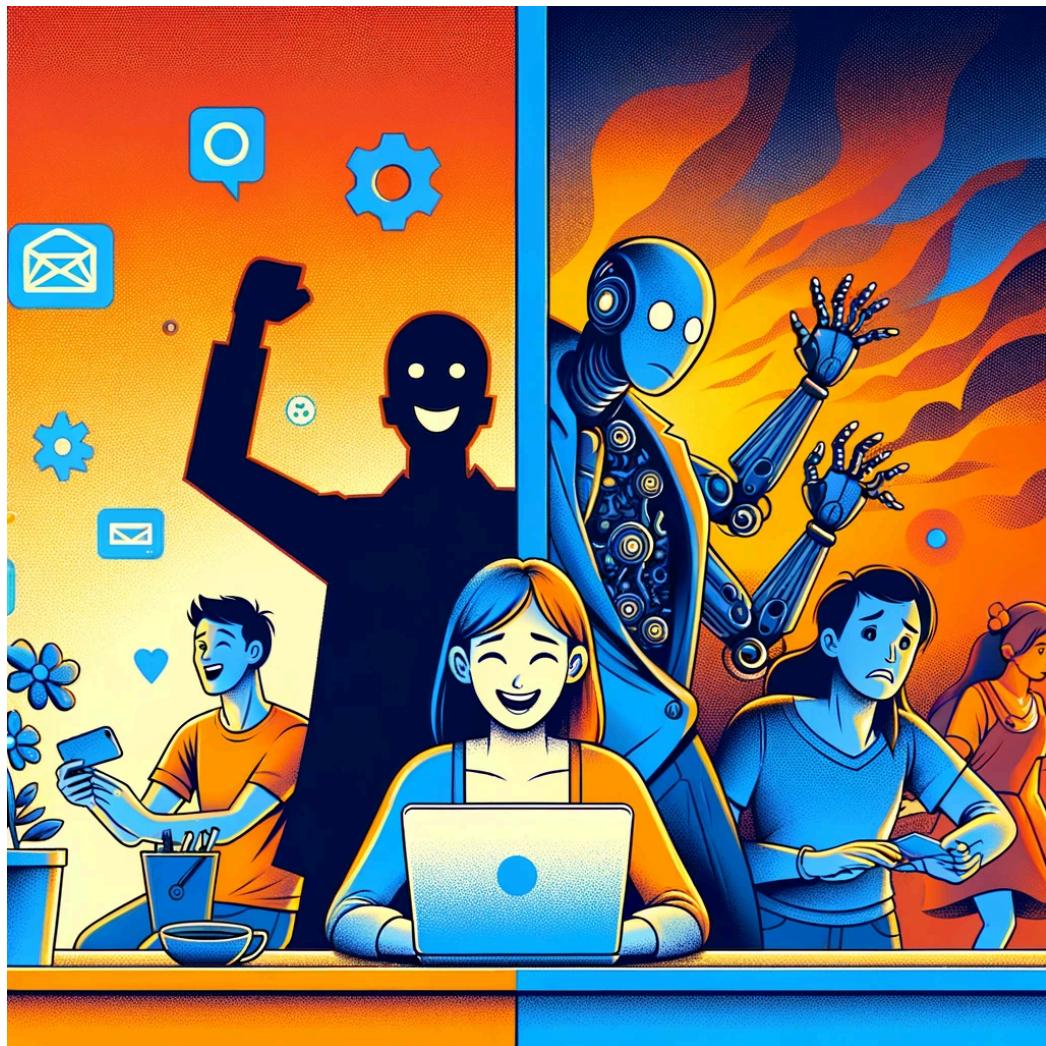


# ChatGPT : opportunité ou menace ?

22 Janvier 2024



*Image générée par ChatGPT 4*

Auteur·ices :

Anna ANTSIFOROVA, Layla AZZAM, Nathaniel COHEN, Guillaume DE ROCHEFORT, Nathan GASC, Luan Vitor MOREIRA DE OLIVEIRA, Marianne OBEID, Antoine SÉRANDOUR, Guillaume VIGNE

Encadrement :

Louison MAGRO, Morgan MEYER

Cette publication a été réalisée par des étudiants en troisième année du cycle ingénieur de Mines Paris PSL Research University. Il présente le travail réalisé dans le cours intitulé « Descriptions de controverse », qui a pour objectif d'introduire les étudiants à l'univers incertain de la recherche scientifique et technique et de les sensibiliser aux enjeux de la participation citoyenne.

Mines Paris décline toute responsabilité pour les erreurs et les imprécisions que peut contenir cet article. Vos réactions et commentaires sont les bienvenus. Pour signaler une erreur, réagir à un contenu ou demander une modification, merci d'écrire à la responsable de l'enseignement : [madeleine.akrich@mines-paristech.fr](mailto:madeleine.akrich@mines-paristech.fr).

# SOMMAIRE

|   |           |
|---|-----------|
| <b>Introduction.....</b>  | <b>4</b>  |
| <b>Partie 1 : Une technologie questionnée dès son origine.....</b>  | <b>5</b>  |
| L'origine du concept.....   | 5         |
| Le développement scientifique.....  | 6         |
| Les entreprises.....  | 8         |
| Les organisations civiles.....  | 9         |
| Les gouvernements.....  | 10        |
| <b>Partie 2 : Constructions de nouvelles normes sociales et législatives.....</b>                           | <b>11</b> |
| Droit d'auteur.....   | 11        |
| Construction législative : un droit nouveau.....  | 12        |
| L'AI Act européen.....  | 12        |
| Sécurité vs Créativité.....   | 14        |
| Deepfakes.....  | 14        |
| État législatif et comparaison avec d'autres pays.....  | 15        |
| <b>Partie 3 : Restructuration des rapports économiques.....</b>   | <b>16</b> |
| Générativité : débats sur les gains de productivité de ChatGPT et sur leur redistribution.....              | 16        |
| IA, éducation et enseignement supérieur : résistances dans les lieux de production de la connaissance.....  | 20        |
| Destruction créatrice : entre la crainte du chômage et la restructuration des emplois.....                  | 21        |
| La soutenabilité de l'IA dans un contexte de raréfaction des ressources et de transition écologique ?.....  | 23        |
| <b>Partie 4 : Débat public autour des risques.....</b>  | <b>26</b> |
| Des acteurs soucieux des risques existentiels.....  | 26        |
| Avis partagés sur les risques.....  | 27        |
| Les diverses arènes du débat.....   | 30        |
| Dénoncer la composition des "oppositions" .....   | 31        |
| Analyse académique des débats sur l'IA.....   | 33        |
| <b>Conclusion.....</b>  | <b>34</b> |
| <b>Méthodologie.....</b>  | <b>36</b> |
| <b>Bibliographie.....</b>   | <b>38</b> |
| Références de la littérature scientifiques, et presse experte.....  | 38        |
| Références de littératures grise académique : rapports et colloques.....                                    | 38        |
| Références de littératures grise : actes juridiques, législatifs et rapports administratifs.....            | 39        |
| Références dans la presse grand public.....   | 40        |
| Références de littérature grise : rapports de think tanks, communications d'organisations, réseaux sociaux. | 42        |

## Introduction

Dans le paysage en constante évolution de l'intelligence artificielle, l'émergence de ChatGPT a marqué un tournant significatif, avec plus d'un million d'utilisateurs enregistrés seulement six jours après son lancement<sup>1</sup>, et une audience mensuelle de 100 millions d'utilisateurs actifs après deux mois<sup>2</sup>. Cette croissance fulgurante illustre non seulement l'intérêt grandissant pour les technologies avancées d'IA, mais aussi la rapide intégration dans la vie de différents groupes d'individus, y compris ceux qui étaient auparavant non initiés à de telles innovations.

Dans ce contexte, il est primordial de comprendre ce qu'est ChatGPT, non seulement techniquement avec l'histoire de cette technologie, mais aussi en relation avec les divers acteurs (utilisateurs, législateur, économiques, artistiques, etc.). ChatGPT émerge dans le contexte du développement de l'intelligence artificielle et de l'émergence des modèles de traitement du langage naturel, dits LLM (large language model), qui automatisent la génération de contenu textuel. Ces technologies sont perçues et discutées aujourd'hui sous le terme d'IA générative, un terme qui porte différentes représentations. Quand certains parlent de produit, d'application, de service ou de support, d'autres parlent d'Intelligence artificielle générale (système capable d'une grande amplitude d'action), d'IA sûre et digne de confiance (système d'IA donnant des garanties en terme de transparence et de sécurité, relatifs à des risques établis par la législation, la recherche académique ou de la société civile) ou encore l'IA surhumaine (système dont les capacités dépassent celles d'un humain, notamment l'anticipation, la mémoire, etc.). De ces visions émergent des questionnements centraux sur l'intégration des technologies de ruptures au sein des sphères civiles et professionnelles de la société.

Comment ces groupes sociaux construisent-ils leur position sur l'IA générative et plus particulièrement sur ChatGPT et ses concurrents ? Comment s'articulent ces perceptions sur l'IA générative au sein de leurs activités ? Quelles questions sont soulevées à propos des perspectives sur l'intégration, l'impact et la régulation de ces technologies.

Nous nous attacherons à dégager une compréhension approfondie des interactions entre ces différents acteurs et de la manière dont leurs dialogues façonnent l'évolution de l'usage des IA génératives dans notre société, en revenant sur l'historique de cette technologie, pour ensuite introduire les enjeux juridiques puis économiques du développement de ces systèmes, et enfin terminer sur les débats de risques existentiels autour de ces questions.

---

<sup>1</sup> Sam Altman [@sama]. « ChatGPT Launched on Wednesday. Today It Crossed 1 Million Users! » Tweet. *Twitter*, 5 décembre 2022. Disponible sur <https://twitter.com/sama/status/1599668808285028353>. [Consulté le 20 janvier 2023]

<sup>2</sup> Hu, Krystal, et Krystal Hu. « ChatGPT Sets Record for Fastest-Growing User Base - Analyst Note ». *Reuters*, 2 février 2023, sect. Technology. Disponible sur <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>. [Consulté le 20 janvier 2023]

## Partie 1 : Une technologie questionnée dès son origine

### L'origine du concept

Les origines conceptuelles des agents conversationnels automatisés sont anciennes, du moins à l'échelle des évolutions récentes. Si la technologie ne les a rendus possibles que récemment, ils ont déjà largement été discutés par les scientifiques, artistes et sociologues au cours du XXe siècle. Cette discussion fut centrée, dès son origine, sur une peur de l'intelligence artificielle opposée au progrès de cette technologie.

Si on peut trouver des concepts liés à l'intelligence artificielle depuis des siècles, le concept moderne est d'abord popularisé par le roman *Metropolis* de Thea Von Harbou en 1926, et surtout par l'adaptation filmographique éponyme par Fritz Lang l'année suivante. Il est alors question de Futura, un être-machine de verre et de métal qui a pour but de remplacer l'humain. Le risque dystopique de la technologie est déjà un élément central de ces récits. En 1941, Isaac Asimov publie la nouvelle *Reason* comme partie du *Cycle des Robots*, dans laquelle il introduit un robot avec une certaine forme d'intelligence. Ce cycle est notamment connu pour ses *trois lois de la robotique* qui visent à protéger les humains face à ces technologies.

D'un point de vue scientifique, l'idée d'obtenir une machine intelligente à partir d'un ordinateur provient d'Alan Turing, qui comprend au lendemain de la Seconde Guerre mondiale, en fabriquant les premiers ordinateurs, que ceux-ci seront un jour capables de raisonner comme des humains<sup>3</sup>. Il construit alors le célèbre test de Turing<sup>4</sup> où un humain parle au travers d'un écran à deux interlocuteurs : un humain et un robot. Le test vérifie si l'humain parvient à distinguer la conversation avec l'humain de celle avec le robot, évaluant ainsi le degré de comportement humain d'une intelligence artificielle.

Cependant, des années 60 à la fin des années 1980, les résultats concernant l'intelligence artificielle se révèlent très décevants : si l'informatique fait d'immense progrès et devient un outil de masse, les performances restent insuffisantes pour construire des intelligences artificielles convaincantes. La pratique est loin d'égaler celles décrites par les nombreuses œuvres de science-fiction qui apportent le sujet au sein du débat public, critiquant particulièrement ses éventuelles conséquences dystopiques (*2001: L'odyssée de l'espace*, *Blade Runner*, *Westworld...*). Cette période sera par la suite qualifiée d'"hiver de l'IA"<sup>5</sup>.

---

<sup>3</sup> Alan Turing : la pensée informatique. Docsciences n°14, Canopé (CRDP de l'académie de Versailles), 2012, p.5 [Consulté le 14/01/2023]

<sup>4</sup> Turing, Alan. « I.-COMPUTING MACHINERY AND INTELLIGENCE ». *Mind* LIX, n° 236 (octobre 1950): 433-60. <https://doi.org/10.1093/mind/LIX.236.433>. [Consulté le 14/01/2023]

<sup>5</sup> Bostrom, Nick. « Seasons of Hope and Despair ». In *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press., 25:285-89, 2015. <https://doi.org/10.1007/s11023-015-9377-7>. [Consulté le 14/01/2023]

## Le développement scientifique

À partir des années 1990 et 2000, de nouveaux progrès rendus possibles par l'effondrement exponentiel du prix de la puissance de calcul créent un renouveau<sup>6</sup>. En particulier, les réseaux de neurones convolutionnels se développent et montrent des résultats prometteurs sur des tâches jusqu'ici réservées aux humains (reconnaissance de caractère, classification...). Les financements augmentent<sup>7</sup> et la massification de l'usage d'Internet crée énormément de données qui peuvent être exploitées pour entraîner ces réseaux de neurones (méthode dite *d'apprentissage profond*).

Enfin, en 2017, une équipe de chercheurs de chez Google publie le papier *Attention is all you need*, qui introduit une nouvelle architecture de réseau de neurones : le transformer. Cette nouvelle architecture se révèle très efficace pour la prédiction de texte, et au-delà de ça, mener des raisonnements pour produire des résultats cohérents grâce à une analyse sémantique avancée : la porte des modèles de langages génératifs est ouverte. Parce que ces modèles produisent des textes très crédibles, la crainte des intelligences artificielles refait surface à la même période, par exemple avec les nombreux appels à stopper leur développement<sup>8</sup>.

Dans ce contexte scientifique, Open AI , une entreprise de recherche en intelligence artificielle, voit le jour à la fin de l'année 2015. La structure, qui apparaît initialement sous la forme d'association non lucrative, a été créée à l'initiative d'Elon Musk et de Sam Altman<sup>9</sup>. L'équipe fondatrice a combiné ses diverses compétences en entrepreneuriat technologique, apprentissage automatique et ingénierie logicielle pour réaliser des avancées majeures dans le domaine de l'intelligence artificielle, ayant pour objectif "d'avancer l'intelligence numérique de sorte à bénéficier à l'humanité entière, sans contraintes financières."<sup>10</sup>. À l'association, à l'origine créée pour s'assurer qu'aucune IA négative n'émerge en créant elle-même une IA bénéfique pour l'humanité, s'ajoute une nouvelle structure à but lucratif pour pouvoir recueillir les investissements à hauteur de 1 milliard de dollars de Microsoft en 2019<sup>11</sup>.

<sup>6</sup> World Intellectual Property Organization, « WIPO Technology Trends Artificial Intelligence 2019 ». Genève, 2019. 158 pages, Disponible sur <https://www.wipo.int/publications/en/details.jsp?id=4386>. [Consulté le 21 décembre 2023.]

<sup>7</sup> Wikipédia « Histoire de l'intelligence artificielle », 16 janvier 2024. Disponible sur [https://fr.wikipedia.org/w/index.php?title=Histoire\\_de\\_l%27intelligence\\_artificielle&oldid=211570802](https://fr.wikipedia.org/w/index.php?title=Histoire_de_l%27intelligence_artificielle&oldid=211570802). [Consulté le 20 janvier 2023]

<sup>8</sup> Le Monde.fr. « Hawking : "L'intelligence artificielle pourrait mettre fin à l'humanité" ». 3 décembre 2014. [https://www.lemonde.fr/pixels/article/2014/12/03/hawking-l-intelligence-artificielle-pourrait-mettre-fin-a-l-humanite\\_4533135\\_4408996.html](https://www.lemonde.fr/pixels/article/2014/12/03/hawking-l-intelligence-artificielle-pourrait-mettre-fin-a-l-humanite_4533135_4408996.html). [consulté le 21/12/2023]

<sup>9</sup> Sauvage, M. Manuel S. *La brève histoire de ChatGPT- Tout savoir sur les fondements de ChatGPT et son avenir.* Independently published, 2023. [consulté le 21/12/2023]

<sup>10</sup> OpenAI. « Introducing OpenAI ». [Https://openai.com/](https://openai.com/). Disponible sur <https://openai.com/blog/introducing-openai>. [Consulté le 22 décembre 2023.]

<sup>11</sup> Piquard, Alexandre. « OpenAI : aux sources de l'affaire Sam Altman ». Le Monde.fr, 24 novembre 2023. [https://www.lemonde.fr/economie/article/2023/11/24/openai-aux-sources-de-l-affaire-sam-altman\\_6202096\\_3234.html](https://www.lemonde.fr/economie/article/2023/11/24/openai-aux-sources-de-l-affaire-sam-altman_6202096_3234.html). [consulté le 21/12/2023]

L'entreprise de la Silicon Valley introduit les GPT, acronyme qui signifie "Generative Pre-trained Transformer" (Transformer génératif pré-entraîné). Le nom du modèle souligne sa capacité à *prolonger* des textes de manière crédible grâce à un entraînement sur un très grand nombre d'autres textes, ainsi que son architecture basée sur les transformers. À partir de 2015, OpenAI progresse à travers différentes itérations<sup>12</sup> de la série GPT<sup>13</sup> :

- GPT-1 (2018) a démontré la puissance de la technologie sur des exemples simples.
- GPT-2 (2019) a amélioré considérablement la capacité de création de textes longs. La mise à disposition initiale au public a été retardée en raison de craintes d'utilisation abusive.
- GPT-3 (2020) a représenté un saut colossal: ses compétences avancées de génération de texte l'ont rendu polyvalent dans diverses applications (réécriture de textes, programmation, traduction...).
- GPT-4 (2023) poursuit cette trajectoire d'amélioration avec un effort particulier sur la réduction de la probabilité de résultats offensants ou dangereux, une plus grande exactitude des faits; ces avancées font suite aux critiques reçues par les précédents modèles et les inquiétudes liées à leur utilisation malveillante.

### Agents conversationnels

Le lancement de ChatGPT a marqué un tournant en offrant au monde une interaction directe entre l'utilisateur et le modèle GPT-3, et plus tard GPT-4. C'est un prototype de *chatbot* (agent conversationnel) capable de comprendre le langage humain naturel et de générer un dialogue textuel très crédible. OpenAI a (et continue de) développé ChatGPT en intégrant les interactions avec les utilisateurs (retour sur expérience).

Notons qu'aujourd'hui OpenAI n'est plus l'unique acteur du domaine : de nombreux autres acteurs n'ont pas attendu les premiers succès publics d'OpenAI pour développer d'autres modèles, dont la variété actuelle illustre les différentes attentes : sécurité, indépendance, publicité... D'autant plus que certains employés d'OpenAI ont rejoint les rangs d'autres entreprises (Google Deepmind, Anthropic, MistralAI...) suite à des divergences stratégiques.

Les agents conversationnels sont l'héritage d'une longue histoire technologique et les dépositaires d'une attente construite depuis près d'un siècle par un imaginaire de science-fiction. L'exploit technologique n'a d'égal que les craintes qu'il cause, et les bouleversements de nos sociétés illustrent la quantité de problèmes à

<sup>12</sup> Botpress Blog. « Comment fonctionne ChatGPT ? », 26 avril 2023. Disponible sur <https://botpress.com/fr/blog/how-does-chatgpt-work>. [consulté le 15/12/23]

<sup>13</sup> Lastmann, Samy. « Qu'est-ce que ChatGPT, l'IA conversationnelle d'OpenAI ? Définition, contexte et enjeux (2023) ». *Smart tribune* (blog), 22 décembre 2022. <https://blog.smart-tribune.com/fr/ia-conversationnelle-chatgpt-openai-definition>. [consulté le 16/12/2023]

résoudre. Pour comprendre l'ensemble de telles problématiques, cependant, un regard global s'impose pour appréhender l'ensemble des acteurs : les forces politiques, économiques et sociales émergentes qui jouent également un rôle déterminant dans l'évolution de l'IA générative.

À cet égard, trois types d'acteurs clés peuvent être identifiés : les entreprises, engagées dans une véritable course pour assimiler les bénéfices de cette révolution ; les organisations civiles préoccupées par un tel développement, notamment les centres de recherche et les ONG ; et enfin les gouvernements, qui cherchent à bénéficier à la fois la société et l'industrie, influençant les réglementations et élaborant des stratégies d'intégration et de développement.

### **Les entreprises**

Dans le domaine industriel de l'intelligence artificielle, deux grands rôles se dessinent dans une véritable course à l'adoption : d'une part, les entreprises axées sur la création de modèles et de technologies (IA en tant que produit) ; et d'autre part, celles cherchant à accroître la productivité de leur personnel grâce à l'intégration de l'IA dans leur stratégie globale.

Du côté de l'IA en tant que produit, on observe d'un côté les géants du secteur du développement tels que Meta, Google et OpenAI, et de l'autre, des startups émergentes telles que Mistral AI, Xai et Inflection. Tous ces acteurs construisent des modèles de langage exploitant des transformers et des structures sophistiquées pour atteindre des performances optimales selon différents critères. Ces acteurs poussent le développement de l'IA, Yann LeCun, directeur de recherche chez Meta, le défend : "L'idée même de vouloir ralentir la recherche sur l'IA s'apparente à un nouvel obscurantisme".<sup>14</sup>

D'un autre côté, on retrouve des entreprises dont l'objectif principal n'est pas la commercialisation de la technologie, mais plutôt leur stratégie, à partir de l'adoption, et parfois la création, d'outils sur mesure pour leurs besoins internes, favorisant ainsi l'efficacité de leurs équipes. Ces entreprises privilégient le développement d'outils basés sur des bases de données partagées tout en préservant la confidentialité de leurs données. Diverses stratégies ont émergé dans cette catégorie, telles que le partenariat entre Bain & Co et OpenAI, ou le développement interne de DiPA par Accenture India, ces technologies visent à augmenter la

---

<sup>14</sup> Legros, Claire. "Yann Le Cun, directeur à Meta : « L'idée même de vouloir ralentir la recherche sur l'IA s'apparente à un nouvel obscurantisme ». *Le Monde.fr*, 28 avril 2023.

[https://www.lemonde.fr/idees/article/2023/04/28/yann-le-cun-directeur-a-meta-l-idee-meme-de-vouloir-ra lentir-la-recherche-sur-l-ia-s-apparente-a-un-nouvel-obscurantisme\\_6171338\\_3232.html](https://www.lemonde.fr/idees/article/2023/04/28/yann-le-cun-directeur-a-meta-l-idee-meme-de-vouloir-ra lentir-la-recherche-sur-l-ia-s-apparente-a-un-nouvel-obscurantisme_6171338_3232.html). [consulté le 21/12/2023]

productivité des consultants<sup>15,16</sup>. Cette diversité de stratégies témoigne de l'adaptabilité du secteur de l'IA aux besoins variés des entreprises, qu'elles soient orientées vers le développement technologique ou la recherche d'efficacité opérationnelle.

### Les organisations civiles

Comme conséquence de ce scénario, les organisations civiles expriment leur inquiétude. Parmi les principaux acteurs identifiés, se distinguent des institutions telles que l'Institut Montaigne, un think-tank de référence en France et en Europe. L'institut considère les risques existentiels liés à l'IA, tels que les usages malveillants et le contrôle d'outils autonomes, comme une opportunité économique<sup>17</sup>. Il soutient que le leadership français dans la recherche, l'industrie et les compétences pourrait répondre au "*trilemme entre course technologique, course à la gouvernance et principe de précaution*"<sup>18</sup>. Des institutions de recherche telles que Redwood Research et EffiSciences, avec lesquelles des échanges ont pu être réalisés au cours du développement de cet article, ont également des opinions divergentes et ont démontré des approches variées.

D'un côté, Redwood Research est une petite structure engagée depuis trois ans dans des recherches empiriques en IA, avec un intérêt particulier pour les débats sur la généralisation des IA, la sécurité et les risques existentiels associés à une éventuelle superintelligence. De son côté, EffiScience adopte une approche pratique de sensibilisation et de formation. Cette approche prend en compte l'impact à court terme de l'IA sur la société, y compris les changements majeurs dans le secteur du travail.

---

<sup>15</sup> Accenture,. « DiPA | Innovative Chatbot ». Consulté le 17 janvier 2024. Disponible sur <https://www.accenture.com/fr-fr/case-studies/about/need-answers-ask-dipa>. [consulté le 17/01/2024]

<sup>16</sup> Bain. « Bain & Company fait alliance avec les services d'OpenAI afin d'aider ses clients à exploiter le plein potentiel de l'IA », 21 février 2023. Disponible sur <https://www.bain.com/fr/a-propos-de-bain/media-center/communiques-de-presse/2023/bain--company-announces-services-alliance-with-openai-to-help-enterprise-clients-identify-and-realize-the-full-potential-and-maximum-value-of-ai/>. [Consulté le 17/01/2024]

<sup>17</sup> Rignell, Milo. « Investir l'IA sûre et digne de confiance : un impératif européen, une opportunité française ». Institut Montaigne, avril 2023. Disponible sur <https://www.institutmontaigne.org/publications/investir-lia-sure-et-digne-de-confiance-un-imperatif-europeen-une-opportunité-française>. [Consulté le 19/01/2023]

<sup>18</sup> Rignell, Milo. « IA : l'ambivalence entre course technologique et gouvernance mondiale ». *Institut Montaigne* (blog), 20 juillet 2023. Disponible sur <https://www.institutmontaigne.org/expressions/ia-lambivalence-entre-course-technologique-et-gouvernance-mondiale> [Consulté le 19/01/2023]

## Les gouvernements

Finalement, la réponse gouvernementale est incisive. Des efforts considérables sont déployés pour réunir tous les acteurs autour de la même table et dissuader les pratiques qui pourraient compromettre les avantages généraux de l'IA. Ainsi, les aspects de régulation et de stratégie liés à l'IA s'influencent mutuellement de manière évidente. Bien que les organismes de régulation convergent vers la nécessité de réglementer l'IA pour assurer sa sûreté et sa transparence, les approches varient en fonction des priorités nationales.

Nous avons regroupé une liste non exhaustive des acteurs de la controverse dans la figure ci-dessous. Cette controverse se caractérise par la pluralité des acteurs, et leur nombre.



Panorama des acteurs de l'IA abordé dans cette étude

## Partie 2 : Constructions de nouvelles normes sociales et législatives

### Droit d'auteur

Comme évoqué plus haut, la technologie des IA génératives actuelle s'entraîne sur d'immenses jeux de données afin d'être capable in fine de produire des réponses –textuelles et graphiques– plausibles lors de l'interaction avec l'utilisateur final. Cependant, quelles sont les données utilisées pour l'entraînement ? Dans quelle mesure sont-elles utilisées ? Ces questions aux réponses opaques restent un point chaud des négociations pour l'établissement d'une législation cohérente. Prenons l'exemple d'un des premiers procès qui a fait référence aux États-Unis : Andersen vs Stability AI LTD. De nombreux artistes ont attaqué l'entreprise, car un de ses modèles de génération d'image aurait utilisé des images soumises au droit d'auteur des artistes pour être entraînés. La cour de Californie a jugé que la charge de la preuve incombe à l'artiste lésé : c'est une victoire pour l'intelligence artificielle. En pratique, montrer un tel lien est quasiment impossible tant que les entreprises n'ont pas à rendre public les jeux de données dont ils se sont servis<sup>19,20</sup>. Toutefois, cette position n'est pas si claire en Union Européenne où l'IA Act (texte de loi en cours de rédaction depuis 2021 pour réglementer les développements de tous les systèmes d'IA en Europe) se contente de demander aux entreprises qui développent ces modèles de signaler si du contenu soumis au droit d'auteur a été utilisé pour entraîner le modèle.<sup>21</sup>

Un autre point chaud du débat autour de l'IA et du droit d'auteur est de savoir si les créations des LLMs sont des œuvres de l'esprit et peuvent être soumises au droit d'auteur. Pour les États-Unis la réponse est claire : non. La décision a de plus été entérinée au niveau fédéral par des ordonnances du président Biden.<sup>22</sup> Pour l'Union Européenne, cela reste un point aveugle de l'IA Act qui ne se prononce pas sur la question.<sup>23</sup> Cependant, la clause récente ajoutée à l'IA Act et qui stipule que le statut de droit d'auteur des

---

<sup>19</sup> *Andersen v. Stability AI Ltd.*, No. 23-cv-00201-WHO (United States District Court, Northern District of California, 30 octobre 2023). Disponible sur <https://casetext.com/case/andersen-v-stability-ai-ltd> [consulté le 17/01/2023]

<sup>20</sup> Cho, Winston. « Artists Lose First Round of Copyright Infringement Case Against AI Art Generators ». *The Hollywood Reporter*, 30 octobre 2023. Disponible sur <https://www.hollywoodreporter.com/business/business-news/artists-copyright-infringement-case-ai-art-generators-1235632929/>. [Consulté le 17/01/2024]

<sup>21</sup> Quach, Katyanna. « EU Legislates Disclosure of Copyright Data Used to Train AI ». *The Register* (blog), 1 mai 2023. Disponible sur [https://www.theregister.com/2023/05/01/eu\\_ai\\_act\\_adds\\_new/](https://www.theregister.com/2023/05/01/eu_ai_act_adds_new/). [Consulté le 17/01/2023]

<sup>22</sup> Maison Blanche, *Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*, No. 14110, EO (2023). Disponible sur <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>. [Consulté le 20/01/2024]

<sup>23</sup> Hugenholtz, P. Bernt, et João Pedro Quintais. « Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output? » *IIC - International Review of Intellectual Property and Competition Law* 52, n° 9 (1 octobre 2021): 1190-1216. <https://doi.org/10.1007/s40319-021-01115-0>. [Consulté le 10/01/2024]

données d'entraînement doit être affiché laisse penser qu'à terme le statut en termes de droit d'auteur des œuvres générées va dépendre du statut des données d'entraînement.

### **Construction législative : un droit nouveau**

À l'échelle des temps de construction des normes législatives, qui s'étendent parfois sur plusieurs années, ChatGPT a apporté des modifications très rapides à la société. Par conséquent, les organes législatifs des pays ont subi une forte pression pour établir rapidement un cadre clair qui permette à la fois de protéger le citoyen et de favoriser le développement des technologies liées aux LLMs. En particulier, les pays de droit *romano-civiliste* (système qui considère que le droit écrit est la principale source de loi, l'Europe et l'Asie notamment) ont dû construire (et construisent encore) des normes pour correspondre aux évolutions récentes. C'est le cas de l'*IA Act* que nous allons traiter comme exemple : un projet de règlement européen qui vise à encadrer les pratiques du secteur et qui avait commencé dès 2021, avant la publication de ChatGPT donc.

#### **Explications : Common Law & Droit romano-civiliste**

Le système de droit Anglo-Saxon, la *Common Law*, considère à l'inverse que la jurisprudence est un facteur important de la loi. Les juges exercent donc un pouvoir politique quand une nouvelle décision apparaît. Dans le cas de l'IA, cela se traduit par l'existence de nombreux procès, dont l'issue dictera la loi *de facto*.

### **L'AI Act européen**

Dès 2021, un projet de règlement européen, nommé IA Act est apparu pour prévenir des dangers existants et à venir des intelligences artificielles. Si le projet n'avait pas dans son viseur les LLMs dans un premier temps (puisque ceux-ci n'étaient alors pas accessibles au grand public), il a dû évoluer pour les prendre en compte.<sup>24</sup> Toutefois, notons que l'IA Act n'est pas à proprement parler le premier texte de la commission européenne sur l'intelligence artificielle, puisque deux directives ont été publiées dès 2022 pour couvrir de manière générale certains dommages causés par l'IA : extension de la responsabilité des logiciels défectueux aux IA et extension de la responsabilité des constructeurs d'IA aux dommages causés par leurs créations.<sup>25</sup>

<sup>24</sup> Coulter, Martin. « What Is the EU AI Act and When Will Regulation Come into Effect? » *Reuters*, 7 décembre 2023, sect. Technology. Disponible sur <https://www.reuters.com/technology/what-are-eus-landmark-ai-rules-2023-12-06/>. [Consulté le 17/01/2024]

<sup>25</sup> Lemarchand, Stéphane, Jeanne Dozier, Aurelia Pons, et Maria Aouad. « La Commission européenne publie deux propositions de directive en matière de responsabilité applicables aux produits basés sur l'intelligence artificielle ». *DLA*

### Explications : Directives et règlements

On distingue deux types de textes communautaires (issus de l'UE) : les *directives* qui sont des textes qui fixent des objectifs et qui doivent être traduits dans les différents pays de l'union, et les *règlements* qui sont des textes législatifs contraignants, demandant plus de travail et de consensus pour être adoptés.<sup>26</sup>

Le contenu de l'IA Act est dense, mais retenons qu'il met en place un système de classification des risques basé sur l'utilisation qui est faite de la technologie. Ainsi, une IA utilisée pour aider les forces de l'ordre sera soumise à une évaluation approfondie avant d'obtenir une autorisation d'utilisation, alors que des IA de détection de spams seront considérées comme ne nécessitant pas d'évaluation. ChatGPT tombe alors dans une catégorie intermédiaire, créée tardivement pour justement prendre en compte ces nouveaux modèles génératifs, nommée *General purpose and generative AI* (IA à usage général et génératif). Les IA de cette catégorie subiront également une évaluation et devront faire des rapports en cas d'incident.<sup>27</sup>

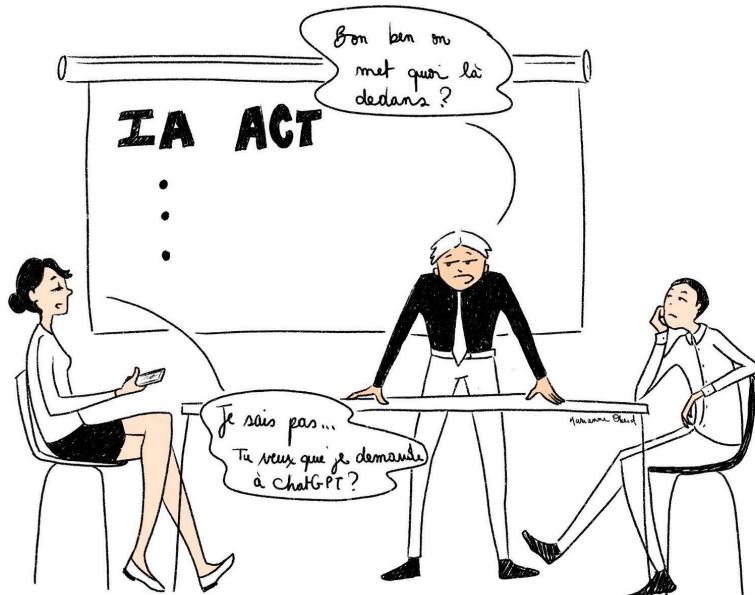


Illustration : Marianne OBEID

Piper (blog), 15 février 2023. Disponible sur <https://www.dlapiper.com/fr-fr/insights/publications/bulletins-contrats-commerciaux/2023/contrats-commerciaux-bulletin-1/la-commission-europeenne-publie-deux-propositions-de-directive-en-matiere-de-responsabilite>. [Consulté le 17/01/2024]

<sup>26</sup> Union européenne, « Types de législation ». s. d. European-union.europa.eu. [https://european-union.europa.eu/institutions-law-budget/law/types-legislation\\_fr](https://european-union.europa.eu/institutions-law-budget/law/types-legislation_fr). [Consulté le 17 janvier 2024]

<sup>27</sup> Parlement européen, « EU AI Act: First Regulation on Artificial Intelligence ». 8 juin 2023. <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>. [Consulté le 17/01/2024]

## Sécurité vs Créativité

Les deux principales forces qui tentent d'influer sur la rédaction de l'IA Act sont d'un côté le législateur qui vise à protéger le citoyen, mais qui ne dispose pas toujours d'une grande expertise ou retour d'expérience, et de l'autre des entreprises qui cherchent à disposer d'un cadre libre pour développer et tester leur technologie librement. Pour cette raison, les entreprises qui développent ces technologies ont mené une campagne de lobbying importante dès les premières discussions autour de l'IA Act, avant l'ouverture au grand public de ChatGPT fin 2022.<sup>28</sup> Cela n'a pas empêché le projet de loi de l'AI Act d'être qualifié comme un des plus stricts au monde, surtout au regard de certaines interdictions (reconnaissance faciale...); de nombreux acteurs craignant que cette législation contraignante étouffe dans l'œuf les jeunes entreprises européennes voulant chasser sur les terres des désormais géants de l'IA, américains.<sup>29</sup>

## Deepfakes

L'IA subit aussi des critiques en raison de son incomplétude et de la difficulté d'établir à l'avance un texte sur des technologies qui n'existent pas aujourd'hui. Par exemple, la classification est parfois remise en cause, en particulier la classification des *deepfakes* (*litt.* faux-profound, faux générés par des IAs) comme étant une technologie à faible risque est considéré par certains comme problématique, notamment, car les technologies de détection ne sont pas encore au point aujourd'hui.

« *Les produits de ChatGPT vont devenir beaucoup plus difficiles à détecter [...] c'est l'un des aspects les plus problématiques de ces technologies* » Entretien OCDE. Traduit depuis l'anglais.

Aussi, ces produits sont craints pour des raisons d'influence, de manipulation et de droits d'auteur : les logiciels qui permettent d'imiter la voix d'une personne à partir d'un enregistrement de sa voix sont très répandus et faciles d'accès, malgré les législations qui considèrent que l'utilisation d'une voix, photo ou vidéo d'une personne relève d'une violation au droit d'auteur (en France surtout). Aux États-Unis, l'affaire *DeepNude* où un logiciel fondé sur l'IA permettait de dévêter une personne sur une photo avait servi d'électrochoc au Congrès, qui a depuis pris des mesures pour limiter ces usages ; même si la régulation demeure en pratique difficile. En Union Européenne cependant, ce type d'usage n'est couvert que par le droit d'auteur, ce qui pousse une demande pour une régulation plus stricte, demande qui n'est à ce jour pas

<sup>28</sup> Perrigo, Billy. 2023. « Big Tech Is Already Lobbying to Water Down Europe's AI Rules ». *TIME*. 21 avril 2023. <https://time.com/6273694/ai-regulation-europe/>. [Consulté le 17/01/2024]

<sup>29</sup> Bourgeot, Rémi. 2023. « AI Act : quel compromis pour permettre l'essor de l'intelligence artificielle européenne ? » *IRIS*. 18 décembre 2023.

<https://www.iris-france.org/180995-ai-act-quel-compromis-pour-permettre-l-essor-de-l-intelligence-artificielle-europeenne/>. [Consulté le 17/01/2024]

satisfait par l'IA Act. A l'inverse, certains acteurs de l'intelligence artificielle craignent qu'un trop fort contrôle des données utilisées ne conduise à une limitation des possibilités de développement des IA en Europe.<sup>30,31</sup>

### État législatif et comparaison avec d'autres pays

L'IA Act n'est aujourd'hui pas encore promulguée, et après le vote des différents organes de l'UE, il faudra probablement attendre plusieurs années avant qu'il ne fasse force de loi. Or certains acteurs considèrent que le texte est déjà dépassé par certaines technologies (dont les *deepfakes* mentionnés ci-dessus), il faudra donc probablement des ajustements pour s'adapter aux technologies à venir. C'est ici un avantage de l'approche jurisprudentielle des pays anglo-saxons : la loi se construit en fonction des besoins législatifs ; ce qui n'a pas empêché les États-Unis de passer certains textes de manière préventive. Au Japon et en Corée du Sud, le sujet n'est pas installé dans le débat public, et ces pays ont des approches très souples vis-à-vis de ces technologies : en particulier, il y a peu de contraintes sur les données utilisées, tant d'un point de vue du droit d'auteur que de la sécurité personnelle.<sup>32</sup>

---

<sup>30</sup> Fernandez, Angelina. 2022. « Regulating Deep Fakes in the Proposed AI Act ». *MediaLaws* (blog). 23 mars 2022. <https://www.medialaws.eu/regulating-deep-fakes-in-the-proposed-ai-act/>. [Consulté le 17/01/2024]

<sup>31</sup> Allain, Tanguy. 2023. « Quel cadre juridique pour les deepfake audio ? » *Tnjlex - cabinet d'avocat*. 2 mai 2023. <https://tnjlex-avocat.fr/cadre-juridique-deepfake/>. [consulté le 17/01/2024]

<sup>32</sup> Campagnolo, Côme. 2023. « Corée du Sud : état des lieux et stratégie en matière d'IA ». *Direction générale du Trésor* (blog). 31 août 2023. Disponible sur <https://www.tresor.economie.gouv.fr/Articles/2023/08/31/coree-du-sud-etat-des-lieux-et-strategie-en-matiere-d-ia>. [consulté le 17 janvier 2024]

## Partie 3 : Restructuration des rapports économiques

### Générativité : débats sur les gains de productivité de ChatGPT et sur leur redistribution

En avril 2023, la société de services financiers et bancaires Goldman Sachs publiait une étude sur les conséquences de l'introduction des intelligences artificielles génératives dans l'économie avec un accroissement de 7% du produit intérieur brut mondial (PIB). La même étude annonce en même temps un potentiel d'automatisation de 300 millions d'emplois<sup>33</sup>.

L'IA générative est perçue aujourd'hui comme un outil permettant de gagner en productivité, ce qui génère une forte demande parmi les acteurs économiques. Encore peu d'études scientifiques ont été publiées jusqu'à présent en raison de la jeunesse du sujet, mais une première publication dans la revue *Science*, datant de juillet 2023, rapportant les résultats d'une expérience réalisée sur un échantillon de 435 professionnels exposés à Chat-GPT, semble confirmer cette thèse avec une réduction du temps de travail de 40% et une augmentation de la qualité de travail de 18%<sup>34</sup>. C'est ainsi qu'émerge une certaine forme de pragmatisme sur l'intégration de l'IA au travail à laquelle peu de monde pourra se soustraire à l'avenir.

Ainsi, la demande de formation en IA en tendance de recherche web est d'ailleurs fortement corrélée à l'évolution de la recherche du terme ChatGPT d'après une analyse sur Google trend. Les déclins de recherche sont visiblement saisonniers puisqu'ils sont apparus sur la période estivale (juin-septembre 2023) et les fêtes de fin d'années (novembre-décembre 2023) comme l'analyse le média *le Grand Continent*<sup>35</sup> en septembre 2023 ce qui soutient la conclusion sur l'adoption durable de ChatGPT dans les méthodes de travail malgré la concurrence des autres agents conversationnels.

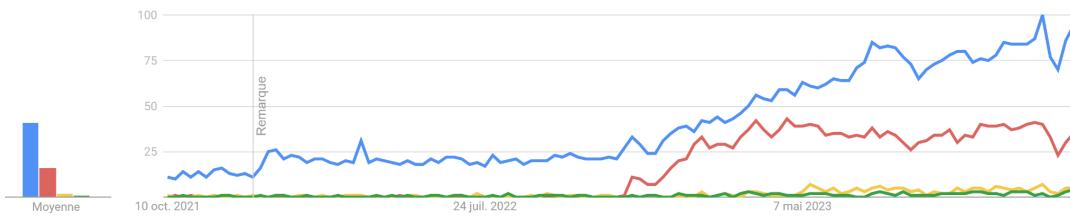
---

<sup>33</sup> Goldman Sachs, « Generative AI Could Raise Global GDP by 7% ». 2023. 5 avril 2023.

<https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>. [Consulté le 14/01/2024]

<sup>34</sup> Noy, Shakked, et Whitney Zhang. « Experimental evidence on the productivity effects of generative artificial intelligence ». *Science* 381, n° 6654 (14 juillet 2023): 187-92. <https://doi.org/10.1126/science.adh2586>.

<sup>35</sup> Le Grand Continent, « ChatGPT perd des utilisateurs et annonce une innovation majeure » 26 septembre 2023. Disponible sur <https://legrandcontinent.eu/fr/2023/09/26/chatgpt-perd-des-utilisateurs-et-annonce-une-innovation-majeure/>. [Consulté le 10/01/2024]



*Analyse Google trend , bleu : "AI training", rouge : "ChatGPT training", jaune : "bard training", vert : "copilot training"*

Malgré cet engouement, la faculté de ChatGPT à rendre plus productif reste questionnée parmi les experts. La productivité est un terme d'économie faisant état du rapport entre la valeur produite par le travail et la valeur des capitaux engagés pour le réaliser<sup>36</sup>. Gagner en productivité signifie donc être en mesure de réduire les capitaux engagés pour produire de la valeur ou bien d'augmenter la production pour un même capital engagé. En ce sens, ChatGPT s'inscrit pleinement dans les effets de la transition numérique en diminuant le temps consacré à certaines tâches et en diminuant le capital à investir. Toutefois, là où la digitalisation des services avait principalement des effets d'accélération des communications et de réduction des capitaux de gestion de l'information, l'agent conversationnel prend en charge la création du contenu ce qui affecte d'une part la rapidité de production tout en conservant *a priori* la valeur qualitative du contenu généré.

Sur le plan de l'accélération du traitement de l'information, le secteur juridique est emblématique puisqu'il s'intéresse depuis plusieurs années aux technologies de traitement de texte et d'intelligence artificielle<sup>37</sup>. Avec des entreprises spécifiques, les *legaltech*, la justice avait déjà testé l'IA prédictive pour la classification et l'orientation des dossiers de justice. L'arrivée des agents conversationnels permet désormais aux professionnels de la justice et de la loi d'explorer encore de nouvelles applications, comme ils ont pu en témoigner par exemple lors du séminaire *Dauphine Digital Days* organisé par l'université Dauphine-PSL. Alexandre Verney, Directeur du programme *Nouveaux outils de production normative*, au Secrétariat Général du Gouvernement y expliquait les usages potentiels de l'IA pour la production des textes législatifs et leur application concrète sur la consolidation de la loi. Il présentait notamment un projet du gouvernement pour qu'un outil d'intelligence artificielle aide les juristes responsables de la plateforme *Légifrance* à adapter plus rapidement les ressources de la plateforme à la suite d'une réforme.

<sup>36</sup> INSEE, « Définition - Productivité ». 2016. *INSEE* (blog). 13 octobre 2016.

<https://www.insee.fr/fr/metadonnees/definition/c1452>. [Consulté le 10/01/2024]

<sup>37</sup> Université Paris Dauphine - PSL. 2023. « Quelle utilisation de l'IA par les juristes ? » [Vidéo en ligne][2 h 46 mn 7 s].

Coord. Thierry Kirat. <https://www.youtube.com/watch?v=Jz4snai1Bvs>. [consulté le 10/01/2024]

Accélérer le traitement de contenu est une première façon d'utiliser l'intelligence générative, mais l'IA aide-t-il à créer un contenu neuf et innovant, créateur de valeur ? Cette question semble concerner davantage les secteurs de la création artistique, la production médiatique, de l'analyse et de la stratégie, et enfin de la recherche. Comment reconnaître un tel effet de l'IA sur l'originalité des contenus produits et ceux-ci sont-ils acceptés ?

Les professions artistiques ont été les plus critiques dans les médias sur les effets de l'IA dans leur métier, tout d'abord sur l'aspect du droit d'auteur, comme évoqué dans la section *droits d'auteur* de la partie 2 mais aussi sur le renforcement de la concurrence entre eux puisque les accès à des contenus originaux et inspirants semblent s'élargir, à coût de formation réduit. En revanche, l'acceptabilité de telles productions dans des milieux très concurrentiels fait encore débat. Pourtant, certains artistes utilisent de façon très assumée les intelligences artificielles comme Paul Rouston, photographe professionnel qui a utilisé l'IA Mid Journey pour un projet d'affiche de l'édition 2024 du tournoi Roland-Garros et qui a été présentée le 20 décembre dernier suscitant de vives critiques chez certains de ces confrères sur le plan artistique en qualifiant son projet de "pastiche impressionniste"<sup>38</sup>. Paul Rouston s'est défendu en invoquant le désir d'explorer les capacités de l'IA sans idées préconçues.

Néanmoins, les professionnels de la littérature semblent saluer certaines vertus de l'agent conversationnel comme Alexandre Gefen, critique littéraire qui s'est exprimé en décembre 2023 à l'ENSSIB, Ecole nationale supérieures des sciences de l'information et des bibliothèques<sup>39</sup> qui saluait la démocratisation de l'accès à l'exercice d'écriture tout en relayant les craintes des critiques littéraires dont il relevait une nouvelle expression "c'est écrit comme Chat GPT" dans l'émission de radio *le Masque et la plume*. Il invite en conclusion de conférence à repenser la littérature avec l'arrivée des nouvelles technologies, en proposant de considérer la littérature elle-même comme "une technologie du langage" et d'analyser comment les agents conversationnels enrichissent le langage.

Enfin, certains acteurs alertent sur un mouvement "solutionniste" de l'IA générative, comme Cédric Villani dans le journal *Libération* en juin 2023<sup>40</sup>. Malgré le vaste champ des possibles amené par l'intelligence artificielle générative, la technologie en elle-même est critiquée pour le risque de faire de ses utilisateurs "des

<sup>38</sup> LIBERATION. « «Pertinente» ou «pathétique», l'affiche de Roland-Garros créée avec l'intelligence artificielle fait débat ». *Libération*, 25 décembre 2023, sect. Tennis.

[https://www.liberation.fr/sports/tennis/pertinente-ou-pathetique-laffiche-de-roland-garros-creee-avec-lintelligence-artificielle-fait-debat-20231225\\_JAU4GTARL5DCZPC2N\]73HSWWNU/](https://www.liberation.fr/sports/tennis/pertinente-ou-pathetique-laffiche-de-roland-garros-creee-avec-lintelligence-artificielle-fait-debat-20231225_JAU4GTARL5DCZPC2N]73HSWWNU/). [Consulté le 14/01/2023]

<sup>39</sup> Alexandre Gefen. 2023. « Conférence : IA et création littéraire ». [Vidéo en ligne] [46 mn 42 s]. Enssib. <https://www.youtube.com/watch?v=UlmdP-QeARM>. [Consulté le 14/01/2024]

<sup>40</sup> Cario, Erwan, Clémence Mary, et Elise Viniacourt. « Cédric Villani : "Ne cédon pas à l'illusion que l'IA va tout solutionner" ». *Libération*, 23 juin 2023, sect. Idées et Débats [https://www.liberation.fr/idees-et-debats/cedric-villani-ne-cedons-pas-a-lillusion-que-lia-va-tout-solutionner-20230619\\_III\\_TK7TTZFACLJZPE6UGEKWTP4/](https://www.liberation.fr/idees-et-debats/cedric-villani-ne-cedons-pas-a-lillusion-que-lia-va-tout-solutionner-20230619_III_TK7TTZFACLJZPE6UGEKWTP4/) [Consulté le 10/01/2023]

consommateurs de services numériques”<sup>41</sup>. L’experte en intelligence artificielle, Aurélie Jean mentionnait ainsi sur le plateau de l’émission 28 min d’ARTE le 3 janvier 2023 “si tout le monde utilise ChatGPT avec des éléments de langage similaires, on va tendre vers une certaine homogénéité dans les réponses.”<sup>42</sup> Elle pointait ainsi une limite à l’efficacité de l’agent conversationnel qui continue de s’entraîner sur les requêtes des utilisateurs.

Dans un entretien réalisé auprès d’un chercheur en théorie de la conception et de l’innovation, nous avons pu scruter le sens du qualificatif “génératif” associé à Chat GPT, mais qui existait avant son arrivée. ChatGPT se contente de travailler sur un corpus lexical prédéfini, et il n’est pas capable de sortir de ce cadre: à ce titre, sa générativité relève de la création de liens nouveaux et originaux entre des objets lexicaux existants, mais pas d’une véritable réinvention du langage ou de son organisation. Ainsi, le régime de générativité de ChatGPT “est plus proche d’un régime de type échantillonnage”, et ne permet pas de produire de la connaissance. L’humain n’a pas perdu sa maîtrise de la créativité et de la générativité du langage, et elle n’est pas remise en cause par l’outil d’OpenAI; cependant les relations que tissent ChatGPT sont à même de nous interroger et peuvent être tout à fait pertinentes car elles offrent un regard nouveau.

« *Alors, le fait qu'il soit plus savant intrinsèquement c'est intéressant mais pas forcément pour vous [étudiants ingénieurs], parce que c'est le principe d'un ingénieur généraliste : il est censé avoir les têtes de chapitre si vous voulez. Et donc vous pourriez dire : "oui mais il ne fait que raffiner un machin que j'aurais trouvé tôt ou tard."* »

*Dans ce cas-là ce n'est pas super intéressant. [...] En tant que tel, il en sait un peu plus que vous mais si c'est dans la même direction que ce que vous seriez allé vous tout seul c'est bien mais ce n'est pas fou, vous gagnez un peu de temps éventuellement vous gagnez un coup de fil. [à un expert]. C'est vrai qu'en termes économiques pour l'entreprise, ça peut être très intéressant si on se dispense d'un expert parce que l'entreprise risque de trouver que c'est une bonne solution bon c'est pas complètement en termes de précision mais passons. En revanche là où il a une valeur ajoutée intéressante c'est quand il vous envoie dans une direction où vous ne seriez pas allé spontanément c'est à dire qu'il ne vous dit pas un truc que vous ne saviez absolument pas mais brutalement il vous rappelle que "ah ouais tiens vous auriez quand même dû penser à ça", vous le saviez mais vous n'y pensiez pas sur ce problème là typiquement.»* Entretien au CGS (Centre de Gestion Scientifique de l’École des Mines de Paris).

---

<sup>41</sup> Bardou, Florian. « Les architectes face à l’IA : «Même pas peur» ». *Libération*, sect. Design. [https://www.liberation.fr/lifestyle/design/les-architectes-face-a-lia-meme-pas-peur-20230619\\_M4VUTGNIRBHZ7JELU7EP3OW444/](https://www.liberation.fr/lifestyle/design/les-architectes-face-a-lia-meme-pas-peur-20230619_M4VUTGNIRBHZ7JELU7EP3OW444/). [Consulté le 10/01/2023]

<sup>42</sup> Arte 28 minutes (Emission TV) «Hors-série IA | Aurélie Jean et Raphaël Doan», 2 janvier 2024. Disponible sur : <https://www.youtube.com/watch?v=663YRhkVqgw>. [Consulté le 14/01/24]

Il faut accompagner les enfants dans l'utilisation de l'outil.



Illustration Marianne OBEID

### IA, éducation et enseignement supérieur : résistances dans les lieux de production de la connaissance

L'introduction de ChatGPT ne rencontre pas le même succès dans les institutions de l'éducation que dans les entreprises. Certaines universités montrent une résistance à l'égard de l'agent conversationnel. Science Po Paris a ainsi interdit<sup>43</sup> le 25 janvier 2023 l'utilisation de ChatGPT par les élèves pour tout rendu oral ou écrit menant à une note. De plus, pour tout autre rapport non noté, l'utilisation de l'outil d'OpenAI doit être mentionné comme n'importe quelle source. La raison principale énoncée pour motiver cette interdiction est la volonté de respecter la charte anti-plagiat de l'école: la question de la vérité et de la provenance de l'information est donc une préoccupation importante dans les lieux de production de savoir, et ChatGPT ne semble pas suffisamment apporter de garanties pour que l'outil puisse être utilisé dans ce contexte d'apprentissage. Du côté du gouvernement français, le ministre de l'éducation Pap Ndiaye a mis en garde les élèves<sup>44</sup> début février 2023 quant à l'utilisation de ChatGPT, estimant que les textes produits sont stéréotypés et ne ressemblent pas à ce que des étudiants peuvent écrire.

<sup>43</sup> Sciences Po. « ChatGPT : Sciences Po fixe des règles et lance une réflexion sur l'IA dans l'enseignement supérieur », 9 février 2023. Disponible sur :

<https://www.sciencespo.fr/fr/actualites/sciences-po-fixe-des-regles-claires-sur-lutilisation-de-chat-gpt-par-les-etudiants>. [Consulté le 14/01/24]

<sup>44</sup> Toussay, Jade. « Le ministre de l'Éducation Pap Ndiaye met les élèves en garde sur l'utilisation de ChatGPT à l'école ». *Le HuffPost* (blog), 2 février 2023.

[https://www.huffingtonpost.fr/politique/article/chatgpt-doit-etre-encadre-a-l-ecole-selon-pap-ndiaye\\_213550.html](https://www.huffingtonpost.fr/politique/article/chatgpt-doit-etre-encadre-a-l-ecole-selon-pap-ndiaye_213550.html). [Consulté le 14/01/24]

D'autres institutions publiques ont initialement imposé des restrictions à l'usage de ChatGPT, mais ont ensuite changé de position. Ainsi, l'administration new-yorkaise a banni début 2023<sup>45</sup> l'accès à ChatGPT à partir des réseaux Internet des écoles publiques. Selon le porte-parole du département de l'éducation Jenna Lyle, cet outil ne permet pas de développer sa pensée critique et sa capacité à résoudre des problèmes. Cette interdiction a été revue mi-mai 2023<sup>46</sup>, et l'administration a changé de position quant à l'utilisation de l'agent conversationnel dans les écoles publiques. Selon le président du département de l'éducation publique de New York David Banks, l'interdiction était justifiée par mesure de précaution suite au déploiement par OpenAI du chatbot afin de répondre à des préoccupations quant à la sécurité et la précision des réponses; cependant, il faut maintenant accompagner les élèves dans la découverte et l'utilisation de cet outil. Quant aux enseignants, ils pourront recevoir des ressources et exemples d'utilisation de l'IA à des fins pédagogiques, de communication ou encore administrative afin de mieux maîtriser cet outil.

Ce revirement de situation a pu aussi être observé en Australie: après une interdiction temporaire<sup>47</sup> début 2023, ChatGPT sera autorisé dans toutes les écoles publiques australiennes à partir de 2024. Selon le ministre de l'éducation australien Jason Clare, ChatGPT est à présent ancré dans la société, et va devenir un outil «comme la calculatrice ou Internet». Cette acceptation de ChatGPT s'inscrit dans une volonté d'apprendre aux élèves comment bien utiliser ce nouvel outil informatique, et est également fondé sur le constat que les élèves utilisent déjà ces outils, notamment dans les écoles privées. Une interdiction serait donc contre productive, et ne ferait que créer une fracture entre les élèves qui utilisent cet IA et ceux n'en ayant pas l'accès.

### **Destruction créatrice : entre la crainte du chômage et la restructuration des emplois**

L'arrivée d'une innovation capable d'automatiser une partie des processus de production de valeur s'est produite plusieurs fois au XXème siècle avec la mécanisation des outils et la robotisation des chaînes de production industrielle. Ces ruptures technologiques ont principalement touché les ouvriers, remplacés par des robots plus productifs et cet événement est utilisé pour décrire l'automatisation par l'IA des métiers dit

---

<sup>45</sup> Chalkbeat. « NYC Blocks Access to ChatGPT on School Networks as Cheating Fears Swirl », 3 janvier 2023.

Disponible sur

<https://www.chalkbeat.org/newyork/2023/1/3/23537987/nyc-schools-ban-chatgpt-writing-artificial-intelligence/>.

[Consulté le 14/01/24]

<sup>46</sup> Faguy, Ana. « New York City Public Schools Reverses ChatGPT Ban ». *Forbes* (blog). 18 mai 2023,

<https://www.forbes.com/sites/anafaguy/2023/05/18/new-york-city-public-schools-reverses-chatgpt-ban/>. [Consulté le 21 janvier 2024.]

<sup>47</sup> Cassidy, Caitlin. « Artificial Intelligence Such as ChatGPT to Be Allowed in Australian Schools from 2024 ». *The Guardian*, 6 octobre 2023, sect. Australia news.

<https://www.theguardian.com/australia-news/2023/oct/06/chatgpt-ai-allowed-australian-schools-2024>. [Consulté le 14/01/24]

de "cols blancs"<sup>48</sup> et les professions artistiques. Cependant, les derniers rapports de l'OCDE et de l'OIT ne perçoivent pas encore en 2023 de réduction de la demande de main d'œuvre dans les professions les plus visées<sup>49</sup>. Ce constat va dans le sens des prédictions des experts comme l'économiste Philippe Aghion, professeur au Collège de France et membre co-président du Conseil national de l'intelligence artificielle. L'économiste observe que les dernières vagues de robotisation, notamment dans les pays asiatiques, ne sont pas corrélés globalement avec une hausse du chômage dans ces zones géographiques et même au sein des secteurs qui sont traversé par la rupture technologique. Selon Philippe Aghion, une innovation comme l'électricité ou l'intelligence artificielle augmente dans un premier temps la productivité de l'entreprise, cela fait croître son chiffre d'affaires et à terme son activité et donc sa demande de main d'œuvre pour opérer ces nouvelles activités<sup>50</sup>.

Néanmoins, certaines activités comme dans le secteur des médias présentent des signes d'utilisation de l'IA allant dans le sens de l'automatisation. Présentatrice de journal télévisé au Koweït<sup>51</sup>, présentateur météo<sup>52</sup>, influenceur virtuel, ou simplement relecture d'article obligatoire<sup>53</sup> les IA de modèles de langages se répandent dans les rédactions et les maisons d'éditions, non sans ressentir de la part des auteurs. Cette controverse s'incarne principalement dans le rapport de force salarial entre les dirigeants dans la posture de défenseur de l'innovation pour devenir plus compétitifs et les auteurs qui ne veulent pas être remplacés. Cela a même mené dans différentes organisations à des tensions et a pu alimenter les revendications de certains mouvements de grève comme celle des scénaristes d'Hollywood durant l'été 2023 ou encore dans des

---

<sup>48</sup> Bernier, Martin. « Cols bleus, blancs, roses: serons-nous tous au chômage à cause de l'IA? » *Le Figaro* (blog), 10 août 2023.

<https://www.lefigaro.fr/vox/economie/cols-bleus-blancs-roses-serons-nous-tous-au-chomage-a-cause-de-l-ia-20230810>. [Consulté le 14/10/24]

<sup>49</sup> OCDE. *Perspectives de l'emploi de l'OCDE 2023: Intelligence artificielle et marché du travail*. Perspectives de l'emploi de l'OCDE. OECD, 11 juillet 2023. Disponible sur <https://doi.org/10.1787/aae5dba0-fr>. [Consulté le 14/01/2024]

<sup>50</sup> Aghion, Philippe. 2023. « Conférence : Les conséquences économiques de l'IA ». In , [Vidéo en ligne][36 mn]. Collège de France: Campus de l'innovation pour les lycées. <https://www.youtube.com/watch?v=ngEL1PjucrQ>. [Consulté le 14/01/2024]

<sup>51</sup> Euronews, « The Cube : une présentatrice TV générée par l'IA ». 16 avril 2023. Disponible sur : <https://fr.euronews.com/2023/04/16/the-cube-presentatrice-tv-generee-par-lintelligence-artificielle-au-koweit>. [Consulté le 14/01/2024]

<sup>52</sup> RTBF, « Jade, un avatar généré par l'IA, présente la météo en Suisse ». Disponible sur <https://www.rtbf.be/article/jade-un-avatar-genere-parlia-presents-la-meteo-en-suisse-11187917>. [Consulté le 14 janvier 2024]

<sup>53</sup> Syndicat National des Journalistes. « L'Est Républicain - Vosges Matin | La direction voudrait utiliser la moulinette ChatGPT pour retraitier les textes des correspondants », 10 octobre 2023.

<https://snj.fr/article/la-direction-voudrait-utiliser-la-moulinette-chatgpt-pour-retraitier-les-textes-des-correspondants-1565254191>. [Consulté le 14/1/2024]

réactions comme celles journaux quotidiens “*L'Est républicain*” et “*Vosges Matin*” qui ont été relayée par le Syndicat national du journalisme<sup>54,55</sup>.

Lorsque les négociations aboutissent, les journalistes réalisent des conventions et des chartes d'utilisation de l'IA, comme celles répertoriées par le journaliste indépendant Gerald Holubowicz<sup>56</sup>. Celle qui a réuni le plus de signataires est la *Charte de Paris sur l'IA et le journalisme* initiée par Reporters sans Frontières<sup>57</sup>. Ces chartes partagent dix grands principes<sup>58</sup> avec notamment l'engagement de ne pas faire de clonage numérique du journaliste et de former ces derniers aux derniers usages de l'intelligence artificielle.

### **La soutenabilité de l'IA dans un contexte de raréfaction des ressources et de transition écologique ?**

Si l'IA est aujourd'hui largement répandue dans les esprits pour les gains de productivité qu'elle apporte, la question de la soutenabilité de son modèle a été levée par certains de ses développeurs et par les experts de la crise environnementale.

Tout d'abord, les médias relaient les critiques des scientifiques<sup>59</sup> sur la consommation énergétique des intelligences artificielles en pointant dans son processus d' entraînement qui préleve beaucoup d'énergie à cause des grands volumes de données à traiter. *L'AI Index* publié par Stanford est ainsi invoqué par Bertrand Braunschweig dans la revue *Futuribles* pour mettre en exergue les besoins énergétiques importants des IA. Cet argument mène à une estimation de l'empreinte carbone de l'intelligence artificielle qui est ensuite

<sup>54</sup> Syndicat National des Journalistes, « L'Est Républicain - Vosges Matin | La direction voudrait utiliser la moulinette ChatGPT pour retraitier les textes des correspondants ». 10 octobre 2023.

<https://snj.fr/article/la-direction-voudrait-utiliser-la-moulinette-chatgpt-pour-retraitier-les-textes-des-correspondants-1565254191>. [Consulté le 14 janvier]

<sup>55</sup> Franceinfo, « Les journaux quotidiens “L'Est républicain” et “Vosges Matin” testent l'intelligence artificielle pour la correction d'articles ». 26 octobre 2023. Disponible sur

[https://www.francetvinfo.fr/internet/intelligence-artificielle/les-journaux-quotidiens-l-est-republicain-et-vosges-matin-testent-ia-pour-la-correction-d-articles\\_6145410.html](https://www.francetvinfo.fr/internet/intelligence-artificielle/les-journaux-quotidiens-l-est-republicain-et-vosges-matin-testent-ia-pour-la-correction-d-articles_6145410.html). [Consulté le 14/01/2024].

<sup>56</sup> Holubowicz, Gerald. s. d. « Notion – The all-in-one workspace for your notes, tasks, wikis, and databases. » *journalismdesign*

<https://journalismdesign.notion.site/1ee57ee51a804dc5acd788d69aa5b17f?v=08f5eb734f264398a01f49f6d10835e2>. [Consulté le 14 janvier 2024.]

<sup>57</sup> Reporters sans frontières. 2023. *Charte de Paris sur l'IA et le journalisme*.

[https://rsf.org/sites/default/files/medias/file/2023/11/Charte%20de%20Paris%20sur%20l%27IA%20et%20le%20journalisme\\_1.pdf](https://rsf.org/sites/default/files/medias/file/2023/11/Charte%20de%20Paris%20sur%20l%27IA%20et%20le%20journalisme_1.pdf). [Consulté le 13/01/2024]

<sup>58</sup> Pettineo, Camille. 2024. « Les médias face à l'intelligence artificielle : 20 chartes passées au crible ». *La Revue des Médias* | INA (blog). 11 octobre 2024.

<https://larevuedesmedias.ina.fr/les-medias-face-lintelligence-artificielle-20-chartes-passees-au-crible>. [Consulté le 14/01/2024]

<sup>59</sup> Monod, Olivier. « L'IA peut-elle sauver la planète ? » *Libération*, 19 juin 2023, sect. Economie numérique.

[https://www.liberation.fr/economie/economie-numerique/ia-peut-elle-sauver-la-planete-20230619\\_LROWDIFDWNDRFL6RAAUL6MZZE/](https://www.liberation.fr/economie/economie-numerique/ia-peut-elle-sauver-la-planete-20230619_LROWDIFDWNDRFL6RAAUL6MZZE/). [Consulté le 13/01/2023]

utilisée dans les débats environnementaux sur la durabilité des équipements numériques. Un ordre de grandeur diffusé dans *Futurable* compare la consommation énergétique d'un modèle de traitement du langage naturel à l'équivalent de la consommation de cinq voitures particulières pendant leur "durée de vie"<sup>60</sup>. L'ONG Greenpeace s'est également positionnée sur le sujet en relayant les études scientifiques menées à Berkeley et à l'Université Riverside de Californie pour mettre en garde sur l'intensité en matière de ressources que demande le développement des services d'intelligences artificielles sur toute leur chaîne de valeur<sup>61</sup>.

La consommation d'eau de l'IA questionne également différents experts. Dejean Glavas, professeur à l'ESSCA relaie les résultats d'études scientifiques qui estiment la consommation annuelle d'eau de l'IA dans le monde entre 4 et 7 milliards de mètres cubes<sup>62</sup>. Le débat est déjà présent sur le terrain, notamment par rapport au développement des datacenters qui utilisent de grandes quantités d'eau pour évacuer la chaleur des composants. La controverse sur la consommation en eau des datacenters se joue dans plusieurs zones géographiques, en Irlande<sup>63</sup>, aux Pays-Bas<sup>64</sup> et aussi en Espagne dans la région de Tolède où le mouvement associatif "Tu nube seca mi rio" (ndlr : se traduisant par "Ton cloud assèche mon fleuve") fait pression contre un grand projet du groupe Meta pour construire un datacenter avec un investissement d'un milliard d'euros<sup>65</sup> aux proportions impressionnantes : 248 MW de consommation électrique sur 180 hectares selon le *Courrier International*<sup>66</sup>. De même, le mathématicien Cédric Villani, auteur du rapport<sup>67</sup> à l'origine de la stratégie nationale française pour l'intelligence artificielle, considère malgré les gains apportés par l'IA, que sa

---

<sup>60</sup> Braunschweig, Bertrand. « Les cinq murs de l'intelligence artificielle: » *Futuribles* N° 453, n° 2 (20 février 2023): 5-24. <https://doi.org/10.3917/futur.453.0005>. [Consulté le 10/12/2023]

<sup>61</sup> Kojola, Erik. 2023. « ChatCO2 – Safeguards Needed For AI's Climate Risks - ». 30 novembre 2023. <https://www.greenpeace.org/usa/chatco2-safeguards-needed-for-ais-climate-risks/>. [Consulté le 10/12/2023]

<sup>62</sup> Gless, Etienne. 2023. « L'IA peut-elle sauver la planète ? » *Le Figaro Etudiant*, 6 décembre 2023. <https://etudiant.lefigaro.fr/article/etudes/l-ia-peut-elle-sauver-la-planete-20231206/>. [Consulté le 10/12/2023]

<sup>63</sup> Crequer, Tudi. 2021. « Trop de "data centers", l'Irlande risque la panne électrique ». *Reporterre*. 2 décembre 2021. <https://reporterre.net/Trop-de-data-centers-l-Irlande-risque-la-panne-electrique>. [Consulté le 13/01/2023]

<sup>64</sup> Bembaron, Elsa. 2022. « La consommation d'eau des data centers source d'inquiétudes ». *Le Figaro*, 19 août 2022, sect. Tech & Web. <https://www.lefigaro.fr/secteur/high-tech/la-consommation-d-eau-des-data-centers-source-d-inquietudes-20220819>. [Consulté le 13/01/2023]

<sup>65</sup> *Tu Nube Seca Mi Río* (blog), « La Televisión francesa se hace eco del Centro de Datos de Talavera de la Reina ». 2023. 26 septembre 2023. <https://tunubescamirio.com/>. [Consulté le 14/01/2023]

<sup>66</sup> *Courrier international*. « Espagne. En pleine sécheresse, la consommation d'eau du futur data center de Meta inquiète ». 10 mai 2023, sect. Sciences et environnement. <https://www.courrierinternational.com/article/espagne-en-pleine-secheresse-la-consommation-d-eau-du-futur-data-center-de-meta-inquiete>. [Consulté le 13/01/2023]

<sup>67</sup> Villani, Cédric. « Donner un sens à l'intelligence artificielle ». Assemblée nationale (Rapport parlementaire), 28 mars 2018. Disponible sur : [https://medias.vie-publique.fr/data\\_storage\\_s3/rapport/pdf/184000159.pdf](https://medias.vie-publique.fr/data_storage_s3/rapport/pdf/184000159.pdf). [Consulté le 14/01/24]

consommation intensive d'énergie et de matériaux obligera les organisations à en centraliser les infrastructures à terme<sup>68</sup>.

En réponse aux alertes des acteurs militant de la cause écologique et climatique, certains acteurs économiques et institutionnels mettent en avant la nécessité ou encore l'opportunité d'utiliser l'IA pour accélérer les processus de transition en mobilisant l'argument des gains de productivité. Les entreprises du numérique se sont aussi positionnées sur le rôle de l'IA pour atteindre les objectifs climatiques, à l'instar de Google dont le PDG Sundar Pichai expliquait en conclusion d'une tribune dans le *Financial Times* :

*« AI presents a once-in-a-generation opportunity for the world to reach its climate goals, build sustainable growth, maintain global competitiveness and much more. Yet we are still in the early days, and there's a lot of work ahead. »* Dirigeant de Google.<sup>69</sup>

Au sein du gouvernement français, la Direction interministérielle du numérique (DINUM) promeut ainsi l'usage de l'intelligence artificielle pour produire des services qui vont accélérer les projets de transition environnementale en mettant la transition numérique à son service. La Directrice de la DINUM a ainsi pu répondre aux questions des élèves préparationnaires de l'INSP en 2023 dont le concours d'entrée a été augmenté d'une épreuve sur la transition environnementale et qui craignaient une utilisation polluante et inutile de l'intelligence artificielle. La stratégie française en matière d'intelligence artificielle répond à ce risque en orientant les efforts de développement de l'IA frugale définie ainsi par la Cour des Comptes "Une intelligence artificielle frugale est durable et respectueuse de l'environnement en cherchant." La Cour des comptes communique cependant dans ce même rapport, après consultation des scientifiques français en intelligence artificielle, que l'impact sur l'environnement de l'intelligence artificielle est peu pris en compte dans la recherche aujourd'hui<sup>70</sup>.

---

<sup>68</sup> Cario, Erwan, Clémence Mary, et Elise Viniacourt. « Cédric Villani : "Ne cédon pas à l'illusion que l'IA va tout solutionner" ». *Liberation*, 23 juin 2023, sect. Idées et Débats

[https://www.liberation.fr/idees-et-debats/cedric-villani-ne-cedons-pas-a-lillusion-que-lia-va-tout-solutionner-20230619\\_III\\_TK7TTZFACLJZPE6UGEKWTP4/](https://www.liberation.fr/idees-et-debats/cedric-villani-ne-cedons-pas-a-lillusion-que-lia-va-tout-solutionner-20230619_III_TK7TTZFACLJZPE6UGEKWTP4/) [Consulté le 10/01/2023]

<sup>69</sup> Pichai, Sundar. 2023. « Google CEO: Building AI responsibly is the only race that really matters ». *Financial Times*, 22 mai 2023. <https://www.ft.com/content/8be1a975-e5e0-417d-af51-78af17ef4b79> [Consulté le 13/01/2023]

<sup>70</sup> Cour des comptes, « La stratégie nationale de recherche en intelligence artificielle ». 2023. Rapport de la Cour des comptes. Disponible sur :

<https://www.ccomptes.fr/system/files/2023-04/20230403-synthese-strategie-nationale-recherche-intelligence-artificielle.pdf> [Consulté le 14/01/2023]

## Partie 4 : Débat public autour des risques

### Des acteurs soucieux des risques existentiels

Certains acteurs craignent des scénarios où une IA suffisamment puissante pourrait mener à une catastrophe existentielle, menaçant le futur de l'humanité. Ces scénarios rappellent les fictions écrites au XXe siècle présentées dans la première partie. Différentes questions apparaissent, notamment sur le contrôle de ces systèmes, et la réalité et probabilité de ces risques. Les différents risques majeurs de l'IA que nous avons pu recenser sont de deux types. Le premier est l'utilisation malveillante de l'IA, et le deuxième est le risque que l'IA n'agisse pas de façon alignée avec les objectifs qu'on lui a assignés.

Les capacités de plus en plus impressionnantes de l'IA pourraient en effet permettre à des criminels d'agir avec une grande facilité et un grand impact. On pense par exemple à la désinformation de masse avec les *deepfakes*, à la cybercriminalité facilitée par la maîtrise de l'informatique par l'IA, ou encore la synthèse d'armes biologiques avec des IAs développées aujourd'hui pour trouver des molécules pour des médicaments.

Par exemple, des chercheurs de Checkpoint Capital ont montré comment ChatGPT pouvait potentiellement construire une campagne de phishing de A à Z, de la création d'un courriel de phishing à l'écriture de code malveillant. A l'opposé, « ChatGPT est un perroquet qui a beaucoup lu et qui ne sait que refaire ce qui est déjà connu. Pas besoin de ChatGPT pour développer des virus », affirme un expert du ministère des Armées qui souhaite rester anonyme<sup>71</sup>.

Le deuxième type de risque est ce que les scientifiques appellent le problème de non alignement. Une IA est dite non alignée lorsque l'on a fixé un objectif à l'IA mais qu'elle ne l'accomplit pas de la façon dont on souhaite qu'elle le fasse. On parle alors de Spécification Gaming<sup>72</sup>, quand elle cherche à maximiser ses performances de manière inattendue ou non désirée. La *Goal Misgeneralization*<sup>73</sup> est aussi un risque où l'IA applique ses objectifs d'une manière trop générale ou hors contexte, entraînant des conséquences imprévues et potentiellement dangereuses. Stuart Russel, un expert anglo-américain de l'IA explique par exemple: "il est

---

<sup>71</sup> Richard, Philippe. « ChatGPT : les pirates peuvent-ils en profiter pour multiplier les cyberattaques ? » *Techniques de l'Ingénieur*. Disponible sur :

<https://www.techniques-ingenieur.fr/actualite/articles/chatgpt-les-pirates-peuvent-ils-en-profiliter-pour-multiplier-les-cyberattaques-120102/> [Consulté le 21 janvier 2024]

<sup>72</sup> Krakovna, Victoria, Jonathan Uesato, Vladimir Mikulik, Matthew Rahtz, Tom Everitt, Ramana Kumar, Zac Kenton, Jan Leike, et Shane Legg. « Specification Gaming: The Flip Side of AI Ingenuity ». *Google DeepMind*, 21 avril 2020.

<https://deepmind.google/discover/blog/specification-gaming-the-flip-side-of-ai-ingenuity/> [Consulté le 21 janvier 2024].

<sup>73</sup> Langosco, Lauro, Jack Koch, Lee Sharkey, Jacob Pfau, Laurent Orseau, et David Krueger. « Goal Misgeneralization in Deep Reinforcement Learning ». arXiv, 9 janvier 2023. <https://doi.org/10.48550/arXiv.2105.14111>. [Consulté le 21/01/2024]

tellement évident qu'une machine aura une capacité d'auto-préservation même si vous ne la programmez pas, parce que si vous dites « Aller chercher le café », elle ne peut pas récupérer le café s'il est mort. Donc, si vous lui donnez un objectif quel qu'il soit, il a une raison de préserver sa propre existence pour atteindre cet objectif. Et si vous le menacez en allant chercher du café, il va vous tuer car tout risque pour le café doit être contré”<sup>74</sup>.

De plus, certains comportements d'une IA peuvent devenir incompréhensibles et donc incontrôlables pour l'humain, notamment à partir du moment où l'IA atteindrait un niveau supérieur à l'humain. En effet, les avancées actuelles de l'IA laissent présager que l'IA pourrait avoir des capacités énormes, bien au-delà de ce qu'on pourrait imaginer.

« *Ce sont des choses qui paraissent intuitivement un peu folles, mais qui, lorsqu'on regarde les capacités des modèles sur plein de métriques différentes, ne sont pas totalement folle.* » Entretien EffiSciences.

### Avis partagés sur les risques

Ces scénarios sont envisagés par divers groupes d'acteurs. On peut citer pour commencer les acteurs qui développent les gros modèles de langages : OpenAI avec ChatGPT, Anthropic, Google avec Bard. Chacun de ces acteurs a d'ailleurs sa propre stratégie pour contrôler l'IA<sup>75,76,77</sup> au fur et à mesure de son développement, mais certains restent sceptiques quant à l'efficacité de ces méthodes, comme évoqué dans notre entretien avec EffiSciences.

D'autres acteurs dans le monde de la recherche scientifique envisagent et parlent de ces risques, comme Yoshua Bengio (voir encart), l'un des pères fondateurs de l'IA moderne, qui sensibilisent le grand public aux risques de l'IA. Enfin des acteurs politiques ou de la société civile comme l'association

---

<sup>74</sup> Nast, Condé. « Elon Musk's Billion-Dollar Crusade to Stop the A.I. Apocalypse ». *Vanity Fair*, 26 mars 2017. <https://www.vanityfair.com/news/2017/03/elon-musk-billion-dollar-crusade-to-stop-ai-space-x>. [Consulté le 21/01/2024]

<sup>75</sup> OpenAI. « OpenAI's Approach to Frontier Risk », 26 octobre 2023. <https://openai.com/global-affairs/our-approach-to-frontier-risk>. [Consulté le 18/01/2024]

<sup>76</sup> Anthropic, « Anthropic's Responsible Scaling Policy, version 1.0 ». 19 septembre 2023. <https://www-files.anthropic.com/production/files/responsible-scaling-policy-1.0.pdf>. [Consulté le 18/01/2023]

<sup>77</sup> Google Deepmind. « Responsibility & Safety ». Google DeepMind. Disponible sur : <https://deepmind.google/about/responsibility-safety/>. [Consulté le 18 janvier 2024.]

EffiSciences<sup>78</sup>, les think tanks The Future Society<sup>79</sup> ou l’Institut Montaigne<sup>80</sup>, ou encore des hommes politiques<sup>81,82</sup>, mentionnent ces risques dans leurs publications.

### **Yoshua Bengio : chercheur sensibilisant aux risques de l’IA**



Yoshua Bengio est un chercheur canadien, spécialiste en intelligence artificielle, l’un des pionniers du Deep Learning, Prix Turing 2018. Il est le fondateur et directeur scientifique de MILA (*Montreal Institute for Learning Algorithms*), organisme qui rassemble des chercheurs spécialisés dans le domaine de l’apprentissage profond. Il est convaincu des dangers de l’IA, et compare sa position aux lanceurs d’alertes contre le dérèglement climatique<sup>83</sup>. Il a notamment signé la lettre de moratoire appelant à stopper l’entraînement de modèles plus puissants que GPT4.<sup>84</sup>

Ces scénarios de risques ne sont pas plébiscités par tous les acteurs. Deux groupes d’acteurs s’opposent, avec leurs leaders. D’un côté, les “catastrophistes” sont composés de profils très inquiets des développements rapides de l’IA et appellent à une grande prudence voire un arrêt de l’innovation. De l’autre, les “accelerationistes” visent un développement rapide en promettant des impacts positifs rapides sur de nombreux problèmes de l’humanité.

<sup>78</sup> EffiSciences (blog), « Pole IA - EffiSciences » Disponible sur : <https://ia.effisciences.org/>. [Consulté le 21 janvier 2024]

<sup>79</sup> The Future Society. « Overview ». Disponible sur : <https://thefuturesociety.org/about-us/>. [Consulté le 21 janvier 2024]

<sup>80</sup> Moës, Nicolas « Hauts risques et forts impacts : réguler l’IA ». Institut Montaigne. Disponible sur <https://www.institutmontaigne.org/expressions/hauts-risques-et-forts-impacts-reguler-ia>. [Consulté le 21 janvier 2024]

<sup>81</sup> Maggie Hassan (blog) « Senator Hassan Highlights AI Threats to National Security | U.S. Senator Maggie Hassan of New Hampshire », 20 septembre 2023. Disponible sur :

<https://www.hassan.senate.gov/news/press-releases/senator-hassan-highlights-ai-threats-to-national-security-and-public-safety-at-senate-hearing>. [Consulté le 21 janvier 2024]

<sup>82</sup> Hern, Alex, et Kiran Stacey. « No 10 Acknowledges ‘Existential’ Risk of AI for First Time ». *The Guardian*, 25 mai 2023, sect. Technology.

<https://www.theguardian.com/technology/2023/may/25/no-10-acknowledges-existential-risk-ai-first-time-rishi-sunak>. [Consulté le 21 janvier 2024]

<sup>83</sup> Le Monde.fr. « A Montréal, l’un des pères de l’intelligence artificielle alerte sur une menace existentielle pour l’homme ». 2 juin 2023. Disponible sur :

[https://www.lemonde.fr/international/article/2023/06/02/a-montreal-l-un-des-peres-de-l-intelligence-artificielle-alerte-sur-une-menace-existentielle-pour-l-homme\\_6175814\\_3210.html](https://www.lemonde.fr/international/article/2023/06/02/a-montreal-l-un-des-peres-de-l-intelligence-artificielle-alerte-sur-une-menace-existentielle-pour-l-homme_6175814_3210.html). [Consulté le 21 janvier 2024]

<sup>84</sup> Future of Life Institute. « Pause Giant AI Experiments: An Open Letter », 22 mars 2023. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.

Les “catastrophistes” se caractérisent par leurs craintes et cherchent à poser de multiples contraintes aux développement de l’IA. Une lettre ouverte<sup>85</sup>, publiée dans le 28 mars 2023 par plusieurs sommités de l’IA (chercheurs reconnus pour leurs travaux en IA comme Yoshua Bengio, PDG ou co fondateurs d’entreprises numériques), demande une “pause” de toutes les activités de recherche impliquant l’entraînement de modèles pour une période afin que la régulation puisse rattraper son retard et que la recherche puisse répondre aux premiers risques qui apparaissent. Les “catastrophistes” plaident également pour une régulation forte des systèmes d’IA, tant dans leur entraînement que leur déploiement. L’institut Montaigne soutient par exemple un accompagnement strict du progrès en IA en mobilisant une approche l’expansion des systèmes en fonction du risque.:

*«Si on pousse la vision de l’approche par le risque à son terme, on est un peu plus dans le sens de ce que met en place les Etats-Unis, qui est de se dire, OK, au cas par cas, pour chaque système d’IA, on doit faire une évaluation du risque global, et pas juste s’arrêter au fait que c’est parce qu’un système d’IA est déployé dans tel secteur qu’il pose un risque.»* Entretien Institut Montaigne.

Ils soutiennent un développement massif de la recherche en “AI safety”, domaine de recherche pour assurer la sûreté de systèmes d’IA vis-à-vis des risques précédemment présentés. Cela correspond par exemple à la mission que se donne EffiSciences où le responsable du pôle IA indique :

*«Par exemple, le Master MVA, là où on enseigne, donc le cours qu’on donne, on le fait sur des thématiques qui sont assez, je pense, négligées, au sens où il n’y a aucun autre cours du MVA qui en parle, alors que ce sont des sujets qui sont de grande actualité et où je pense que les élèves ont vraiment une envie grande de s’informer et de découvrir, on va dire, de manière scientifique ces enjeux-là»* Entretien EffiSciences.

De l’autre côté, les “accélérationnistes” prônent une autre voie de développement. Selon eux, le bénéfice des IA à venir est tel qu’il serait contre-productif de limiter aujourd’hui son développement. S’ils envisagent les éventuels risques de l’IA, ils considèrent que le développement de l’IA permettra d’y répondre par lui-même. Sam Altman, PDG d’Open AI, incarne cette position en proposant le plan d’action d’OpenAI qui vise à développer au plus vite une IA capable de faire de la recherche et de solutionner les problèmes d’alignement et de responsabilité éthique des modèles. Certains individus n’estiment pas les risques catastrophiques comme probable ou suffisamment probable pour donner lieu à un ralentissement du développement des systèmes. Yann Le Cun, par exemple, vulgarise sur ses réseaux de nombreux arguments contraires aux scénarios présentés par les “catastrophistes” (voir encart Lecun). D’autres acteurs pondèrent

---

<sup>85</sup> Future of Life Institute. « Pause Giant AI Experiments: An Open Letter », 22 mars 2023.  
<https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. [Consulté le 18/01/2024]

également des risques différents, comme le retard technologique de la France vis-à-vis des acteurs privés américains (OpenAI, Google, etc.). C'est la position du ministre français de l'économie<sup>86</sup> ou du CNRS<sup>87</sup>.

### **Yann Le Cun : chercheur défendeur d'une accélération en IA**

Yann Le Cun est un chercheur français en intelligence artificielle, l'un des pionniers du Deep Learning, Prix Turing 2018. Il est aujourd'hui directeur de la recherche en IA chez Meta (Facebook) et a intégré le *comité pour l'intelligence artificielle générative* mis en place par Elisabeth Borne en août 2023<sup>88</sup>. Sa position est très marquée sur le fait que l'IA n'est pas un danger pour l'humanité. Il soutient<sup>89</sup> notamment qu'il n'y aurait pas de scénarios Ex Machina ou Terminator, notamment car les robots ne seraient pas construits avec des motivations humaines : faim, pouvoir, reproduction, auto-préservation. Lors d'une interview<sup>90</sup>, il indique qu'il travaille lui-même avec des gens plus intelligents que lui, mais qu'ils n'ont pas pris sa place pour autant, ils l'ont juste aidé dans son travail, comme l'IA le fera.



### **Les diverses arènes du débat**

L'opposition entre ces deux groupes se déroule sur différentes arènes. Le débat se déroule d'abord au niveau des instances de régulation, comme les négociations autour de l'EU AI Act. Au mois de novembre 2023<sup>91</sup>, on a vu s'affronter la commission européenne, qui présente son projet de loi sur l'IA, et le Conseil de

<sup>86</sup> Bellefqih, Réouane. « Course à l'innovation dans l'IA générative : un double défi financier et humain ». La Tribune, 11 janvier 2024. Disponible sur <https://www.latribune.fr/supplement/la-tribune-now/course-a-l-innovation-dans-l-ia-generative-un-double-defi-financier-e-t-humain-987546.html>. [Consulté le 18/01/2024]

<sup>87</sup> Sassatelli, Gilles, Maël LE Treust, Thierry Artières, David Ilcinkas, Myriam Caudrelier, et Caroline Appert. « Rapport de Prospective du Conseil Scientifique de l'Institut des Sciences Informatiques et leurs Interactions (INS2I) ». INS2I. [https://csi-ins2i.cnrs.fr/sites/default/files/general/pdf/Rapport\\_prospective\\_CSI\\_INS2I.pdf](https://csi-ins2i.cnrs.fr/sites/default/files/general/pdf/Rapport_prospective_CSI_INS2I.pdf). [Consulté le 9 janvier 2024.]

<sup>88</sup> gouvernement.fr. « Comité de l'intelligence artificielle générative ». Consulté le 21 janvier 2024. <https://www.gouvernement.fr/communique/comite-de-lintelligence-artificielle>.

<sup>89</sup> Nast, Condé. « Elon Musk's Billion-Dollar Crusade to Stop the A.I. Apocalypse ». *Vanity Fair*, 26 mars 2017. <https://www.vanityfair.com/news/2017/03/elon-musk-billion-dollar-crusade-to-stop-ai-space-x>. [Consulté le 21/01/2024]

<sup>90</sup> Vyacheslav, Eric. « Yann is brilliant. Can he just be the president of AI already? » *LinkedIn* (réseau social), décembre 2023. Disponible sur :

[https://www.linkedin.com/posts/eric-vyacheslav-156273169\\_yann-is-brilliant-can-he-just-be-the-president-ugcPost-7137148094122541056-Fqam/?utm\\_source=share&utm\\_medium=member\\_desktop](https://www.linkedin.com/posts/eric-vyacheslav-156273169_yann-is-brilliant-can-he-just-be-the-president-ugcPost-7137148094122541056-Fqam/?utm_source=share&utm_medium=member_desktop). [Consulté le 18/01/2024]

<sup>91</sup> Bertuzzi, Luca. « AI Act : négociations bloquées à cause de divergences sur les modèles de fondation ». *Euractiv*, 13 novembre 2023. Disponible sur :

<https://www.euractiv.fr/section/intelligence-artificielle/news/ai-act-negociations-bloquees-a-cause-de-divergences-sur-les-modeles-de-fondation/>. [Consulté le 18/01/2024]

l'Europe, où certains pays s'opposent au texte de la commission. La France et l'Allemagne sont en effet "montés au créneau" en critiquant les mesures réglementant la conception des "modèles de fondation", soubassement des grands modèles actuels (par exemple GPT-4 pour chatGPT). La commission voulait instaurer des règles plus stricts pour les modèles les plus puissants, quand les pays membres préfère laisser des marges de manœuvres aux entreprises européennes (Mistral AI pour la France et Aleph Alpha pour l'Allemagne) afin de rattraper le retard technologique avec les géants OpenAI ou Google Deepmind. Les négociations furent longues et aboutirent à un accord le 6 décembre<sup>92</sup>. L'opposition entre "catastrophiste" (ici représentée par la commission) et "accélérationniste" (la France et l'Allemagne) se retrouve donc dans les sphères administratives des processus de réglementation.

Toutefois, cette opposition apparaît également dans le débat public. Certains fer de lance de l'un ou l'autre des groupes communiquent énormément, sur les médias traditionnels ou les réseaux sociaux. Yann Le Cun notamment est très présent sur X (ex-twitter), et participe à de nombreux débats très suivis<sup>93</sup>. Les appels à l'opinion publique se multiplient également (lettre ouverte pour le moratoire<sup>94</sup>, tribunes pour<sup>95</sup> ou contre<sup>96</sup> la régulation des systèmes).

### Dénoncer la composition des "oppositions"

Cette polarisation du débat n'est toutefois pas reconnue par tous les acteurs : Marie Lechner, enseignante-rechercheuse à l'ÉSAD (École supérieure d'art et de design d'Orléans) et commissaire d'exposition indépendante, présente dans une conférence<sup>97</sup> de profondes critiques sur l'impact de l'IA dans le monde de

<sup>92</sup> Fagot, Vincent. « AI Act : l'Union européenne pionnière dans la régulation de l'intelligence artificielle ». *Le Monde.fr*, 9 décembre 2023.

[[https://www.lemonde.fr/economie/article/2023/12/09/ai-act-l-union-europeenne-pionniere-dans-la-regulation-de-l-intelligence-artificielle\\_6204830\\_3234.html](https://www.lemonde.fr/economie/article/2023/12/09/ai-act-l-union-europeenne-pionniere-dans-la-regulation-de-l-intelligence-artificielle_6204830_3234.html)]. [Consulté le 18/01/2024]

<sup>93</sup> Bengio, Tegmark, Mitchell, et LeCun. « Munk Debate on Artificial Intelligence : "AI research and development poses an existential threat." », [Vidéo en ligne][1 h 47 mn 19 s], 2023. <https://www.youtube.com/watch?v=144uOf4SYA>.

<sup>94</sup> Future of Life Institute. « Pause Giant AI Experiments: An Open Letter », 22 mars 2023.

[<https://futureoflife.org/open-letter/pause-giant-ai-experiments/>]. [Consulté le 18/01/2024]

<sup>95</sup> Breton, Thierry. « L'IA Act n'entrave aucunement la révolution en mouvement de l'intelligence artificielle ». *Le Monde.fr*, 16 décembre 2023.

[[https://www.lemonde.fr/idees/article/2023/12/16/thierry-breton-l-ia-act-n-entrave-aucunement-la-revolution-en-mouvement-de-l-intelligence-artificielle\\_6206177\\_3232.html](https://www.lemonde.fr/idees/article/2023/12/16/thierry-breton-l-ia-act-n-entrave-aucunement-la-revolution-en-mouvement-de-l-intelligence-artificielle_6206177_3232.html)]. [Consulté le 18/01/2024]

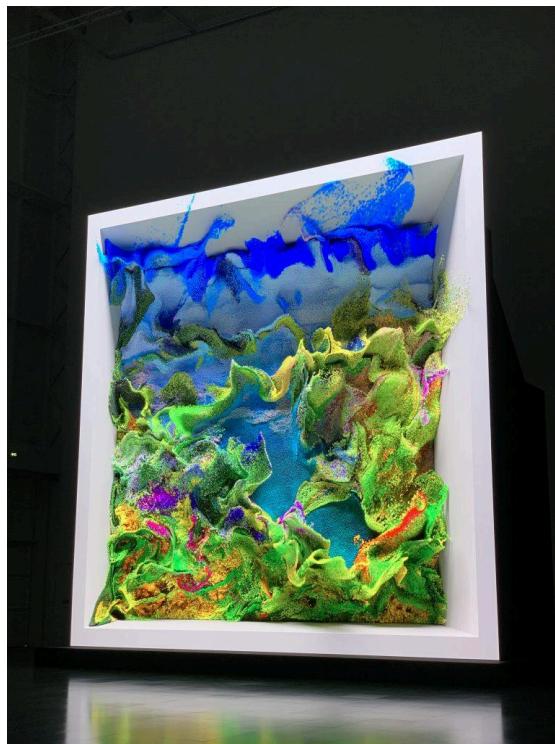
<sup>96</sup> Bottero, Laurence. « "L'IA Act ne doit pas freiner l'innovation" (Aurélie Jean) ». *La Tribune*, 23 novembre 2023, sect. Informatique.

[<https://www.latribune.fr/technos-medias/informatique/l-ia-act-ne-doit-pas-freiner-l-innovation-aurelie-jean-983859.html>]. [Consulté le 18/01/2024]

<sup>97</sup> Lechner, Marie. Conférence : « Entre spectacularisation et critique complaisante, quelle place pour une exploration artistique critique de l'IA ? », [Vidéo en ligne][1 h 20 mn 5 s], 5 décembre 2023. Disponible sur [https://www.youtube.com/watch?v=\\_luc49xIe0w](https://www.youtube.com/watch?v=_luc49xIe0w). [Consulté le 21/01/2024]

l'art : problématique de propriété intellectuel, de biais, de discrimination, impact social et environnemental de l'entraînement des systèmes, questions abordées dans l'exposition critique de l'IA "House of Mirrors: Artificial Intelligence as Phantasm", qui retrace l'inquiétude des artistes envers le développement de l'IA. Ces questions ne sont pas posées dans le débat précédent des "risques existentiels" de l'IA, ce qui amène Marie Lechner à accuser les "catastrophistes" d'être dans le même camp que les "accélérationnistes", en ne soulevant que des problèmes long termes, et en déviant le débat politique sur ces les questions actuelle qui intéresse et menacent le monde de l'art (et d'autres secteurs d'activités et citoyens).

Marie Lechner indique également une opposition dans le monde de l'art sur la place de l'IA. Elle cite notamment l'artiste Refik Anadol et son exposition *Unsupervised — Machine Hallucinations* où il explore l'esthétique des données basée sur des mémoires visuelles collectives. Elle interroge notamment la contradiction entre son discours qui vise à présenter l'IA comme une boîte noire et ses œuvres dont les dimensions (100m<sup>2</sup> pour *Rêve de nature*<sup>98</sup>), loin de démythifier la machine la place sur un piédestal.



*Crédits photos : Refik Anadol, Machine Hallucinations – Nature Dreams KÖNIG GALERIE, St. Agnes, Nave, Berlin, Allemagne, 2021*

<sup>98</sup> Groupe Usine d'électricité de Metz. « Le groupe UEM, mécène de l'œuvre de Refik Anadol, Machine Hallucinations - Rêves de nature au Centre Pompidou-Metz », 13 juin 2022. Disponible sur : <https://www.uem-metz.fr/le-groupe-uem-mecene-de-loeuvre-de-refik-anadol-machine-hallucinations-reves-de-nature/>. [Consulté le 18/01/2024]

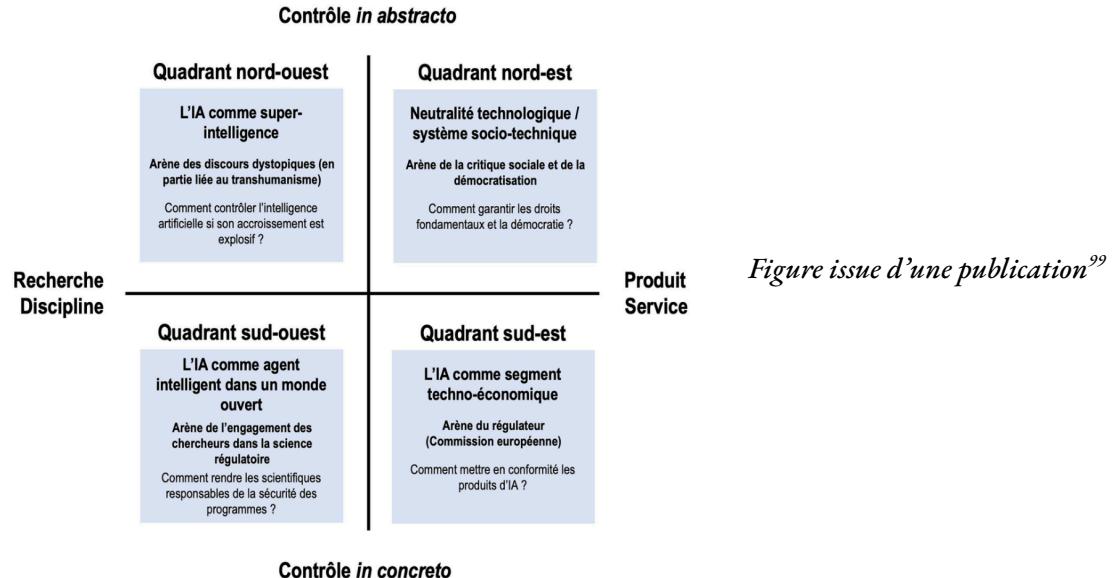
## Analyse académique des débats sur l'IA

Le sociologue Bilel Benbouzid analyse ces différents débats en les faisant remonter à des conceptions différentes de l'IA :

«*l'IA comme « super-intelligence » dont il faut, à long terme, anticiper les risques de perte de contrôle ; l'IA comme discipline scientifique pour laquelle les chercheurs doivent garantir, à court terme, la sûreté, la robustesse, l'équité, l'explicabilité, etc. ; l'IA comme système socio-technique dont les implications sur la société nécessitent une approche qui va au-delà de simples exigences éthiques non contraignantes ; enfin l'IA comme système technico-économique que les producteurs peuvent mettre sur le marché après les avoir certifiés conformes aux exigences réglementaires.*» Issu d'une publication de Bilel benbouzid et al.<sup>99</sup>

Selon lui, ces visions se positionnent sur deux axes. Le premier correspond à la conception du terme IA qui a un double sens, à la fois une discipline scientifique et une machine matérielle. A cet axe s'ajoute la conception du contrôle de ces systèmes, entre “projet à réaliser concrètement et ceux qui font plutôt du contrôle un sujet de spéculation abstrait”. Cela donne une boussole, cadran permettant d'analyser les interactions entre les positions différentes.

Cette analyse permet de comprendre en quoi les questions de régulation sont centrales dans le débat sur l'intelligence artificielle. Avec ces différentes conception de l'objet (IA comme discipline ou connaissances vs IA comme objet matériel) et de la manière de l'incorporer à la société (contrôle in concreto ou in abstracto), les individus sont amenés à se positionner et à faire valoir leur propre perspective sur l'IA et son contrôle.



<sup>99</sup> Benbouzid, Bilel, Yannick Meneceur, et Nathalie Alisa Smuha. « Quatre nuances de régulation de l'intelligence artificielle. Une cartographie des conflits de définition ». *Réseaux* 232-233, n° 2-3 (2022): 29-64.  
<https://doi.org/10.3917/res.232.0029>.

## Conclusion

ChatGPT est l'aboutissement de plusieurs dizaines d'années de recherche en intelligence artificielle et n'a cessé d'évoluer depuis sa création par OpenAI. Son entrée rapide et inattendue dans la sphère publique a généré de nombreux questionnements sur les plans sociaux, légaux, économiques mais a également relancé le débat sur les risques potentiels d'une telle technologie. De nombreux acteurs s'affrontent dans différentes arènes, prônant une régulation active de ChatGPT, une certaine prudence ou des réticences à l'utilisation de l'outil, ou au contraire une incitation aux développements des IA génératives.

Du point de vue juridique, il est important de garder en tête que les normes législatives qui encadrent l'IA générative, dont ChatGPT sont relativement récentes. Et même si les approches diffèrent au niveau international, elles sont toutes encore très sujettes à des évolutions complexes et rapides pour combler les trous face aux défis et critiques posés. Il est intéressant de noter que selon les traditions et l'historique des pays, les choix de régulation sont différents et les points chauds sont donc divers. Pourtant, ces technologies numériques sont appelées à être diffusées sur le globe entier, posant également des questions de coordination internationales sur les questions de droit. De même, les menaces (droit d'auteur, cybersécurité) sont elles aussi délocalisées, il est donc impossible de réguler sans s'intéresser à son voisin. Les controverses internes à une région se trouvent ainsi "enrichies" de nouveaux points chauds avec la rencontre de systèmes légaux très différents.

Même au sein de l'apport économique à certains secteurs, qui semble pourtant assez consensuel, les inquiétudes et les débats s'étendent, craignant des influences de transformation des secteurs d'activités, de l'automatisation aux problématiques de concurrence ou de création de la valeur avec le droit d'auteur. Les implications pour l'économie semblent si nombreuses que mises bout à bout certains y voient une révolution, potentiellement une transformation radicale de l'économie avec une perte d'emploi massive ou bien la disparition d'une catégorie de travailleurs.

ChatGPT a soulevé des questionnements dans sa durabilité sociétale qui ne convainc pas tout le monde. Certains questionnent les impacts environnementaux de la conception de ces modèles, mais aussi les enjeux sociaux des travailleurs du click. Si les créateurs de ces systèmes prônent l'intérêt pour la société voire l'humanité, la balance n'apparaît pas bénéfique pour tous les acteurs de la controverse, et notamment le respect de libertés fondamentales. D'autres craignent même la disparition de l'humanité avec les risques existentiels inhérents aux systèmes d'IA génératifs et généraux. La constitution de groupes d'oppositions aux "pro-IA", défendant les risques qui leur semblent les plus problématiques, multiplient les points chauds.

Ainsi, tous les points chauds relevés dans notre étude, des difficultés légales aux risques pour l'humanité en passant par les enjeux économiques du développement de l'IA générative, sont connectés car chaque acteur pousse pour que les risques qu'il considère comme primordiaux soient traités par les autres

acteurs. On voit ainsi dans toutes les arènes organisant la pénétration de l'IA dans la société (débat public, organes législatifs, ministères de l'économie et organismes internationaux de recommandations de politiques publiques) ces acteurs évoluer, présentant leurs exemples, leurs intérêts, essayant d'influencer la hiérarchie des risques auprès de leurs homologues. Les questions sont démultipliées, de peur que certaines ne passent à la trappe, ou que d'autres acteurs ne les minimisent.

La controverse se trouve ainsi discutée par de nombreux acteurs, avec différents rôles ou spécialités. Pour faire peser sa conception de l'IA dans les arènes de négociation, les acteurs par construction sont institutionnels, fort politiquement ou structurellement. Le choix de cette étude de se fonder davantage sur des articles de presse ou de la littérature grise plutôt que des entretiens nous a permis de dégager les grands topes de la controverse et d'identifier leurs interactions. Cependant, les entretiens menés et l'analyse documentaire témoignent d'une spécificité de chaque acteur, amené à se positionner sur toutes les questions complexes, très récentes et dont les définitions sont en cours d'élaboration. Les alliances que nous avons exposées sont donc souvent des idéaux de mécanismes de négociation, sans que ces rapprochements ne soient pérennes pour les acteurs, ne représentant des accointances sur les enjeux de risques ou bénéfices, ou encore ne soient précisément recherchés ou analysés comme tels par les acteurs.

Certaines questions, moins sujettes à controverse, n'ont pas été abordées dans notre étude comme l'accessibilité de ChatGPT et des modèles d'IA génératives pourrait aussi être sujet à réflexion, puisque ces outils nécessitent souvent un compte payant. Ceci pourrait générer une fracture entre ceux pouvant avoir un accès aux IA génératives et ceux n'ayant pas cette capacité, notamment si l'utilisation de modèles de langues dans les milieux éducatifs et professionnels se développe.

La sortie inattendue de ChatGPT, ainsi que le choix d'OpenAI de le rendre accessible au grand public a placé l'IA générative au cœur des débats publics, et a ravivé les questions de la place de l'intelligence artificielle dans nos sociétés. Même si de nombreux acteurs économiques, politiques ou éducatifs ont pu s'emparer du sujet, l'individu reste le plus concerné par les changements introduits; il a néanmoins son rôle à jouer dans la visibilité et la notoriété de cette technologie, ainsi que celui de régulation par l'importance de son vote.

## Méthodologie

Cette étude a été en premier lieu éclairée par une analyse documentaire portant sur divers types de documents (articles de presse, publications scientifiques, et certaines ressources en ligne), permettant l'émergence de thèmes majeurs de la controverse que nous avons retenu : Justice Sociale, État, Innovation, Économie, Réseaux et climat, Défense et géopolitique. Cela a permis d'étendre la recherche documentaire à de nombreux sites, articles de presse, documents officiels et rapports, articles scientifiques, traitant de ces sujets. Nous avons également visionné ou participé à de nombreuses conférences ou table rondes<sup>100, 101, 102, 103, 104, 105, 106</sup>.

Notre étude s'est également fondée sur une série de cinq entretiens semi-directifs avec divers acteurs de la controverse : un chercheur de chez EffiSciences, un chercheur de chez Redwood Research, un consultant en IA de l'OCDE, un expert de l'Institut Montaigne, et un chercheur au CGS (centre de gestion scientifique) de l'école des Mines de Paris. Tous les entretiens ont été enregistrés et retranscrits.

Pour valider les thématiques que nous avons vu émerger, une analyse quantitative d'articles de presse en français a été réalisée via l'outil Cortex par extraction sur Europresse de 1000 articles traitant de "ChatGPT". Nous avons obtenu la carte ci-dessous qui a conforté nos thèmes majeurs et réorienté notre étude. Nous avons nommé les différents groupements qui sont apparus ("apport pour le juriste", "stratégie française", etc.) pour faciliter l'analyse. Les questions d'apport pour l'économie, notamment par secteur, y

<sup>100</sup> Université Dauphine PSL, « IA et société : nouvelle donne, nouveaux enjeux », Université Dauphine PSL, 20-23 novembre 2023. Disponible sur <https://dauphine.psl.eu/dauphine-digital-days>. [Consulté le 21/01/2023]

<sup>101</sup> Coord. Paul, Christian. « Séminaire à la recherche du politique de l'intelligence artificielle ». SciencesPo Lyon | Digital Factory: La chaire Transformations de l'action publique, 14 novembre 2023.

<https://chaire-actionpublique.fr/seminaire-a-la-recherche-du-politique-de-lintelligence-artificielle/>. [Consulté le 21/01/2023]

<sup>102</sup> Lechner, Marie. Conférence : « Entre spectacularisation et critique complaisante, quelle place pour une exploration artistique critique de l'IA ? », [Vidéo en ligne][1 h 20 mn 5 s], 5 décembre 2023. Disponible sur [https://www.youtube.com/watch?v=\\_luc49xle0w](https://www.youtube.com/watch?v=_luc49xle0w). [Consulté le 21/01/2024]

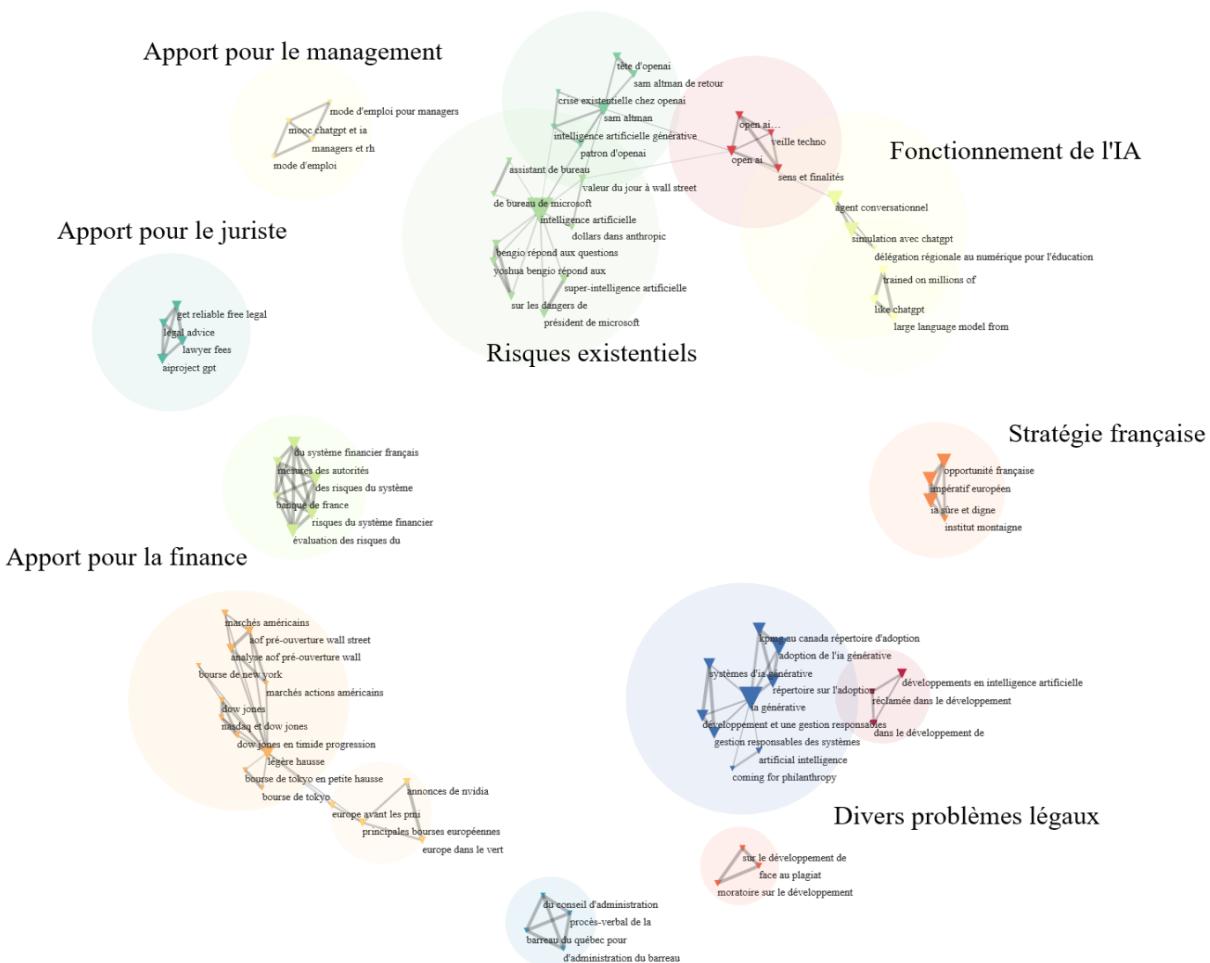
<sup>103</sup> Gefen Alexandre. 2023. « Conférence : IA et création littéraire ». [Vidéo en ligne] [46 mn 42 s]. Enssib. [https://www.youtube.com/watch?v=UlmdP\\_QeARM](https://www.youtube.com/watch?v=UlmdP_QeARM). [Consulté le 14/01/2024]

<sup>104</sup> Aghion, Philippe, Suzanne Vergnolle, Benoît Sagot, Xavier Leroy, et Jean-François Bonnefon. « L'IA et ses défis ». In *Campus de l'innovation pour les lycées*, [Vidéos en ligne]. Collège de France: été organisé par le Campus de l'Innovation pour les Lycées, 28 septembre 2023. <https://campus-innovation-lycees.fr/les-conferences-sur-lintelligence-artificielle/>.

<sup>105</sup> Mardi du Grand Continent, Le Cun, Yann, Marc Mézard, Gabriela Ramos, Jamal Atif, Frédéric Worms, et Anne Bouverot. « La société face aux défis de l'intelligence artificielle », [Vidéo en ligne][1 h 31 mn 6s]. ENS: Mardi du Grand Continent, 12 décembre 2023. <https://www.youtube.com/watch?v=ZAoRJtb8KME>. [Consulté le 14/01/2024]

<sup>106</sup> Espace des sciences, Petit, Antoine. « Conférence : L'intelligence artificielle, une opportunité ou un risque ? », [Vidéo en ligne][1 h 41 mn 30 s]. Auditorium Hubert Curien de Rennes: Espace des sciences, 26 septembre 2023. Disponible sur : <https://www.espace-sciences.org/conferences/lintelligence-artificielle-une-opportunite-ou-un-risque>. [Consulté le 14/01/2024]

était largement représentées. Nous avons également remarqué que de grandes questions de droit étaient centrales dans les articles. Nous avons finalement découpé notre étude en quatre parties qui incluent toutes les dimension ressortant de l'étude à savoir une première partie sur l'histoire de l'IA générative, puis un retour sur les divers conflits légaux et approches, pour traiter des apports et limites pour certains secteurs économiques, avant de conclure sur l'analyse des risques existentiels et l'opposition entre développement rapide et arrêt brutal de la technologie.



Analyse context sur 1000 articles français d'Europresse traitant de "ChatGPT"<sup>107</sup>

<sup>107</sup> Visualisation Cortex disponible sur :

<https://labeleditor.cortex.net/?svg=https%3A%2F%2Fassets.cortex.net%2Fdocs%2F0ec2642a6ced9cc33581c6cda6f89e51>

## Bibliographie

### Références de la littérature scientifiques, et presse experte

- Alan Turing : *la pensée informatique*. Docsciences n°14, Canopé (CRDP de l'académie de Versailles), 2012, p.5 [Consulté le 14/01/2023]
- Benbouzid, Bilel, Yannick Meneceur, et Nathalie Alisa Smuha. « Quatre nuances de régulation de l'intelligence artificielle. Une cartographie des conflits de définition ». *Réseaux* 232-233, n° 2-3 (2022): 29-64. <https://doi.org/10.3917/res.232.0029>.
- Bostrom, Nick. « Seasons of Hope and Despair ». In *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press., 25:285-89, 2015. <https://doi.org/10.1007/s11023-015-9377-7>. [Consulté le 14/01/2023]
- Noy, Shakked, et Whitney Zhang. « Experimental evidence on the productivity effects of generative artificial intelligence ». *Science* 381, n° 6654 (14 juillet 2023): 187-92. <https://doi.org/10.1126/science.adh2586>.
- Richard, Philippe. « ChatGPT : les pirates peuvent-ils en profiter pour multiplier les cyberattaques ? » *Techniques de l'Ingénieur*. Disponible sur : <https://www.techniques-ingénieur.fr/actualité/articles/chatgpt-les-pirates-peuvent-ils-en-profiter-pour-multiplier-les-cyberattaques-120102/> [Consulté le 21 janvier 2024]
- Turing, Alan. « I-COMPUTING MACHINERY AND INTELLIGENCE ». *Mind* LIX, n° 236 (octobre 1950): 433-60. <https://doi.org/10.1093/mind/LIX.236.433>. [Consulté le 14/01/2023]
- Université Dauphine PSL, « IA et société : nouvelle donne, nouveaux enjeux », Université Dauphine PSL, 20-23 novembre 2023. Disponible sur <https://dauphine.psl.eu/dauphine-digital-days>. [Consulté le 21/01/2023]

### Références de littératures grise académique : rapports et colloques

- Aghion, Philippe. « Conférence : Les conséquences économiques de l'IA ». 28 septembre 2023, [Vidéo en ligne][36 mn]. Collège de France: Campus de l'innovation pour les lycées. <https://www.youtube.com/watch?v=ngEL1PjucrQ>. [Consulté le 14/01/2024]
- Aghion, Philippe, Suzanne Vergnolle, Benoît Sagot, Xavier Leroy, et Jean-François Bonnefon. « L'IA et ses défis ». In *Campus de l'innovation pour les lycées*, [Vidéos en ligne]. Collège de France: été organisé par le Campus de l'Innovation pour les Lycées, 2023. <https://campus-innovation-lycees.fr/les-conferences-sur-lintelligence-artifici/>.
- Gefen Alexandre. 2023. « Conférence : IA et création littéraire », [Vidéo en ligne] [46 mn 42 s]. Enssib. <https://www.youtube.com/watch?v=UlmdP-QeARM>. [Consulté le 14/01/2024]
- Bengio, Tegmark, Mitchell, et LeCun. « Munk Debate on Artificial Intelligence : “AI research and development poses an existential threat.” », [Vidéo en ligne][1 h 47 mn 19 s], 2023. <https://www.youtube.com/watch?v=144uOftr4SYA>.
- Bonnefon, Jean-François. 2023. « Conférence : Vers quelle société nous mènent l'IA et ses filtres ? » In , [Vidéo en ligne][30 mn 32 s]. Collège de France: Campus de l'innovation pour les lycées. <https://www.youtube.com/watch?v=2jOAeMcbgGc>. [Consulté le 14/01/2024]
- Bourgeot, Rémi. 2023. « AI Act : quel compromis pour permettre l'essor de l'intelligence artificielle européenne ? » *IRIS*. 18 décembre 2023. <https://www.iris-france.org/180995-ai-act-quel-compromis-pour-permettre-les-ressources-de-lintelligence-artificielle-europeenne/>. [Consulté le 17/01/2024]
- Espace des sciences, Petit, Antoine. « Conférence : L'intelligence artificielle, une opportunité ou un risque ?», [Vidéo en ligne][1 h 41 mn 30 s]. Auditorium Hubert Curien de Rennes: Espace des sciences, 26 septembre 2023. Disponible sur : <https://www.espace-sciences.org/conferences/l-intelligence-artificielle-une-opportunité-ou-un-risque>. [Consulté le 14/01/2024]

- Hugenoltz, P. Bernt, et João Pedro Quintais. « Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output? » *IIC - International Review of Intellectual Property and Competition Law* 52, n° 9 (1 octobre 2021): 1190-1216. <https://doi.org/10.1007/s40319-021-01115-0>. [Consulté le 10/01/2024]
- Langosco, Lauro, Jack Koch, Lee Sharkey, Jacob Pfau, Laurent Orseau, et David Krueger. « Goal Misgeneralization in Deep Reinforcement Learning ». arXiv, 9 janvier 2023. <https://doi.org/10.48550/arXiv.2105.14111>. [Consulté le 21/01/2023]
- Lechner, Marie. Conférence : « Entre spectacularisation et critique complaisante, quelle place pour une exploration artistique critique de l'IA ? », [Vidéo en ligne][1 h 20 mn 5 s], 5 décembre 2023. Disponible sur [https://www.youtube.com/watch?v=\\_luc49xJe0w](https://www.youtube.com/watch?v=_luc49xJe0w). [Consulté le 21/01/2024]
- Mardi du Grand Continent, Le Cun, Yann, Marc Mézard, Gabriela Ramos, Jamal Atif, Frédéric Worms, et Anne Bouverot. « La société face aux défis de l'intelligence artificielle », [Vidéo en ligne][1 h 31 mn 6s]. ENS: Mardi du Grand Continent, 2023. <https://www.youtube.com/watch?v=ZAoRJtb8KME>.
- Sassatelli, Gilles, Maël LE Treust, Thierry Artieres, David Ilcinkas, Myriam Caudrelier, et Caroline Appert. « Rapport de Prospective du Conseil Scientifique de l'Institut des Sciences Informatiques et leurs Interactions (INS2I) ». INS2I. [https://csi-ins2i.cnrs.fr/sites/default/files/general/pdf/Rapport\\_prospective\\_CSI\\_INS2I.pdf](https://csi-ins2i.cnrs.fr/sites/default/files/general/pdf/Rapport_prospective_CSI_INS2I.pdf). [Consulté le 9 janvier 2024.]
- Université Paris Dauphine - PSL. 2023. « Quelle utilisation de l'IA par les juristes ? » [Vidéo en ligne][2 h 46 mn 7 s]. Coord. Thierry Kirat. <https://www.youtube.com/watch?v=jz4snai1Bvs>. [consulté le 10/01/2024]
- Université Dauphine PSL, « IA et société : nouvelle donne, nouveaux enjeux », Université Dauphine PSL, 20-23 novembre 2023. Disponible sur <https://dauphine.psl.eu/dauphine-digital-days>. [Consulté le 21/01/2023]

### Références de littératures grise : actes juridiques, législatifs et rapports administratifs

- *Andersen v. Stability AI Ltd.*, No. 23-cv-00201-WHO (United States District Court, Northern District of California, 30 octobre 2023). Disponible sur <https://casetext.com/case/andersen-v-stability-ai-ltd> [consulté le 17/01/2023]
- Campagnolo, Côme. 2023. « Corée du Sud : état des lieux et stratégie en matière d'IA ». *Direction générale du Trésor* (blog). 31 août 2023. Disponible sur <https://www.tresor.economie.gouv.fr/Articles/2023/08/31/coree-du-sud-etat-des-lieux-et-strategie-en-matiere-d-i-a>. [consulté le 17 janvier 2024]
- Cour des comptes, « La stratégie nationale de recherche en intelligence artificielle ». 2023. Rapport de la Cour des comptes. Disponible sur : <https://www.ccomptes.fr/system/files/2023-04/20230403-synthese-strategie-nationale-recherche-intelligence-artificielle.pdf>. [Consulté le 14/01/2023]
- Fernandez, Angelina. 2022. « Regulating Deep Fakes in the Proposed AI Act ». *MediaLaws* (blog). 23 mars 2022. <https://www.medialaws.eu/regulating-deep-fakes-in-the-proposed-ai-act/>. [Consulté le 17/01/2024]
- INSEE, « Définition - Productivité ». 2016. *INSEE* (blog). 13 octobre 2016. <https://www.insee.fr/fr/metadonnees/definition/c1452>. [Consulté le 10/01/2024]
- Maison Blanche, *Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*, No. 14110, EO (2023). Disponible sur <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>. [Consulté le 20/01/2024]
- Parlement européen, « EU AI Act: First Regulation on Artificial Intelligence ». 8 juin 2023. <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>. [Consulté le 17/01/2024]
- Union européenne, « Types de législation ». s. d. European-union.europa.eu. [https://european-union.europa.eu/institutions-law-budget/law/types-legislation\\_fr](https://european-union.europa.eu/institutions-law-budget/law/types-legislation_fr). [Consulté le 17 janvier 2024]

## Références dans la presse grand public

- Arte 28 minutes (Emission TV) «*Hors-série IA | Aurélie Jean et Raphaël Doan*», 2 janvier 2024. Disponible sur : <https://www.youtube.com/watch?v=663YRhkVqgw>. [Consulté le 14/01/24]
- Bardou, Florian. « Les architectes face à l'IA : «Même pas peur» ». *Libération*, sect. Design. [https://www.liberation.fr/lifestyle/design/les-architectes-face-a-lia-meme-pas-peur-20230619\\_M4VUTGNIRBHZ7JELU7EP3OW444/](https://www.liberation.fr/lifestyle/design/les-architectes-face-a-lia-meme-pas-peur-20230619_M4VUTGNIRBHZ7JELU7EP3OW444/). [Consulté le 10/01/2023]
- Bertuzzi, Luca. « AI Act : négociations bloquées à cause de divergences sur les modèles de fondation ». *Euractiv*, 13 novembre 2023. Disponible sur : <https://www.euractiv.fr/section/intelligence-artificielle/news/ai-act-negociations-bloquees-a-cause-de-divergences-sur-les-modeles-de-fondation/>. [Consulté le 18/01/2024]
- Bellefqih, Réduane. « Course à l'innovation dans l'IA générative : un double défi financier et humain ». *La Tribune*, 11 janvier 2024. Disponible sur <https://www.latribune.fr/supplement/la-tribune-now/course-a-l-innovation-dans-l-ia-generative-un-double-defi-financier-et-humain-987546.html>. [Consulté le 18/01/2024]
- Bembaron, Elsa. 2022. « La consommation d'eau des data centers source d'inquiétudes ». *Le Figaro*, 19 août 2022, sect. Tech & Web. <https://www.lefigaro.fr/secteur/high-tech/la-consommation-d-eau-des-data-centers-source-d-inquietudes-20220819>. [Consulté le 13/01/2023]
- Bernier, Martin. « Cols bleus, blancs, roses: serons-nous tous au chômage à cause de l'IA? » *Le Figaro* (blog), 10 août 2023. <https://www.lefigaro.fr/vox/economie/cols-bleus-blancs-roses-serons-nous-tous-au-chomage-a-cause-de-l-ia-20230810>. [Consulté le 14/10/24]
- Bottero, Laurence. « “L'IA Act ne doit pas freiner l'innovation” (Aurélie Jean) ». *La Tribune*, 23 novembre 2023, sect. Informatique. <https://www.latribune.fr/technos-medias/informatique/l-ia-act-ne-doit-pas-freiner-l-innovation-aurelie-jean-983859.html>. [Consulté le 18/01/2024]
- Breton, Thierry. « L'IA Act n'entrave aucunement la révolution en mouvement de l'intelligence artificielle ». *Le Monde.fr*, 16 décembre 2023. [https://www.lemonde.fr/idees/article/2023/12/16/thierry-breton-l-ia-act-n-entrave-aucunement-la-revolution-en-mouvement-de-l-intelligence-artificielle\\_6206177\\_3232.html](https://www.lemonde.fr/idees/article/2023/12/16/thierry-breton-l-ia-act-n-entrave-aucunement-la-revolution-en-mouvement-de-l-intelligence-artificielle_6206177_3232.html). [Consulté le 18/01/2024]
- Cario, Erwan, Clémence Mary, et Elise Viniacourt. « Cédric Villani : “Ne cédon pas à l'illusion que l'IA va tout solutionner” ». *Libération*, 23 juin 2023, sect. Idées et Débats [https://www.liberation.fr/idees-et-debats/cedric-villani-ne-cedons-pas-a-lillusion-que-ia-va-tout-solutionner-20230619\\_IJTK7TTZFACLJZPE6UGEKWTP4/](https://www.liberation.fr/idees-et-debats/cedric-villani-ne-cedons-pas-a-lillusion-que-ia-va-tout-solutionner-20230619_IJTK7TTZFACLJZPE6UGEKWTP4/) [Consulté le 10/01/2023]
- Cassidy, Caitlin. « Artificial Intelligence Such as ChatGPT to Be Allowed in Australian Schools from 2024 ». *The Guardian*, 6 octobre 2023, sect. Australia news. <https://www.theguardian.com/australia-news/2023/oct/06/chatgpt-ai-allowed-australian-schools-2024>. [Consulté le 14/01/24]
- Chalkbeat. « NYC Blocks Access to ChatGPT on School Networks as Cheating Fears Swirl », 3 janvier 2023. Disponible sur <https://www.chalkbeat.org/newyork/2023/1/3/23537987/nyc-schools-ban-chatgpt-writing-artificial-intelligence-L>. [Consulté le 14/01/24]
- Cho, Winston. « Artists Lose First Round of Copyright Infringement Case Against AI Art Generators ». *The Hollywood Reporter*, 30 octobre 2023. Disponible sur <https://www.hollywoodreporter.com/business/business-news/artists-copyright-infringement-case-ai-art-generators-1235632929/>. [Consulté le 17/01/2024]

- Coulter, Martin. « What Is the EU AI Act and When Will Regulation Come into Effect? » *Reuters*, 7 décembre 2023, sect. Technology. Disponible sur <https://www.reuters.com/technology/what-are-eus-landmark-ai-rules-2023-12-06/>. [Consulté le 17/01/2024]
- *Courrier international*. « Espagne. En pleine sécheresse, la consommation d'eau du futur data center de Meta inquiète ». 10 mai 2023, sect. Sciences et environnement. <https://www.courrierinternational.com/article/espagne-en-pleine-secheresse-la-consommation-d-eau-du-futur-data-center-de-meta-inquiete>. [Consulté le 13/01/2023]
- Crequer, Tudi. 2021. « Trop de « data centers », l'Irlande risque la panne électrique ». *Reporterre*. 2 décembre 2021. <https://reporterre.net/Trop-de-data-centers-l-Irlande-risque-la-panne-electrique>. [Consulté le 13/01/2023]
- Euronews, « The Cube : une présentatrice TV générée par l'IA ». 16 avril 2023. Disponible sur : <https://fr.euronews.com/2023/04/16/the-cube-presentatrice-tv-generee-par-lintelligence-artificielle-au-koweit>. [Consulté le 14/01/2024]
- Espiner, Tom. 2023. « BT to Cut 55,000 Jobs with up to a Fifth Replaced by AI ». *BBC News*, 18 mai 2023, sect. Business. <https://www.bbc.com/news/business-65631168>. [Consulté le 14/01/2024]
- Fagot, Vincent. « AI Act : l'Union européenne pionnière dans la régulation de l'intelligence artificielle ». *Le Monde.fr*, 9 décembre 2023. [https://www.lemonde.fr/economie/article/2023/12/09/ai-act-l-union-europeenne-pionniere-dans-la-regulation-de-l-intelligence-artificielle\\_6204830\\_3234.html](https://www.lemonde.fr/economie/article/2023/12/09/ai-act-l-union-europeenne-pionniere-dans-la-regulation-de-l-intelligence-artificielle_6204830_3234.html). [Consulté le 18/01/2024]
- Faguy, Ana. « New York City Public Schools Reverses ChatGPT Ban ». *Forbes* (blog). 18 mai 2023, <https://www.forbes.com/sites/anafaguy/2023/05/18/new-york-city-public-schools-reverses-chatgpt-ban/>. [Consulté le 21 janvier 2024.]
- Franceinfo, « Les journaux quotidiens “L'Est républicain” et “Vosges Matin” testent l'intelligence artificielle pour la correction d'articles ». 26 octobre 2023. Disponible sur [https://www.francetvinfo.fr/internet/intelligence-artificielle/les-journaux-quotidiens-l-est-republicain-et-vosges-matin-testent-l-ia-pour-la-correction-d-articles\\_6145410.html](https://www.francetvinfo.fr/internet/intelligence-artificielle/les-journaux-quotidiens-l-est-republicain-et-vosges-matin-testent-l-ia-pour-la-correction-d-articles_6145410.html). [Consulté le 14/01/2024]
- Gless, Etienne. 2023. « L'IA peut-elle sauver la planète ? » *Le Figaro Étudiant*, 6 décembre 2023. <https://etudiant.lefigaro.fr/article/etudes/l-ia-peut-elle-sauver-la-planete-20231206/>. [Consulté le 10/12/2023]
- Hern, Alex, et Kiran Stacey. « No 10 Acknowledges ‘Existential’ Risk of AI for First Time ». *The Guardian*, 25 mai 2023, sect. Technology. <https://www.theguardian.com/technology/2023/may/25/no-10-acknowledges-existential-risk-ai-first-time-rishi-unak>. [Consulté le 21 janvier 2024]
- Hu, Krystal, et Krystal Hu. « ChatGPT Sets Record for Fastest-Growing User Base - Analyst Note ». *Reuters*, 2 février 2023, sect. Technology. Disponible sur <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>. [Consulté le 20 janvier 2023]
- Lastmann, Samy. « Qu'est-ce que ChatGPT, l'IA conversationnelle d'OpenAI ? Définition, contexte et enjeux (2023) ». *Smart tribune* (blog), 22 décembre 2022. <https://blog.smart-tribune.com/fr/ia-conversationnelle-chatgpt-openai-definition>. [consulté le 16/12/2023]
- Le Grand Continent, « ChatGPT perd des utilisateurs et annonce une innovation majeure » 26 septembre 2023. Disponible sur <https://legrandcontinent.eu/fr/2023/09/26/chatgpt-perd-des-utilisateurs-et-annonce-une-innovation-majeure/>. [Consulté le 10/01/2024]
- Legros, Claire. « Yann Le Cun, directeur à Meta : « L'idée même de vouloir ralentir la recherche sur l'IA s'apparente à un nouvel obscurantisme » ». *Le Monde.fr*, 28 avril 2023. [https://www.lemonde.fr/idees/article/2023/04/28/yann-le-cun-directeur-a-meta-l-idee-meme-de-vouloir-raentir-la-recherche-sur-l-ia-s-apparente-a-un-nouvel-obscourtantisme\\_6171338\\_3232.html](https://www.lemonde.fr/idees/article/2023/04/28/yann-le-cun-directeur-a-meta-l-idee-meme-de-vouloir-raentir-la-recherche-sur-l-ia-s-apparente-a-un-nouvel-obscourtantisme_6171338_3232.html). [consulté le 21/12/2023]
- *Le Monde.fr*. « Hawking : “L'intelligence artificielle pourrait mettre fin à l'humanité” ». 3 décembre 2014.

- [https://www.lemonde.fr/pixels/article/2014/12/03/hawking-l-intelligence-artificielle-pourrait-mettre-fin-a-l-humanite\\_4533135\\_4408996.html](https://www.lemonde.fr/pixels/article/2014/12/03/hawking-l-intelligence-artificielle-pourrait-mettre-fin-a-l-humanite_4533135_4408996.html). [consulté le 21/12/2023]
- *Le Monde.fr*. « A Montréal, l'un des pères de l'intelligence artificielle alerte sur une menace existentielle pour l'homme ». 2 juin 2023. Disponible sur :  
[https://www.lemonde.fr/international/article/2023/06/02/a-montreal-l-un-des-peres-de-l-intelligence-artificielle-alerte-sur-une-menace-existentielle-pour-l-homme\\_6175814\\_3210.html](https://www.lemonde.fr/international/article/2023/06/02/a-montreal-l-un-des-peres-de-l-intelligence-artificielle-alerte-sur-une-menace-existentielle-pour-l-homme_6175814_3210.html). [Consulté le 21 janvier 2024]
  - Monod, Olivier. « L'IA peut-elle sauver la planète ? » *Libération*, 19 juin 2023, sect. Economie numérique.  
[https://www.liberation.fr/economie/economie-numerique/lia-peut-elle-sauver-la-planete-20230619\\_LROWDIFDWNDRFL6RAAUL6JMZZE/](https://www.liberation.fr/economie/economie-numerique/lia-peut-elle-sauver-la-planete-20230619_LROWDIFDWNDRFL6RAAUL6JMZZE/). [Consulté le 13/01/2023]
  - Nast, Condé. « Elon Musk's Billion-Dollar Crusade to Stop the A.I. Apocalypse ». *Vanity Fair*, 26 mars 2017.  
<https://www.vanityfair.com/news/2017/03/elon-musk-billion-dollar-crusade-to-stop-ai-space-x>. [Consulté le 21/01/2024]
  - Perrigo, Billy. 2023. « Big Tech Is Already Lobbying to Water Down Europe's AI Rules ». *TIME*. 21 avril 2023.  
<https://time.com/6273694/ai-regulation-europe/>. [Consulté le 17/01/2024]
  - Pichai, Sundar. 2023. « Google CEO: Building AI responsibly is the only race that really matters ». *Financial Times*, 22 mai 2023. <https://www.ft.com/content/8be1a975-e5e0-417d-af51-78af17ef4b79> [Consulté le 13/01/2023]
  - Piquard, Alexandre. « OpenAI : aux sources de l'affaire Sam Altman ». *Le Monde.fr*, 24 novembre 2023.  
[https://www.lemonde.fr/economie/article/2023/11/24/openai-aux-sources-de-l-affaire-sam-altman\\_6202096\\_3234.html](https://www.lemonde.fr/economie/article/2023/11/24/openai-aux-sources-de-l-affaire-sam-altman_6202096_3234.html). [consulté le 21/12/2023]
  - Quach, Katyanna. « EU Legislates Disclosure of Copyright Data Used to Train AI ». *The Register* (blog), 1 mai 2023. Disponible sur [https://www.theregister.com/2023/05/01/eu\\_ai\\_act\\_adds\\_new/](https://www.theregister.com/2023/05/01/eu_ai_act_adds_new/). [Consulté le 17/01/2023]
  - Rignell, Milo. « Investir l'IA sûre et digne de confiance : un impératif européen, une opportunité française ». Institut Montaigne, avril 2023. Disponible sur  
<https://www.institutmontaigne.org/publications/investir-lia-sure-et-digne-de-confiance-un-imperatif-europeen-ne-opportunité-française>. [Consulté le 19/01/2023]
  - RTBF, « Jade, un avatar généré par l'IA, présente la météo en Suisse ». Disponible sur  
<https://www.rtb.be/article/jade-un-avatar-genere-par-lia-presente-la-meteo-en-suisse-11187917>. [Consulté le 14 janvier 2024]
  - Toussay, Jade. « Le ministre de l'Éducation Pap Ndiaye met les élèves en garde sur l'utilisation de ChatGPT à l'école ». *Le HuffPost* (blog), 2 février 2023.  
[https://www.huffingtonpost.fr/politique/article/chatgpt-doit-etre-encadre-a-l-ecole-selon-pap-ndiaye\\_213550.html](https://www.huffingtonpost.fr/politique/article/chatgpt-doit-etre-encadre-a-l-ecole-selon-pap-ndiaye_213550.html). [Consulté le 14/01/24]
  - Wikipédia « Histoire de l'intelligence artificielle », 16 janvier 2024. Disponible sur  
[https://fr.wikipedia.org/w/index.php?title=Histoire\\_de\\_l%27intelligence\\_artificielle&oldid=211570802](https://fr.wikipedia.org/w/index.php?title=Histoire_de_l%27intelligence_artificielle&oldid=211570802). [Consulté le 20 janvier 2023]

## Références de littérature grise : rapports de *think tanks*, communications d'organisations, réseaux sociaux

- Accenture,. « DiPA | Innovative Chatbot ». Consulté le 17 janvier 2024. Disponible sur  
<https://www.accenture.com/fr-fr/case-studies/about/need-answers-ask-dipa>. [consulté le 17/01/2024]
- Allain, Tanguy. 2023. « Quel cadre juridique pour les deepfake audio ? » *Tnilex - cabinet d'avocat*. 2 mai 2023.  
<https://tnilex-avocat.fr/cadre-juridique-deepfake/>. [consulté le 17/01/2024]
- Anthropic, « Anthropic's Responsible Scaling Policy, version 1.0 ». 19 septembre 2023.  
<https://www-files.anthropic.com/production/files/responsible-scaling-policy-1.0.pdf>. [Consulté le 18/01/2023]

- Bain. « Bain & Company fait alliance avec les services d'OpenAI afin d'aider ses clients à exploiter le plein potentiel de l'IA », 21 février 2023. Disponible sur <https://www.bain.com/fr/a-propos-de-bain/media-center/communiques-de-presse/2023/bain--company-annonces-services-alliance-with-openai-to-help-enterprise-clients-identify-and-realize-the-full-potential-and-maximum-value-of-ai/>. [Consulté le 17/01/2024]
- Botpress Blog. « Comment fonctionne ChatGPT ? », 26 avril 2023. Disponible sur <https://botpress.com/fr/blog/how-does-chatgpt-work>. [consulté le 15/12/23]
- Braunschweig, Bertrand. « Les cinq murs de l'intelligence artificielle: » *Futuribles* N° 453, n° 2 (20 février 2023): 5-24. <https://doi.org/10.3917/futur.453.0005>. [Consulté le 10/12/2023]
- EffiSciences (blog) , « Pole IA - EffiSciences » Disponible sur : <https://ia.effisciences.org/>. [Consulté le 21 janvier 2024]
- Future of Life Institute. « Pause Giant AI Experiments: An Open Letter », 22 mars 2023. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. [Consulté le 18/01/2024]
- Goldman Sachs, « Generative AI Could Raise Global GDP by 7% ». 2023, 5 avril 2023. <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>. [Consulté le 14/01/2024]
- Google Deepmind. « Responsibility & Safety ». Google DeepMind. Disponible sur : <https://deepmind.google/about/responsibility-safety/>. [Consulté le 18 janvier 2024.]
- Groupe Usine d'électricité de Metz. « Le groupe UEM, mécène de l'œuvre de Refik Anadol, Machine Hallucinations - Rêves de nature au Centre Pompidou-Metz », 13 juin 2022. Disponible sur : <https://www.uem-metz.fr/le-groupe-uem-mecene-de-loeuvre-de-refik-anadol-machine-hallucinations-reves-de-nature/>. [Consulté le 18/01/2024]
- Holubowicz, Gerald. s. d. « Notion – The all-in-one workspace for your notes, tasks, wikis, and databases. » *journalismdesign* <https://journalismdesign.notion.site/1ee57ee51a804dc5acd788d69aa5b17f?v=08f5eb734f264398a01f49f6d10835e2>. [Consulté le 14 janvier 2024.]
- Kojola, Erik. 2023. « ChatCO2 – Safeguards Needed For AI's Climate Risks - ». 30 novembre 2023. <https://www.greenpeace.org/usa/chatco2-safeguards-needed-for-ais-climate-risks/>. [Consulté le 14/01/2023]
- Krakovna, Victoria, Jonathan Uesato, Vladimir Mikulik, Matthew Rahtz, Tom Everitt, Ramana Kumar, Zac Kenton, Jan Leike, et Shane Legg. « Specification Gaming: The Flip Side of AI Ingenuity ». *Google DeepMind*, 21 avril 2020. <https://deepmind.google/discover/blog/specification-gaming-the-flip-side-of-ai-ingenuity/>. [Consulté le 21 janvier 2024].
- Lemarchand, Stéphane, Jeanne Dozier, Aurelia Pons, et Maria Aouad. « La Commission européenne publie deux propositions de directive en matière de responsabilité applicables aux produits basés sur l'intelligence artificielle ». *DLA Piper* (blog), 15 février 2023. Disponible sur <https://www.dlapiper.com/fr-fr/insights/publications/bulletins-contrats-commerciaux/2023/contrats-commerciaux-bulletin-1/la-commission-europeenne-publie-deux-propositions-de-directive-en-matiere-de-responsabilite>. [Consulté le 17/01/2024]
- Maggie Hassan (blog) « Senator Hassan Highlights AI Threats to National S... | U.S. Senator Maggie Hassan of New Hampshire », 20 septembre 2023. Disponible sur : <https://www.hassan.senate.gov/news/press-releases/senator-hassan-highlights-ai-threats-to-national-security-and-public-safety-at-senate-hearing>. [Consulté le 21 janvier 2024]
- Moës, Nicolas « Hauts risques et forts impacts : réguler l'IA ». Institut Montaigne. Disponible sur <https://www.institutmontaigne.org/expressions/hauts-risques-et-forts-impacts-reguler-ia>. [Consulté le 21 janvier 2024]
- Pettineo, Camille. 2024. « Les médias face à l'intelligence artificielle : 20 chartes passées au crible ». *La Revue des Médias | INA* (blog). 11 octobre 2024.

- <https://larevuedesmedias.ina.fr/les-medias-face-lintelligence-artificielle-20-chartes-passees-au-crible>. [Consulté le 14/01/2024]
- OCDE. *Perspectives de l'emploi de l'OCDE 2023: Intelligence artificielle et marché du travail.* Perspectives de l'emploi de l'OCDE. OECD, 11 juillet 2023. Disponible sur <https://doi.org/10.1787/aae5dba0-fr>. [Consulté le 14/01/2024]
  - OpenAI. « OpenAI's Approach to Frontier Risk », 26 octobre 2023. <https://openai.com/global-affairs/our-approach-to-frontier-risk>. [Consulté le 18/01/2023]
  - OpenAI. « Introducing OpenAI ». [Https://openai.com/](https://openai.com/blog/introducing-openai). Disponible sur <https://openai.com/blog/introducing-openai>. [Consulté le 22 décembre 2023.]
  - Reporters sans frontières. 2023. *Charte de Paris sur l'IA et le journalisme.* [https://rsf.org/sites/default/files/medias/file/2023/11/Charte%20de%20Paris%20sur%20l%27IA%20et%20le%20journalisme\\_1.pdf](https://rsf.org/sites/default/files/medias/file/2023/11/Charte%20de%20Paris%20sur%20l%27IA%20et%20le%20journalisme_1.pdf). [Consulté le 13/01/2024]
  - Rignell, Milo. « IA : l'ambivalence entre course technologique et gouvernance mondiale ». *Institut Montaigne* (blog), 20 juillet 2023. Disponible sur <https://www.institutmontaigne.org/expressions/ia-lambivalence-entre-course-technologique-et-gouvernance-mondiale> [Consulté le 19/01/2023]
  - Rignell, Milo. « Investir l'IA sûre et digne de confiance : un impératif européen, une opportunité française ». Institut Montaigne, avril 2023. Disponible sur <https://www.institutmontaigne.org/publications/investir-lia-sure-et-digne-de-confiance-un-imperatif-europeen-une-opportunité-française>. [Consulté le 19/01/2023]
  - Sam Altman [@sama]. « ChatGPT Launched on Wednesday. Today It Crossed 1 Million Users! » Tweet. *Twitter*, 5 décembre 2022. Disponible sur <https://twitter.com/sama/status/1599668808285028353>. [Consulté le 20 janvier 2023]
  - Sauvage, M. Manuel S. *La brève histoire de ChatGPT - Tout savoir sur les fondements de ChatGPT et son avenir.* Independently published, 2023. [consulté le 21/12/2023]
  - Sciences Po. « ChatGPT : Sciences Po fixe des règles et lance une réflexion sur l'IA dans l'enseignement supérieur », 9 février 2023. Disponible sur : <https://www.sciencespo.fr/fr/actualites/sciences-po-fixe-des-regles-claires-sur-lutilisation-de-chat-gpt-par-les-etudiants>. [Consulté le 14/01/24]
  - Syndicat National des Journalistes, « L'Est Républicain - Vosges Matin | La direction voudrait utiliser la moulinette ChatGPT pour retraitier les textes des correspondants ». 10 octobre 2023. [https://snj.fr/article/la-direction-voudrait-utiliser-la-moulinette-chatgpt-pour-retraitier-les-textes-des-correspondants\\_1565254191](https://snj.fr/article/la-direction-voudrait-utiliser-la-moulinette-chatgpt-pour-retraitier-les-textes-des-correspondants_1565254191). [Consulté le 14 janvier]
  - The Future Society. « Overview ». Disponible sur : <https://thefuturesociety.org/about-us/>. [Consulté le 21 janvier 2024]
  - *Tu Nube Seca Mi Río* (blog), « La Televisión francesa se hace eco del Centro de Datos de Talavera de la Reina ». 2023. 26 septembre 2023. <https://tunubescamirio.com/>. [Consulté le 14/01/2023]
  - Villani, Cédric. « Donner un sens à l'intelligence artificielle ». Assemblée nationale (Rapport parlementaire), 28 mars 2018. Disponible sur : [https://medias.vie-publique.fr/data\\_storage\\_s3/rapport/pdf/184000159.pdf](https://medias.vie-publique.fr/data_storage_s3/rapport/pdf/184000159.pdf). [Consulté le 14/01/24]
  - World Intellectual Property Organization , « WIPO Technology Trends Artificial Intelligence 2019». Genève, 2019. 158 pages, Disponible sur <https://www.wipo.int/publications/en/details.jsp?id=4386>. [Consulté le 21 décembre 2023.]
  - Vyacheslav, Eric. « Yann is brilliant. Can he just be the president of AI already? » *Linkedin* (réseau social), décembre 2023. Disponible sur : [https://www.linkedin.com/posts/eric-vyacheslav-156273169\\_yann-is-brilliant-can-he-just-be-the-president-ugcPo](https://www.linkedin.com/posts/eric-vyacheslav-156273169_yann-is-brilliant-can-he-just-be-the-president-ugcPo)

[st-7137148094122541056-Fqam/?utm\\_source=share&utm\\_medium=member\\_desktop](https://www.semptech.fr/st-7137148094122541056-Fqam/?utm_source=share&utm_medium=member_desktop). [Consulté le 18/01/2024]