalités visant à assurer la transparence vis-à-vis des utilisateurs

- ▶ Le règlement européen sur l'intelligence artificielle (RIA) impose notamment aux fournisseurs de systèmes d'IA de veiller à ce que leurs systèmes soient conçus et développés de manière à ce que **les individus qui les utilisent soient informés qu'ils interagissent avec un système d'IA, et ce dès la première interaction** (article 50). Cette transparence est également requise pour les contenus audio ou vidéo constituant un hypertrucage.
- ▶ **Les plateformes et les moteurs de recherche en ligne fournissent aux utilisateurs** de leurs services **des outils visant à accroître la transparence de l'information qu'ils voient**, afin qu'ils puissent exercer leur esprit critique de manière éclairée, notamment en présence de contenus générés par l'IA. Par exemples, certains d'entre eux :
 - ▶ **mettent à disposition des utilisateurs des fonctionnalités permettant d'en « savoir plus »** sur une information (ex : *Meta*, *Google Search*) ;
 - ▶ **permettent aux créateurs d'apposer une mention de l'usage de l'IA** sur leurs contenus, en conformité avec le RSN et le RIA ;

- ▶ **aident les utilisateurs à comprendre pourquoi un contenu leur est recommandé** (ex : *TikTok*), en mentionnant dans leurs conditions générales d'utilisation (GCU) des informations claires sur les paramètres utilisés dans leurs systèmes de recommandation, ainsi que les options dont ils disposent pour modifier ou influencer ces principaux paramètres. L'objectif de cette mesure prévue par l'article 27 du RSN est de donner aux utilisateurs la possibilité de modifier de manière aisée et à tout moment l'option de recommandation qui leur est appliquée, lorsque plusieurs options sont disponibles. L'article 38 du RSN prévoit en outre pour les très grandes plateformes et les très grands moteurs de recherche en ligne l'obligation d'introduire une option de recommandation qui ne repose pas sur du profilage ;

- ▶ **mettent en place des étiquettes de « contenu non vérifié »** lorsque les vérificateurs de faits ne peuvent pas conclure si un contenu constitue ou non de la désinformation, notamment lorsqu'il est généré par l'IA (ex : *TikTok*) ;

- ▶ **donnent l'opportunité aux utilisateurs de fournir des éléments de contexte sur les informations** qu'ils voient (ex : « Community Notes » de X), même si cette approche comporte des limites qu'il est important de prendre en compte pour exercer son esprit critique de manière libre et éclairée :
 - des groupes organisés d'utilisateurs peuvent coordonner leurs efforts pour influencer le contenu des notes et manipuler le dispositif ;
 - la manière dont les notes sont jugées utiles ou non et les critères qui amènent à les rendre publiques peuvent s'avérer opaques ;
 - certains utilisateurs peuvent ne pas avoir les compétences nécessaires pour évaluer correctement la véracité des informations, tandis que d'autres peuvent publier des notes fondées sur des informations incomplètes ou erronées.

Partie 4 : Des conseils pour exercer son esprit critique de manière éclairée

La détection de contenu généré par l'IA générative (texte, image, audio) est de plus en plus importante à mesure que les technologies se développent. Par ailleurs, dans un environnement où l'information circule rapidement et dans lequel les opinions peuvent être influencées par des informations incorrectes ou manipulées, **il est nécessaire d'exercer son esprit de manière libre et éclairé en ayant de bons réflexes.**

- **Vérifier les sources et recouper les informations** : une information (texte, image, vidéo, etc.) provenant d'une seule source a plus de chance d'être erronée, manipulée, mais également biaisée (ex : orientation politique, commerciale, etc.). Recouper les informations, en s'appuyant notamment sur des sources faisant autorité, permet de vérifier si une information est vraie et de renforcer sa crédibilité (ex : faire appel à des vérificateurs de faits tels que *Les Surligneurs* pour la désinformation juridique et/ou à des sources d'autorité, comme la presse et les médias).
- **Utiliser des outils d'IA dédiés à la vérification des faits ou des contenus** : afin de prévenir la propagation de la désinformation, des outils permettant de rapprocher les utilisateurs des plateformes des vérificateurs de faits ont été développés. Certaines initiatives, comme l'IA *Vera* de l'ONG française *LaReponse.tech*, permettent de façon très simple de vérifier la véracité d'une information en contactant l'IA via *WhatsApp* ou par téléphone (09.74.99.12.95), à l'image d'une relation humain-humain. D'autres outils permettent également d'évaluer la probabilité qu'une image ou vidéo ait été manipulée ou générée par l'IA²².
- **S'appuyer sur les ressources pédagogiques existantes** : le site du CLEMI, [les 8 conseils de l'AFP pour repérer les images artificielles](#) ou encore le guide contre la désinformation de la Délégation à l'information et à la communication de la Défense (DICOd) du ministère des Armées permettent de fournir des bonnes pratiques visant à détecter un contenu manipulé.
- **Consulter les initiatives citoyennes et collaboratives** visant à créer un nouveau rapport à l'information (« [CitizenFake & What The Fake](#) », « [LightHouse Report](#), [BellingCat](#) », etc.).
- **Être conscient de ses biais cognitifs** : cette prise de conscience permet de reconnaître quand un jugement est influencé par des préférences personnelles, des croyances ou des émotions, plutôt que par des faits objectifs. Cela pousse à adopter une approche plus réfléchie et critique face à l'information, et aide à éviter les pièges de contenus manipulateurs qui exploitent ces biais (par exemple, à travers des titres sensationnalistes).

²² l'IA Vera de l'ONG française [LaReponse.tech](#)

À noter que ces bonnes pratiques peuvent se combiner, afin de dépasser les limites propres à chacune : par exemple, la recherche de sources fiables peut être entravée par un accès limité à l'information, tandis que l'analyse logique peut parfois négliger des contextes sociaux ou émotionnels importants.

Enfin, il est également possible de **participer de manière proactive à la lutte contre les contenus synthétiques et manipulés par l'IA en prenant part à des projets collaboratifs** tels que celui porté par Campus cyber, Viginum et la Région académique Île-de-France : «the Osint project». Ce dernier vise à créer un outil de détection des contenus générés par des IA (textes et images), utilisés à des fins de désinformation. Cet appel au grand public vise à recueillir l'expertise du plus grand nombre, dont les contributions peuvent être ajoutées à la [page suivante](#).

Quel impact environnemental pour l'IA ?

Si l'IA et le récent accroissement de ses puissances de calcul et de collecte des données sont envisagés par certains comme ayant un potentiel impact bénéfique pour l'environnement (ex : un impact sur les systèmes dédiés à l'environnement ou sur l'amélioration de l'efficacité énergétique des systèmes), il n'empêche que **l'IA consomme de grandes quantités de ressources qu'il s'agisse d'électricité, d'eau ou encore de métaux rares.**

Ainsi, si l'on manque de données récentes et fiables portant sur la consommation réelle de l'IA, l'Agence internationale de l'Énergie (AIE) prévoit néanmoins « une **multiplication par 10 de la consommation d'électricité du secteur de l'IA entre 2023 et 2026** »²³.

Face à cette projection, il est important de prendre conscience de l'impact environnemental de son utilisation de l'IA et de faire le choix d'une utilisation ciblée et raisonnée. Une ressource produite par le CNNum en partenariat avec plusieurs autres partenaires sera publiée prochainement et reviendra plus en détails sur ces questions.



²³ Pour une intelligence artificielle durable, Ministère de la transition écologique et Ministère de l'aménagement des territoires, 11 Février 2025.

Bibliographie

Rapports :

- Académie des Technologies,
[IA générative & mésinformation, 2025.](#)
- Arcom,
[Conclusions de la mission de l'Arcom sur l'impact de l'intelligence artificielle dans les domaines de la création et de l'information, 2024.](#)
- États généraux de l'information,
[Protéger et développer le droit à l'information : une urgence démocratique, 2024.](#)
- États généraux de l'information,
[Fiches thématiques produites par le groupe « Espace informationnel et innovation technologique des États généraux de l'information », 2024.](#)
- Viginum,
<https://www.sgdsn.gouv.fr/publications/defis-et-opportunités-de-l'intelligence-artificielle-dans-la-lutte-contre-les->

Sites web :

- AFP Factuel,
[ressources sur l'intelligence artificielles.](#)
- Newsguard,
[Centre de suivi de l'IA.](#)
[Les publications d'AI Forensics.](#)

