



**iea**

Institut pour  
l'Éducation Augmentée

*Promouvoir l'alliance entre l'éducation et les nouvelles technologies*

# QUESTIONS ET RÉPONSES SUR L'IA.

Entretien avec **BERNARD OURGHANLIAN**

« Enjeux scientifiques, sociétaux et éthiques »

## **Institut pour l'éducation augmentée.**

L'Institut pour l'Education Augmentée est dédié à la promotion de l'alliance entre l'éducation et les nouvelles technologies, et en particulier à l'implémentation de l'intelligence artificielle dans les processus éducatifs pour renouer avec l'idéal méritocratique de notre système éducatif.

Notre site internet pour retrouver nos actions et travaux :

<https://institut-education-augmentee.org/>

Contact :

[sg@institut-education-augmentee.fr](mailto:sg@institut-education-augmentee.fr)

Contact presse :

[rp@institut-education-augmentee.fr](mailto:rp@institut-education-augmentee.fr)



# SOMMAIRE

<b>Sommaire</b>	<b>3</b>
<b>Remerciements</b>	<b>5</b>
<b>Bernard Ourghanlian</b>	<b>7</b>
<b>Introduction</b>	<b>9</b>
ChatGPT et IA générative : pour paraphraser Shakespeare, est-ce « beaucoup de bruit pour rien » ou bien se passe-t-il vraiment quelque chose de nouveau ?	9
Quelles sont les principales évolutions technologiques qui ont rendu possibles les IA génératives ?	15
<b>Questions philosophiques</b>	<b>21</b>
Faut-il voir dans l'IA le risque d'un remplacement de l'intelligence humaine par la machine ?	21
L'IA remet-elle en cause la façon dont nous définissons l'intelligence ?	27
Faut-il craindre de l'IA qu'elle nous prive petit à petit de notre autonomie intellectuelle ?	30
<b>Questions sociologiques et sociétales</b>	<b>34</b>
A travers l'apparition des deepfakes, on voit poindre le risque de la désinformation et de la manipulation de l'opinion à grande échelle : que faut-il faire pour atténuer ce risque ?	34

On parle beaucoup de l'impact de l'IA sur l'avenir du travail et sur les risques de disparition de millions d'emplois : que faut-il en penser et que faut-il faire ? 40

L'IA ne risque-t-elle pas d'accroître les écarts de richesse avec des personnes encore plus riches et d'autres encore plus pauvres ? 50

Quel est l'impact de l'IA sur le réchauffement climatique ? 56

### **Questions sur l'IA et l'éducation 64**

L'arrivée de ChatGPT dans le monde de l'éducation a engendré énormément de questionnements : peut-on mettre l'intelligence artificielle au service de l'éducation et, si oui, comment ? 64

### **Questions sur le futur de l'IA 71**

Quelle est ta vision sur le futur de l'IA ? Verra-t-on rapidement l'arrivée d'une AGI ? D'une IA forte ? 71

Quelles sont les principaux défis que la recherche doit encore adresser pour faire progresser l'IA ? 81

### **En guise de conclusion 88**

# REMERCIEMENTS

« Chères lectrices, chers lecteurs,

*C'est avec une profonde reconnaissance que nous saluons le travail remarquable de **Bernard Ourghanlian** sur les enjeux scientifiques, sociétaux et éthiques de l'intelligence artificielle. Une discussion riche et stimulante a émergé entre Bernard – qui fut également enseignant – et moi-même, autour d'un thème qui nous tient particulièrement à cœur : l'intelligence artificielle en éducation.*

*Ce dialogue, à la croisée de la technique, de la philosophie et du sens critique, s'est nourri de la rigueur intellectuelle et de l'humilité de Bernard. **Un grand merci à Bernard.** Il résulte de ses réflexions un document analytique d'une grande richesse, qui constitue une contribution précieuse à l'exploration des impacts de l'intelligence artificielle sur notre société.*

*Depuis sa création, l'**Institut pour l'Éducation Augmentée** s'engage à promouvoir une réflexion ouverte, inclusive et rigoureuse sur les transformations induites par les technologies émergentes. Notre ambition est de mieux comprendre et anticiper les effets de l'IA, en particulier dans le domaine sensible et fondamental de l'éducation.*

*Nous sommes convaincus que, bien utilisée, l'intelligence artificielle peut devenir un levier puissant pour améliorer les processus éducatifs. Mais cette conviction s'accompagne d'une conscience aigüe des défis éthiques, sociaux et institutionnels qu'elle soulève.*

*C'est autour de cette tension féconde – entre promesses et risques, entre mythe et réalité – que notre réflexion s'est construite.*

**L'éducation augmentée par l'IA (EDiA)** est un sujet complexe, exigeant prudence, modestie et lucidité. Entre logiques mercantiles, inerties institutionnelles et attentes sociétales, il nous faut poser la question essentielle : comment intégrer ces technologies de manière juste, éclairée et humaine ?

*Ce document s'inscrit pleinement dans cette démarche. Il nous invite à penser l'IA dans toutes ses dimensions, à en interroger les usages, à en anticiper les dérives, mais aussi à en explorer les potentialités pour une éducation plus équitable, plus inclusive, et plus centrée sur l'élève et les équipes éducatives.*

*L'intelligence artificielle générative représente une opportunité unique pour repenser notre rapport au savoir, au sein de l'école républicaine comme dans l'ensemble du tissu éducatif. Elle peut contribuer à renforcer la qualité de l'enseignement, à personnaliser les parcours, et à mieux accompagner chaque élève dans son développement – à condition de rester fidèles à notre devoir de bienveillance envers les générations futures.*

*Ensemble, construisons une vision équilibrée, éclairée et ambitieuse de l'avenir de l'éducation augmentée, dans une perspective constructive, adogmatique et collective. La complexité des enjeux appelle une intelligence partagée, interdisciplinaire et engagée.*

*Bonne lecture à toutes et à tous,*

*Julien Cartier »*

# BERNARD OURGHANLIAN

Bernard Ourghanlian a commencé sa carrière comme enseignant en mathématique au collège et au lycée puis comme enseignant-chercheur à l'université d'Orsay où il a enseigné les probabilités et les statistiques en IUT et en Master 2. Il est docteur en mathématiques et est l'auteur de plusieurs publications spécialisées relatives aux statistiques et à l'informatique, notamment d'un ouvrage de référence sur les microprocesseurs Alpha.

Puis, il a participé à la conception et à l'implémentation des premières techniques d'imagerie numérique par tomographie et par résonance magnétique nucléaire au sein de la société Thomson CGR.

Il a ensuite rejoint Digital Equipment Corporation (DEC) où il fut le Directeur Technique de Digital France ; il y a participé activement à la conception et au développement de l'architecture des microprocesseurs Alpha et des différents éléments logiciels système permettant d'en assurer le support.

Par la suite, il a rejoint Microsoft France qu'il a rejoint en 1999 et quitté en 2023 pour partir à la retraite. Il y a eu la responsabilité de la conduite globale de la stratégie de Microsoft en France, en matière de technologie et de sécurité. Il a également eu en charge la gestion des risques technologiques de l'entreprise, notamment

en matière de cybersécurité. Bernard Ourghanlian a été membre du comité exécutif de Microsoft France depuis le début de l'année 2009. Il a été également administrateur Numeum et de l'Université de Paris Saclay.

Pendant cette période, il a représenté Microsoft France sur le plan technique auprès des principaux responsables gouvernementaux et des grandes agences gouvernementales, des directions informatiques, des comités de standardisation ou de normalisation ainsi qu'auprès du monde de l'éducation et de la recherche. Il a également travaillé en liaison étroite avec Microsoft Corporation, avec Microsoft Research et les différents groupes produits et a participé activement à l'évolution des produits et des solutions de Microsoft. Il a été fondateur et membre du comité de direction du centre de recherche commun qui a été créé entre Inria et Microsoft Research.

Bernard Ourghanlian a animé pendant de nombreuses années des cours et des séminaires au sein de grandes écoles et d'universités françaises, notamment à l'ESME-Sudria et à Centrale Paris.



## CHAPITRE 1

# INTRODUCTION

***ChatGPT et IA générative : pour paraphraser Shakespeare, est-ce « beaucoup de bruit pour rien » ou bien se passe-t-il vraiment quelque chose de nouveau ?***

Avant d'entamer cet échange, je voudrais - et ce n'est pas de la fausse-modestie - partager avec vous l'effroi qui est le mien devant l'immensité d'un tel sujet : rien que sur les aspects scientifiques, selon [l'Artificial Intelligence Index Report 2024 de l'université de Stanford](#), entre 2010 et 2022, le nombre total de publications sur l'IA a presque triplé, passant d'environ 88 000 en 2010 à plus de 240 000 en 2022. Cela fait 660 papiers de recherche publiés chaque jour... Et je ne parle pas des très nombreux papiers publiés en sociologie, philosophie, sans compter les rapports gouvernementaux, la presse, les blogs,... Inutile de vous dire que je suis très loin de les avoir tous lus... Devant cette insondable somme de connaissances, l'humilité est de rigueur...

L'intelligence artificielle n'a pas simplement fait irruption dans nos vies ; elle a, pour ainsi dire, déplacé l'horizon même de notre rapport au savoir, à la technique et au monde. En témoignent les chiffres : une semaine après son lancement fin novembre 2022, ChatGPT, produit d'OpenAI, avait déjà conquis un million

d'utilisateurs. Deux mois plus tard, ce chiffre atteignait les 100 millions, marquant une vitesse d'adoption inédite dans l'histoire des technologies numériques. Pour mémoire, TikTok avait mis neuf mois à franchir ce seuil, Instagram trente. Cette accélération révèle moins une fascination passagère qu'un basculement profond dans notre manière d'interagir avec la connaissance. Pour la première fois en septembre 2024, ChatGPT a dépassé les 3 milliards de visites. En février 2025, OpenAI indiquait avoir franchi la barre des 400 millions d'utilisateurs hebdomadaires.

Il faut souligner que ChatGPT ne se laisse pas aisément enfermer dans la catégorie d'« outil ». Contrairement aux technologies instrumentales traditionnelles, qui s'inscrivent dans une logique de fonctionnalité bien définie (rédiger, calculer, communiquer), ChatGPT se présente comme un dispositif polymorphe, aux usages aussi multiples qu'imprévisibles : assistance à la rédaction, soutien à la programmation, aide à la recherche, génération de contenus créatifs, synthèse d'articles, jeux de langage poétiques,... Cette versatilité fonctionnelle en fait non pas un outil parmi d'autres, mais un vecteur de transformation structurelle des usages, ce qui justifie son assimilation à une *General Purpose Technology* (GPT) - à entendre cette fois dans son acception économique.

Les *General Purpose Technologies*, de la machine à vapeur à l'électricité en passant par les technologies de l'information, sont de ces innovations rares qui transforment en profondeur les structures économiques, sociales et culturelles. L'intelligence artificielle générative, en tant qu'itération contemporaine de cette

dynamique, ne saurait être réduite à une simple évolution technologique. Elle engage une mutation anthropologique.

Certes, l'IA ne date pas de ChatGPT. Depuis plusieurs décennies, des systèmes dits intelligents œuvrent en silence dans nos routines numériques : moteurs de recommandation, reconnaissance faciale, traduction automatique, tri automatique du courrier à la Poste,... Mais jusqu'alors, l'IA agissait dans une certaine opacité, opérant dans les coulisses de nos existences. Le déploiement de l'IA générative marque une rupture : pour la première fois, le grand public interagit de manière directe avec une entité computationnelle qui semble répondre, créer, dialoguer. L'IA quitte les marges pour devenir interlocutrice.

Ce renversement n'est pas sans évoquer la célèbre dialectique hégélienne du maître et de l'esclave selon laquelle le maître profite du travail de l'esclave pendant que ce dernier transforme le monde à tel point que le premier ne le reconnaît plus. Longtemps, l'IA a agi comme un serviteur invisible, alimenté par nos données, agissant sans reconnaissance. Aujourd'hui, elle se présente comme un « sujet technique », comme si le serviteur accédait soudain à une forme d'autonomie dans le dialogue. La dépendance se reconfigure, les hiérarchies vacillent, et c'est peut-être cela qui, plus que la technique elle-même, dérange et fascine.

Historiquement, le concept d'« intelligence artificielle » naît en 1956, lors de la [conférence fondatrice de Dartmouth](#), sous l'impulsion de Claude Shannon, John McCarthy, Nathaniel

Rochester et Marvin Minsky,. Le terme s'inscrit dans une filiation philosophique plus ancienne, en particulier celle d'Hippolyte Taine dans *De l'Intelligence*, qui dès 1870 envisageait l'intelligence humaine comme une « machine » donc comme un mécanisme potentiellement répliquable. Dès lors, l'histoire de l'IA s'écrit comme une alternance de promesses et de déceptions, de printemps féconds et d'hivers austères. Le printemps actuel de l'IA, amorcé par les progrès fulgurants du deep learning, a été consacré par le prix Turing décerné en 2018 à Yoshua Bengio, Geoffrey Hinton et Yann LeCun, figures emblématiques de cette révolution.

Mais l'IA ne transforme pas seulement les pratiques sociales ou les imaginaires : elle bouleverse les disciplines scientifiques elles-mêmes. En témoigne la reconnaissance récente des travaux de John J. Hopfield (américain) et Geoffrey E. Hinton (canado-britannique) qui ont reçu le prix Nobel de Physique pour leurs contributions exceptionnelles dans le domaine de la science des réseaux neuronaux et de l'intelligence artificielle, ou ceux de David Baker, professeur à l'Université de Washington à Seattle, et de Demis Hassabis et John Jumper, de DeepMind qui ont reçu le prix Nobel de Chimie grâce à leurs travaux sur la conception et la prédiction des structures protéiques à partir des séquences d'acides aminés - travaux rendus possibles par des architectures neuronales avancées.

Face à ces mutations, deux attitudes coexistent. D'un côté, une tentative de domestication de l'inédit, visant à intégrer l'IA dans nos cadres conceptuels familiers : on la réduit à un simple

automate linguistique, un « [perroquet approximatif](#) » selon le ministre délégué au numérique de l'époque, Jean-Noël Barrot, aujourd'hui ministre de l'Europe et des Affaires Etrangères. De l'autre, une dramatisation prospective : tribunes alarmistes, appels à un moratoire sur le développement des modèles plus puissants que GPT-4, invocation d'un possible basculement civilisationnel. Sans parler des références innombrables à la science fiction et à des films cultes tels que « 2001 : L'Odyssée de l'espace », « Blade Runner », « Terminator », « A.I. Intelligence artificielle » ou bien encore « Matrix » ou « Her »,...

Ainsi, la [célèbre pétition](#) appelant à une pause dans la course à l'Intelligence Artificielle Générale qui est un sujet dont nous reparlerons certainement, signée par des figures aussi diverses qu'Elon Musk, Yuval Noah Harari, Yoshua Bengio ou Steve Wozniak, soulève des questions légitimes, bien que controversées. Faut-il craindre l'automatisation massive de nos métiers, y compris ceux qui donnent sens à notre humanité ? Devons-nous laisser les IA saturer l'espace informationnel de contenus invérifiables ? Jusqu'où sommes-nous prêts à déléguer à des systèmes algorithmiques des fonctions de jugement, de décision, voire de gouvernamentalité ?

Cependant, cette tribune est elle-même critiquable sur plusieurs plans :

- Elle néglige un certain nombre d'enjeux concrets déjà bien présents : biais algorithmiques, discriminations, impacts environnementaux,...

- Elle reconduit, sans distance, les discours de l'industrie technologique sur une IA omnipotente, au lieu de considérer ses limites actuelles.
- Elle passe aussi sous silence un élément fondamental : la concentration du pouvoir technologique entre les mains de quelques géants du numérique, avec ses implications politiques, géopolitiques et éthiques.

Parmi les critiques les plus virulentes de cette pétition, l'article d'Emily Bender : [Policy makers: Please don't fall for the distractions of #Alhype](#) : « *Décideurs politiques : ne perdez pas votre temps avec les fantasmes des technophiles qui disent : « Oh NON, on construit quelque chose de trop puissant.» Écoutez plutôt ceux qui étudient comment les entreprises (et les gouvernements) utilisent la technologie (et les discours sur l'IA) pour concentrer et exercer leur pouvoir. »*

En somme, ce qui est en jeu aujourd'hui, ce n'est pas seulement la performance d'un système technique, mais bien la manière dont nous souhaitons organiser nos sociétés face à l'émergence de nouveaux agents cognitifs artificiels. Sommes-nous en train d'ouvrir une ère de coopération inédite entre l'humain et la machine ? Ou risquons-nous de perdre, dans cette accélération, le contrôle sur nos propres systèmes ? L'intelligence artificielle n'est pas qu'un objet technique. Elle est un miroir tendu à notre condition humaine.

## ***Quelles sont les principales évolutions technologiques qui ont rendu possibles les IA génératives ?***

Les progrès spectaculaires de l'intelligence artificielle au cours de la dernière décennie ne doivent pas être interprétés comme le fruit d'une seule découverte technologique isolée. Ils s'inscrivent plutôt dans une écologie de l'innovation, c'est-à-dire un système complexe où plusieurs dimensions techniques, théoriques et matérielles interagissent de manière cumulative et exponentielle. Trois facteurs apparaissent comme déterminants dans cette trajectoire évolutive : l'abondance des données, l'accélération matérielle, et l'ingéniosité algorithmique.

### 1. L'abondance des données ou la matière première de l'intelligence computationnelle

À l'ère du numérique, les données sont devenues le substrat de l'intelligence artificielle, de la même manière que l'expérience constituait la base de l'empirisme classique. L'apprentissage automatique - et en particulier l'apprentissage profond - repose sur des corpus massifs d'informations : textes, images, vidéos, sons, données biométriques... autant de fragments du réel traduits en langage mathématique. Ces données permettent non seulement d'entraîner des modèles performants, mais aussi de faire émerger des corrélations et des représentations que l'intuition humaine ne saurait soupçonner.

Selon [EPOCH.AI](#), la taille des ensembles de données utilisés pour former les grands modèles de traitement du langage a crû à un rythme moyen de 2,9 fois par an depuis 2010, une dynamique qui reflète une sorte d'« inflation informationnelle ». Toutefois, cette expansion n'est pas infinie. Il est d'ores et déjà projeté que les données textuelles de haute qualité pourraient être épuisées avant 2026, et que celles relatives à la vision pourraient l'être entre 2030 et 2060. Nous atteignons donc une limite ontologique : le monde n'est pas infiniment descriptible, ou du moins pas indéfiniment dans des formats exploitables par les architectures actuelles.

Cette perspective soulève une question philosophique profonde : l'intelligence artificielle, pour persister dans son développement, devra-t-elle inventer ses propres réalités, par la génération de données synthétiques, ou redéfinir le cadre de son apprentissage via des paradigmes plus économes comme l'apprentissage actif ou par transfert ? L'« humain dans la boucle » (*human-in-the-loop*) comme autre solution possible suggère ici une dialectique inédite entre cognition humaine et apprentissage machine.

## 2. La puissance de calcul ou la matérialité de l'intelligence

La deuxième dimension fondamentale des progrès de l'IA repose sur le socle matériel, souvent invisible, de l'infrastructure computationnelle. Les CPU, GPU et autres TPU incarnent cette matérialité silencieuse, mais décisive, qui rend possible l'actualisation des potentiels algorithmiques. Leur évolution est

exponentielle : les capacités de calcul mobilisées pour entraîner les modèles les plus récents ont augmenté d'un facteur compris entre 4 et 5 par an depuis 2010 ([EPOCH.AI](#)). Ce rythme effréné témoigne d'un transfert énergétique massif vers la cognition machinique, mais il interroge également sur la soutenabilité environnementale et éthique de cette trajectoire.

Les innovations telles que les TPU (Tensor Processing Units) de Google ou encore les derniers GPU de la société NVIDIA permettent de traiter des masses de données à des vitesses inouïes, mais elles imposent en retour une dépendance aux ressources énergétiques et matérielles qui reconfigure la géopolitique du numérique. L'intelligence computationnelle ne flotte pas dans l'abstraction ; elle repose sur une économie des ressources physiques, sur des chaînes d'approvisionnement mondialisées, sur des infrastructures dont les implications sont éminemment politiques.

### 3. Les architectures de modèles ou la forme de l'intelligence

Enfin, le cœur des transformations actuelles réside dans la manière dont l'intelligence artificielle structure sa pensée - ou plus précisément, ses processus d'inférence et de modélisation. L'introduction en 2017 de l'architecture Transformer dans le papier désormais célèbre « [Attention is all you need](#) » et publié par Google Brain, constitue un moment paradigmatique. Elle ne se contente pas d'améliorer la performance : elle modifie la logique même du traitement informationnel. En abandonnant la

linéarité des modèles séquentiels classiques (de type RNN - *Réseaux de Neurones Récurrents* ou LSTM et LSTM - *Long Short-Term Memory*), les Transformers comme les GPT (*Generative Pretrained Transformer*) permettent une modélisation globale et contextuelle du langage. Ils capturent les relations entre les éléments d'un texte à longue distance, favorisant une compréhension plus fine et une génération plus cohérente.

La particularité des Transformers réside dans leur capacité à se focaliser sur les relations entre les mots d'une phrase, indépendamment de leur position.

- *Self-attention* (auto-attention) : Ce mécanisme est au cœur des Transformers et permet au modèle de pondérer l'importance de chaque mot en fonction des autres mots de la séquence. Par exemple, dans une phrase comme « Le chat noir dort sur le canapé », le modèle peut « se focaliser » sur le lien entre « chat » et « dort », et ignorer des éléments moins pertinents dans le contexte (comme « canapé »). Le mécanisme de *self-attention* calcule une matrice d'attention qui détermine les relations entre chaque paire de mots, permettant au modèle de saisir le contexte avec une grande précision.
- *Multi-head attention* : Cette approche consiste à diviser l'attention en plusieurs « têtes », où chaque tête se focalise sur un aspect différent de la relation entre les mots. Cela permet d'analyser des relations variées dans une phrase et d'obtenir une compréhension contextuelle plus

fine. Par exemple, une tête pourrait se concentrer sur les liens syntaxiques (comme sujet-verbe), tandis qu'une autre pourrait capter des relations sémantiques.

Les Transformers se distinguent des modèles séquentiels antérieurs, comme les RNN et les LSTM, par leur capacité à traiter simultanément tous les éléments d'une séquence au lieu de les traiter un par un. Cette approche présente plusieurs avantages :

- **Parallélisation** : Grâce à la structure non séquentielle des Transformers, les modèles peuvent traiter tous les mots d'une phrase en parallèle, ce qui accélère considérablement l'entraînement. En comparaison, les RNN devaient traiter les mots de manière séquentielle, rendant le processus plus lent
- **Capture des dépendances à longue distance** : Dans les phrases longues, les RNN et LSTM avaient du mal à capturer les relations entre des mots éloignés, car les informations se « diluaient » au fil de la séquence. Le mécanisme de *self-attention* des Transformers, quant à lui, permet de modéliser efficacement ces relations, quelle que soit la distance entre les mots.

Les modèles de langage de grande taille (LLM - *Large Language Model*) comme GPT-3 ou GPT-4 ne sont pas seulement capables de produire des textes. Ils incarnent une polyvalence cognitive émergente, puisqu'ils peuvent, à partir d'une simple consigne en langage naturel, rédiger, traduire, résumer, programmer, interpréter. En cela, ils deviennent des instruments de médiation

cognitive, des interfaces entre l'intention humaine et l'espace des possibles numériques.

Paradoxalement, plus ces modèles deviennent performants, plus la puissance de calcul nécessaire pour atteindre une performance donnée diminue, à un rythme estimé à un facteur de trois par an ([EPOCH.AI](#)). Cela signifie que l'intelligence artificielle ne progresse pas seulement par accumulation, mais aussi par raffinement, par une forme de sublimation algorithmique.

Les progrès de l'IA ne sont pas uniquement des faits techniques. Ils dessinent les contours d'une épistémologie nouvelle, où la production de savoir ne repose plus seulement sur la déduction humaine ou l'observation empirique, mais sur la mise en relation massive de données à travers des architectures computationnelles. Ce qui est en jeu, ce n'est pas uniquement ce que l'IA peut faire, mais ce que nous faisons de ce qu'elle fait, c'est-à-dire la manière dont nous intégrons cette intelligence exogène dans notre propre vision du monde.

Les trois dimensions décrites - données, matériel, architectures - ne sont pas indépendantes : elles forment un triptyque dont l'équilibre conditionne l'avenir de l'intelligence artificielle. À travers elles, c'est une nouvelle figure de la rationalité qui se dessine, algorithmique, distribuée, évolutive - et peut-être, à terme, autonome.

## CHAPITRE 2

# QUESTIONS PHILOSOPHIQUES

### ***Faut-il voir dans l'IA le risque d'un remplacement de l'intelligence humaine par la machine ?***

« On peut imaginer que les machines finiront par se mesurer à l'homme dans tous les domaines purement intellectuels », écrivait Alan Turing en 1950, dans ce qui demeure l'un des actes fondateurs de la pensée computationnelle. Cette intuition précoce – prophétique en un sens – précède de peu la naissance institutionnelle de l'intelligence artificielle en tant que champ de recherche, officialisée à Dartmouth en 1956, comme nous venons de le voir. Mais cette déclaration va bien au-delà d'une simple conjecture technique : elle exprime un renversement anthropologique latent, selon lequel les capacités que nous pensions exclusivement humaines pourraient un jour être partagées – voire surpassées – par des entités artificielles.

L'IA est née d'un rêve prométhéen, d'un désir de créer un double intellectuel de l'homme. Si les technologies ont évolué, ce rêve, lui, persiste comme un mythe opératoire au cœur des imaginaires contemporains. À chaque progrès significatif, il ressurgit, souvent accompagné d'angoisses existentielles : serons-nous un jour

remplacés ? Deviendrons-nous obsolètes ? Lorsqu'on aborde la question du supposé « remplacement » de l'intelligence humaine par la machine, cette idée s'inscrit souvent dans une représentation collective d'une évolution progressive, presque naturelle, où l'humain se verrait progressivement relégué dans une logique quasi-darwinienne, perçu comme moins efficace face à des dispositifs techniques toujours plus performants. Cette vision traduit moins une réalité objective qu'une angoisse contemporaine liée à notre rapport aux technologies, à la redéfinition des compétences, et au statut de l'humain dans des systèmes sociotechniques en mutation. Elle met en lumière une dynamique anthropologique où l'homme, en tant qu'être façonné par ses outils, en vient à redouter d'être supplanté par ceux-ci. Le philosophe et journaliste allemand, Gunther Anders, dans *L'Obsolescence de l'homme* (1956), identifie cette inquiétude avec acuité. Selon lui, la honte ressentie par l'homme moderne ne réside pas tant dans sa faiblesse que dans la perfection croissante de ses propres créations. Nous sommes confrontés à une asymétrie temporelle : nous restons biologiquement ce que nous étions au paléolithique, tandis que nos technologies évoluent à une cadence vertigineuse. Anders nomme cela la « honte prométhéenne » – ce malaise de se sentir inadéquat face à ses propres produits techniques.

Mais cette honte ne survient que si la machine devient notre semblable. Car nul ne rougit devant un marteau ou une paire de ciseaux, mais on peut être troublé par une IA capable de produire des poèmes en alexandrins, de résoudre des problèmes

complexes ou de diagnostiquer une maladie mieux qu'un médecin. Le problème ne réside pas uniquement dans la puissance de la machine, mais dans sa ressemblance cognitive apparente. C'est ici qu'intervient une hypothèse structurante de l'histoire de l'IA : celle de l'existence d'un « [système physique symbolique](#) », formulée par Allen Newell et Herbert A. Simon. Selon eux, un système capable de manipuler des symboles de manière appropriée possède, en principe, les moyens nécessaires et suffisants pour manifester une forme d'intelligence générale. Cette thèse implique deux choses : que l'intelligence humaine est réductible à un traitement symbolique – une opération syntaxique, pour ainsi dire – et que, de ce fait, une machine pourrait, en principe, en reproduire l'essence.

Cette idée plonge ses racines dans une tradition philosophique longue : Hobbes, pour qui le raisonnement n'est qu'un calcul ; Leibniz, qui rêvait d'un [calculus ratiocinator](#) universel ; Hume, qui fragmentait la perception en « [impressions atomiques](#) » ; Kant, enfin, pour qui toute expérience est médiée par des structures formelles de [l'entendement](#). L'IA s'inscrit dans cette filiation : penser serait calculer, percevoir serait classifier, comprendre serait corrélérer.

Mais de quelle « intelligence » s'agit-il ici ? Le plus souvent, d'une capacité à résoudre des problèmes : planifier, inférer, diagnostiquer, optimiser. Cette définition fonctionnelle – ou instrumentale – de l'intelligence, bien qu'opérationnelle, reste partielle. Elle fait de la cognition un ensemble de fonctions décomposables, modélisables et, par extension, simulables. Ce

réductionnisme a alimenté le fonctionnalisme cognitif, selon lequel une machine reproduisant les fonctions cognitives d'un cerveau humain pourrait être dite « intelligente ».

Or, l'histoire de l'IA a montré les limites de cette hypothèse. Les tentatives d'agrèger les fonctions du cerveau (vision, langage, calcul, mémoire, etc.) n'ont pas spontanément donné naissance à une intelligence globale. Les décennies 1970-1990, marquées par les hivers de l'IA, en témoignent : les promesses initiales se sont heurtées à des murs techniques et conceptuels.

Pourtant, l'émergence récente de l'IA générative constitue une rupture épistémologique. Non pas tant par sa nouveauté conceptuelle que par ses effets culturels et sociaux. Contrairement aux IA classiques orientées vers la classification ou la décision, les modèles génératifs (comme les LLMs) produisent de nouveaux contenus – textes, images, sons, vidéos, codes, musique – avec une aisance déconcertante. Quatre caractéristiques leur confèrent une aura quasi-magique :

1. La simplicité d'usage : l'utilisateur interagit par le langage naturel, abolissant la barrière technique entre lui et la machine. Cela crée une illusion de dialogue, de compréhension mutuelle, de réciprocité cognitive.
2. La rapidité d'exécution : quelques secondes suffisent là où l'humain mobiliserait heures ou jours. Cette compression du temps cognitif accentue le sentiment de décalage anthropotechnique.

3. La vraisemblance du résultat : les contenus produits ne sont pas seulement fonctionnels, ils sont esthétiquement et logiquement cohérents. Ils « ressemblent » à ce que produirait un humain.
4. La performance objective : certains modèles surpassent des humains dans des tâches complexes (examens de droit, diagnostics médicaux), même si ces performances doivent encore être rigoureusement contextualisées.

Ces propriétés favorisent l'automatisation de tâches auparavant réputées irréductiblement humaines, notamment créatives, interprétatives ou discursives. En cela, l'IA générative agit comme un miroir déformant de nos propres compétences : elle imite sans comprendre, génère sans intention, produit sans conscience. Mais cette absence de subjectivité n'annule pas son efficacité : c'est même peut-être ce qui la rend redoutable.

Ainsi se dessine une nouvelle figure de la honte prométhéenne : non plus face à la machine- outil, mais face à la machine-auteur, à la machine-interlocuteur, à la machine-penseur. L'angoisse n'est plus seulement celle du remplacement, mais celle de la dissolution des frontières de l'humain. Si la machine peut écrire, diagnostiquer, composer, alors que reste-t-il de notre spécificité ?

L'IA générative nous force à repenser ce que nous entendons par intelligence. Elle exhibe les formes extérieures de la pensée – langage, structure, enchaînements logiques – sans pour autant posséder d'intériorité, d'intentionnalité, ni d'expérience vécue. Elle n'est pas un sujet, mais un processus. Et c'est peut-être là

l'essentiel : l'IA ne nous remplace pas en tant que conscience, mais elle nous conteste en tant que fonction.

À nous, désormais, de décider si nous voulons défendre une conception de l'intelligence ancrée dans la subjectivité et la réflexivité, ou si nous acceptons qu'une intelligence sans conscience puisse, fonctionnellement, nous égaler – voire nous dépasser. Ce choix est autant ontologique qu'éthique. Car il ne s'agit pas seulement de savoir ce que la machine peut faire, mais de déterminer ce que nous voulons que l'homme reste.

## ***L'IA remet-elle en cause la façon dont nous définissons l'intelligence ?***

L'émergence rapide de l'intelligence artificielle ne constitue pas seulement une révolution technologique : elle interroge en profondeur notre conception de l'intelligence elle-même. Historiquement définie par des critères humains tels que la raison, le langage, l'adaptation ou la créativité, l'intelligence se voit aujourd'hui concurrencée par des systèmes artificiels capables de réaliser des tâches cognitives complexes. Cette situation génère une véritable blessure narcissique : et si l'intelligence humaine n'était qu'un cas particulier d'une compétence plus générale, potentiellement « artificialisable » ?

Traditionnellement, la philosophie occidentale, depuis Aristote jusqu'à Descartes, a associé l'intelligence à la raison, à la conscience et au langage. Or les IA modernes, notamment les grands modèles de langage, montrent qu'il est possible d'émuler certaines formes de raisonnement, de dialogue, voire d'invention, sans conscience ni subjectivité. Cela remet en cause la présupposition que penser suppose un sujet pensant.

La sociologie montre que la notion d'intelligence est socialement construite et historiquement située. L'école valorise certaines compétences cognitives (logique, mémorisation, abstraction) au détriment d'autres (intuition, sens pratique, émotion). L'IA amplifie cette tendance en automatisant la pensée formelle, rendant visibles nos propres biais de définition : ce que nous appelons

« intelligent » est souvent ce qui est efficace dans un cadre donné.

Du point de vue anthropologique, l'humain a longtemps été pensé comme un « animal rationnel ». L'apparition d'êtres non vivants capables de performances cognitives remplace la question : « Qu'est-ce que l'homme ? » par « Qu'est-ce qu'un esprit ? » ou même : « Qu'est-ce qu'un agent ? » L'écart entre intelligence biologique et computationnelle oblige à repenser l'évolution non comme un sommet, mais comme une bifurcation.

Actuellement, les IA sont spécialisées, statistiques et privées d'intentionnalité. Elles réussissent certaines tâches (traduire, diagnostiquer, raisonner sur des données), mais à l'intérieur de cadres imposés. En ce sens, elles sont intelligentes dans un sens opérationnel, mais n'ont pas d'intelligence au sens fort (compréhension, volonté, subjectivité).

Pour que l'IA devienne véritablement intelligente (et peut-être consciente), plusieurs conditions semblent nécessaires :

- Autonomie dialogique : capacité à poser ses propres questions (méthode socratique) ;
- Capacité critique : aptitude à remettre en cause ses propres modèles et valeurs ;
- Intégration sensorimotrice : lien entre perception, action et langage, à l'image de l'animal ou de l'enfant humain ;
- Finalité propre : orientation vers des buts internes, non dictés par l'humain.

Freud disait que Copernic nous a délogés du centre du monde, Darwin du sommet du vivant, et lui-même du contrôle de notre esprit. L'IA suggère que l'intelligence elle-même n'est plus notre propriété exclusive. C'est une blessure d'autant plus profonde qu'elle touche notre identité moderne fondée sur la rationalité.

Plutôt que de tenter de maintenir une définition rigide de l'intelligence humaine, cette blessure pourrait nous amener à valoriser d'autres dimensions : la vulnérabilité, l'éthique, la créativité poétique, le relationnel. Si l'IA nous dépasse en logique, alors c'est peut-être dans l'émotion, la mémoire incarnée, la transmission symbolique que réside notre singularité.

L'IA ne détruit pas la notion d'intelligence ; elle l'étend, la déplace et nous force à en redéfinir les contours. En cela, elle constitue une opportunité théorique, mais aussi existentielle : penser l'intelligence comme une compétence distribuée, partagée, peut nous libérer d'un humanisme centré sur la domination cognitive, et ouvrir à une anthropologie relationnelle, critique et post-humaniste.

## ***Faut-il craindre de l'IA qu'elle nous prive petit à petit de notre autonomie intellectuelle ?***

L'expérience contemporaine de l'interaction avec une intelligence artificielle - que ce soit via une interface conversationnelle comme ChatGPT, ou à travers la génération d'images ou de sites web - ne se réduit pas à un simple usage utilitaire. Elle suscite souvent l'impression, difficile à formuler mais persistante, qu'« il se passe quelque chose » : une reconfiguration silencieuse de notre rapport à la technique, une métamorphose de notre manière même d'agir et de penser. Cette intuition trouve un écho puissant dans la pensée de Bruno Latour, qui propose dans ses travaux sur la médiation technique une vision relationnelle et dynamique de l'humain et de la technologie.

Selon Latour, lorsqu'un agent humain se heurte à une limite dans la réalisation d'une finalité - un objectif ou une tâche - il peut s'engager dans un « [détour](#) » en s'alliant à un tiers, ici un non-humain, une entité technique, l'IA. Celle-ci n'a pas de fin propre, au sens intentionnel ou psychologique du terme, mais elle est configurée pour exécuter des fonctions précises (générer du texte, produire des images, résoudre des problèmes, ...). De cette association naît ce que Latour appelle un « tiers hybride », un nouveau type d'agent, fruit de la médiation entre l'humain et la machine. L'étudiant utilisant ChatGPT n'est pas simplement un étudiant augmenté : il devient le vecteur d'une finalité émergente qui n'existait ni chez l'un, ni chez l'autre pris isolément.

Ce couplage entre l'humain et la machine nous engage dans une

forme de coévolution, où les technologies ne sont pas de simples instruments neutres, mais bien des médiateurs qui reconfigurent nos habitudes de pensée et nos critères d'action. Cette relation, que l'on peut qualifier de quasi-symbiotique, soulève des questions fondamentales sur la nature de la cognition humaine à l'ère de l'intelligence artificielle. En ce sens, les travaux récents de chercheurs de l'Université Catholique du Sacré-Cœur (Italie), publiés dans [Nature Human Behaviour](#), prolongent l'intuition latourienne : l'IA pourrait fonctionner comme une extension cognitive de notre cerveau.

Ce prolongement cognitif engendre, selon ces chercheurs, l'émergence d'un nouveau mode de raisonnement, qu'ils nomment le « Système 0 », en référence et en distinction aux deux systèmes décrits par [Daniel Kahneman](#). Là où le système 1 est rapide, intuitif, et émotionnel, et le système 2 plus lent, réfléchi et logique, le système 0 externalise le raisonnement lui-même en le confiant à une instance algorithmique. L'IA n'interprète pas les données ; elle les manipule, les ordonne, les optimise. La signification, quant à elle, reste l'apanage de l'humain, seul capable d'assigner une valeur ou un sens aux sorties générées. Il s'agit donc d'une délégation partielle de la pensée, qui soulage certes certaines charges cognitives, mais au prix d'un possible affaiblissement de notre autonomie intellectuelle.

Ainsi, une [étude récente de l'université de Carnegie Mellon et de Microsoft](#) semble indiquer que l'IA a un impact négatif sur la pensée critique. Selon cette étude, l'intégration croissante de l'intelligence artificielle générative dans les environnements de

travail transforme en profondeur les dynamiques cognitives, les rapports au savoir et les formes de compétence. Cette mutation ne se limite pas à un changement d'outils, mais touche aux fondements mêmes de la culture du travail intellectuel. D'un point de vue sociologique, elle illustre un déplacement des pratiques professionnelles vers des formes de délégation automatisée de la réflexion, entraînant une homogénéisation des réponses, une réduction de l'initiative individuelle et un affaiblissement des compétences critiques.

Du côté anthropologique, cette évolution soulève des questions sur la transformation de l'humain dans son rapport à la technique : l'IA ne se contente pas de prolonger les capacités humaines, elle les reconfigure. La pensée critique, autrefois centrale dans les rituels cognitifs des « travailleurs du savoir », tend à être externalisée, fragilisant ainsi les formes d'apprentissage, de jugement personnel et d'adaptation culturelle. Ce phénomène de convergence mécanisée signale une possible mutation de l'humain vers une dépendance accrue aux systèmes algorithmiques, réduisant la diversité interprétative et symbolique propre aux sociétés humaines.

Face à ces constats, des pistes sont explorées pour « réhumaniser » l'usage de l'IA : faire en sorte que les technologies ne remplacent pas la pensée, mais qu'elles l'encadrent, la provoquent et la nourrissent. Il s'agit alors moins d'optimiser les performances que de préserver les conditions de subjectivation dans un monde de plus en plus médiatisé par des intelligences non humaines.

Ce danger n'est pas sans rappeler les craintes antiques formulées par Platon dans son *Phèdre*. Par la bouche de Socrate, Platon critique l'écriture - invention du dieu égyptien Thot - en la décrivant comme un faux remède à l'oubli. Loin de renforcer la mémoire, l'écriture risquerait, selon le roi Thamous, d'induire une illusion de savoir : elle nous dispense de l'effort de penser en profondeur, et rompt la transmission vivante du savoir. L'analogie est saisissante : de même que l'écriture externalise la mémoire, l'IA externalise une partie du raisonnement. Dans les deux cas, la technologie rend quelque chose possible, tout en fragilisant l'exercice même qu'elle prétend améliorer.

Victor Hugo écrivait dans *Notre-Dame de Paris* que l'imprimerie avait « tué » l'architecture, en lui substituant un nouveau médium pour l'expression symbolique de l'humanité. En trois mots,

Victor Hugo explicitait le fait que l'imprimerie allait tuer la sculpture, que le texte reproductible à l'infini remplacerait l'imagier sculpté au portail des cathédrales pour servir à l'éducation du peuple.

L'IA pourrait bien accomplir un geste similaire : moins un remplacement qu'un déplacement profond de nos structures mentales. Elle ne se contente pas de nous aider à penser ; elle transforme la nature même de la pensée.

## CHAPITRE 3

# QUESTIONS SOCIOLOGIQUES ET SOCIÉTALES

***A travers l'apparition des deepfakes, on voit poindre le risque de la désinformation et de la manipulation de l'opinion à grande échelle : que faut-il faire pour atténuer ce risque ?***

L'émergence des *deepfakes* (« hypertrucage » est le Terme proposé par l'Office québécois de la langue française) - ces contenus générés par intelligence artificielle qui imitent à la perfection voix, visages et comportements humains - représente moins une rupture technologique qu'un basculement ontologique : elle interroge ce que signifie encore, à l'ère numérique, le vrai, le réel, le témoignage, la confiance. Ce qui est en jeu ici n'est pas simplement la vérité factuelle, mais les conditions mêmes de la croyance partagée, sur lesquelles repose toute vie sociale.

Philosophiquement, ce phénomène nous replonge dans les questionnements platoniciens de l'illusion : dans *La République*, le « mythe de la caverne » décrit des hommes enchaînés, prenant des ombres pour des réalités. Les *deepfakes* ne sont pas de simples

ombres numériques ; ils sont des artefacts sophistiqués, générés par des réseaux de neurones profonds, capables de manipuler les signaux sensoriels là où, jadis, la vérité passait par la vérifiabilité perceptive. Le problème n'est donc pas que les images mentent - elles l'ont toujours fait - mais qu'elles deviennent indiscernables du réel, affaiblissant ainsi la distinction entre perception et simulation.

Scientifiquement, les *deepfakes* sont le produit d'un progrès impressionnant en apprentissage non supervisé, notamment via les GANs (*Generative Adversarial Networks*), qui opposent deux systèmes d'IA - l'un générant des contenus, l'autre tentant de les détecter comme faux. Cette dynamique est, par nature, évolutive : chaque amélioration dans la détection renforce, dialectiquement, la capacité de dissimulation. Nous assistons donc à une forme de « course aux armements » cognitivo-algorithmique, qui ne saurait être résolue uniquement par des moyens techniques.

C'est ici que l'analyse sociologique devient décisive. Comme le soulignait Niklas Luhmann, les sociétés modernes fonctionnent sur la base de [systèmes de confiance](#) : envers les institutions, les médias, les experts. Or, la circulation incontrôlée des *deepfakes* ne produit pas uniquement de la désinformation ; elle fragilise la structure même de cette confiance, en semant le doute non seulement sur ce qui est faux, mais aussi sur ce qui est vrai. La conséquence la plus grave

n'est pas l'adhésion à une fausseté, mais le cynisme généralisé, ce

que certains sociologues nomment la post-vérité, où toute information devient suspecte.

Dès lors, que faire ? Trois axes d'action s'imposent, me semble-t-il :

1. Un travail épistémologique sur l'éducation à l'image : il ne s'agit plus simplement de « repérer un fake » mais de former à une lecture critique des productions médiatiques, en incluant les logiques de production, les biais algorithmiques, les effets de cadrage. Cela suppose une refonte des compétences informationnelles, intégrant la philosophie des médias et l'éthique de la technologie.
2. Une régulation sociotechnique : il faut élaborer un cadre juridique et éthique à la fois national et international, qui responsabilise les créateurs et diffuseurs de contenus générés par IA, sans pour autant brider la recherche ou la liberté d'expression. Cela nécessite, comme le propose le sociologue Bruno Latour, de « [politiser les artefacts](#) », c'est-à-dire de reconnaître que la technique n'est jamais neutre, et qu'elle doit être intégrée dans les débats démocratiques.
3. Un renforcement des contre-pouvoirs cognitifs : des initiatives collectives de vérification des faits, des consortiums de journalistes, d'experts techniques, de philosophes, doivent œuvrer à rétablir des points de stabilité dans le chaos informationnel. L'objectif n'est pas de revenir à un âge d'or de la vérité objective - qui n'a jamais vraiment existé - mais de restaurer un espace de

discussion rationnelle, où les désaccords sont possibles parce qu'ils se fondent sur un socle partagé.

Sur le plan technique, de nombreux travaux ont été entrepris pour tenter de détecter les *deepfakes*, par exemple [ceux du MIT sur les vidéos](#). D'autres approches font appel, par exemple, au tatouage numérique qui permet de savoir si une image est authentique ou générée par une IA. Le tatouage numérique est une technique qui consiste à insérer de manière invisible ou faiblement perceptible des informations dans un contenu numérique (image, vidéo, audio, texte) afin d'en authentifier la source, protéger les droits d'auteur, ou tracer sa diffusion. Il s'apparente à une signature cachée ou à un marqueur intégré, un peu comme un filigrane invisible.

Le principe général du tatouage numérique est le suivant : tout d'abord on injecte un signal ou une séquence codée (ex. : identifiant, date, signature chiffrée) dans les données :

- dans les pixels d'une image ou d'une vidéo (en modifiant légèrement leur intensité) ; les coefficients de transformation (comme ceux issus d'une transformation de Fourier ou DCT, utilisés dans JPEG ou MPEG) ;
- les fréquences audio dans un fichier son.

Cette insertion est faite de manière à être robuste (résister à la compression, au recadrage, etc.) mais invisible pour l'œil ou l'oreille humaine.

Plus tard, un algorithme permet d'extraire ou de détecter la

marque sans nécessairement disposer du contenu original.

Certains tatouages sont fragiles (ils disparaissent à la moindre altération), d'autres sont robustes (ils survivent à certaines manipulations).

Le tatouage numérique est certainement l'une des stratégies les plus prometteuses contre les contenus générés ou modifiés par IA. Il peut servir à :

- Authentifier l'origine : En tatouant les contenus générés par des modèles IA (images, vidéos, voix), on peut signaler leur origine artificielle. Ainsi, OpenAI, Google ou Meta travaillent sur des systèmes qui marquent toutes les images produites par leurs IA.
- Assurer la traçabilité : Permet de remonter à la source d'un *deepfake* circulant sur Internet, ce qui peut être utile pour des vérifications dans les médias ou devant la justice.
- Aider à la détection automatisée : les systèmes peuvent scanner des flux en ligne à la recherche de contenus tatoués ou non tatoués. Ainsi, l'absence de tatouage pourrait suggérer un contenu suspect.

Cependant, cette promesse se heurte à plusieurs limites structurelles. D'abord, le tatouage peut être supprimé ou corrompu par des transformations numériques (compression, recadrage, filtres), rendant son efficacité incertaine. Ensuite, sa valeur dépend de son adoption généralisée : si seules quelques entreprises l'intègrent, des contenus produits par des IA non

éthiques échapperont à tout marquage. Enfin, le tatouage soulève des questions éthiques : qui contrôle les informations tatouées ? Peut-on tracer un contenu sans porter atteinte à la vie privée de son auteur ou de ses utilisateurs ? La frontière entre protection et surveillance devient floue.

En outre, le tatouage numérique ne résout pas la racine anthropologique du problème : notre tendance à faire confiance aux apparences. Si les individus ne sont pas éduqués à l'incertitude numérique et à la pensée critique, aucun dispositif technique ne pourra suffire. Le risque est alors celui d'une délégation aveugle de la véracité aux machines - paradoxe ultime d'un monde où l'intelligence critique humaine est progressivement désactivée au profit d'automatismes techniques.

La difficulté de la tâche a conduit AWS, Facebook, Microsoft, le [comité directeur pour l'intégrité des médias du Partenariat sur l'IA](#) et des universitaires à s'associer pour créer le *Deepfake Detection Challenge* (DFDC). L'objectif de ce défi est d'inciter les chercheurs du monde entier à développer de nouvelles technologies innovantes permettant de détecter les *deepfakes* et les médias manipulés.

En somme, atténuer le risque des *deepfakes* ne signifie pas simplement détecter la falsification, mais repenser nos écologies de la vérité : non plus comme un donné, mais comme une construction collective, toujours fragile, toujours à défendre.

## ***On parle beaucoup de l'impact de l'IA sur l'avenir du travail et sur les risques de disparition de millions d'emplois : que faut-il en penser et que faut-il faire ?***

L'essor de l'intelligence artificielle réactualise, dans l'espace public comme dans le débat intellectuel, une ancienne angoisse collective : celle de la disparition du travail humain. Lorsque des acteurs d'Hollywood, des entreprises de services ou des travailleurs indépendants redoutent l'éviction par la machine, ce n'est pas seulement une crainte économique qu'ils expriment, mais un trouble profond quant à la place de l'humain dans un monde technicisé. Cette préoccupation légitime nous fait nous souvenir, en écho, de la fameuse devise « Vivre libre en travaillant ou mourir en combattant. » qui fut celle, en 1831, des canuts, ouvriers lyonnais de la soie, qui se rassemblèrent et détruisirent les métiers mécaniques Jacquard qui commençaient à prendre leur place dans les usines. Ceci renvoie à une interrogation anthropologique sur le statut du travail dans nos sociétés : activité de subsistance, mais aussi de reconnaissance sociale, de construction identitaire et d'intégration symbolique.

Or, l'automatisation permise par l'IA opère un double mouvement. D'un côté, elle remplace certaines tâches et menace des emplois – c'est l'effet d'éviction. De l'autre, elle augmente la productivité, stimule la demande et génère de nouveaux besoins, de nouveaux rôles – c'est l'effet d'entraînement. Ce balancement entre destruction et création interroge non seulement l'économie, mais la capacité des structures sociales à redistribuer les gains, à

accompagner les transitions, et à reformuler les trajectoires professionnelles dans un monde mouvant.

Dans les récits dominants de la modernité industrielle, la technologie est souvent perçue comme un moteur ambivalent : elle détruit des formes d'emploi, bouleverse les équilibres sociaux, mais finit par engendrer de nouvelles fonctions, de nouveaux besoins, de nouvelles figures du travailleur. Ce schéma, hérité d'une lecture schumpétérienne du capitalisme, repose sur une foi implicite dans la capacité de l'innovation à se régénérer en permanence, en redistribuant les rôles économiques au sein de la société. Ainsi, après les bouleversements de la mécanisation, c'est dans l'extraction du charbon que s'est cristallisée une nouvelle demande de force de travail – preuve que chaque rupture technique appelle, tôt ou tard, une reconfiguration sociale.

Mais l'irruption de l'intelligence artificielle met à l'épreuve cette vision cyclique et relativement optimiste du progrès. En automatisant des tâches jusqu'ici considérées comme relevant de la compétence humaine la plus spécifique – le raisonnement, l'analyse, la création – l'IA semble s'attaquer non seulement aux métiers, mais aux fondements même de notre distinction entre l'humain et la machine. Ce déplacement ne concerne plus seulement l'économie, mais les représentations collectives : il fragilise une anthropologie implicite dans laquelle la valeur humaine était associée à la complexité, à l'intuition, à la réflexion critique.

Dès lors, la question n'est plus seulement de savoir si l'IA va, à terme, créer plus d'emplois qu'elle n'en détruit, mais de comprendre comment elle recompose les formes de l'activité humaine, les hiérarchies symboliques du travail, et les conditions de possibilité d'une vie sociale où chacun peut encore se reconnaître dans ce qu'il fait. Le défi est autant ontologique que sociétal : que devient l'humain dans un monde où ses fonctions cognitives sont partagées – voire surpassées – par des artefacts techniques ? Et comment penser collectivement la justice sociale à l'heure où l'innovation n'a plus pour horizon une redistribution automatique des rôles, mais une redéfinition de ce que signifie « contribuer » à la société ?

Les premières données empiriques, notamment en France et aux États-Unis, montrent que l'adoption de l'IA par les entreprises peut, dans certains cas, favoriser l'emploi. Ainsi, en France, l'enquête réalisée annuellement par l'Insee qui permet d'étudier les effets de l'adoption de l'IA par les entreprises en France (cf. [IA : Notre Ambition Pour La France](#)). On constate que l'emploi total des entreprises ayant adopté l'IA augmente davantage que dans les entreprises ne l'ayant pas adoptée, alors que ces deux groupes suivaient une tendance antérieure similaire. L'effet résulte principalement de la création de nouveaux emplois, plutôt qu'un maintien plus important d'emplois existants. On constate aussi qu'il n'y a pas d'effets différenciés sur les emplois occupés par des hommes par rapport à ceux occupés par des femmes : l'adoption de l'IA a des effets comparables pour l'emploi masculin et féminin.

Mais cette dynamique est inégalement répartie : certains métiers, certaines classes sociales, voire certains genres sont plus exposés que d'autres. Il ne s'agit donc pas d'un simple processus technique, mais bien d'une transformation sociale qui rejoue les inégalités préexistantes tout en en créant de nouvelles. L'automatisation devient alors un miroir grossissant de nos rapports de pouvoir, de nos hiérarchies professionnelles et des rôles sociaux différenciés.

L'IA générative introduit une rupture supplémentaire : elle rend substituables non plus seulement les tâches routinières, mais également des activités jusque-là perçues comme hautement humaines – créatives, réflexives, relationnelles. Ce glissement nourrit une remise en question du monopole cognitif de l'humain et, par ricochet, de la valeur sociale attribuée à certaines professions.

Une autre [étude](#) montre que l'arrivée de ChatGPT a eu un effet négatif sur l'emploi et la rémunération des travailleurs américains *free-lance* (travail rémunéré à la tâche). L'étude s'intéresse plus précisément à une plateforme proposant des missions ciblant des tâches de petite ou moyenne envergure au sein de plusieurs métiers (saisie de données, conception graphique, développement de logiciels, marketing, etc.) et met en évidence une baisse du nombre d'emplois et des revenus, y compris pour les actifs ayant une expérience plus importante, suite à la mise en service de ChatGPT en novembre 2022.

En résumé, ces premières études suggèrent que l'effet de

productivité domine en moyenne pour les salariés en entreprise, tandis que l'effet d'éviction semble être plus important pour les individus indépendants devant effectuer pour l'essentiel des tâches plus facilement remplaçables par l'IA.

Une autre approche plus prospective fondée sur les principales tâches composant l'économie a été utilisée par l'[OIT](#). Cette étude conclut que, dans l'ensemble du monde et y compris dans les pays développés, le nombre d'emplois ayant un potentiel *d'amélioration par l'IA* (13,4 %) est bien plus élevé que celui ayant un potentiel *de remplacement par l'IA* (5,1 %). L'étude note par ailleurs des inégalités de genre : 3,5 % des emplois principalement tenus par des femmes ont un potentiel de remplacement, contre 1,6 % des emplois principalement tenus par des hommes. Une autre [étude du Fonds Monétaire International](#) adoptant une méthodologie relativement proche aboutit à des chiffres plus élevés : 60 % des emplois seraient fortement exposés à l'IA, environ la moitié de ces emplois pouvant profiter d'un potentiel *d'amélioration par l'IA* tandis que l'autre moitié ferait face à un potentiel *de remplacement par l'IA*.

On constate que les deux approches convergent vers des conclusions similaires : au total, le déploiement de l'IA dans l'économie devrait avoir un effet global positif sur le nombre d'emplois. Les prévisions catastrophistes sur la fin du travail ne sont pas plus crédibles que les prévisions du même type faites dans le passé. Mais, la complexité des situations et des métiers ne permet pas de fournir une réponse uniforme et globale aux défis de l'IA. Des études plus précises sont nécessaires pour aborder la

variété des secteurs, des chaînes de valeur et prendre en considération les statuts et leurs évolutions.

Dans ce contexte, des propositions comme le revenu universel émergent comme réponses à la fois économiques et philosophiques. Il ne s'agit plus seulement d'indemniser une perte, mais de repenser le lien entre travail, revenu et dignité. [L'expérience menée par Sam Altman](#), le PDG et cofondateur d'OpenAI (depuis des années, Sam Altman s'inquiète publiquement de la nécessité d'un revenu universel à mesure que l'IA élimine les emplois traditionnels tout en créant d'énormes réserves de richesses détenues par quelques-uns), aux États-Unis, bien que sujette à débat, met en lumière des effets sociétaux qui dépassent la seule dimension monétaire : développement de l'autonomie, diminution des comportements à risque, mobilité choisie, sortie de situations d'abus. Ces résultats esquissent une anthropologie du « travail libéré » où la sécurité économique devient le socle d'une réinvention des existences.

En somme, l'intelligence artificielle ne pose pas seulement la question du remplacement de l'homme par la machine ; elle pose celle du contrôle social du progrès. Qui décide des usages de la technique ? Qui bénéficie de ses fruits ? Quelle vision de la vie bonne oriente nos choix collectifs ? La société qui s'esquisse ne sera pas seulement le produit de nos innovations, mais le reflet de nos valeurs et de notre capacité à redéfinir les finalités du travail, du lien social et de la solidarité.

Comme on vient de le voir, l'IA bouleverse les fondements du

travail, de l'éducation et de la citoyenneté. À mesure qu'elle s'infiltré dans les sphères productives et cognitives, elle redéfinit les compétences nécessaires à l'insertion professionnelle et à la participation active à la vie démocratique. Face à cette transformation structurelle, la formation devient un enjeu central. Il ne s'agit plus seulement de former « aux outils », mais de refonder des politiques publiques capables d'accompagner l'ensemble de la population dans cette mutation technologique, culturelle et sociale. L'enjeu est double : éviter l'exclusion numérique et préparer des individus capables de cohabiter avec des machines intelligentes.

La première dimension d'une politique publique de formation à l'ère de l'IA concerne l'école. Aujourd'hui, les systèmes éducatifs restent majoritairement fondés sur des modèles hérités de l'ère industrielle : transmission descendante, standardisation des savoirs, évaluation de la mémorisation. L'IA exige un basculement vers une éducation à la fois plus critique, plus créative et plus adaptative.

Il convient d'intégrer dans les programmes scolaires une littératie numérique avancée, qui dépasse la simple maîtrise technique des outils pour inclure la compréhension des modèles algorithmiques, des biais des données, et des enjeux éthiques associés. Mais cette approche ne doit pas être réservée à une élite technique. Elle doit devenir un socle commun, au même titre que la lecture ou les mathématiques, car l'IA n'est pas une spécialité : c'est une infrastructure sociale.

Par ailleurs, les savoirs disciplinaires doivent être hybridés : sciences humaines et sociales, philosophie, éthique et sociologie doivent accompagner l'enseignement du code, de la donnée et de l'automatisation. Ce croisement est nécessaire pour former des citoyens capables non seulement d'utiliser les IA, mais aussi de les questionner. L'enjeu n'est pas de produire une génération d'ingénieurs, mais une génération consciente.

La deuxième dimension d'un tel renouveau des politiques publiques est de former tout au long de la vie.

L'arrivée de l'IA sur le marché du travail annonce une reconfiguration majeure des emplois. Certains métiers seront automatisés, d'autres transformés, et de nouveaux rôles émergeront. Dans ce contexte, l'un des piliers des politiques publiques doit être la formation continue. Cela suppose une rupture avec les logiques actuelles de reconversion, trop souvent ponctuelles, inadaptées et inégalement accessibles.

Le pouvoir politique doit mettre en place une stratégie nationale de reconversion à grande échelle, articulée autour de plusieurs leviers :

- Des droits individuels à la formation étendus et garantis tout au long de la vie active, y compris pour les travailleurs précaires ou indépendants.
- Des plateformes publiques de formation en ligne, gratuites, collaboratives et certifiantes, qui permettent à chacun d'apprendre à son rythme.
- Des partenariats entre l'État, les universités et les

entreprises pour identifier les besoins en compétences émergentes et adapter les programmes en temps réel.

Cette politique doit viser non seulement l'employabilité, mais aussi l'autonomie cognitive : former les individus à coopérer avec les IA, à superviser leurs décisions, à combler leurs limites - plutôt qu'à les remplacer ou à s'en remettre passivement à elles.

Dernier volet du renouveau des politiques publiques : l'éducation populaire et l'inclusion afin de lutter contre la fracture cognitive.

En effet, une autre priorité des politiques publiques est d'éviter une fracture cognitive entre une minorité technophile capable de comprendre et d'influencer les IA, et une majorité reléguée à un rôle de consommateurs passifs. Cette fracture serait non seulement sociale, mais aussi démocratique.

Pour y faire face, il faut renforcer les politiques d'éducation populaire et de médiation numérique, à travers des dispositifs de proximité : bibliothèques, centres sociaux, maisons de quartier, médiathèques, où des ateliers d'initiation à l'IA, de débat public et d'analyse critique pourraient être organisés. Ces espaces devraient accueillir aussi bien les personnes âgées que les jeunes en décrochage, les travailleurs sans diplôme que les citoyens curieux.

Il s'agit ici de démocratiser la culture numérique, en rappelant que comprendre l'IA n'est pas un luxe, mais un droit. Car dans un monde où les décisions sont de plus en plus codées dans des algorithmes opaques, ne pas comprendre, c'est subir.

Face à l'essor de l'IA, l'éducation et la formation ne peuvent être des variables d'ajustement : elles doivent devenir le cœur des politiques publiques. Non pas dans une logique de rattrapage technologique, mais dans une perspective d'émancipation collective. Former à l'IA, c'est moins former aux machines qu'à l'humain que nous voulons rester. Cela suppose de sortir d'une vision utilitariste de l'éducation pour en faire un instrument de souveraineté cognitive, sociale et démocratique. Ce n'est qu'à cette condition que l'intelligence artificielle deviendra un outil d'augmentation collective - et non une force de marginalisation silencieuse.

## ***L'IA ne risque-t-elle pas d'accroître les écarts de richesse avec des personnes encore plus riches et d'autres encore plus pauvres ?***

L'intelligence artificielle (IA) suscite des interrogations quant à son impact sur les inégalités économiques. Les analyses économiques, scientifiques, sociologiques et anthropologiques convergent pour souligner que, sans régulation adéquate, l'IA pourrait exacerber les disparités de richesse.

Sur le plan économique, selon le [Fonds Monétaire International](#) (FMI), l'IA pourrait transformer environ 40 % des emplois à l'échelle mondiale, avec un impact plus prononcé dans les pays à revenu élevé, où environ 60 % des emplois pourraient être affectés. Cette transformation pourrait accentuer les inégalités salariales, touchant principalement les classes moyennes.

Les travailleurs hautement qualifiés, capables de collaborer efficacement avec l'IA, pourraient voir leur productivité et leurs revenus augmenter. En revanche, les travailleurs moins qualifiés risquent de se retrouver cantonnés à des emplois précaires, mal rémunérés et difficilement automatisables, tels que les services à la personne.

Sur le plan scientifique, les systèmes d'IA, s'ils sont entraînés sur des données biaisées, peuvent perpétuer ou amplifier les discriminations existantes. Par exemple, des algorithmes de recrutement ou d'octroi de prêts peuvent défavoriser certaines populations en fonction de critères tels que le genre ou l'origine

ethnique. De plus, le développement et le déploiement de l'IA sont souvent dominés par quelques grandes entreprises technologiques, concentrant ainsi les bénéfices économiques et le pouvoir décisionnel entre les mains d'une minorité.

Sur le plan sociologique, l'IA pourrait renforcer la fracture numérique, excluant les individus ou les communautés n'ayant pas accès aux technologies ou aux compétences nécessaires pour les utiliser. Cette exclusion pourrait aggraver les inégalités sociales, en limitant l'accès à l'éducation, à l'emploi ou aux services publics. Par ailleurs, l'automatisation de certaines tâches pourrait entraîner une dévalorisation de certains métiers, affectant l'estime de soi et la reconnaissance sociale des travailleurs concernés.

Dans une perspective anthropologique ce questionnement encourage à une redéfinition du travail et du lien social. Ainsi l'IA remet en question la place du travail dans nos sociétés. Si certaines tâches sont automatisées, cela pourrait libérer du temps pour d'autres activités, mais aussi entraîner une perte de sens pour ceux dont l'identité est fortement liée à leur profession. Cette redéfinition du travail pourrait modifier les structures sociales et les relations interpersonnelles.

L'IA présente donc des opportunités considérables, mais aussi des risques significatifs en matière d'inégalités économiques et sociales. Pour éviter une exacerbation des disparités, il est essentiel de mettre en place des politiques publiques visant à :

- Garantir un accès équitable aux technologies et aux

formations numériques.

- Assurer la transparence et l'équité des algorithmes.
- Redistribuer les gains de productivité générés par l'IA de manière inclusive.

En adoptant une approche multidisciplinaire et en impliquant l'ensemble des acteurs de la société, il est possible de faire de l'IA un levier de progrès partagé, plutôt qu'un facteur d'exclusion. Mais ceci risque fort de relever de l'utopie sans une réelle prise de conscience partagée à l'échelle mondiale.

L'émergence de l'intelligence artificielle générative (IA), notamment des grands modèles de langage (LLMs), a suscité un débat intense sur son impact potentiel sur les inégalités socioéconomiques. L'article [\*Generative AI and the Future of Inequality\*](#), publié par Nathan Wilmers du *Massachusetts Institute of Technology* (MIT), propose une analyse approfondie de la manière dont l'IA générative pourrait reconfigurer le marché du travail et influencer la répartition des revenus. Le présent essai synthétise les arguments et les résultats clés de ce travail, en explorant les manières multiples par lesquelles l'IA générative pourrait à la fois atténuer et aggraver les inégalités existantes.

Wilmers avance que l'IA générative pourrait jouer un rôle de « nivellement des compétences », en particulier au bénéfice des travailleurs les moins qualifiés dans différents secteurs. En améliorant leurs capacités, les outils d'IA pourraient réduire l'écart de performance entre les meilleurs employés et les autres. Cette démocratisation des compétences pourrait entraîner une

réduction des inégalités salariales, remettant en question les systèmes méritocratiques traditionnels qui justifient souvent les écarts de revenus. Par exemple, des tâches auparavant réservées à des professionnels hautement formés ou diplômés pourraient devenir accessibles à un plus large éventail de travailleurs, diminuant ainsi la prime associée aux diplômes prestigieux.

Par ailleurs, contrairement aux révolutions technologiques précédentes qui ont surtout touché le travail manuel, l'IA générative menace aujourd'hui les professions intellectuelles telles que les métiers du droit, de la médecine ou de la gestion. En automatisant des tâches comme l'analyse de données, la rédaction de rapports ou le diagnostic médical, l'IA pourrait diminuer la demande de main-d'œuvre hautement qualifiée. Ce déplacement pourrait entraîner une redéfinition de la valeur économique accordée à certaines compétences, voire une remise à plat des hiérarchies traditionnelles au sein des organisations.

De plus, la généralisation de l'IA générative risque d'éroder les justifications méritocratiques qui sous-tendent de nombreuses institutions éducatives et professionnelles. Si les outils d'IA permettent à un plus grand nombre d'individus d'accomplir des tâches complexes, l'importance des parcours scolaires d'élite et des compétences rares pourrait décliner. Cette évolution pourrait remettre en cause les processus de sélection fondés sur le capital culturel ou les réseaux sociaux, en favorisant des critères plus inclusifs en matière de recrutement et de promotion.

Malgré son potentiel d'égalisation, l'IA générative comporte aussi plusieurs risques d'accentuation des inégalités. Wilmers met en évidence plusieurs mécanismes :

- Déplacement du travail : L'automatisation des tâches par l'IA pourrait pousser certains professionnels à se reconverter dans des emplois moins qualifiés, augmentant la concurrence et exerçant une pression à la baisse sur les salaires dans ces secteurs.
- Disparités entre entreprises : Les entreprises capables d'intégrer efficacement l'IA pourraient creuser l'écart avec leurs concurrentes, consolidant leur position et exacerbant les inégalités de performance et de revenus.
- Surveillance accrue : Les technologies d'IA pourraient renforcer le contrôle sur les salariés, accroître la pression au travail et encourager des formes d'emploi précaires ou instables.

Enfin, la transformation induite par l'IA générative sur les marchés du travail pourrait encourager l'émergence de nouvelles formes de mobilisation collective. Wilmers suggère que les préoccupations liées aux pertes d'emplois et à la compression des salaires pourraient susciter un regain de syndicalisation, y compris chez les cols blancs traditionnellement éloignés de ces dynamiques. Les syndicats pourraient jouer un rôle clé pour négocier les conditions d'intégration de l'IA, défendre les droits des travailleurs, et garantir que les bénéfices de l'innovation soient partagés équitablement.

L'analyse de Nathan Wilmers propose une vision nuancée de l'impact de l'IA générative sur les inégalités sociales et économiques. Si cette technologie ouvre des perspectives de démocratisation des compétences et de réduction de certaines inégalités, elle risque aussi de renforcer les déséquilibres existants si son déploiement n'est pas accompagné de politiques publiques adéquates. Il est donc crucial que les décideurs, les éducateurs et les acteurs économiques anticipent ces défis, en investissant dans la formation continue, en favorisant des pratiques de recrutement inclusives, et en instaurant des régulations qui assurent une transition technologique juste et équilibrée.

## ***Quel est l'impact de l'IA sur le réchauffement climatique ?***

L'émergence des grands modèles de langage (LLM), tels que GPT-4, ne représente pas seulement un progrès technologique : elle cristallise un tournant décisif dans nos relations à la technique, à l'énergie, à l'environnement, et plus largement à notre organisation sociale. Derrière la puissance calculatoire se cache une réalité matérielle aux conséquences multiples, qui oblige à repenser les dynamiques d'innovation à l'aune de la soutenabilité.

L'illusion de l'immatérialité des technologies de l'IA masque un lourd tribut énergétique. L'entraînement de modèles de grande ampleur repose sur des centres de données nécessitant des millions de GPU, mobilisés pendant des semaines voire des mois. Le cas emblématique de GPT-3 - environ 552 tonnes de CO<sub>2</sub> émises pour un seul entraînement - montre l'ampleur de cette consommation. Les chiffres supposés pour GPT-4 sont encore plus vertigineux : jusqu'à 15 000 tonnes de CO<sub>2</sub>, soit des niveaux comparables à l'impact annuel de villes entières.

Cette demande énergétique croissante traduit une nouvelle logique [extractiviste](#) : après les ressources naturelles et humaines, c'est désormais la capacité électrique globale qui devient l'enjeu stratégique. Le projet « Stargate », porté par Microsoft et OpenAI, en est le symbole : 100 milliards de dollars pour créer un super-cluster d'entraînement de modèles dédié à l'IA, nécessitant une puissance de 5 GW, soit l'équivalent d'une

centrale nucléaire. Ces infrastructures sont si énergivores qu'elles exigent parfois leur propre production électrique, mettant sous pression les réseaux nationaux. Des pays comme l'Irlande ou Singapour imposent désormais des limites à l'implantation de nouveaux centres de données.

Les géants technologiques, tout en affichant des engagements vers la neutralité carbone (Microsoft promettant même un bilan négatif), peinent à tenir leurs promesses. Les rapports récents révèlent une envolée des émissions : +48 % pour Google depuis 2019, +30 % pour Microsoft depuis 2020. L'augmentation parallèle de la consommation électrique - parfois doublée en quatre ans - traduit une dépendance croissante à des ressources non renouvelables, souvent au détriment des objectifs climatiques globaux.

Cet écart entre les discours et la réalité révèle une tension fondamentale entre progrès technologique et soutenabilité écologique. À l'heure où les transitions énergétiques sont cruciales, l'IA semble suivre un chemin inverse, accélérant la consommation de ressources sans intégrer suffisamment les impératifs planétaires.

Dans l'imaginaire collectif, l'intelligence artificielle est souvent perçue comme une technologie immatérielle, flottant dans les nuages numériques, offrant à l'humanité des solutions efficaces et innovantes. Pourtant, comme le souligne la chercheuse Kate Crawford dans son ouvrage *Atlas of AI* (2021) et son article publié dernièrement : [\*Generative AI and Creative Work: Narratives,\*](#)

*Values, and Impacts* (2024), cette vision est une illusion dangereuse. Loin d'être immatérielle, l'IA repose sur des infrastructures matérielles massives et sur des dynamiques de pouvoir qui concentrent richesses et contrôle entre les mains de quelques acteurs. Comprendre l'impact de l'IA exige ainsi de replacer cette technologie dans son ancrage matériel et son contexte sociopolitique global. Ainsi, Kate Crawford déconstruit l'idée selon laquelle l'intelligence artificielle serait « propre » ou « virtuelle ». À chaque étape de son existence – de l'extraction des matières premières jusqu'aux Datacenters –, l'IA s'inscrit dans une chaîne d'exploitation écologique :

- **Extraction des ressources** : La fabrication des serveurs, des puces et des infrastructures de l'IA repose sur des métaux rares (cobalt, lithium, terres rares) dont l'extraction détruit des écosystèmes fragiles et engendre des conditions de travail souvent inhumaines.
- **Consommation énergétique** : L'entraînement des grands modèles de langage (comme GPT-3, GPT-4) demande des quantités astronomiques d'énergie. Par exemple, un seul modèle peut émettre autant de CO<sub>2</sub> que plusieurs voitures sur toute leur durée de vie.
- **Pression sur l'eau** : Les Datacenters utilisent des quantités colossales d'eau pour refroidir leurs serveurs, contribuant localement à des stress hydriques dans des régions déjà vulnérables.

Crawford insiste : l'IA est une technologie extractive, tout comme l'était historiquement l'exploitation coloniale des ressources naturelles. Elle prolonge donc une logique d'épuisement du vivant pour nourrir des promesses de progrès numérique. Sur le plan politique et économique, Kate Crawford démontre que l'essor de l'IA ne diffuse pas le pouvoir ; au contraire, il le concentre :

- Oligopole technologique : Quelques grandes entreprises (Microsoft, Amazon, Google, Meta, OpenAI) détiennent une avance massive en matière de données, de calcul et de capital. Cela leur permet de façonner les standards techniques, les usages, et même les régulations à leur avantage.
- Capture de l'État : Ces entreprises influencent les politiques publiques à travers le lobbying, les partenariats public-privé et leur pouvoir économique. Cela limite la capacité des États à réguler efficacement des technologies qui deviennent pourtant essentielles à la souveraineté.
- Asymétrie d'information : Le développement de l'IA repose sur la captation et l'exploitation de données personnelles à grande échelle. Ce système crée un déséquilibre radical : les individus sont transparents aux yeux des algorithmes, tandis que les mécanismes internes de l'IA restent opaques et protégés par le secret industriel.

Ainsi, loin d'incarner un progrès partagé, l'IA, telle qu'elle se développe aujourd'hui, renforce des dynamiques d'inégalités économiques, politiques et sociales. L'accès aux outils de l'IA, leur conception et leurs effets sont déterminés par une minorité de centres de pouvoir mondialisés.

En mettant en lumière ces dimensions souvent ignorées, Kate Crawford appelle à repenser profondément notre rapport à l'IA. Elle invite à :

- Démythifier l'IA, en la considérant non comme une entité autonome et neutre, mais comme une production sociale, ancrée dans des réalités matérielles et politiques.
- Demander des comptes aux acteurs dominants : par des mécanismes de transparence, de régulation internationale, de justice environnementale.
- Promouvoir des alternatives locales, éthiques et durables, notamment en développant des IA sobres, open source, et respectueuses des droits fondamentaux.

C'est une véritable question politique qui est posée : à quelles conditions pouvons-nous imaginer une intelligence artificielle qui serve l'humanité sans accélérer l'épuisement de ses ressources et la concentration des pouvoirs ?

À travers son analyse puissante et documentée, Kate Crawford nous oblige à regarder l'IA non pas comme une utopie technique détachée du monde, mais comme une extension des réalités matérielles, écologiques et politiques existantes. Loin de promettre automatiquement un monde meilleur, l'IA, dans son

développement actuel, exacerbe les crises environnementales et sociales. En prendre conscience est la première étape pour penser une alternative éthique et soutenable.

L'IA ne transforme pas seulement notre environnement naturel : elle reconfigure nos valeurs, nos priorités et notre rapport au monde. Elle impose une anthropologie de la vitesse, de l'accélération, de l'optimisation - une logique qui rappelle celle du capitalisme technicien décrite par Jacques Ellul ou Bernard Stiegler. Dans cette perspective, l'IA devient un moteur d'aliénation énergétique, capturant des ressources collectives pour servir des objectifs privés, souvent opaques. Ce modèle pose aussi des questions de justice : à qui profite cette énergie ? À quelles fins sociales est-elle mobilisée ? L'investissement de plusieurs dizaines de milliards dans l'entraînement de modèles fermés contraste avec les besoins criants en matière d'éducation, de santé ou d'accès à l'énergie pour les populations marginalisées. L'IA pourrait ainsi renforcer les inégalités, non seulement économiques, mais aussi environnementales et symboliques.

Face à ces défis, plusieurs voies s'ouvrent. L'une d'elles consiste à repenser les usages de l'IA à travers le prisme de la sobriété et de l'efficacité. Les petits modèles linguistiques (SLM), par exemple, montrent qu'il est possible d'obtenir des performances significatives à un coût énergétique bien moindre. Leur spécialisation, leur accessibilité et leur compatibilité avec les énergies renouvelables en font des alternatives crédibles aux LLM.

D'autres initiatives, comme l'optimisation des architectures (*Sparse Transformers*, distillation de connaissances) ou les méthodes d'entraînement adaptatif, visent à réduire le gaspillage computationnel. Le recours aux énergies renouvelables, ou encore à du matériel spécialisé plus économe, comme les puces optiques analogiques, contribue également à faire émerger une IA plus durable.

Enfin, l'IA elle-même peut être mobilisée au service de l'environnement. Des recherches prometteuses utilisent les modèles génératifs pour identifier des matériaux capables de capturer efficacement le CO<sub>2</sub>, comme les structures métallogéniques (MOF). En explorant des milliards de configurations moléculaires, l'IA permet d'accélérer la recherche vers des solutions concrètes à la crise climatique. C'est d'ailleurs l'un des paradoxes de l'IA : plus on l'utilise et plus les modèles mis en œuvre sont de grande taille, plus l'impact environnemental engendré est important mais l'IA, à travers par exemple la mise au point de nouveaux matériaux, peut permettre de lutter contre le réchauffement climatique. On peut même imaginer de remonter dans le temps en capturant le CO<sub>2</sub> contenu dans l'atmosphère et en le [séquestrant](#). Un tel paradoxe renforce indéniablement la tentation du « solutionnisme technologique », concept popularisé en 2014, sous la plume du chercheur américain d'origine biélorusse Evgeny Morozov dans son ouvrage « *Pour tout résoudre, cliquez ici* ».

L'intelligence artificielle représente une bifurcation majeure dans notre histoire technologique. Mais si elle veut être un vecteur d'émancipation, elle ne peut faire l'impasse sur sa matérialité. Ce que nous appelons « intelligence artificielle » aujourd'hui est le produit d'une chaîne énergétique, industrielle et politique complexe. Il ne s'agit donc pas seulement d'innover, mais de choisir dans quelle direction orienter cette innovation.

Il est temps de construire une intelligence artificielle située, réflexive et écologiquement consciente - non pas seulement performante, mais juste et soutenable. Cette transition ne pourra se faire qu'en intégrant des approches pluridisciplinaires, en impliquant la société civile, et en sortant d'un imaginaire technosolutionniste qui confond croissance exponentielle et progrès humain.

## CHAPITRE 4

# QUESTIONS SUR L'IA ET L'ÉDUCATION

***L'arrivée de ChatGPT dans le monde de l'éducation a engendré énormément de questionnements : peut-on mettre l'intelligence artificielle au service de l'éducation et, si oui, comment ?***

Depuis les balbutiements de l'informatique jusqu'à l'irruption récente de modèles conversationnels comme ChatGPT, l'éducation n'a cessé d'être traversée par la question du rapport entre savoir, technique et humanité. Chaque avancée technologique – de l'imprimerie à Internet – a été à la fois porteuse d'enthousiasme et d'inquiétude : peur de l'oubli des savoirs traditionnels et espoir d'une diffusion démocratisée du savoir. L'intelligence artificielle s'inscrit dans ce continuum, mais avec une intensité inédite. Peut-on dès lors mettre cette nouvelle puissance au service de l'éducation sans en perdre le sens profond ? Et si oui, comment ?

L'une des principales promesses de l'intelligence artificielle en éducation est certainement la personnalisation des

apprentissages à travers un tutorat personnalisé. Là où l'enseignement traditionnel peine à s'adapter aux rythmes et styles cognitifs divers, l'IA, par ses capacités d'analyse de données et d'adaptation en temps réel, offre la perspective d'un accompagnement véritablement différencié. Elle peut également libérer les enseignants de certaines tâches répétitives, leur redonnant du temps pour se consacrer à la médiation humaine, la créativité pédagogique, et le lien social.

Dans cette optique, l'IA n'est pas un substitut, mais un outil au service d'une éducation plus humaine. À condition, toutefois, de penser son intégration de manière critique, sans céder aux sirènes de la substitution technologique.

L'usage non encadré des IA génératives présente cependant un risque majeur : celui d'une délégation massive de l'effort intellectuel. Pourquoi s'efforcer de comprendre, de rédiger, d'analyser, si une machine peut produire des réponses instantanées et convaincantes ? Ceci est, décrit avec un exemple très parlant, dans le [billet](#) de Franck Jeannin, le président et cofondateur d'ALGOSUP aux côtés d'Éric Larchevêque, le cofondateur de Ledger « Depuis un moment déjà, j'observe les signes avant-coureurs : les examens oraux où certains étudiants récitent par cœur un texte qu'ils ne comprennent manifestement pas, au contenu totalement générique et émaillé de mots qui ne font pas partie de leur vocabulaire. Ou encore ces rapports écrits où l'étudiant ne prend même pas la peine de relire la production de ChatGPT et copie/colle la totalité du texte, y compris le *prompt* (la question posée à l'IA). » Sa conclusion est des plus

éloquente : « Moralité : l'IA est un outil formidable pour ceux qui sont capables de s'en passer. »

Pour répondre à un tel défi, l'approche pédagogique doit évoluer en mettre davantage l'accent sur le processus plutôt que sur le seul résultat final ; valorisant des activités où l'élève coconstruit ses savoirs avec l'IA, au lieu de lui demander des réponses toutes faites ; formant à l'usage critique de l'IA, comme on forme à la lecture critique des sources ou à l'esprit scientifique.

Ainsi, l'enjeu n'est pas d'interdire l'IA, mais de l'intégrer comme un partenaire d'apprentissage réfléchi, au même titre que le livre ou Internet avant elle.

Depuis Platon et son célèbre mythe de l'invention de l'écriture (dans le "Phèdre"), la peur que les technologies affaiblissent la mémoire, la réflexion et l'autonomie humaine est récurrente. Socrate lui-même voyait dans l'écriture un risque : celui de substituer un savoir mort à la véritable sagesse.

L'IA ravive cette vieille interrogation : l'outil nous libère-t-il ou nous asservit-il ? La réponse dépend moins de l'outil que de l'usage que nous en faisons. Comme le souligne Bernard Stiegler, philosophe contemporain de la technique, tout outil est un « *pharmakon* » : à la fois poison et remède mais aussi bouc-émissaire. Il appartient à l'éducation de faire de l'IA un remède – un moyen d'augmenter la capacité critique et réflexive des élèves – plutôt qu'un poison qui atrophie l'esprit.

Du point de vue sociologique, l'intégration de l'IA en éducation pose plusieurs défis majeurs :

- Inégalités d'accès : Tous les élèves n'ont pas les mêmes ressources pour bénéficier des outils d'IA, risquant d'accentuer les fractures sociales existantes.
- Transformation du rôle des enseignants : Ceux-ci deviennent de moins en moins de simples transmetteurs de savoir, et de plus en plus des médiateurs, des guides critiques au sein d'un environnement saturé d'informations.
- Réinvention des dynamiques éducatives : L'apprentissage devient potentiellement plus horizontal, plus collaboratif, mais aussi plus fragile si le lien humain est négligé.

L'expérience de [Khanmigo](#), l'assistant IA de la Khan Academy, illustre bien ces enjeux : en agissant, non comme un fournisseur de réponses, mais comme un « Socrate numérique », il encourage l'élève à raisonner, à persévérer, à s'interroger. « Je ne peux rien apprendre à personne, je ne peux que les faire réfléchir ». Cette citation attribuée à Socrate, est au cœur de la méthode d'enseignement qu'il a créée, la méthode socratique. Au cœur de cette méthode se trouve une pratique sociale d'interrogation, fondée sur l'examen critique des énoncés par le biais du dialogue. Cette démarche, plutôt que de produire immédiatement des affirmations alternatives, repose sur un processus de déconstruction : elle fragilise les certitudes en les soumettant à l'épreuve de la discussion collective. D'un point de vue sociologique et anthropologique, cette méthode peut être comprise comme un rituel de mise à l'épreuve des normes de

pensée partagées, où le savoir émerge moins comme possession individuelle que comme produit d'une interaction sociale visant à révéler les limites des représentations admises.

Pour cela, on peut utiliser un prompt ([instruction](#) selon le [Journal Officiel](#)) socratique comme dans l'exemple suivant :

Imaginons que je demande à ChatGPT : « peux-tu me résoudre l'équation :  $2x^2-8x+3=0$  ».

Dans une utilisation normale de ChatGPT, ce dernier me fournira la réponse en me donnant les principales étapes de la résolution et les 2 racines de cette équation du second degré. Si, par compte, je fais précéder la même question par le prompt suivant : « Agis comme un professeur socratique : aide-moi à résoudre cette équation en me posant des questions pour que je comprenne. », ChatGPT ne me fournira pas directement la réponse à ma question mais commencera par le dialogue suivant :

**ChatGPT** : « Très bien. Prenons l'équation suivante ensemble :  $2x^2-8x+3=0$ .. Commençons par la base. Quelles sont les différentes formes d'une équation quadratique que tu connais, et à quoi servent-elles dans la résolution ? » Moi : «  $ax^2+bx+c=0$  »

**ChatGPT** : « Excellent ! Tu as bien identifié la forme générale de l'équation quadratique. Dans cette forme, quels sont les coefficients a, b et c pour notre équation  $2x^2-8x+3=0$  ? » Moi : «  $a=2$ ,  $b=-8$  et  $c=3$  »

**ChatGPT** : « Très bien, c'est exact :  $a=2$   $b=-8$ ,  $c=3$ . Maintenant, pour résoudre cette équation, on peut utiliser la formule du discriminant :

$$\Delta=b^2-4ac$$

Saurais-tu calculer la valeur de  $\Delta$  avec les valeurs que tu viens de donner ? »

...

On voit qu'une telle approche permet d'utiliser ChatGPT comme un tuteur et non comme un fournisseur de solution...Un tel idéal pédagogique demande cependant une conception exigeante et une supervision humaine constante. Cette approche est voisine d'une approche originale proposée par un chercheur de DeepMind dans un article : [Boundless Socratic Learning With Language Games](#) qui a proposé de fonder l'intelligence artificielle future non sur des données sensorielles, mais sur le langage seul, à travers une méthode inspirée de Socrate. L'idée : faire émerger une IA capable de raisonner et s'améliorer par elle-même en dialoguant avec ses propres productions linguistiques, sans recours à des observations du monde réel.

Cette approche s'ancre dans la tradition platonicienne, où le langage est le vecteur principal de la pensée. Comme dans les dialogues de Platon, l'IA avancerait par réfutation (*elenchos*), explorant ses propres contradictions pour progresser, avec « l'ignorance savante » comme moteur cognitif. Si elle aboutissait, cette méthode bouleverserait les rapports au savoir, au travail

scientifique et aux processus éducatifs. Elle annoncerait une ère où les machines pourraient innover sans données empiriques, remettant en cause le monopole humain sur la découverte rationnelle.

Trois conditions sont nécessaires pour une telle mise en place : des boucles de rétroaction efficaces (pour s'auto-corriger), une couverture conceptuelle large (pour généraliser), et une capacité de calcul suffisante. La clé : articuler des jeux de langage (au sens de Wittgenstein), chacun doté de règles internes, permettant à l'IA de s'auto-évaluer. Le vrai défi reste la mise en place d'un mécanisme méta-linguistique capable de coordonner ces jeux pour nourrir l'apprentissage.

L'irruption de l'IA dans le monde éducatif est à n'en pas douter une opportunité historique : celle de repenser l'acte d'enseigner et d'apprendre à l'ère des machines intelligentes. Mais elle est aussi un test : saurons-nous préserver l'essence de l'éducation – le développement autonome de l'esprit humain – face à la tentation de la facilité technologique ?

La réponse ne sera ni purement technique ni purement pédagogique : elle sera fondamentalement éthique et politique. Former les esprits à utiliser l'IA, non pour se soustraire à l'effort de penser, mais pour aller plus loin dans l'exigence critique et créative, telle est la mission nouvelle d'une éducation digne de ce nom à l'ère de l'IA.

## CHAPITRE 5

# QUESTIONS SUR LE FUTUR DE L'IA

***Quelle est ta vision sur le futur de l'IA ? Verra-t-on rapidement l'arrivée d'une AGI ? D'une IA forte ?***

Tout d'abord, l'Intelligence Artificielle Générale (IAG) constitue un objet de recherche aussi fascinant qu'insaisissable. Au croisement de la science cognitive, de l'informatique théorique, de la philosophie de l'esprit et de l'ingénierie computationnelle, elle désigne un horizon technologique : celui de systèmes artificiels capables de comprendre, apprendre, raisonner et s'adapter de manière générale, indépendamment de contextes ou de tâches spécifiques.

Mais cette généralité reste difficile à cerner. Peut-on en fournir une définition rigoureuse ? Cette question n'est pas seulement lexicale : elle engage des considérations ontologiques, épistémologiques et méthodologiques. Définir l'IAG, c'est aussi prendre position sur la nature de l'intelligence, sur les limites de la computation, et sur la possibilité même de simuler la cognition humaine.

Dans une approche fonctionnelle - majoritaire dans les milieux industriels et en IA appliquée - l'IAG est définie comme la

capacité, pour un système, de réaliser l'ensemble des tâches cognitives humaines, ou du moins un sous-ensemble suffisamment large pour permettre la généralisation. Elle se caractérise généralement par : la polyvalence adaptative, l'apprentissage par transfert, la résilience cognitive en contexte inconnu.

Mais cette définition, si elle est pragmatique, reste behavioriste : elle évalue l'intelligence à l'aune de la performance observable, sans se prononcer sur les mécanismes internes, ni sur la nature de la compréhension mobilisée.

Certains chercheurs privilégient aussi une définition comparative : l'IAG est une IA dont les capacités égalent ou dépassent celles de l'intelligence humaine dans toutes ses dimensions. Cette définition s'appuie sur une anthropomorphisation implicite, posant l'humain comme métrique de référence. Elle soulève des enjeux quantitatifs (niveau de performance), qualitatifs (plasticité cognitive, émotion, conscience, créativité). Mais elle suppose une connaissance stable de ce qu'est l'intelligence humaine – or celle-ci est multidimensionnelle, incarnée, culturellement située.

Également, dans une perspective formelle, des travaux comme ceux de Marcus Hutter ([AIXI](#)) ou Legg & Hutter ([intelligence universelle](#)) cherchent à définir l'IAG comme un agent capable d'optimiser ses performances dans une grande classe d'environnements, indépendamment de leur structure a priori. Ces modèles proposent une formalisation élégante, mais

nécessitent des ressources irréalistes (omniscience bayésienne, capacité de calcul infini) et reposent sur des hypothèses idéalisées. Ils offrent néanmoins un cadre théorique utile pour évaluer les limites fondamentales de l'apprentissage et de la généralisation computationnelle.

Enfin, une tradition issue de la [philosophie de l'esprit](#) postule qu'une véritable intelligence générale requiert plus que des performances fonctionnelles : elle suppose la présence d'une compréhension sémantique, d'intentionnalité, voire de conscience phénoménale.

Des arguments célèbres comme : la « [chambre chinoise](#) » (Searle) ou le « [problème difficile de la conscience](#) » (Chalmers), soulignent les limites d'une approche purement syntaxique. Dans ce cadre, une IAG sans subjectivité ne serait qu'une simulation d'intelligence – non une intelligence authentique.

Plutôt qu'une définition unique et universelle, il semble plus pertinent d'adopter une approche pluraliste et contextuelle, distinguant :

- IAG faible : généralité fonctionnelle sans subjectivité ;
- IAG forte : intelligence dotée d'intentionnalité, d'auto-modélisation, voire de conscience ;
- IAG instrumentale : systèmes multi-domaines capables de raisonnement abstrait dans des environnements complexes.

Ce découplage sémantique permettrait de clarifier les débats, d'orienter les recherches, et de mieux anticiper les implications éthiques et sociétales.

Définir l'Intelligence Artificielle Générale engage donc bien plus qu'un choix de mots : c'est adopter une posture vis-à-vis de ce que signifie penser, comprendre, apprendre. Il n'existe pas de définition neutre, mais seulement des définitions situées, reflétant nos conceptions de l'humain, de la cognition et de la technique.

C'est peut-être dans cette indétermination même que réside la richesse du concept : l'IAG n'est pas seulement un objet technique à définir, mais un miroir tendu à notre propre intelligence.

Pour en revenir à la question de l'émergence de l'IAG et de son calendrier, les avis divergent fortement, divisant chercheurs et penseurs en trois grands courants : les « bientôtistes », les

« tardistes » et les « jamaisistes ». Chacun présente des arguments convaincants, fondés sur les tendances technologiques, les perspectives philosophiques et des hypothèses sur la nature même de l'intelligence.

## 1. Le scénario du « Bientôt » (IAG dans 5 à 20 ans)

Les optimistes – ou peut-être les alarmistes – soulignent les progrès fulgurants de l'apprentissage automatique au cours de la dernière décennie. Le saut des premiers systèmes de traitement du langage naturel aux modèles de langage actuels comme GPT-4, Claude ou Gemini est frappant. Ces systèmes génèrent un langage fluide, écrivent du code, raisonnent en mathématiques et en logique, et montrent même des formes rudimentaires d'utilisation d'outils et de mémoire.

Les partisans de cette vision estiment que l'IAG n'est pas un rêve lointain, mais une question d'échelle et d'affinage. La *scaling hypothesis* ou « hypothèse d'échelle » suggère que l'intelligence émerge naturellement avec l'augmentation de la taille et de la complexité des modèles. Cette hypothèse désigne l'idée selon laquelle les performances des modèles d'IA, en particulier des modèles d'apprentissage profond, continuent à s'améliorer de manière prévisible lorsqu'on augmente leur taille (nombre de paramètres), la quantité de données d'entraînement, et la puissance de calcul disponible.

Autrement dit, « plus gros = meilleur » – à condition de disposer des ressources nécessaires. Cette hypothèse repose sur des résultats empiriques, notamment observés dans le développement des grands modèles de langage (LLMs), où l'on constate que : (a) les erreurs diminuent selon une loi de puissance à mesure que l'on passe à l'échelle, (b) des capacités émergentes (comme le raisonnement, la traduction ou le codage) apparaissent à certaines échelles de taille ou de données. Ces

capacités émergentes ont été décrites dans un article de recherche publié par Sébastien Bubeck et son équipe : « [Sparks of Artificial General Intelligence: Early experiments with GPT-4](#) ». Sébastien a d'ailleurs travaillé au sein du laboratoire de recherche commun entre l'Inria et Microsoft Research que j'avais contribué à créer en 2005.

Cette hypothèse est au cœur de la stratégie adoptée par OpenAI, Google DeepMind, Anthropic et d'autres laboratoires, qui misent sur des modèles toujours plus massifs pour franchir de nouveaux seuils cognitifs.

Certains estiment que nous observons déjà des « proto-IAG » dans les systèmes actuels, dotés de capacités multimodales, d'une auto-amélioration limitée, et d'une généralité croissante.

Ils notent aussi que les entreprises d'IA se lancent dans la course à l'IAG avec des investissements colossaux. OpenAI, DeepMind, Anthropic affichent explicitement cet objectif. Les capacités de calcul, les données d'entraînement et l'efficacité algorithmique augmentent de manière exponentielle. Dans cet élan, quelques percées supplémentaires en mémoire, planification ou raisonnement dirigé par des buts pourraient suffire à franchir le seuil.

Mais cet optimisme est nuancé par la prudence. Malgré leur polyvalence, les systèmes actuels restent étroits. Ils manquent de compréhension véritable, d'ancrage dans le monde réel et de capacité à généraliser solidement hors des distributions d'entraînement. Leur raisonnement est fragile, et ils peinent avec

la pensée abstraite, le bon sens ou la morale. L'IAG pourrait nécessiter bien plus qu'un simple changement d'échelle : peut-être un changement de paradigme complet.

## 2. Le scénario du « Plus Tard » (IAG dans 50 à 100 ans ou plus)

Ceux du camp « plus tard » acceptent la possibilité de l'IAG mais estiment que les trajectoires actuelles sont insuffisantes. Ils soulignent que l'intelligence générale n'est pas une somme d'aptitudes étroites ; elle implique l'apprentissage par transfert, la modélisation flexible du monde, l'expérience incarnée, la compréhension émotionnelle et la cognition sociale. Ces aspects de l'intelligence sont profondément enracinés dans la biologie, le contexte et l'évolution.

Ce courant soutient que notre compréhension scientifique fondamentale est encore lacunaire. Nous ne savons pas comment créer des systèmes qui comprennent la causalité, maintiennent une identité persistante ou forment des objectifs de manière autonome. Beaucoup de chercheurs estiment que les modèles de langage sont des imitateurs impressionnants mais superficiels – capables d'imiter un discours intelligent sans en saisir le sens. Des critiques comme Noam Chomsky affirment que l'IA manque de véritable sémantique ; elle traite la forme, pas le fond.

L'analogie avec les débuts de l'aviation est éclairante : il a fallu des siècles de théorie, d'expérimentation et d'échec avant que

l'humanité ne maîtrise le vol. L'IAG, étant bien plus complexe et moins comprise que le vol, pourrait nécessiter des décennies d'avancées théoriques – en neurosciences, en sciences cognitives, voire en informatique quantique – avant que de réels progrès ne soient réalisés.

De plus, l'intégration est un défi majeur. Une IAG devrait intégrer perception, action, raisonnement, mémoire, émotion et apprentissage dans un agent cohérent et orienté vers un but. Les systèmes actuels sont largement modulaires et disjoints. Créer une architecture cognitive unifiée pourrait prendre bien plus de temps qu'on ne l'anticipe.

### 3. Le scénario du « Jamais »

Les voix les plus sceptiques avancent que l'IAG est un mirage – un concept fondé sur une anthropomorphisation excessive et des hypothèses erronées sur l'intelligence. Selon elles, l'intelligence n'est pas un algorithme abstrait mais le produit de la biologie humaine, de la culture, des émotions et de l'expérience vécue. Les machines peuvent simuler certains aspects de l'intelligence, mais ne peuvent être intelligentes comme le sont les humains.

D'un point de vue philosophique, la conscience, l'intentionnalité et la compréhension pourraient ne pas être répliquables de façon computationnelle. L'argument de la « chambre chinoise » de John Searle dont nous avons déjà parlé suggère qu'une machine peut manipuler des symboles (syntaxe) sans comprendre leur sens

(sémantique). De même, le « problème difficile de la conscience » dont nous venons également de parler demeure irrésolu : même si l'on peut simuler un comportement intelligent, peut-on créer une conscience de soi ou une expérience subjective ?

Certains sceptiques pensent que l'IAG n'est pas seulement lointaine – elle est incohérente dans sa conception. Ils estiment que les chercheurs en IA confondent performance sur des benchmarks et véritable intelligence, et que l'idée d'une machine dotée d'une généralité humaine est une erreur de catégorie. D'autres invoquent le « problème du cadre » : la difficulté d'anticiper toutes les conséquences pertinentes dans des situations réelles – une tâche que les humains gèrent sans effort, mais que les machines trouvent inabordable.

Il y a aussi une inquiétude pragmatique : l'IAG pourrait s'avérer indésirable ou trop risquée à poursuivre. Le problème de l'alignement – s'assurer que les objectifs de l'IAG correspondent aux valeurs humaines – reste irrésolu. Certains estiment que l'humanité pourrait volontairement interrompre le projet IAG avant qu'il n'aboutisse, par crainte des risques existentiels qu'il pose.

L'IAG émergera-t-elle bientôt, plus tard, ou jamais ? Chaque position en dit plus sur nos hypothèses que sur l'avenir lui-même. Le camp du « bientôt » mise sur l'élan technologique et l'échelle computationnelle ; le camp du « plus tard » fait confiance à la complexité et à l'inconnu ; le camp du « jamais » s'appuie sur le scepticisme philosophique et l'unicité de la cognition humaine.

En vérité, le calendrier de l'IAG dépend d'inconnues encore insoupçonnées : des percées que nous ne pouvons pas encore prédire, des obstacles que nous n'avons pas encore rencontrés, et peut-être des paradigmes que nous n'avons même pas encore imaginés. À mesure que nous avançons, la question pourrait évoluer de « pouvons-nous créer l'IAG ? » à « devons-nous la créer ? » – et comment faire en sorte que ce que nous créons reste en accord avec nos valeurs, notre éthique et notre humanité même.

## **Quelles sont les principaux défis que la recherche doit encore adresser pour faire progresser l'IA ?**

Aujourd'hui, le modèle dominant basé sur les transformers massifs (LLMs type GPT, Gemini, Claude) montre des limites fondamentales :

- Coût computationnel extrême (entraînement, inférence).
- Inefficacité à long terme : les LLM peinent à généraliser sur des tâches complexes sans surdimensionnement (paramètres, données).
- Mémorisation inefficace : ils n'ont pas de mémoire dynamique structurée, donc doivent « réapprendre » chaque fois, ce qui est coûteux et imprécis.
- Faible interprétabilité : leur fonctionnement interne est très difficile à analyser, ce qui nuit à la fiabilité en contexte critique.

Beaucoup de chercheurs appellent donc aujourd'hui à repenser l'architecture fondamentale de l'IA au lieu de simplement augmenter la taille des modèles.

Ainsi, MAMBA ([Mamba: Linear-Time Sequence Modeling with Selective State Spaces](#), 2024) est une nouvelle architecture très prometteuse, conçue pour modéliser des séquences longues beaucoup plus efficacement que les transformers. Parmi les caractéristiques clés de MAMBA, on peut retenir :

- Abandon de l'auto-attention : Mamba utilise des *state space models* (SSMs) améliorés pour traiter les séquences

en temps linéaire, sans dépendre d'une attention quadratique.

- Mémoire structurée : MAMBA simule un état interne qui évolue au fil du temps, beaucoup plus proche des mécanismes biologiques (cerveau humain, systèmes dynamiques).
- Applications : traitement du texte long, audio, biologie computationnelle, séries temporelles complexes.
- Avantages : efficacité mémoire exceptionnelle, calcul parallèle possible, meilleure généralisation sur séquences longues.

En résumé, MAMBA ouvre la voie à des IA plus légères, capables de traiter des données continues (audio, texte, vidéo) avec moins de paramètres et plus de stabilité.

D'une manière générale et en complément de cette approche novatrice, les *Small Language Models* (SLMs) constituent une réponse au gigantisme des LLMs. En voici les principales caractéristiques :

- Taille : de 100 millions à 1 milliard de paramètres (contre 100B+ pour les LLMs).
- Optimisation : meilleure utilisation des données, architectures légères (*quantization*, *sparsity*), *fine-tuning* spécifique.
- Performance : pour des tâches ciblées (par exemple service client, traduction spécifique, recherche

documentaire), les SLMs atteignent des résultats proches de LLMs, à coût énergétique très réduit.

- Applications : edge computing (environnements mobiles, et Internet des Objets), cloud localisé, usages industriels ou souverains (modèles internes d'entreprises, Etats).

Exemples de SLMs récents :

- Phi-4-Mini [Empowering innovation: The next generation of the Phi family](#) : modèle de 3,8 milliards de paramètres, conçu pour la vitesse et l'efficacité
- TinyLlama : LLM allégé open source.
- DistilGPT : compression de GPT pour l'inférence rapide.

On peut donc penser que le futur de l'IA n'est probablement pas dans l'inflation continue de la taille des modèles. Il est plus probablement dans l'optimisation :

- architectures plus biologiquement plausibles (MAMBA, RWKV, State Space Models) ;
- modèles plus petits, plus spécialisés (SLMs) ;
- IA frugale capable de traiter des séquences longues efficacement et durablement.

On pourrait donc imaginer une sorte de progression prévisionnelle (2025-2030) de l'évolution potentielle des grands modèles de langage dans un futur proche (encore une fois, sous toute réserve et avec beaucoup de modestie) :

## 2025 :

- Consolidation des SLM (Small Language Models) ;
- Multiplication des IA spécialisées, localisées, optimisées pour des usages industriels et personnels ;
- Premiers déploiements commerciaux de MAMBA et d'autres State Space Models. 2026 :
- Maturité des modèles séquentiels efficaces ;
- Adoption dans la biologie computationnelle, les systèmes d'énergie, la surveillance environnementale ;
- Montée des IA embarquées sur mobiles et terminaux légers.

## 2027 :

- Hybridation Transformer + MAMBA ;
- Création d'architectures hybrides : maintien de l'attention locale + modélisation détendue via SSM ;
- Progrès dans la gestion de très longs contextes (> 1 million de tokens). 2028 :
- Accélération de la neuromorphisation ;
- Premiers systèmes d'IA exploitant massivement les puces neuromorphiques pour IA frugale ;
- L'edge AI (calcul local sans cloud) devient majoritaire dans

plusieurs secteurs industriels.

2029 :

- Déploiement de mémoires dynamiques adaptatives ;
- IA capables de mémoriser, oublier et généraliser sur des bases adaptatives ;
- Montée des assistants personnels « cognitifs » intégrés à la vie quotidienne. 2030 :
- Vers une IA « organique » ;
- Modèles inspirés de processus biologiques (métabolisme computationnel, évolution adaptative) ;
- Fusion IA multimodale + adaptative + sobre avec la mise en place d'un nouveau paradigme de systèmes intelligents distribués.

Il est à noter que Yann Lecun le directeur scientifique de l'IA de Meta et prix Turing a vision complètement différente du futur de l'IA. Ainsi, il affirme ne plus s'intéresser aux grands modèles de langage. Il estime que ces derniers s'amélioreront simplement de façon incrémentale, en utilisant toujours davantage de données et de puissance de calcul. Il s'intéresse plutôt à relever les quatre défis suivants qui lui semblent plus importants :

- Aider les machines à comprendre le monde physique ;
- Donner aux machines une mémoire durable ;

- Enseigner aux machines à raisonner ;
- Enseigner aux machines à planifier.

Il pense en effet que les LLM, qui prédisent des mots ou des « tokens », ne sont pas la meilleure façon de comprendre le monde réel, car les tokens sont limités et ne peuvent pas tout capturer, comme les détails des vidéos ou des images.

Lecun a ainsi introduit des « modèles du monde », différents des LLM. Ces modèles, construits à l'aide d'architectures prédictives d'intégration conjointe (JEPA - *Joint Embedding Predictive Architectures*), apprennent des concepts abstraits sur le monde (comme des images ou des vidéos) et effectuent des prédictions dans cet espace abstrait, et non à partir de données brutes.

Lecun a ainsi prédit que, d'ici 3 à 5 ans, les modèles JEPA progresseront à petite échelle, puis se développeront pour atteindre une intelligence artificielle avancée qui impliquera des machines capables de raisonner et de planifier à l'aide de modèles abstraits du monde, et pas seulement de mots. C'est ainsi que les humains et les animaux pensent, sans recourir au langage. La vision de Yann Lecun est, en effet, celle d'une IA apprenant à la manière d'un bébé ou un d'animal, en observant le monde.

Selon lui, c'est ainsi que l'IA doit fonctionner, de cette manière abstraite, pour véritablement comprendre le monde. Pour plus de précisions sur sa vision, il peut être utile de consulter le billet de blog : « [V-JEPA: The next step toward Yann LeCun's vision of](#)

*advanced machine intelligence (AMI)* »

Cette controverse démontre, s'il en était besoin, que le futur de l'IA n'est pas encore écrit, loin d'en faut...

## CHAPITRE 6

# EN GUISE DE CONCLUSION

Le développement de l'intelligence artificielle connaît une accélération remarquable depuis la fin des années 2010, portée par les progrès du deep learning, la disponibilité massive de données, et les capacités de calcul croissantes. À l'horizon de cinq à dix ans, il est attendu que l'IA s'intègre de manière toujours plus systémique aux infrastructures économiques, sociales, culturelles et scientifiques. Cette évolution soulève à la fois des promesses et des défis techniques, politiques, environnementaux et éthiques majeurs.

L'IA, particulièrement sous sa forme générative, transformera radicalement le travail intellectuel avec l'automatisation des tâches cognitives routinières (rédaction, codage de base, création de contenus standardisés), le déplacement des compétences vers la supervision, l'orchestration de systèmes d'IA, et la capacité à interpréter et contextualiser les productions automatiques. Certains secteurs (médias, services juridiques, finance) connaîtront ainsi une reconfiguration profonde de leurs modèles économiques.

Le coût du développement des IA de pointe (pré-entraînement, infrastructure, *fine-tuning*) entraînera un resserrement du nombre

d'acteurs capables de développer des modèles souverains. Ceci pose clairement le problème de la dépendance vis-à-vis des grands fournisseurs d'IA (Amazon AWS, Microsoft Azure, Google Cloud, etc.). La souveraineté numérique - notamment depuis l'évènement de Donald Trump - est déjà devenue une préoccupation géostratégique majeure un peu partout dans le monde, alimentant des investissements publics massifs.

Parmi les défis critiques qu'il convient de relever dans les années à venir :

### 1. Soutenabilité environnementale

Les systèmes d'IA consomment des ressources énergétiques et naturelles considérables (cf. les travaux de [Kate Crawford](#) dont nous venons de parler). À moyen terme, la tension entre innovation IA et transition écologique s'intensifiera. La recherche sur :

- Les modèles frugaux,
- L'IA distribuée,
- Les matériels spécialisés (ASICs, *neuromorphic computing*), sera décisive pour réduire l'empreinte environnementale.

### 2. Régulation éthique et gouvernance

Avec la généralisation de l'IA dans les domaines sensibles,

plusieurs enjeux deviendront centraux :

- Transparence algorithmique pour accéder à des explications claires sur les décisions automatisées.
- Justice algorithmique afin de prévenir les biais discriminatoires dans les modèles.
- Responsabilité légale pour définir la chaîne de responsabilité entre développeurs, intégrateurs et utilisateurs.

Les débats autour des droits des données, de la protection de la vie privée, et des droits d'auteur face à l'IA générative sont déjà devenus des enjeux juridiques majeurs.

### 3. Risques systémiques

À mesure que l'IA sera impliquée dans les infrastructures critiques (énergie, santé, transport), les risques d'interdépendance et de défaillance systémique augmenteront. Des stratégies de « redondance intelligente » et d'audit en temps réel des modèles seront nécessaires pour limiter les conséquences d'erreurs ou de dérives.

À court et moyen terme, l'IA ne se présente pas seulement comme un outil d'optimisation ; elle redessine les contours des sociétés humaines, du travail, de l'économie et de la culture. Ce futur sera marqué par une tension fondamentale entre accélération technologique et nécessité de soutenabilité et de

justice.

La gouvernance démocratique des systèmes d'IA, la réinvention de modèles économiques inclusifs, et l'adaptation des cadres légaux et culturels seront les conditions nécessaires pour que l'IA ne soit pas seulement un amplificateur d'inégalités, mais un levier d'innovation éthique et durable.

À l'ère d'une accélération technologique sans précédent, l'interaction entre l'humain et l'intelligence artificielle révèle des tensions profondes entre progrès technique et autonomie individuelle. Loin de se limiter à une simple amélioration fonctionnelle de nos sociétés, l'IA tend à instaurer une nouvelle forme de dépendance cognitive et décisionnelle. Le phénomène de sursocialisation technologique, où les comportements humains se moulent aux logiques algorithmiques, interroge radicalement notre rapport à la liberté, à la pensée critique et à l'expérience concrète du réel.

La facilité promise par l'automatisation masque une dépossession progressive du pouvoir de décision, au profit d'architectures techniques opaques. Ce glissement, loin d'être anodin, s'inscrit dans une dynamique historique où les idéaux progressistes sont récupérés par les logiques de marché et de contrôle systémique. Le progressisme technologique contemporain ne semble plus viser l'émancipation individuelle, mais plutôt l'intégration fluide de l'humain dans un environnement numérique automatisé.

Ainsi, face à l'illusion d'un progrès ininterrompu, il devient essentiel de repenser notre rapport aux technologies émergentes

afin de ne pas se laisser bercer par un certain « solutionnisme technologique ». Préserver l'autonomie humaine nécessite d'établir des limites claires aux usages de l'IA, de réhabiliter la capacité critique, et de redonner à l'expérience directe et à l'engagement actif leur rôle central. Le véritable progrès ne réside pas dans la soumission volontaire à l'IA, mais dans la réaffirmation de l'humain comme sujet libre et créateur face aux dispositifs techniques. L'IA est la dernière d'une longue série de technologies qui soulèvent des questions quant à leur impact sur la qualité de la pensée humaine, une lignée qui inclut l'écriture (à laquelle Socrate s'opposait), l'imprimerie (à laquelle Trithème s'opposait), les calculatrices (à laquelle les professeurs d'arithmétique s'opposaient) et Internet. Ceci nous renvoie au rôle capital de l'éducation car, comme le disait fort justement Confucius il y a 2500 ans : « Apprendre sans penser est inutile mais penser sans apprendre est dangereux. » Le rôle des enseignants, des éducateurs, des pédagogues, des parents n'a jamais été plus critique : apprendre à prendre de la distance, à remettre en cause des réponses toutes faites, apprendre à apprendre sont des missions essentielles que chacun doit s'approprier. « Chose vraiment surprenante (...) c'est de voir des millions de millions d'hommes, misérablement asservis, et soumis tête baissée, à un joug déplorable, non qu'ils soient contraints par une force majeure, mais parce qu'ils sont fascinés et, pour ainsi dire, ensorcelés par le seul nom d'un, qu'ils ne devraient redouter, puisqu'il est seul, ni chérir, puisqu'il est, envers eux tous, inhumain et cruel » nous disait Etienne de la Boétie dans son « Discours de

la servitude volontaire ». Cette soumission volontaire à l'intelligence artificielle n'est pas une fatalité et c'est à chacun d'entre nous qu'il convient de prendre la distance souhaitable.

Nous devons donc être conscients des risques, mais aussi des potentialités si riches de cette technologie. En tant que société, il nous revient de décider comment nous voulons gérer cette évolution et nous devons le faire de manière réfléchie et responsable. En construisant des systèmes d'IA alignés avec nos valeurs humaines et en anticipant les défis éthiques et sociétaux, nous pourrions peut-être créer un futur où l'IA et l'humanité coexistent harmonieusement, plutôt que de se livrer à une lutte pour la domination.

L'histoire de l'IA (et la nôtre) ne se fera pas sans nous. Elle est faite de discontinuités, et de crises. Aucun spécialiste sérieux ne sait vraiment de quoi l'avenir de l'IA sera fait car il reste tant de chose à faire par nous, les humains. Résister au mythe du progrès linéaire, ce n'est pas fermer les yeux sur notre pouvoir d'action et notre responsabilité. Même si cela implique d'admettre, pour une fois, que l'on ne sait pas ; le fameux « Je sais que je ne sais rien » de Socrate. Car quand il s'agit de déterminer la direction dans laquelle on veut collectivement aller, pour reprendre le propos d'Edgar Morin dans « Comment vivre en temps de crise » : « c'est dans la mesure où il y a incertitude qu'il y a [...] possibilité d'action, de décision, de changement, de transformation »

# Institut pour l'éducation augmentée

# QUESTIONS ET

# RÉPONSES SUR L'IA.

Entretien avec **BERNARD OURGHANLIAN**  
« Enjeux scientifiques, sociétaux et éthiques »



**Directeur de publication :** Julien Cartier

**Auteurs :** Bernard Ourghanlian et Julien Cartier

*Cette publication est la propriété exclusive de l'Institut pour l'Education Augmentée, association loi 1901, Siret :*

*93364722400015 ; toute reproduction, même partielle, qui n'aurait pas été autorisée par l'association ou les auteurs de ce document est réputée interdite.*