

# **La société face aux avancées des sciences et des techniques**

## **Le cas de l'intelligence artificielle et de la génétique**

Par Virginie Courtier-Orgogozo<sup>1</sup> et Laurence Devillers<sup>2</sup>

*article publié dans la revue Futuribles le 26 décembre 2023*

lien : <https://www.futuribles.com/la-societe-face-aux-avancees-des-sciences-et-des-techniques/>

1. Directrice de recherche CNRS (Centre national de la recherche scientifique), spécialisée sur les mécanismes de l'évolution, la génétique et le forçage génétique ; professeure associée à l'École polytechnique ; professeure attachée à la chaire Biodiversité et écosystèmes du Collège de France ; et conseillère scientifique de Futuribles International.

2. Professeure et chercheuse, spécialiste des interactions humain-machine ; enseignante à Sorbonne Université et membre du LIMSI (Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur) / CNRS ; membre du CNPEN (Comité national pilote d'éthique du numérique).

**Les avancées scientifiques et technologiques font régulièrement la une des médias, comme en témoignent, par exemple, les nombreux articles et commentaires autour des progrès de l'intelligence artificielle (IA), en particulier depuis le lancement du célèbre ChatGPT fin 2022. Cependant, au-delà des anecdotes sur la façon dont peut être utilisé un agent conversationnel comme ChatGPT et sans contester les questions soulevées par ses usages au quotidien, les individus sont-ils bien et suffisamment informés sur cette nouvelle technologie qu'est l'IA ? Disposent-ils d'une vision claire et précise leur permettant d'envisager les enjeux multiples et majeurs qui s'y rapportent ? De même s'agissant des avancées en matière de génétique : sait-on précisément où en est la recherche et quelles sont les applications potentielles qui pourraient en découler dans les années à venir (y compris grâce à l'IA) ?**

**Dans ces deux domaines scientifiques majeurs pour notre avenir, comme dans de nombreux autres, nous sommes certes informés, mais parfois de manière parcellaire, incomplète, et sans un certain nombre de précautions pourtant indispensables. Face aux idées reçues et pour réconcilier les individus avec une science qu'ils ont parfois tendance à regarder avec défiance, il est essentiel de leur fournir des éléments de compréhension précis, de leur dire ce que l'on sait (ou pas) des technologies en cours de développement, de leur montrer les enjeux, les limites, le tout de manière la plus pédagogique possible.**

**C'est ainsi qu'ils pourront disposer des clefs nécessaires pour se faire une opinion documentée et non mystifiée des enjeux sociétaux qui se rapportent à ces technologies. Dans cet article, les chercheuses Virginie Courtier-Orgogozo et Laurence Devillers explorent ces enjeux pédagogiques, au travers des exemples des progrès de la génétique et de l'IA, et proposent des pistes « pour mieux armer le grand public face aux avancées scientifiques et technologiques ». S.D.**

Dans un monde de plus en plus complexe et technologique, où les avancées sont de plus en plus rapides, le grand public peine parfois à saisir les enjeux et les éléments clefs pour comprendre et utiliser au mieux les nouvelles technologies. Après avoir explicité les raisons générales de ce décrochage, et l'intérêt d'une société bien formée aux technologies, nous présenterons quelques exemples de compréhension erronée des systèmes d'intelligence artificielle (IA) et des biotechnologies en génétique, qui peuvent conduire la société à prendre de mauvaises décisions. Puis nous proposerons quelques pistes pour mieux armer le grand public face aux avancées scientifiques et technologiques.

## **Un décrochage manifeste**

Si dans certaines régions du monde, le manque d'éducation du grand public aux nouvelles technologies peut s'expliquer par un manque de ressources ou même par la volonté délibérée de certains gouvernements autoritaires de tenir la population à l'écart afin de mieux contrôler la diffusion de l'information et de prévenir les mobilisations, dans les pays riches à régime démocratique, les raisons de ce décrochage sont autres et multiples.

D'une part, l'inertie institutionnelle et le manque de compréhension des enjeux peuvent ralentir les politiques et les programmes d'éducation aux nouvelles technologies. Ce phénomène est d'autant plus important que les technologies évoluent de plus en plus vite. D'autre part, certaines entreprises technologiques peuvent parfois hésiter à éduquer le public sur les failles de sécurité ou les préoccupations liées à la vie privée de leurs produits, afin de protéger leurs intérêts commerciaux et de ne pas nuire à leur image de marque. Il est parfois avancé également que certaines technologies actuelles sont trop complexes pour être comprises par le citoyen lambda. Nous récusons cette idée et pensons au contraire qu'il est tout à fait possible aujourd'hui de mieux former le grand public aux nouvelles technologies et enjeux associés, en améliorant la vulgarisation, les moyens de communication et la prise de conscience de chacun.

Face à la désinformation, aux manipulations de l'information et aux informations contradictoires, une personne ayant un esprit critique et un savoir avancé saura mieux réagir, s'adapter aux changements technologiques en cours, participer activement à la société et prendre des décisions éclairées sur des questions touchant à sa vie privée, telles que le choix entre plusieurs traitements médicaux potentiels, la sécurité en ligne, un vote électoral ou la protection de la vie privée. Une meilleure compréhension des sciences et technologies qui nous entourent permettra également de réduire l'anxiété et de vivre de manière plus sereine. De plus, une société bien éduquée aux sciences est mieux préparée à participer activement à des débats publics sur des questions scientifiques et technologiques, telles que la réglementation de nouvelles technologies ou la gestion des ressources naturelles. Enfin, de nombreux problèmes mondiaux, comme le changement climatique, la crise de la biodiversité, la sécurité alimentaire ou l'accès à l'eau potable, exigent vraisemblablement des solutions plurielles comportant une part à la fois scientifique et technologique. Une population bien éduquée est plus en mesure de pouvoir contribuer à la résolution de ces problèmes. Une des voies essentielles pour nos sociétés est donc de mieux former les citoyens et les politiques aux nouvelles technologies et aux enjeux associés.

## **Dans le domaine de l'IA**

### ***L'IA est-elle intelligente ?***

L'intelligence artificielle (IA) est un ensemble de théories, d'algorithmes et de logiciels qui ont pour objectif de simuler certaines capacités cognitives de l'être humain, comme les capacités d'apprentissage, de perception, d'aide à la décision et de génération. Pour l'instant, les programmes d'IA sont plutôt performants pour la perception et la génération, mais ils n'ont pas de capacités de raisonnement semblables à celles d'un être humain pour décider. Ils ne forment pas des concepts, ne comprennent pas ce qu'ils produisent et n'appliquent pas de logique au sens humain. Ils sont, en revanche, capables d'apprendre et de généraliser, certes de façon très différente des êtres humains.

Il existe de nombreux types de modèles utilisés en apprentissage machine (machine learning en anglais) comme les arbres de décision, la classification bayésienne, les réseaux de neurones, etc. L'apprentissage machine à partir d'une structure de type réseau de neurones artificiels, cherche à imiter de façon très simplifiée notre fonctionnement cérébral. Warren McCulloch et Walter Pitts

proposent le premier neurone artificiel en 1943, c'est un modèle très simplifié du neurone biologique. L'apprentissage de la machine repose sur l'idée de plasticité cérébrale décrit par la « règle de Hebb » en 1947, c'est-à-dire la capacité des neurones à modifier de façon durable l'efficacité de leur connexion synaptique représentée par des poids (des réels) dans les réseaux artificiels. Le premier réseau capable d'apprendre des fonctions complexes est le perceptron multicouche avec l'algorithme de rétropropagation du gradient d'erreur ; il date de 1987. Les dernières avancées qui ont bouleversé les performances obtenues en apprentissage machine sont dues principalement à l'augmentation des capacités de calcul, à l'accès à de gigantesques ensembles de données et à des algorithmes de plus en plus sophistiqués, comme les algorithmes neuronaux d'apprentissage profond ou *deep learning* (2010) et les *transformers*<sup>1</sup> (2017).

Deux types majeurs d'algorithmes neuronaux existent : les systèmes d'IA prédictive et ceux d'IA générative. L'utilisation du deep learning pour la reconnaissance d'images de chats est un exemple de système d'IA prédictive. En effet, ce type d'IA est formé sur des données existantes (dans ce cas, des images de chats) et utilise ces informations pour prédire ou classer de nouvelles données (par exemple, déterminer si une nouvelle image contient un chat ou non). Dans ce contexte, l'IA prédictive est utilisée pour prédire la présence d'un chat dans une image donnée. Pour entraîner ce type d'IA, l'ingénieur en apprentissage automatique écrit un programme qui implémente un réseau de neurones. Les paramètres du modèle, les poids du réseau de neurones, sont initialement définis de manière aléatoire. Pendant la phase d'apprentissage, ces paramètres sont ajustés en fonction des données d'entraînement fournies (cette phase peut être très longue avec un énorme corpus) pour minimiser l'erreur entre les prédictions du modèle et les valeurs cibles. Dans notre exemple, les données d'entraînement sont des corpus d'images de chats et d'images sans chat, et les valeurs cibles, les classes « chat » ou « autres ». L'apprentissage profond permet à une IA d'apprendre à partir de grandes quantités de données, bien au-delà de ce qu'un cerveau humain pourrait traiter. Ces IA, grâce au traitement de ces masses de données, peuvent résoudre des problèmes complexes comme la détection précoce de tumeur cancéreuse dans des radios qui seraient autrement impossibles à détecter par l'œil humain. Cet apprentissage laborieux n'a en tout cas rien à voir avec la façon dont un enfant apprend très rapidement, par exemple en touchant ou en goûtant.

**Les systèmes d'IA générative**<sup>2</sup> sont des approches différentes de production du texte, des images, des programmes, etc. Ils répondent à des invites ou requêtes (appelés *prompts*) en produisant de nouvelles données — par exemple la séquence de mots la plus probable après le prompt —, à partir de caractéristiques communes apprises sur un corpus de données de textes de très grande taille. Ces systèmes se servent de modèles de fondation qui permettent de produire un résultat présentant un certain degré de similarité avec les données d'apprentissage qui ont servi à le construire. Un modèle de fondation (*foundation model*), appellation introduite par l'université de Stanford, est un modèle de grande taille fondé sur une architecture de réseau de neurones profond, entraîné par apprentissage autosupervisé sur une grande quantité de données non annotées. Ces systèmes utilisent l'hypothèse distributionnelle, qui stipule que le sens d'un mot peut être déduit de son contexte, c'est-à-dire des mots voisins. Les entités du modèle de fondation (appelées *tokens*)

---

1 Transformeur (ou modèle autoattentif) : modèle d'apprentissage profond utilisé principalement dans le domaine du traitement automatique des langues. Voir Vaswani Ashish et alii, « Attention Is All You Need », ArXiv, Cornell University, 12 juin 2017 (mis à jour le 2 août 2023). URL : <https://arxiv.org/abs/1706.03762>. Consulté le 30 novembre 2023.

2 « Systèmes d'intelligence artificielle générative : enjeux d'éthique », Avis n° 7, CNPEN, 30 juin 2023. URL : [https://www.ccne-ethique.fr/sites/default/files/2023-09/CNPEN\\_avis7\\_06\\_09\\_2023\\_web-rs2.pdf](https://www.ccne-ethique.fr/sites/default/files/2023-09/CNPEN_avis7_06_09_2023_web-rs2.pdf). Consulté le 30 novembre 2023.

peuvent être des suites de caractères qui ne sont pas forcément des mots de la langue mais seulement des parties de mot. Le modèle génératif apprend à prédire l'entité suivante dans une phrase, puis l'entité suivante et ainsi de suite, un peu comme la fonction de complétion automatique sur un clavier.

Il est important de noter que les systèmes d'intelligence artificielle, qui sont fondés sur des techniques d'apprentissage machine, génèrent des modèles qui expriment des corrélations statistiques entre les éléments des données (comme des parties de mot ou des parties d'image) utilisées pour leur entraînement. Les représentations internes de ces modèles sont cependant difficiles à interpréter par les humains.

Les systèmes d'IA générative comme le *chatbot*<sup>3</sup> ChatGPT utilisent une combinaison des trois techniques d'apprentissage statistique à différentes étapes. Tout d'abord, l'apprentissage autosupervisé est utilisé pour produire les modèles de fondation sans annotation a priori. Ensuite, l'apprentissage supervisé est utilisé pour affiner ces modèles en les entraînant sur des données spécifiques et en filtrant certains résultats. Enfin, l'apprentissage par renforcement est utilisé pour optimiser les performances du système en sélectionnant les meilleurs résultats. Dans la méthode RLHF (*Reinforcement Learning with Human Feedback*), l'apprentissage par renforcement est utilisé pour aligner les résultats avec les préférences exprimées par les annotateurs humains lors de la phase supervisée<sup>4</sup>. L'objectif est de faire en sorte que ces réponses soient en accord avec les valeurs humaines, même si les systèmes ne comprennent pas la signification de ces valeurs.

Les IA génératives ne distinguent pas ce qui est possible de ce qui est impossible, ni ce qui est vrai de ce qui est faux. Elles peuvent produire des sorties combinant des assertions vraies et des assertions fausses, des résultats inattendus, parfois qualifiés d'hallucinations, ou conduire à des émergences de comportements. La vérification de ce qui est vrai ou faux n'est pas toujours facile, d'autant plus que le modèle de fondation, par construction, ne produit aucune référence aux sources. Des méthodes pour attribuer des sources aux textes générés sont soit un module spécial inclus dans le modèle (comme dans Bing de Microsoft), soit un moteur de recherche supplémentaire (comme le modèle Alfred de la *start-up* française LightOn). Les résultats produits par les systèmes d'IA générative sont souvent indiscernables de ceux produits par un humain.

Ces IA génératives sont très puissantes et peuvent être adaptées à différentes applications en peu de temps. Elles sont multitâches, multimodales et multilingues, et peuvent générer des images, de la musique, des paroles, des programmes ou des synthèses de textes. La formation de ces modèles consomme beaucoup de ressources et peut prendre plusieurs semaines sur d'énormes centres de calcul. C'est un domaine de recherche très actif pour rendre ces processus plus efficaces et respectueux de l'environnement.

### **Perception de l'IA : fantasmes et réalité**

Devant les progrès impressionnants de l'IA, on peut être tenté de la considérer comme une intelligence quasi humaine, mais l'IA ne peut pas expliquer les résultats qu'elle produit et elle reste dépendante des données sur lesquelles elle se fonde. Actuellement, dans le domaine public,

---

<sup>3</sup> Un chatbot est un agent conversationnel. Voir « Agents conversationnels : enjeux d'éthique », Avis n° 3, CNPEN, 15 septembre 2021. URL : [https://www.ccne-ethique.fr/sites/default/files/2022-02/Avis\\_n°3\\_agents\\_conversationnels\\_enjeux\\_d%27éthique.pdf](https://www.ccne-ethique.fr/sites/default/files/2022-02/Avis_n°3_agents_conversationnels_enjeux_d%27éthique.pdf). Consulté le 30 novembre 2023.

<sup>4</sup> Cf. le recours aux modérateurs kenyans par la société OpenAI. Voir Perrigo Billy, Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic », Time Magazine, 18 janvier 2023. URL : <https://time.com/6247678/openai-chatgpt-kenya-workers/>. Consulté le 30 novembre 2023.

l'intelligence artificielle passe pour une entité intelligente au sens humain du terme, pour un outil technologique qui détient la vérité. Cette idée monopolise la plupart des questionnements associés à l'IA et camoufle les vrais enjeux : l'accroissement de la désinformation, des inégalités et l'épuisement de la planète. Comment l'IA façonne-t-elle notre compréhension de nous-mêmes et de nos sociétés ?

Face à la montée rapide de l'IA, plusieurs personnes — dont le chercheur Yoshua Bengio qui a reçu le prix Turing — ont appelé à un moratoire volontaire sur son développement, afin de fournir plus de temps à la société pour évaluer les risques associés<sup>5</sup>. Mais un tel moratoire semble irréaliste dans la situation actuelle : les avantages potentiels de l'IA sont trop importants et nos sociétés sont trop compétitives, tant sur le plan interne qu'entre elles. Comme le montre très bien Kate Crawford<sup>6</sup>, l'IA est une technologie d'extraction : des minéraux extraits de la terre à la main-d'œuvre tirée des travailleurs de l'information à bas salaire, en passant par les données extraites de chaque action et interaction avec les machines.

Internet est un réseau planétaire qui pourrait alimenter une évolution vers une gouvernance non démocratique. L'IA a le potentiel de manipuler l'information et de centraliser le pouvoir. Les entreprises technologiques utilisent d'ailleurs aujourd'hui cette capacité de l'IA pour remodeler le monde, en s'en cachant à peine. L'Américain Sam Altman, PDG et cofondateur d'OpenAI (à l'origine de ChatGPT), projette de développer une intelligence artificielle générale (IAG) qui aurait les mêmes capacités qu'un « humain médian », voire des capacités supérieures, et pourrait ainsi être embauchée pour n'importe quel emploi. Par cette vision, Sam Altman semble clairement afficher son intention de remplacer de nombreux humains dans différents métiers par la technologie.

Le grand public a découvert ces derniers mois la puissance de ChatGPT, cet agent conversationnel développé par OpenAI, grâce auquel on gagne souvent beaucoup de temps en cherchant des informations. Pour les chercheurs, le lancement de ChatGPT n'est qu'une étape dans une longue histoire de recherches sur l'IA. Le modèle de langage interne de ChatGPT est fondé sur un système d'IA générative de type *transformer* proposé par Google en 2017. Mais pour le grand public, c'est une véritable rupture. Chacun a pu se rendre compte de la puissance de cet outil qui simule notre langage. Le fait que ChatGPT puisse produire de l'audio, du récit, de l'image, de la vidéo, et également du code informatique, oblige à reconnaître la puissance des systèmes d'IA générative. Face à ChatGPT, les individus se sentent à la fois démunis, enthousiastes et inquiets pour le futur, qu'il s'agisse du devenir de leur métier ou de l'éducation scolaire de leurs enfants.

L'IA générative pourrait certes rendre un grand nombre de services dans de nombreux métiers. Sur des tâches administratives un peu pénibles, elle fait gagner du temps, même s'il faut souvent vérifier. Il est évident qu'il sera difficile de s'en passer. Pour autant, ChatGPT est loin d'être parfait. C'est une machine incapable de raisonner dans la temporalité, comme elle est incapable de raisonner dans notre espace 3D. Ses performances informationnelles ne doivent pas occulter la réalité des capacités de ses systèmes. Ce que la machine apprend reste assez abstrait mais n'a rien à voir avec notre compréhension du monde. La machine est performante parce qu'elle manipule beaucoup de données ; plus elle utilise de données, plus elle donne l'impression de

---

5 Piquard Alexandre, Elon Musk et des centaines d'experts réclament une « pause » dans le développement de l'intelligence artificielle », Le Monde, 29 mars 2023. URL : [https://www.lemonde.fr/economie/article/2023/03/29/elon-musk-et-des-centaines-d-experts-reclament-une-pause-dans-le-developpement-de-l-ia\\_6167461\\_3234.html](https://www.lemonde.fr/economie/article/2023/03/29/elon-musk-et-des-centaines-d-experts-reclament-une-pause-dans-le-developpement-de-l-ia_6167461_3234.html). Consulté le 30 novembre 2023. (Voir également Soupizet Jean-François, « Un appel à la pause dans le développement de l'IA », Futuribles, n° 455, juillet-août 2023, p. 99-103 [NDLR].)

6 Crawford Kate, Atlas of AI, New Haven : Yale University Press, 2021.

répondre de manière intelligente. Nos façons d'apprendre sont très différentes ; les enfants apprennent en touchant et en expérimentant, alors que l'IA apprend statistiquement sans raisonnement, en observant les répétitions.

La vitesse du traitement algorithmique des masses d'informations numérisées fait cependant de ChatGPT un outil vertigineux pour des citoyens de plus en plus livrés à la pression du temps et à des rythmes de vie accélérés. Depuis Turing qui voulait créer des machines pensantes, depuis les sciences cognitives ou la cybernétique..., avons-nous vraiment évolué dans la compréhension des IA dans la société ? Comme les machines parlent aujourd'hui de mieux en mieux, on leur associe des connaissances, une conscience, des affects, à l'image des robots qui se substituent aux amoureux absents. Beaucoup de gens ont des sentiments d'anxiété ou de peur face à ces machines. Pour limiter les inquiétudes, il semble important d'aborder l'IA de manière plus pragmatique : en parler à l'école, expérimenter, s'approprier les outils, les questionner, apprendre à comprendre comment ils fonctionnent.

### ***Potentiels et limites de l'IA à plus long terme***

Les scientifiques ont compris le potentiel de l'IA pour changer la science. Mais rares sont ceux qui ont une vision à long terme, intégrant à la fois les dépenses d'énergie et de matière, ainsi que les conséquences sociétales du développement de ces recherches. L'IA peut déjà coder des programmes informatiques, dessiner des images et prendre des notes mais, bien plus que cela, l'IA promet de bouleverser l'économie de la connaissance. En médecine, l'IA permet d'accélérer les processus de recherche et d'invention de nouveaux médicaments pour les maladies négligées car trop rares, des vaccins contre de nouveaux virus pour lesquels il fallait jusque-là des années pour trouver des solutions. L'IA est également une aide précieuse pour le médecin, par exemple pour prédire le devenir d'un cancer agressif ou pour diagnostiquer des cardiopathies valvulaires. Cependant, il est nécessaire de prendre en compte, dans l'équation de l'innovation, les dépenses d'énergie et de matière nécessaires pour mener tous ces travaux et pour résoudre des problèmes allant de la maladie au changement climatique. Pouvons-nous imaginer que les systèmes d'IA puissent continuellement réanalyser les données et mettre à jour les résultats sur les maladies, ou les changements environnementaux, au fur et à mesure qu'ils se produisent, mais cela sans y intégrer les coûts d'énergie et de matière pour la société ?

L'application de l'IA aux structures tridimensionnelles des protéines<sup>7</sup> bénéficie de grandes bases de données. Cette technologie est proche de celle utilisée par les générateurs d'images d'IA les plus populaires, comme DALL-E. Les scientifiques ont passé des décennies et dépensé des milliards de dollars pour caractériser expérimentalement les structures 3D de nombreuses protéines, et regrouper toutes ces informations en une base de données massive qu'il est possible de donner à étudier à l'IA. L'IA est aujourd'hui utilisée pour concevoir la structure 3D de nouvelles protéines. Les chercheurs les fabriquent ensuite en laboratoire pour voir si elles fonctionnent réellement. La forme d'une protéine détermine souvent son efficacité. Cependant, il est loin d'être certain que tous les domaines scientifiques bénéficieront de la même manière de l'IA. Le manque de données et la complexité de différents domaines les rendent plus difficiles à étudier que les protéines à l'aide de l'IA. Cependant, des systèmes sont activement développés dans les domaines de la génétique, de l'étude du climat, de la physique des particules.

Les 100 dernières années de recherche scientifique ont été marquées par une grande sérendipité — de nombreuses découvertes sont le fruit du hasard. L'IA apparaît aujourd'hui comme

---

<sup>7</sup> Tunyasuvunakool Kathryn, « Highly Accurate Protein Structure Prediction for the Human Proteome », Nature, vol. 596, n° 7873, p. 590-596, 2021. URL : <https://www.nature.com/articles/s41586-021-03828-1>. Consulté le 30 novembre 2023.

un accélérateur pour faire avancer la recherche médicale, spatiale, mais également optimiser les dépenses d'énergie. Les progrès récents dans le développement du modèle d'IA générative (*Large Language Model* / LLM pour un modèle de fondation sur le texte) révèlent son étonnante capacité dans diverses tâches de calcul sur du texte, des images, etc. Dans le domaine de la psychologie par exemple, l'utilisation d'agents conversationnels utilisant des IA génératives est testée pour détecter les capacités fondamentales de la théorie de l'esprit à attribuer des états mentaux tels que les croyances, les émotions, les désirs, etc. Ce genre d'outil nécessite des normes et une loi sur l'IA pour être déployé afin d'éviter la manipulation cognitive.

### ***Enjeux d'éthique et de transparence***

Il est important de reconnaître les avantages potentiels de l'utilisation de modèles de langage d'IA dans la rédaction et la recherche scientifique, concernant notamment la bibliographie et la préparation de manuscrits. L'IA est peut-être un bon outil pour nous permettre de faire de nouvelles découvertes scientifiques à partir de la littérature scientifique, mais il y aura toujours nécessité de tester ensuite les idées scientifiques issues de l'IA pour savoir si elles sont valables ! De plus, des études montrent qu'il est possible de créer de toutes pièces des articles scientifiques frauduleux, très ressemblants à des originaux. Si, de prime abord, un article peut sembler convaincant, il faut être vigilant pour détecter certaines failles. En effet, des experts mettent souvent en avant la présence de certaines inexactitudes sémantiques et des erreurs dans les références.

En conséquence, il faut rappeler l'importance d'une vigilance accrue et de meilleures méthodes de détection pour lutter contre l'utilisation potentiellement abusive de l'IA dans la recherche scientifique. Les transformations technologiques sont en train d'arriver dans tous les domaines, de la vie privée à la vie politique en passant par la vie professionnelle, en mettant en lumière de nombreux problèmes éthiques. Le lien entre la vitesse de l'innovation technologique, le temps pour la réflexion sociétale et les intérêts économiques est au centre du problème éthique. Ce décalage est susceptible de générer pendant plusieurs années des tensions. Il est nécessaire de mener des recherches sur ces IA génératives ouvertes. Alors que l'impact sociétal des modèles de fondation augmente, la transparence de ces modèles diminue, et ils pourraient devenir tout aussi opaques que les plates-formes de médias sociaux. Il est essentiel d'inverser cette tendance : la transparence est une condition préalable essentielle à la responsabilité publique, à l'innovation scientifique et à une gouvernance efficace.

Les grandes entreprises américaines qui produisent des IA génératives comme ChatGPT d'OpenAI, Bing de Microsoft, Bard de Google, Llama de Meta, etc. sont bien connues du grand public. Les entreprises montantes en France comme MistralAI ou Light On (Alfred) se spécialisent dans le développement de modèles de langage à grande échelle. En 2021, LightOn a lancé PAGnol (où PAG signifie pré-apprentissage génératif), le plus grand modèle de langue française, puis Alfred une IA conversationnelle et un LLM en source ouverte pour les entreprises. En 2023, ils ont lancé également Paradigm, une plate-forme d'IA générative pour les grandes entreprises.

L'impact environnemental du développement extrêmement rapide du numérique implique de mesurer le coût énergétique, et plus généralement l'empreinte environnementale des systèmes d'IA générative et des modèles de fondation, afin de les inscrire dans la transition écologique. Ces exigences ne sont pas respectées par la plupart des modèles de fondation actuels. Des chercheurs de l'université de Stanford ont constaté que « les fournisseurs de modèles de fondation ne rendent pas compte de manière cohérente de l'utilisation de l'énergie, des émissions, de leurs stratégies de mesure des émissions et des mesures prises pour atténuer les émissions <sup>10</sup> ».

---

8 Bommasani Rishi et alii, « Do Foundation Model Providers Comply with the Draft EU AI Act? », HAI (Human-centered Artificial Intelligence), université de Stanford, 2021. URL :

Un laboratoire de recherche à but non lucratif consacré à l'intelligence artificielle générative vient d'être lancé en France, en novembre 2023, par le fondateur de Free (Xavier Niel), avec une mise de fonds à hauteur de 300 millions d'euros<sup>9</sup>. Ce laboratoire appelé Kyutai (« sphère » en japonais) est financé à parts égales par Xavier Niel, Éric Schmidt, l'ancien directeur général de Google, et Rodolphe Saadé, le patron de la compagnie de transport maritime CMA-CGM (Compagnie maritime d'affrètement-Compagnie générale maritime). Souhaitons que cette initiative soit le début d'une ambition de développement positif et responsable de l'IA générative, prenant en compte les facteurs économiques, mais aussi écologiques, sociétaux et éthiques.

## **Dans le domaine de la génétique**

### ***Les gènes n'ont pas tout pouvoir***

Concernant la génétique, généralement le grand public a tendance à penser que tout est écrit dans les gènes et à attribuer un pouvoir trop grand aux facteurs génétiques<sup>10</sup>. Par exemple, quand dans les années 2004-2006, la société californienne Genetic Savings & Clone a proposé de recréer par clonage des chats identiques à partir de matériel génétique<sup>11</sup>, plusieurs personnes — fortunées — qui avaient un animal de compagnie exceptionnel mais vieillissant ont sollicité ses services. Même si la compagnie a suscité de nombreuses questions éthiques<sup>12</sup> et a dû mettre la clef sous la porte suite à des problèmes financiers en 2006, la question de savoir si un chat cloné allait forcément être identique au chat initial ne s'est pas posée dans l'esprit du grand public. Or, chaque être vivant est le résultat à la fois de son génome et de l'environnement dans lequel il a vécu jusqu'à présent. L'environnement est défini ici au sens large comme tous les paramètres autres que la séquence d'acide désoxyribonucléique (ADN) : l'alimentation, les expositions à des agents pathogènes chimiques ou naturels, l'environnement social et culturel, l'éducation, le stress, etc. Les caractéristiques d'un individu ne sont pas définies uniquement par la séquence de son ADN.

En fait, l'ADN seul est une molécule inerte. Ce n'est que dans un environnement particulier, au sein d'une cellule et en contact avec toute une machinerie moléculaire que l'ADN peut transmettre des informations qui contribueront à faire émerger des propriétés particulières visibles à l'échelle de l'individu. Certains effets potentiels de l'ADN ne s'observent que dans des conditions environnementales particulières. Par exemple, des humains porteurs de variants génétiques associés à une faible capacité à synthétiser la vitamine D peuvent présenter une santé osseuse et immunitaire diminuée dans des environnements où l'exposition au soleil est limitée, mais ces effets de l'ADN sont « guéris » en vivant dans des régions plus ensoleillées ou grâce à des suppléments de vitamine D. Dans de nombreux cas, la génétique n'impose pas un destin inexorable : il est bien souvent possible d'aller à l'encontre de la génétique en modifiant certains paramètres de l'environnement. Tout le monde a à l'esprit des maladies monogéniques (causées par des mutations dans un seul gène) telles que certaines myopathies et autres maladies génétiques neuromusculaires, que les médias mettent en lumière chaque année, en France, au moment du Téléthon. Mais hormis ces

---

<https://crfm.stanford.edu/2023/06/15/eu-ai-act.html>. Consulté le 30 novembre 2023.

9 Marin Jérôme, « Xavier Niel lance Kyutai, un laboratoire d'intelligence artificielle à but non lucratif », CaféTech, 20 novembre 2023. URL : <https://cafetech.fr/2023/11/20/xavier-niel-lance-kyutai-un-laboratoire-dintelligence-artificielle-a-but-non-lucratif/>. Consulté le 30 novembre 2023.

10 Munnich Arnold, Programmé mais libre. Les malentendus de la génétique, Paris : Plon, 2016.

11 Page d'accueil de l'entreprise Genetic Savings & Clone (archivée sur Internet Archive) : <https://web.archive.org/web/20060429204721/http://savingsandclone.com/index.html>. Consultée le 30 novembre 2023.

12 « Dead Cats to Remain a Memory », Wired, 12 octobre 2006. URL :

<https://www.wired.com/2006/10/dead-cats-to-remain-a-memory/>. Consulté le 30 novembre 2023.

quelques maladies particulières, qui sont peu influencées par l'environnement, la plupart des troubles et des caractéristiques de chaque être humain sont déterminés par une combinaison de plusieurs gènes, et fortement influencés par l'environnement et les conditions de vie. Le champion de ski finlandais Eero Mäntyranta, qui a gagné sept médailles olympiques dans les années 1960, avait une mutation particulière dans son génome, permettant à son sang de bien mieux transporter l'oxygène<sup>13</sup>. Le fait que la plupart des autres membres de sa famille, aussi porteurs de la même mutation, n'aient pas été champions olympiques montre bien l'influence de l'environnement. Ce que nous sommes est influencé à la fois par notre ADN, par notre environnement, et par les interactions entre notre ADN et notre environnement.

Ainsi, le matériel génétique que nous possédons ouvre un ensemble de possibilités et, selon l'environnement que rencontrera notre génome au fil de notre vie, certaines de ces possibilités se développeront tandis que d'autres ne verront pas le jour. La croyance erronée et cependant répandue en la toute-puissance des gènes a profité au développement de certaines firmes, comme 23andMe, à partir de 2007 aux États-Unis. 23andMe proposait à l'époque, pour 99 dollars US, une analyse du génome et une estimation des risques de développer plus d'une centaine de maladies. Mais l'agence américaine des produits alimentaires et médicamenteux, la Food and Drug Administration (FDA), a qualifié cette publicité de « mensongère » en 2013 et 23andMe a dû renoncer à ces tests génétiques prédictifs. Aujourd'hui, 23andMe vend surtout aux particuliers des tests concernant la généalogie et a obtenu l'autorisation de détecter seulement une dizaine de maladies, pour lesquelles le lien entre génétique et prédisposition est clair.

### ***La part de l'inné et de l'acquis***

Même si la question revient souvent, il est généralement impossible de savoir, pour une personne donnée, quelle est la part de l'inné et de l'acquis dans tel comportement ou telle caractéristique physique. Les gènes et l'environnement sont entremêlés et nécessaires tous les deux, de manière différente, à l'apparition de telle ou telle caractéristique. Notons aussi que les facteurs génétiques sont plus aisément détectés que les facteurs environnementaux : les séquences d'ADN représentent des données faciles à classer et à analyser car ce sont des suites de lettres, alors que les facteurs environnementaux sont quantitatifs et bien plus divers, ils se prêtent donc moins bien aux analyses.

La recherche des causes génétiques se fait par association statistique entre certains gènes et certains traits de caractère (couleur des yeux, maladie, etc.). Prenons comme analogie le concours du meilleur croissant au beurre de France. Pour savoir quelle est la part de la recette, du cuisinier et du contexte sociétal qui explique le premier prix au concours du meilleur croissant, il faudrait pouvoir tester toutes les combinaisons possibles de ces trois paramètres. Or, nous n'avons pas accès à toutes ces possibilités. De même, les associations génétiques sont effectuées à partir d'un certain ensemble d'individus. Les résultats obtenus ne sont valables que pour la population qui a été étudiée, au moment où l'étude a été réalisée, et il n'est pas certain que les résultats obtenus s'appliquent aussi à une autre population.

Enfin, les différences génétiques identifiées ne sont pas forcément la cause : les mécanismes sous-jacents peuvent tout aussi bien être environnementaux. Par exemple, dans une société imaginaire où les enfants aux yeux bleus n'auraient pas le droit d'aller à l'école, l'illettrisme et la

---

13 La Chapelle Albert (de), Traskelin Ann-Liz et Juvonen Eeva, « Truncated Erythropoietin Receptor Causes Dominantly Inherited Benign Human Erythrocytosis », PNAS (Proceedings of the National Academy of Sciences of the United States of America), vol. 90, n° 10, 15 mai 1993, p. 4495-4499. URL : <https://www.pnas.org/doi/abs/10.1073/pnas.90.10.4495>. Consulté le 30 novembre 2023.

tendance à errer dans les rues seraient associés aux gènes de couleur des yeux <sup>14</sup>. Quand certaines particularités génétiques ont un lien très fort avec certains facteurs culturels, il n'est pas possible de distinguer la part de l'environnement et la part génétique dans ces études d'association.

### ***La difficulté à trouver les causes des maladies***

Il est difficile pour le grand public de manipuler des probabilités et des risques. Notre esprit a naturellement tendance à simplifier et à ne pas prendre en compte la variabilité intrinsèque d'un phénomène que l'on observe. De plus, ce qui intéresse une personne est généralement son propre cas, ou bien les personnes de son entourage, mais pas ce qui se passe à l'échelle d'une population. Par exemple, si un patient non fumeur atteint d'un cancer du poumon souhaite savoir comment il l'a contracté, son médecin va pouvoir, au mieux, identifier des mutations à l'origine de ce cancer, mais il ne va pas être en mesure d'identifier la ou les cause(s) exacte(s) : est-ce l'exposition aux particules fines, ou une autre cause ?

En effet, l'identification des facteurs de risque est basée sur l'étude de toute une population, sur laquelle sont réalisées diverses mesures statistiques. Cette approche holistique permet d'estimer par exemple le nombre de décès et de cas de cancer du poumon par an à Grenoble, ainsi que de calculer quel impact pourrait avoir une réduction du transport routier sur cette mortalité<sup>15</sup>. Mais malheureusement, ces méthodes « populationnelles » ne permettent pas d'identifier les causes individuelles du cancer pour chaque patient. Il y a, ici encore, un décalage entre ce que la science peut apporter et ce que la personne lambda voudrait pouvoir tirer de la science.

### ***Nos connaissances en génétique restent limitées***

Grâce à la mise au point des ciseaux moléculaires CRISPR-Cas9<sup>16</sup> en 2012, il est aujourd'hui possible de modifier facilement et à bas coût la séquence d'ADN de la plupart des espèces, et d'y introduire les mutations souhaitées<sup>17</sup>. Cela laisse penser — autre idée reçue fautive — que l'on peut faire tout ce que l'on veut avec les ciseaux moléculaires CRISPR. En effet, le grand public a tendance à considérer ces ciseaux moléculaires comme une technologie de modification génétique sans nuance, ce qui peut susciter des inquiétudes quant à la création de bébés / animaux / plantes « sur mesure ».

En réalité, on connaît encore très mal les bases génétiques de certains traits de caractère. Pour la plupart des changements que l'on pourrait souhaiter apporter (des plantes qui résistent à la sécheresse et au réchauffement climatique et qui croissent bien, des plantes qui produisent des ratios de nutriments adaptés aux besoins humains, etc.), on ne sait pas si telle ou telle mutation va

---

14 Foucart Stéphane, En biologie, les “bons” et “mauvais” gènes font un inquiétant retour, alimentant les théories raciales », Le Monde, 20 janvier 2023. URL : [https://www.lemonde.fr/idees/article/2023/01/20/en-biologie-les-bons-et-mauvais-genes-font-un-inquietant-retour-alimentant-les-theories-racialistes\\_6158611\\_3232.html](https://www.lemonde.fr/idees/article/2023/01/20/en-biologie-les-bons-et-mauvais-genes-font-un-inquietant-retour-alimentant-les-theories-racialistes_6158611_3232.html). Consulté le 30 novembre 2023.

15 Slama Rémy, « Causes et conditions extérieures des maladies et de la santé », Leçon inaugurale du Collège de France, 31 mars 2022. URL : <https://www.college-de-france.fr/fr/agenda/lecon-inaugurale/causes-et-conditions-exterieures-des-maladies-et-de-la-sante-0>. Consulté le 30 novembre 2023.

16 Acronyme de Clustered Regularly Interspaced Short Palindromic Repeats, faisant référence à une technique de manipulation de l'ADN réalisée avec la nucléase Cas9 (pour CRISPR associated protein 9).

17 Zhang F., « Development of CRISPR-Cas Systems for Genome Editing and Beyond », Quarterly Reviews of Biophysics, vol. 52, e6, 2019. URL : <https://doi.org/10.1017/S0033583519000052>. Consulté le 30 novembre 2023.

effectivement permettre de les obtenir. Souvent, les mutations introduites ont, en plus de leurs effets attendus, des effets non prévus : augmentation de la létalité, diminution du rendement, etc. De plus, leurs effets peuvent varier selon les conditions environnementales.

De nombreuses recherches sont encore à mener pour mieux comprendre les causes génétiques, ce qui ne nous empêche pas de commencer à utiliser les outils d'IA pour certaines pathologies et maladies génétiques.

### **Génétique et IA : un exemple avec le traitement du cancer**

Ces dernières années, l'IA a permis une avancée majeure dans le diagnostic de certains cancers. En utilisant cette technologie, des cancers comme ceux du poumon, du sein ou de la prostate peuvent être dépistés de façon précoce. L'IA permet de détecter des tumeurs sur des radios, alors que le cancer est au début de son développement sur les images radio. Elle permet de faire gagner du temps aux radiologues (qui peuvent utiliser ce temps pour effectuer d'autres tâches plus complexes telles que des vérifications) et surtout détecter des tumeurs que l'œil humain a du mal à percevoir.

En décembre 2022, une collaboration entre l'Institut Curie et Ibex Medical Analytics a permis de démontrer la performance, la fiabilité et la mise en application clinique d'un algorithme d'IA capable de diagnostiquer les cancers du sein lors de biopsies mammaires, qui sont des prélèvements de tissus. L'intelligence artificielle, nommée Galen Breast, a été développée dans l'objectif de réduire les erreurs et d'améliorer la qualité du diagnostic. Ces travaux, dont les résultats sont détaillés dans la revue *NPJ Breast Cancer*<sup>18</sup>, font état de la toute première mise en œuvre d'une IA dans le cadre d'une utilisation clinique. Cette IA n'a pas pour vocation de remplacer les médecins. C'est un outil qui fournit des informations qui permettent de détecter l'avancée de différents types de cancer du sein et ainsi d'affiner la précision du diagnostic. « Le système est surtout développé pour être un appui aux pathologistes, pour vérifier qu'ils n'ont pas manqué quelque chose lors de sa lecture, ainsi que pour automatiser le traitement de certains examens chronophages et fastidieux comme le taux de lymphocytes dans la tumeur », explique Suzette Delalogue, oncologue médicale, spécialiste du cancer du sein et directrice du programme de prévention du cancer Interception de l'institut Gustave Roussy<sup>19</sup>.

Une nouvelle application de l'IA en génétique est en train de voir le jour : la prédiction de l'effet des mutations sur les individus. En septembre, les scientifiques du laboratoire de recherche de Google DeepMind ont publié un modèle d'intelligence artificielle capable de prédire si des mutations génétiques ponctuelles, c'est-à-dire des changements d'une seule lettre de l'ADN, sont susceptibles de déclencher une pathologie<sup>20</sup>. Leur modèle a été évalué et validé en utilisant d'autres bases de données cliniques non incluses dans leurs données d'entraînement. Il montre que sur les 70 millions de mutations observées, environ 30 % pourraient entraîner une maladie. Nous sommes encore loin de pouvoir prédire l'effet de telle ou telle mutation sur tel ou tel organe — ici seul l'effet global sur l'organisme (délétère, bénin ou inconnu) est prédit, mais c'est une première

---

18 Sandbank Judith et alii, « Validation and Real-World Clinical Application of an Artificial Intelligence Algorithm for Breast Cancer Detection in Biopsies », *NPJ Breast Cancer*, vol. 8, article n° 129, 2022. URL : <https://www.nature.com/articles/s41523-022-00496-w.pdf>. Consulté le 30 novembre 2023.

19 Voir la page : <https://www.gustaveroussy.fr/interception>. Consultée le 30 novembre 2023.

20 Cheng Jun et alii, « Accurate Proteome-wide Missense Variant Effect Prediction with AlphaMissense », *Science*, vol. 381, n° 6664, 19 septembre 2023. URL : <https://www.science.org/doi/10.1126/science.adg7492>. Consulté le 30 novembre 2023.

avancée vers de meilleures prédictions. Ce travail pourrait améliorer le diagnostic des maladies génétiques rares et contribuer à la découverte de nouveaux gènes associés à des maladies.

Les systèmes d'IA peuvent également permettre de trouver de nouveaux traitements thérapeutiques pour soigner certaines maladies génétiques et notamment des cancers. L'IA est capable de prendre en compte des milliers de données (interactions physiques entre molécules, voies de signalisation cellulaires, physiologie et régulations hormonales, associations entre certaines mutations et risques de développer certaines maladies, effets secondaires déjà connus de certains médicaments, etc.) et d'en extraire une liste de molécules candidates qui peuvent éventuellement aider à soigner certaines pathologies. Des essais cliniques sont ensuite nécessaires pour tester ces médicaments potentiels.

L'emploi de l'IA permet de réduire le temps et le coût de la phase de recherche exploratoire des meilleures pistes pharmaceutiques. Aujourd'hui, l'IA se révèle donc être un outil intéressant pour soutenir ces recherches et les rendre plus rapides et plus efficaces. L'IA va indubitablement conduire à une évolution importante des métiers de la santé. Il est nécessaire, de plus en plus rapidement, de former les actifs à ces nouveaux outils tout en évoquant les risques éthiques. Il est également urgent de former le grand public à ces enjeux de connaissances et de soins.

### **Comment armer le grand public ?**

L'IA et la génétique sont en train de transformer la société, le travail et aussi tout particulièrement le domaine scientifique, ouvrant la voie à des découvertes sans doute sans précédent. L'IA en science ou en médecine dépend d'algorithmes complexes et de vastes quantités de données, parfois sensibles. Afin d'être à l'abri d'un usage malveillant de l'IA, il est important d'adopter une approche éthique et responsable pour exploiter pleinement le potentiel de cette technologie révolutionnaire, tout en préservant la confidentialité et l'intégrité des informations scientifiques. Il est aussi important de discuter des finalités des technologies, et de les rendre plus transparentes et compréhensibles de façon à mettre en place certains garde-fous.

Trois aspects sont indispensables : la loi, les normes et l'éthique. Au niveau européen, l'AI Act (la loi sur l'IA), mise en place par la Commission européenne le 21 avril 2021, votée au Parlement européen en juin 2023 et qui devrait être finalisée en décembre 2023<sup>21</sup>, se concentre sur l'identification des applications qui présentent un risque et nécessitent une réglementation. Trois niveaux différents de risque (inacceptable, haut, modéré) ont été considérés. L'AI Act propose donc d'offrir une protection, en imposant « des obligations à certains systèmes d'IA en raison des risques spécifiques qu'ils présentent ». Des normes vont accompagner le déploiement de la loi sur l'IA pour aligner les technologies sur ces obligations 24<sup>22</sup>. Tout système d'IA est produit par de multiples agents, IA et humains : les données annotées qui servent à l'apprentissage viennent par exemple des humains, et les paramètres d'apprentissage des IA sont décidés par des humains. La régulation doit suivre les progrès rapides des IA et doit pouvoir s'adapter, au fil de l'évolution de la compréhension des risques. Il est également nécessaire de se doter de comités d'éthique du numérique, comme la France l'a fait pour la bioéthique.

Former le grand public à ces enjeux d'innovation mais aussi d'éthique est fondamental. Un obstacle majeur réside dans le fait que les non-scientifiques considèrent souvent la science comme définitive et fournissant une vérité immuable, plutôt que comme un processus de découverte qui s'autocorrige. Il est normal d'avoir des discussions et des débats sur ce que les connaissances

---

21 Ce numéro a été bouclé le 4 décembre 2023 (NDLR).

22 Voir « Artificial Intelligence », Cenelec. URL : <https://www.cenelec.eu/areas-of-work/cenelec-topics/artificial-intelligence/>. Consulté le 30 novembre 2023.

scientifiques nous apportent dans telle ou telle situation, et sur ce qu'il est préférable de faire. Il est important de prendre en compte une approche pluridisciplinaire et de considérer les différents aspects de la question. Quelles sont les implications qui découlent de cette observation ? Quelles sont les questions sociales et comportementales pertinentes ? Souvent, ceux qui attisent la méfiance à l'égard de la science se focalisent sur certains aspects et en oublient d'autres, ce qui peut conduire à la subversion des preuves. La question du port du masque et celle de la vaccination pendant la pandémie de Covid-19 en sont des exemples frappants. La science n'est pas seulement constituée de faits bruts et de données, ce sont aussi des modes de pensée et de raisonnement, des façons de voir le monde et de le comprendre, qui peuvent être remis en question quand de nouvelles observations sont faites.

Le rôle de l'École est essentiel pour apprendre à apprendre. À l'école, on peut inciter les enfants au questionnement sur la part de l'inné et de l'acquis, et leur montrer avec divers exemples (les muscles qui grossissent en allant à la salle de sport, le volume et la physiologie d'une tique qui sont totalement modifiés après un repas de sang, etc.) que la génétique n'explique pas tout et que l'environnement est un facteur important à prendre en compte dans l'origine des différences entre individus. De même, il est important de dépasser la tentation de plagiat facilitée par l'arrivée d'outils comme ChatGPT, en les invitant à comprendre ces objets et à réfléchir aux usages de l'IA plutôt qu'à être de simples consommateurs <sup>25</sup><sup>23</sup>. Il ne s'agit pas de dire que l'IA est un bien ou un mal, mais de se demander si elle peut avoir des incidences négatives et d'étudier les risques, et en même temps d'en comprendre le pouvoir d'innovation.

L'Union européenne a beaucoup à gagner en élaborant un cadre pour l'utilisation de l'IA et de la génétique qui soit associée à la confiance et au respect des droits humains et de l'état de droit. Au-delà même de l'Europe, il faudrait imaginer un équivalent du GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat) pour l'IA : un outil de régulation et de coopération à l'échelle mondiale.

© *futuribles* no 458 . janvier-février 2024

---

23 Voir l'exemple des « capsules Éthique du numérique pour les enfants », Fondation Blaise Pascal. URL : <https://www.fondation-blaise-pascal.org/nos-actions/les-projets-de-la-fondation/capsules-ethique-du-numerique-pour-les-enfants/>. Consulté le 30 novembre 2023.