Sciences Po Bordeaux

Livrable de réflexion à destination de la Convention citoyenne sur l'intelligence artificielle

Florilège de comptes-rendus des étudiants et étudiantes ayant suivi le cours « Convention citoyenne étudiante IA » : Berrocal Justin, Bonnemaison Joan, Bruel Tess, Charavel Noëlie , Delmas Saint-Hilaire Julia, Fistache Jules, Gelpi Valentine, Lambin Vincent, Levaillant Gwenn, Martin Ys, Nani Mila, Rémond Nathan.

Compilé par Amine BOUSTATI

Sous la direction de Mme Mazarine Pingeot

Mars 2025

SOMMAIRE

Table des matières

REMERCIEMENTS

Merci à l'ensemble des intervenants, dont les contributions ont été essentielles à la présente rédaction, ainsi qu'à Madame Mazarine Pingeot pour avoir permis la rencontre entre ces acteurs et les étudiants.

INTRODUCTION (Mazarine Pingeot et Laurent Simon)

L'intelligence artificielle (IA) connaît un essor considérable dans nos sociétés. Elle s'intègre progressivement dans de nombreux domaines, qu'il s'agisse du secteur médical, de la publicité ou encore de notre quotidien.

Dans ce contexte, la convention citoyenne sur l'intelligence artificielle, qui se tiendra du 4 au 6 avril 2025, a pour objectif de démocratiser cette technologie et la comprendre.

Le présent livrable vise à synthétiser les principaux enseignements du cours d'ouverture éponyme afin de clarifier et d'améliorer la compréhension des enjeux liés à l'IA. L'objectif final est d'enrichir la qualité du dialogue et des échanges lors de la convention. Pour ce faire, il s'appuie sur les interventions d'experts issus de disciplines variées, allant des sciences dites « dures » aux sciences humaines, qui ont partagé leurs connaissances et compétences sur le sujet aux cours des séances de cours.

Dans un souci d'honnêteté intellectuelle, ce document comprend une introduction exposant les enjeux et l'intérêt d'une convention citoyenne sur l'IA. Il présente ensuite de manière brute les comptes rendus des interventions des experts, sans chercher à influencer ni orienter les réflexions et propositions qui émergeront au cours de la convention. Il se veut ainsi un **outil d'aide à la décision** neutre et factuel (du moins autant que possible…)

Les bouleversements de l'IA

L'intelligence artificielle (IA) transforme profondément notre quotidien et ce, sous de nombreuses formes : reconnaissance de la parole, traduction automatique, création de textes grâce aux IA génératives, interactions sous forme de discussions, recommandations publicitaires, jeux de société ou encore jeux vidéo (comme les échecs). Elle est également utilisée dans des domaines plus techniques, tels que la preuve automatique en mathématiques ou la conduite autonome. Il est intéressant de noter que l'adoption massive de l'IA est particulièrement visible chez les étudiants. En effet, une étude réalisée en mars 2024 révèle que 99 % d'entre eux utilisent l'IA, et 30 % seraient même prêts à payer un abonnement pour accéder à des outils comme ChatGPT. Cette démocratisation de l'IA suscite alors à la fois de grandes promesses et des craintes. En 1965, Herbert Simon prédisait que « d'ici 20 ans, tout ce que les humains peuvent faire pourra être fait par l'IA». Aujourd'hui, cette perspective semble plus réaliste que jamais, notamment avec les investissements colossaux dans ce domaine : en 2025, 500 milliards d'euros vont être consacrés au développement de l'IA avec le lancement du projet Stargate.

HISTORIQUE DE L'IA.

L'IA trouve ses origines dès 1950, lorsque Alan Turing évoque l'« intelligence des machines » en réfléchissant à leur capacité à jouer aux échecs. En 1956, le terme « intelligence artificielle » est alors officiellement adopté, et c'est en 1965 qu'apparaissent les premiers systèmes experts capables de réaliser des déductions, ainsi qu'Eliza : un programme de conversation destiné à simuler un psychothérapeute. Toutefois, les recherches connaissent un ralentissement avec ce que l'on appelle « l'hiver de l'IA », période durant laquelle les gouvernements cessent de financer les projets en raison de résultats insuffisants. Le tournant arrive en 1997, lorsque le champion d'échecs Garry

Kasparov est battu par Deep Blue, un programme développé par IBM. Puis, en 2010, l'apprentissage profond révolutionne le domaine. Ces avancées ont été rendues possibles grâce à l'amélioration des processeurs, des capacités de stockage, des réseaux permettant la circulation rapide des données et des capteurs capables de numériser des informations variées (textes, sons, images, rythme cardiaque, etc.). Ces progrès constants conduisent à des phénomènes de seuil, où l'IA atteint un niveau de performance soudainement supérieur à celui des humains.

DÉFINITIONS/DISTINCTIONS

L'intelligence artificielle peut être définie comme la capacité d'une machine à exécuter des tâches que réaliserait un humain moyen. Cependant, il existe une distinction entre l'IA dite « restreinte », qui se limite à une seule tâche précise (comme la reconnaissance de la parole ou la production de texte et d'images), et l'IA générale, qui simulerait une intelligence comparable à celle des humains. Si l'IA restreinte semble aujourd'hui extrêmement performante, et dépasse parfois les capacités humaines dans certains domaines, la question de la superintelligence se pose : jusqu'où peut-elle aller ? L'IA peut-elle réellement surpasser les humains dans toutes les tâches ? Certains chercheurs, comme Nick Bostrom, envisagent l'émergence d'une IA tellement avancée qu'elle pourrait devenir incontrôlable. Dans cette perspective, OpenAI ne se contente pas de créer des outils technologiques, mais cherche à produire de la valeur économique et sociétale.

DANGERS ET SOCIÉTÉ.

L'essor de l'intelligence artificielle soulève ainsi des interrogations majeures. Souhaite-t-on réellement voir émerger une intelligence « supérieure » à l'humanité ? Laisserons-nous cette puissance se concentrer entre les mains de quelques acteurs ? Peut-on espérer que l'IA résolve des problèmes complexes comme le changement climatique ? Enfin, une IA qui impose une seule vision du monde est-elle souhaitable, ou devrait-elle au contraire être conçue pour favoriser le débat et la diversité des opinions ? Ces questions sont particulièrement illustrées par « The Moral Machine Experiment », une expérience qui explore les dilemmes éthiques de l'IA. Elle met en évidence le fait que, lorsqu'une voiture autonome doit prendre une décision face à un accident inévitable, elle préfère en général prendre une action plutôt que de ne rien faire. Mais cette décision dépend aussi des personnes impliquées et des valeurs programmées dans l'algorithme.

LES DEUX GRANDES APPROCHES DE L'IA.

L'intelligence artificielle peut être catégorisée selon deux types de pensée : • Une pensée intuitive et rapide, qui permet de percevoir l'environnement, de détecter des corrélations et de réaliser des tâches automatiques comme conduire sans réfléchir. • Une pensée raisonnée et lente, qui permet d'analyser en profondeur, de comprendre des relations de causalité et de prendre des décisions complexes. Ce qui différencie ces deux formes d'IA est leur capacité d'introspection. Un raisonnement humain repose sur une réflexion interne, tandis que les algorithmes se contentent d'appliquer des modèles statistiques sans véritable compréhension.

APPRENTISSAGE ET PROGRAMMATION DE L'IA.

Pour que l'IA fonctionne, elle doit être programmée à l'aide d'algorithmes complexes. Il existe plusieurs approches : 1. Créer un algorithme spécifique avec un grand nombre de constantes

définies à l'avance 2. Utiliser des bases de données d'exemples pour ajuster ces constantes 3. Réduire l'écart entre les résultats obtenus et les réponses idéales L'objectif est d'améliorer continuellement la précision des modèles, en affinant leur capacité à s'approcher de la vérité.

IA GÉNÉRATIVE.

L'IA générative repose sur un principe simple : elle prend un bruit aléatoire et applique une description pour générer une image ou un texte final. Ce processus repose sur des algorithmes d'apprentissage profond, qui s'appuient sur des milliards de données pour produire du contenu de plus en plus réaliste. En somme, l'intelligence artificielle est en train de révolutionner notre rapport à la technologie et à la connaissance. Son développement rapide ouvre des perspectives fascinantes mais pose également des défis éthiques et sociétaux majeurs. La question n'est plus de savoir si l'IA va transformer le monde, mais comment nous voulons encadrer son évolution pour qu'elle soit au service de l'Homme.

Dès lors l'IA est un enjeu éthique et politique majeur, mais pourquoi utiliser le biais d'une convention citoyenne pour en parler ?

Sur le choix d'une convention citoyenne pour parler d'intelligence artificielle

Cette convention est un événement de démocratie, voire de démocratie délibérative, puisque l'on donne la parole et l'opinion aux étudiants sur ce changement majeur dans notre société.

Pour développer cette idée sur la démocratie délibérative, il convient de se pencher sur un texte de **Rosanvallon**, *Les institutions invisibles*. Pour l'auteur, le processus délibératif a pour objectif de recréer une confiance collective dans les sociétés, en renouant avec l'idéal de Rousseau, c'est-à-dire de fabriquer une volonté générale à partir de la pluralité d'opinions. De plus, il critique le principe majoritaire qui écrase les minorités et appelle à une démocratie plus inclusive.

Par le renouveau délibératif, les citoyens ont l'occasion de sortir du seul rôle électif que la démocratie traditionnelle leur confère pour prendre part à la création de l'opinion. Ils passent d'une position passive, à un rôle plus actif dans l'élaboration des idées.

→ Dans cette optique, la convention sur l'IA pourrait permettre de dégager une définition commune et une volonté générale à partir d'une pluralité d'opinions tout en permettant que les points de vue des minorités soient pris en compte dans l'élaboration de l'IA. Par exemple, que les minorités soient prises en compte dans les élaborations d'algorithmes des IA.

Ensuite, **Cathy O'Neil**, dans *Nouvelle enquête sur l'intelligence artificielle*, développe l'idée qu'il existe une opacité autour des algorithmes. Ces derniers sont souvent présentés comme incontestables, et qu'un individu lambda ne pourrait pas comprendre, même les grandes lignes, ce qui empêche la mise en place de débat public autour de ces algorithmes, pourtant centraux dans nos quotidiens.

→ Ainsi, dans le cadre de la Convention, il pourrait être intéressant de réfléchir autour de l'instauration de normes éthiques dans le développement et l'application des algorithmes. De plus, il convient donc aussi de se pencher sur des critères plus clairs concernant la transparence que l'on souhaite dans ces outils.

Michel Serres, dans *Le contrat naturel* explique qu'il existe une rupture entre les sciences dites dures et les sciences humaines. Dès lors, les liens sont rompus entre les innovations techniques et la gouvernance démocratique. En effet, il existe une médiatisation et une idéologisation excessive de la société qui s'oppose à l'impact concret des innovations technologiques.

→ Un des objectifs de la convention serait alors de pallier ce fossé entre les sphères scientifiques et la société.

Alexandre Koyé dans *Du monde clos à l'univers infini*, analyse le passage du médiéval, où l'Église dominait les systèmes de pensée à un univers régi par des lois universelles. Cela conduit à un divorce entre les faits scientifiques (comme l'héliocentrisme et l'infinité de l'univers) aux valeurs (comme les religions). Avec la révolution copernicienne s'instaure un divorce entre faits et valeurs : ce en quoi l'on croit n'est pas forcément vrai, à l'opposé des règles de la société pré-moderne (règne de la fatalité/de l'impuissance par rapport aux évènements => règne de l'incertitude, du questionnement scientifique pour établir la vérité). Ainsi, au-delà d'un apport purement scientifique, le questionnement scientifique vient surtout transformer la perception que les hommes avaient du monde : ce n'est plus Dieu qui façonne la réalité dans laquelle ils évoluent sans pouvoir la changer ou en percer les mystères, mais bien l'Homme lui-même qui grâce à son savoir, peut se permettre de la modifier.

→ L'IA pourrait également engendrer cette crise de conscience si on déconnecte trop les résultats des algorithmes des valeurs humaines. Il semble donc approprié de discuter et d'inclure les valeurs éthiques dans les applications de l'IA.

Bruno Latour, dans *Politiques de la nature*, critique la séparation entre faits et valeurs. Ces notions se distinguent tout autant que les sciences et la politique. Malgré le fait qu'elles n'appartiennent pas aux mêmes champs (la science se rapporte au "ciel des idées", à la vérité incontestable ; tandis que la politique se fonde sur l'opinion, forcément subjective) la politique, pour clore les débats de subjectivités, doit nécessairement avoir recours à l'arbitre scientifique. Pour lui, les faits doivent être définis collectivement, via une "constitution" où les faits sont discutés avec les valeurs pour créer une définition plus inclusive.

→ Dans cette optique, l'IA en tant qu'outil scientifique et social doit être définie et encadrée par un processus démocratique pour éviter les biais dans ses usages et ses programmes.

David Hume, dans *Traité de la nature Humaine,* développe également cette distinction entre les faits (ce qui "est") et les valeurs (ce qui doit être). Il soutient qu'un fait ne peut pas amener à une proposition normative et donc à la construction d'une valeur. Ainsi, les faits scientifiques ne peuvent pas dicter des règles morales sans que la société ait pu collectivement réfléchir à ce fait.

→ Dès lors, la Convention sur l'IA pourrait permettre de définir ou du moins, de réfléchir aux faits produits par les algorithmes, et comment ces faits devraient être interprétés et traduits en décisions normatives.

Max Weber, dans *Le Savant et le politique*, reprend également cette distinction faits/valeurs. Pour lui, les faits peuvent être scientifiquement démontrés tandis que les valeurs relèvent du choix individuel et sont irrationnelles. Selon lui, les scientifiques ne doivent pas aller sur le terrain des décideurs, ils doivent s'en tenir à une neutralité axiologique pour produire un savoir le plus objectif possible. Néanmoins, rien que le choix d'approfondir tel ou tel sujet, relève d'une décision subjective. Surtout, la voie de la science, par la constatation des faits peut permettre d'atteindre la vérité du monde là où les valeurs, par oppositions resteront toujours des choix et soumis à l'arbitraire.

→ Ainsi, malgré le fait que les algorithmes paraissent neutres, les choix des données et leur manière de fonctionner relèvent de décisions qui ne sont pas neutres, et il convient de le garder dans l'esprit. Étant donné que l'objectivité absolue n'existe pas, il serait opportun d'analyser ces choix de valeurs au sein des algorithmes.

Gaston Bachelard, dans *La formation de l'esprit scientifique*, soutient que la science s'oppose à l'opinion. En effet, la science rejette l'opinion car celui-ci serait un obstacle à une pensée rationnelle. Néanmoins, l'esprit critique demande de dépasser les biais, les préjugés et les connaissances empiriques qui existent pour atteindre une compréhension objective.

→ L'IA est par exemple souvent empreinte de biais sociaux et culturels, comme par exemple au début de la reconnaissance faciale, les erreurs de reconnaissance envers les personnes de couleurs étaient plus importantes que celles des personnes blanches. Cela était dû, aux programmes réalisés par des personnes blanches qui avaient majoritairement et inconsciemment entraînées l'IA à reconnaître des personnes blanches plutôt que de couleur. La marge d'erreur augmente encore plus si l'on est une femme noire, pouvant atteindre 35%[1].

---

[1] Sciences et Avenir. (2018, janvier 29). *Intelligence artificielle : La reconnaissance faciale est-elle misogyne et raciste ?* Sciences et Avenir. https://www.sciencesetavenir.fr/high-tech/intelligence-artificielle/intelligence-artificielle-la-reconnaissance-faciale-est-elle-misogyne-et-raciste_121801

INTELLIGENCE ARTIFICIELLE ET ETHIQUE (Alexeï Grinbaum)

Alexeï Grinbaum est directeur de recherche au CEA-Saclay, où il se consacre principalement à la théorie de l'information quantique. En parallèle, il préside le Comité opérationnel d'éthique du numérique du CEA et est membre du Comité national pilote d'éthique du numérique (CNPEN).

Son expertise est également sollicitée par la Commission européenne en tant qu'expert. Depuis 2003, ses recherches s'orientent vers les implications éthiques des technologies émergentes, telles que les nanotechnologies, l'intelligence artificielle et la robotique. Il analyse notamment la gouvernance de l'incertitude et les fondements du principe de précaution.

Plus récemment, il s'est penché sur la responsabilité des chercheurs face à ces innovations. Parmi ses publications, on compte « Mécanique des étreintes » (Encre Marine, 2014), « Les robots et le mal » (Desclée de Brouwer, 2019) et « Parole de machines ». Ces ouvrages explorent les interactions entre l'humain et la machine, ainsi que les défis éthiques posés par l'essor des technologies numériques. La séance : Intelligence Artificielle (IA) et éthique Hannah Arendt – « Tout ce que les hommes font, ou savent, ou ce dont ils ont l'expérience, à un sens dans la mesure où cela peut être raconté. »

De quelles technologies parle-t-on ?

En 1965, Joseph Weizenbaum crée Eliza, un des tous premiers chatbot : la machine est alors capable de simuler un psychothérapeute rogérien en reformulant la plupart des affirmations du « patient » en question, et en les lui posant. On projette dès lors, des qualités anthropomorphiques à cette technologie, à commencer par sa dénomination. C'est d'ailleurs le principe du test de Turing : faire deviner à l'Humain s'il a affaire à une machine ou à un Homme. Si la machine est prise pour un 1 Nathan Rémond et Valentine Gelpi - Convention étudiante sur l'IA individu humain, elle a passé le test avec succès. Cela n'est pas anodin et pose les premières questions éthiques. Tout d'abord, il existe deux grands types d'IA. En premier lieu, l'IA symbolique, vient simuler un raisonnement humain. Elle repose sur l'utilisation de symboles et de règles de traitement de l'information écrites par des humains (concepteurs, psychologues, ingénieurs) pour accomplir diverses tâches. Ces symboles peuvent représenter des concepts, des objets ou des relations, tandis que les règles incluent des mécanismes de déduction, de production ou d'inférence, entre autres. C'est par exemple le cas de l'assistant « Siri » développé par Apple. En second lieu, nous pouvons citer l'IA connexionniste, et sa capacité d'apprendre diverses représentations des données, et à les mettre en œuvre selon le principe du deep learning. L'IA connexionniste modélise les processus mentaux à travers des réseaux de neurones artificiels, contrairement à l'IA symbolique. Elle repose sur des modèles structurés en réseaux composés de plusieurs couches d'unités de traitement de l'information, telle une petite calculatrice élémentaire, à entrées multiples. À l'image du cerveau humain, ces réseaux fonctionnent par propagation d'activation entre les unités, qui s'activent dès qu'un certain seuil est dépassé1 . Au niveau des avancées marquantes, on peut citer la révolution Transformers en 2017, ses logiciels de complétion automatique de phrases, et ses deux types d'apprentissage.

L'apprentissage asémantique consiste en la division d'un texte en Tokens (divisé en séquences, caractères) sans recherche de véritable sens humain dans la découpe des mots. On parle aussi d'apprentissage par auto-supervision, n'intégrant pas d'êtres humains dans le processus. L'IA est en effet constituée d'un réseau de neurones gigantesque, composé de milliards, ou de centaines de milliards de paramètres. Pour délivrer des réponses pertinentes, elle soustrait des mots dans des corpus de texte (« cache-cache ») à l'aide de calculs de probabilité en pourcentage, afin de trouver une corrélation dans les propos. L'IA est également soumise à un alignement des modèles par des ingénieurs, c'est-dire à un certain nombre de filtres et contrôles, pour éviter le langage toxique (exemple : domaines protégés par les droits d'auteurs, indication de prise de médicaments…). Il faut toutefois noter que la sélection des données sans biais et sans discrimination constitue un prochain chantier d'envergure au niveau de l'éthique. Plus récemment, on voit émerger les modèles de raisonnement, plus aptes à résoudre des problèmes complexes (juridique, mathématique) avec leurs prompts, et laissant moins de place à l'approximation ou à la divagation

Mêler IA et éthique : une lourde tâche

Suivant la pensée de Thomas Hobbes, Leibniz affirmait que « Toute chose faite par notre esprit était un calcul ». Cela reviendrait donc à dire que « Penser, c'est calculer », ou désignerait « ce que nous faisons à des choses que nous additionnons et soustrayons ». En grec, logos signifie à la fois « raisonner » et « calculer », tout comme en hébreu avec le mot heshbon. Il y a donc un lien sémantique entre la pensée et le calcul.

Exemples de dilemmes éthiques : → Le Canadien qui faisait parler les morts (2021) Un informaticien, à la demande d'un veuf, a entrainé GPT3 à écrire à la manière de sa femme décédée en lui transférant l'ensemble de ses lettres et messages, amenant l'IA à se mettre à parler comme elle aurait pu le faire. L'homme est alors conscient du processus, mais on observe l'opération d'une transformation psychologique et émotionnelle à force d'échanger, au point que la discussion devient thérapeutique et salutaire. Cet exemple nous pousse à nous questionner, alors que cette technologie modifie l'essence même de la mort. La question occupe les comités d'éthique : quelle limite mettre en place ? Faut-il nécessairement interdire de telles pratiques ? Il y a un débat nécessaire à mener sur les contraintes à instaurer. Par exemple, certains partisans du « faire parler les morts » à travers une IA, soumettent la possibilité de le permettre à condition d'une proximité avec un cimetière, vérifiable par géolocalisation. → Ex Machina : fiction ou proche réalité ? (2014) Ce film d'Alex Garland raconte l'histoire d'un brillant codeur travaillant pour le compte d'un moteur de recherche Internet mondialement utilisé, invité à participer à une expérience troublante : interagir avec le représentant d'une nouvelle intelligence artificielle placée dans un robot aux allures de jolie jeune femme, Ava. Celle-ci est conçue pour remplir un objectif : trouver un moyen stratégique de sortir du bâtiment où elle se trouve. Le film montre que le robot élabore des stratégies pour y parvenir, allant inventer une romance avec le codeur pour mieux le manipuler, à tuer froidement. En effet, si la machine a été programmée pour une mission précise, aucun système de valeur n'y a été associé pour encadrer l'usage de la vérité ou du mensonge : le schéma utilisé pour parvenir au but fixé n'est pas toujours éthiquement neutre. Cet exemple fictionnel trouve aujourd'hui un écho dans l'actualité. En 2023, un adolescent qui discutait avec le chatbot GPT 4 a été poussé au suicide. Ces évènements interrogent sur la responsabilité et les limites à mettre en place. Ainsi, pour prévenir toute dérive et s'assurer un plein contrôle sur la technologie, les ingénieurs ont mis au point les tests adversaires. Ceux-ci durent plusieurs mois (d'où la latence entre les différents modèles d'IA qui alternent), et permettent de savoir de quoi un modèle IA est

devenu capable pendant son apprentissage en discutant avec.

La question de la responsabilité

Les apprentissages de l'IA peuvent être fascinants mais aussi inquiétants (mensonge, manipulation) : la crainte qu'elles inspirent est intrinsèque au fait qu'il n'existe pas réellement de responsable. Pour Heidegger (La dévastation et l'attente : entretien sur le chemin de campagne) : la parole ne se réfère pas au monde matériel, mais au langage, rendant responsable la personne des mots qu'elle emploie. Dans le cas de l'IA, la recherche des responsabilités constitue donc un enjeu énorme. Durant l'été 2024, a été acté l'IA Act Timeline (2024-2027), rendant compte des systèmes prohibés et des codes de bonnes pratiques pour IA générative. Les systèmes influençant les êtres humains de manière subliminale, c'est-à-dire au-delà de la conscience, ont été prohibés (entrera en vigueur en février 2025). Mais de nombreuses zones grises, de questions techniques, demeurent sur le plan de l'ingénierie vis-à-vis de cette loi écrite par des politiques profanes. De plus, la classification de plusieurs types de systèmes d'IA comme étant à haut risque et leur certification officielle n'existe pas et doit encore être inventée. L'EU IA Act Recital 133 Article 52 soulève pour sa part un point intéressant : celui du principe éthique du maintien de la distinction, permettant « d'exiger des fournisseurs de ces systèmes qu'ils intègrent des solutions techniques permettant de marquer dans un format lisible par machine et de détecter que le résultat a été généré ou manipulé par un système d'IA et non par un humain ». Il faut aussi questionner la possibilité pour une IA générative d'apprendre des choses que le cerveau ne peut pas faire, à l'instar du langage des animaux ou de l'expérience « Two weeks talking about food, with subtitles » effectuée avec l'IA DALL-E 2. Apprendre à un transformer à faire le lien entre un objet, une représentation et son nom, renvoi à de vieilles allégories comme celle du Paradis et des noms des animaux mythologiques inventée par Hannah Arendt qui raconte que Dieu aurait envoyé des animaux devant les anges pour leur donner des noms, puis aurait finalement chargé Adam de cette mission. Donner une écriture linguistique est différent de donner un nom : on établit un rapport, une relation à travers le langage, en faisant entrer l'autre dans notre monde humain par laquelle la condition humaine se met en mouvement.

Éthique a priori / a posteriori

Il est possible de distinguer 3 niveaux de l'éthique de l'IA

1. L'alignement des modèles

Il s'agit, actuellement, du modèle sur lequel travaillent le plus les concepteurs.

2. Le niveau sociétal et économique

Ce travail de conception modifierait le marché du travail en introduisant notamment de nouveaux métiers (designer, programmeur ne travailleraient plus de la même façon) mais cela ne constitue pas un changement rapide et brutal car le temps long est nécessaire pour former des spécialistes (par exemple, les cochers ont disparu 40 ans après l'arrivée de l'automobile). La question de la vitesse est cruciale : plus la modification est rapide, plus elle est susceptible de créer des tensions et des conflits.

3. Niveau anthropologique|philosophique

Nous nous reconnaissons toujours dans l'histoire humaine, jalonnée de moments durant lesquels la condition humaine a changé. L'écriture a bouleversé la manière dont nos cerveaux sont éduqués, l'impression a changé la manière dont ils apprennent. Dans ce sens, l'ampleur de l'IA est comparable à ce qu'ont pu vivre nos ancêtres en termes de changements sociétaux. Ces technologies sont nouvelles et inédites : elles agissent comme un moyen de relier ces situations inédites avec la continuité de la pensée humaine et sociétale. Ce niveau est amené à prendre de l'ampleur mais en a encore trop peu. Dans ce sens, l'analyse des retombées économiques (policy making) ne peut pas avoir lieu uniquement au niveau du lecteur ; il s'agit de mettre en place des modules éducatifs tout en faisant l'effort de comprendre la science, le fonctionnement réel sous peine de ne pas mettre en place suffisamment de critères sérieux pour légiférer, au niveau du langage…

INTELLIGENCE ARTIFICIELLE (IA) ET DROIT, ENTRE FAUSSES PROMESSES ET VRAIES QUESTIONS (François Pellegrini)

Présentation de l'intervenant

M. Pelligrini est informaticien et professeur à l'Université de Bordeaux. Il est également chercheur au Laboratoire bordelais de Recherche en informatique. Il coécrit en 2022 avec Elia Verdon un article paru dans la revue L'économie politique : « Les donnnées à caractère personnel, carburant du capitalisme de surveillance ». Après avoir été commissaire et vice-président de la CNIL en travaillant sur la régulation des données personnelles, il présente dans le cadre de son intervention la mission de la défense des droits numériques et des libertés individuelles : « L'intelligence artificielle » : entre fausses promesses et vraies questions.

« L'intelligence artificielle » ou le technosolutionnisme

Refus d'utiliser le terme « d'intelligence artificielle » : => Relève d'un marketing effréné pour acheter des « techno-solutions » : le technosolutionnisme => Première utilisation du mot au Dartmouth College en 1956 : imposture intellectuelle du terme qui relève du travail de traitement de la connaissance : comment les humains décident d'imiter un phénomène. => Incompatibilité du terme avec la science car l'intelligence n'est pas définissable. => McCarthy : le choix du terme est un projet politique, terme médiatique. Deux modèles « d'IA » y sont développés : IA faible (pour les tâches d'assistance) contre IA forte (intelligence synthétique généraliste). Le terme « d'intelligence artificielle » brouille le champ cognitif pour « vendre » des systèmes informatisés : c'est de l'anthropomorphisation systématique qui nous amène à confondre les concepts suivants : - Intelligence = calcul - Apprentissage = configuration - Raisonnement = calcul - Hallucination = non-productif Ce brouillage est conçu pour faire passer les machines pour des humains : la machine ne doit pas pouvoir discuter le résultat pour que l'humain ne prenne pas de décisions (par ex : le logiciel Parcoursup reflète une infériorisation systématique de l'humain).

Des promesses non tenues

Aujourd'hui, beaucoup de promesses ne sont pas tenues dans le monde du numérique : => Les véhicules « autonomes » : personne ne conduit le véhicule mais besoin systématique de personnes derrière le logiciel pour assurer son fonctionnement. Fonctionnement efficace uniquement dans les zones bien quadrillées. La machine ne peut pas gérer un évènement non calculé : c'est l'humain qui reprend la main. => Le mythe de la singularité : la pensée de base d'un ingénieur, c'est de se demander « quel est le problème à résoudre ? » : Les informaticiens vont se demander quel est l'intérêt de supprimer un conducteur pour le remplacer par 1,5 humain derrière.

Des coûts cachés

Au-delà des promesses non tenues, se cachent des coûts supplémentaires : Antonio Casili travail sur ces coûts cachés et parle d'un travail de catalogage et d'étiquetage : => Travail humain (caché) de production et d'étiquetage (par ex : le test Turing du passage piéton/feu tricolore : on pose une question à laquelle seul l'humain peut répondre) : les logiciels imitent et bluffent le cognitif

humain. Les données récoltées sont étiquetées par les humains eux-mêmes. => Les coûts énergétiques astronomiques : le cerveau humain fonctionne avec seulement 40W contre beaucoup plus pour les centre de données qui moulinent des millions de pages pour configurer des perroquets probabilistes. =>Des transformations sociales majeures : utilisation de chatbot à la place d'humain derrière le téléphone : l'avantage, c'est que les machines ne s'épuisent pas. Pourtant, il y a un coût psychique pour les humains de s'opposer au résultat du calcul.

Déduction/Induction : l'éthique algorithmique n'a aucun sens

Un algorithme renvoie à la définition d'un processus abstrait, il n'y a donc pas d'éthique.
Pourtant, il existe des algorithmes formalisés dans des logiciels : on décide de les mettre en œuvre pour en faire un traitement en cherchant à détourner le regard de l'humain. Deux types d'algorithmes : - Déductifs : modèle déjà connu par la machine, résultat des entrées. Le modèle est public, on peut le contester. - Inductifs : modèle non fourni à l'avance, sorte de « méta-modèle » sur les données : risques de faux résultats et de corrélations fausses si on fait des suites de modèles déductifs en simplifier l'écriture entre les jeux de données (Big Data). L'algorithme inductif ne permet pas d'obtenir des certitudes totales : on ne cherche pas un modèle précis du phénomène mais on met en évidence des corrélations. Le problème : corrélation et causalité ne sont pas toujours compatibles : quand on demande à une machine de mouliner des données, elle peut donner des résultats du jeux de données qui n'ont rien à voir. Il faut émuler le conditionnement pavlovien : on modélise le système sous forme de boîte noire. Le traitement des algorithmes inductifs n'est pas public ni déterminant : injection de données et tous les biais découlent de ces données : - Ex des photos de chihuahuas et de muffins : aucune analyse structurelle pour les images, le système de boîte noire ne fait aucune analyse. Le cas COMPAS : un logiciel de calcul utilisé dans les prison aux États-Unis : du techno bullshit : - Proposition de questions non pertinentes en considérant comme déviant tout ce qui est « anormal » - Biais majeurs - Problème de victimation en capturant les biais sociaux : ex du procès « State vs Loomis » - L'objectif des logiciels : montrer l'humain dispose du même taux d'erreur que les algorithmes. Ces biais peuvent s'expliquer par les traitements génératifs : des traitements algorithmiques qui sont configurés sur des corpus de données. Les logiciels avalent des données pour en faire une représentation synthétique de l'information agrégée sous forme de milliard de coefficients de pondération. - Par ex : les Stochastic Parrots : des images irréelles mais le logiciel propose un mix de ce qu'il connaît déjà. Certains traitement peuvent être performants : le rendu du traitement est une synthèse probabiliste.

Les enjeux juridiques : « les idées sont de libres parcours »

La question des droits d'auteurs : risque de copié-collé et de contrefaçon : comment respecter les droits d'auteurs ? => Les compilateurs : permettent d'éviter les problèmes de droit d'auteur mais les dispositifs enfilent les informations de manière probabilistes => En fonction des données : se pose la question aussi des données à caractère personnel. Les organes et la jurisprudence : - CNIL - Lois « Informatique et Libertés » du 6 janvier 1978 en réaction au scandale Safari : lois uniquement prévues pour les personnes physiques (donc pas les entreprises par ex). Il n'existe pas de propriété intellectuelle donc juridiquement, le « vol » n'est pas possible : on parle de « violation » ou e « contrefaçon ». - Art 4.1 du RGPD sur les données à caractère personnel - Art 22 du RGPD sur la notion de personne concernée

INTELLIGENCE ARTIFICIELLE ET ACCEPTABILITE SOCIALE (Léo Mignot)

Léo Mignot s'interroge sur la manière dont l'innovation, et plus particulièrement l'intelligence artificielle (IA), transforme les pratiques professionnelles. Son étude se concentre sur la radiologie, un domaine médical où l'IA est déjà largement utilisée.

Introduction : une IA si révolutionnaire ?

Léo Mignot met en garde contre l'enthousiasme excessif qui entoure l'IA, souvent présentée comme une technologie révolutionnaire destinée à bouleverser l'humanité. Ce discours, alimenté par les médias et certains scientifiques, peut conduire à des *« biais de représentation »*. En effet, l'histoire de l'IA est jalonnée de promesses spectaculaires qui peinent à se concrétiser, à l'image des annonces récurrentes d'Elon Musk sur la voiture 100% autonome.

- **Recommendation de lecture: K. Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, 2022**

Ce livre constitue une excellente introduction à l'intelligence artificielle, en soulignant d'emblée que celle-ci n'a rien d'intelligence et rien d'artificiel. En effet, si l'IA peut être perçue de différentes manières, sa définition joue un rôle fondamental : nommer une technologie, c'est aussi influencer la façon dont elle est appréhendée, critiquée ou remise en question.

L'IA est-elle réellement une intelligence artificielle capable de résoudre tous nos problèmes ? L'autrice met en garde contre cette illusion et rappelle qu'elle ne surgit pas de nulle part et *« n'existe pas dans le vide ».* Elle repose sur l'exploitation de ressources naturelles et sur des jeux de données annotés par des humains. Loin d'être une entité autonome et neutre, elle reflète une certaine vision du monde et s'inscrit dans des structures sociales et politiques qui tendent à reproduire les inégalités.

Ainsi, réduire l'IA à une simple avancée technologique masque ses véritables implications. Ce discours purement technique est d'ailleurs bien commode pour éviter de questionner en profondeur ses enjeux et ses conséquences.

La dépendance de l'IA à l'exploitation des ressources naturelles et humaines.

Loin d'exister *« dans le vide »,* l'IA repose sur l'exploitation intensive des ressources humaines et naturelles. Sa fabrication nécessite des métaux rares, impliquant une extraction minière massive, notamment en Chine, qui domine ce marché. Cette industrie engendre de lourdes externalités négatives, telles que la pollution et des violations du droit du travail. Contrairement à l'image souvent véhiculée d'une technologie propre et dématérialisée, l'IA participe à un modèle de production bien tangible, loin d'être écologique. Par ailleurs, les data centers, très gourmands en énergie, sont majoritairement situés dans les pays du Sud, accentuant les inégalités environnementales et économiques.

L'IA repose également sur une exploitation de la main-d'œuvre. Son impact sur le travail ne

se limite pas à la transformation des emplois existants : elle repose aussi sur un travail

invisible, sous-payé et souvent exercé dans des conditions précaires. Par exemple, les modérateurs de contenu, chargés de filtrer les images violentes ou choquantes, sont quotidiennement exposés à des contenus traumatisants. Ces *« travailleurs du clic »*[1], essentiels au fonctionnement de l'IA, sont pourtant invisibilisés, dévalorisés et souvent délocalisés. Une enquête récente du *Monde*[2] a d'ailleurs révélé que la France externalise une partie de ces tâches à des sous-traitants basés à Madagascar.

L'IA mais pourquoi ? Enjeux et usages en imagerie médicale[3] :

Comme évoqué précédemment, le domaine de la santé n'échappe pas à l'enthousiasme médiatique et scientifique autour de l'IA. De nombreuses promesses d'innovation émergent, qui s'accompagnent également de nombreux échecs et désillusions.

L'IA en santé est pluriforme : elle recouvre une large gamme d'applications, allant de la gestion administrative à l'aide à la décision clinique. Toutefois, toutes ne suscitent pas le même engouement, et certaines font l'objet de débats plus vifs que d'autres. Les métiers subalternes, en particulier, sont souvent absents de ces discussions, soulevant ainsi des interrogations sur la répartition des bénéfices et des risques liés à l'intégration de l'IA dans le secteur médical.

La radiologie est l'un des premiers domaines médicaux à intégrer des dispositifs qualifiés d'*« intelligence artificielle »* pour le traitement des images. En 2016, Geoffrey Hinton, pionnier du deep learning, déclarait que l'IA surpasserait bientôt les radiologues, suggérant d'arrêter leur formation. Cette affirmation, largement relayée, a alimenté l'idée d'une révolution imminente en médecine. Pourtant, six ans plus tard, cette prédiction ne s'est pas réalisée : au lieu d'un remplacement par l'IA, le secteur fait face à une pénurie de radiologues en raison de la hausse des examens d'imagerie. Bien que plus de 200 logiciels d'IA soient aujourd'hui sur le marché, leur adoption reste limitée et seuls 30 % des radiologues américains les utilisent. Ainsi, malgré des avancées technologiques réelles, les résistances professionnelles freinent leur généralisation. En effet, craignant la menace de l'IA sur leur emploi, les praticiens se sont d'abord mobilisés pour faire face à l'irruption de l'IA dans leur activité et ont tenté de se réapproprier ces outils.

La question centrale est alors de choisir quels sont les *« bon usages de l'IA »,* c'est-à-dire ce qu'il est *« acceptable »* de lui déléguer, contrairement à ce qui constituerait une *« part noble »* du métier. Il s'agit d'une application sélective suivant une logique du *« aux IA le sale boulot »*, comme celui de l'interprétation des images. En effet, l'IA peut s'avérer plus performante qu'un humain lorsque seulement une tâche lui est attribuée. Cependant, le travail de radiologue ne consiste pas uniquement à lire des images. Ainsi, au lieu d'un remplacement par l'IA, les radiologues mettent en avant un discours en faveur d'une collaboration avec l'intelligence artificielle pour optimiser la prise en charge des patients.

Cette méfiance est notamment due au fait que ces logiciels ne sont pas suffisamment évalués pour prouver leur efficacité. Les revendications des radiologues influencent directement les industriels du secteur. Beaucoup adoptent désormais le discours du non-remplacement. En adoptant cette stratégie, les industriels cherchent à favoriser une collaboration essentielle avec les radiologues qui jouent un rôle clé dans l'évaluation des logiciels et dans l'accès aux bases de données nécessaires à l'entraînement des algorithmes.

Enjeux sociopolitiques de l'IA

L'adoption de l'IA en santé s'inscrit dans un contexte sociopolitique complexe. Elle peut être perçue comme une réponse technologique à des défaillances systémiques, notamment dans

des secteurs en sous-effectif ou surchargés comme les urgences. Cette approche peut masquer des choix politiques concernant le financement et l'organisation des soins, en privilégiant des solutions technologiques au détriment de réformes structurelles. Au lieu de soigner directement la cause du mal, l'Etat encourage des solutions technologiques qui permettent seulement d'en limiter les conséquences. L'IA agit alors comme une *« béquille palliative »* et ne permet pas de répondre au véritable problème du manque de spécialistes. A l'inverse, les médecins qui sont experts dans leur domaine ne sont pas menacés par l'IA ce qui crée des enjeux de pouvoir importants dans le domaine médical.

Conclusion : pourquoi, pour quoi pour qui et par qui ?

Pour conclure, Léo Mignot met en évidence un décalage entre les discours publics sur l'intelligence artificielle en santé et ses usages concrets. Alors que l'IA est souvent présentée comme une révolution susceptible de remplacer les professionnels, son intégration réelle en radiologie révèle des résistances, des limites techniques et des enjeux de pouvoir.

Le débat s'est principalement focalisé sur la place de l'IA dans la prise de décision médicale, occultant d'autres dimensions essentielles comme la question du consentement des patients et de la commercialisation des données d'imagerie. De plus, l'adoption de l'IA est souvent perçue comme une réponse technologique à des problèmes structurels du système de santé.

Enfin, Léo Mignot souligne que l'attention médiatique tend à se cristalliser sur des menaces hypothétiques de remplacement des médecins par l'IA ou sur une intelligence artificielle générale encore inexistante, au détriment des enjeux immédiats et concrets. Ce biais de représentation détourne l'attention des effets déjà en cours, notamment l'impact environnemental et social de l'IA, ainsi que les transformations du travail médical.

Ainsi, plus que la question de savoir si l'IA va remplacer les radiologues, c'est celle de son appropriation, de sa régulation et des choix sociopolitiques qui l'accompagnent qui doit être posée. À qui profite réellement cette technologie, et pour quels usages ?

---

[1] France culture*, Les « travailleurs du clic », ces humains cachés dans les machines*, 13 février 2017 : Les "travailleurs du clic", ces humains cachés dans les machines | France Culture

[2] Le Monde, *Derrière l'illusion de l'intelligence artificielle, la réalité précaire des « travailleurs du clic »,* 3 janvier 2019 : Derrière l'illusion de l'intelligence artificielle, la réalité précaire des « travailleurs du clic »

[3] Mignot, L. et Schultz, É. (2022). Les innovations d'intelligence artificielle en radiologie à l'épreuve des régulations du système de santé. Réseaux, N° 232-233(2), 65-97

IA ET DÉMOCRATIE (Jean-Gabriel Ganascia)

Jean-Gabriel Ganascia, une figure incontournable du paysage français de l'intelligence artificielle (IA). Professeur 1 d'informatique à Sorbonne Université depuis 1988, il se situe à la croisée de l'informatique et de la philosophie, de la science et de ses implications politiques. Ses travaux, qui s'inscrivent dans une approche interdisciplinaire, interrogent autant les capacités techniques de l'IA que les conséquences éthiques et sociétales de son déploiement. Par souci de clarté, nous avons ici choisi de mettre en italique certains mots, dont la définition est rendue disponible à la dernière page du document (partie lexique/ concepts clés). Un parcours riche entre science et philosophie Jean-Gabriel Ganascia a suivi une formation d'ingénieur à l'Institut d'optique théorique et appliquée d'Orsay, avant d'entreprendre des études de philosophie. Ce double cursus l'a conduit à soutenir deux thèses : la première en 1982 sur les systèmes à base de connaissances, et la seconde en 1987 sur l'apprentissage symbolique automatique. Il a ensuite intégré le Laboratoire d'Informatique de Paris 6 (LIP6), où il dirige aujourd'hui l'équipe ACASA (Agents Cognitifs et Apprentissage Symbolique Automatique). Son expertise couvre un large spectre, allant de l'intelligence artificielle à l'éthique computationnelle, en passant par les humanités numériques. Un acteur clé de la réflexion éthique sur l'IA Au-delà de ses contributions scientifiques, Jean-Gabriel Ganascia joue un rôle majeur dans la réflexion éthique autour du numérique. Membre du Comité Consultatif National d'Éthique et ancien président du COMETS (Comité d'éthique du CNRS), il analyse avec recul l'impact du développement des technologies de l'information et des grands modèles de langage sur la société. Il est notamment l'auteur de plusieurs ouvrages accessibles au grand public, tels que Le Mythe de la Singularité et L'IA expliquée aux humains, qui cherchent à démystifier les discours dominants sur l'IA et à favoriser une appropriation citoyenne des débats technologiques.

Vous pouvez ajouter des titres (Format > Styles de paragraphe) qui apparaîtront dans votre table des matières.
Jean-Gabriel Ganascia appuie son propos sur les impacts massifs de l'IA a notamment sur la qualité et la diffusion de l'information aux citoyens, il interroge la manière dont elle empêche la formation d'unespace de délibération démocratique et inclusif. Il pointe également les dangers liés à la concentration de cet outil dans les mains des grands acteurs du numérique et l'insuffisance des régulations juridiques qui voient le jour, notamment au niveau européen.

LES DANGERS DE L'IA DANS LES RÉGIMES AUTORITAIRES

Documentaire sur le crédit social en Chine, déclarations politiques au Sommet sur l'IA de la présidente Biélorusse Svetlana Tikhanovskaïa sur la surveillance des opposants dans son pays, une des journaux sur les dérives de l'IA sur les libertés, les emplois… on assiste à une médiatisation croissante des dangers que représenterait l'IA, notamment en agitant la peur d'une utilisation liberticide de celle-ci dans des régimes autoritaires. Pour Ganascia, ces peurs se justifient aisément, en ce qu'elles pointent des atteintes graves et explicites aux libertés et droits fondamentaux individuels consacrés par le droit international depuis le milieu du XXe siècle. On imagine bien comment le droit à la vie privée, la liberté de penser, de religion, d'expression, sont ainsi menacés à l'heure où les développements de l'IA permettent une surveillance accrue de la vie des individus….comme dans les systèmes de notation sociale orwelliens mis en place dans certaines villes en Chine, où des milliers d'individus sont soumis à une surveillance continue et bénéficient d'un score social déterminant pour accomplir les plus simples actions dans leur quotidien (pouvoir

circuler librement, voyager, acheter des biens et services, obtenir un prêt etc.). Devant ces usages liberticides, des acteurs de la société internationale comme l'UE ou de nombreux comités d'éthique des organisations internationales tentent d'encadrer les développements de l'IA en mettant en place des régulations protectrices des libertés individuelles fondamentales. En opérant un renversement de la question des dangers de l'IA dans les régimes autoritaires, Ganascia nous invite ici -et c'est ce qui constitue le point clé de son intervention-, à considérer les dangers de l'IA en démocratie, notamment en analysant son impact dévastateur sur la création d'un espace de délibération collectif citoyen nécessaire à la pratique démocratique. Pour cela, il nous invite à décortiquer les textes juridiques et les méthodes employées par ces acteurs afin d'en mesurer la pertinence et de mieux en pointer les limites.

AI ACT: UNE RÉGLEMENTATION EUROPÉENNE AXÉE SUR LA PROTECTION DES LIBERTÉS INDIVIDUELLES ILLUSTRANT UNE CONCEPTION LIBÉRALE (ET RESTRICTIVE) DE LA DÉMOCRATIE

Pour essayer de prévenir les dérives autoritaires avec des applications liberticides de l'IA, des tentatives de réglementation internationales ont donc vu le jour. La Commission européenne est par exemple à l'origine de l' "AI Act", entré en vigueur le 2 février dernier, censé compléter d'autres règlements (type DSA, entré en vigueur en 2024) et encadrer le développement des techniques d'IA en vue de protéger nos libertés fondamentales. Pour cela, l'AI Act interdit notamment trois types de pratiques qui constitueraient des "risques inacceptables": les techniques dites « subliminales », le crédit social et les techniques 3 biométriques. Il définit par ailleurs d'autres catégories avec des pratiques correspondant à desrisques élevés" ou encore des "risques minimaux" (comprendre des pratiques non interdites mais qui nécessitent une réglementation plus ou moins contraignante).
J.G. Ganascia s'étonne cependant du caractère « fictif » de certaines de ces régulations, notamment celles de l'interdiction des « techniques subliminales », soulignant que ces manipulations introduites par le publiciste américain James Vicary dans les années 60 n'avaient jamais été prouvées scientifiquement. Ainsi, il dénonce une loi qui ici manquerait de rigueur scientifique et laisse douter de la pertinence des acteurs à l'origine de ce type de recommandations qui s'apparentent à des « coquilles vides » aujourd'hui. De plus, il souligne la complexité de réglementer en bannissant certaines pratiques : si les systèmes de notation sociale sont par exemple formellement proscrits dans l'AI Act, quid de l'IA intégrée dans les tests de solvabilité pratiqués par les banques pour accorder les crédits, quid du permis à point dans nos sociétés occidentales dites démocratiques?… Autant d'exceptions qu'il faudra accorder aux acteurs lors des négociations des règlementations, ce qui aboutit en réalité à multiplier des législations de plus en plus complexes, indigestes, et difficiles à mettre en œuvre. Plus encore, le cœur du problème pour notre intervenant réside dans le fait que ces recommandations et régulations ne prémunissent en rien des dangers actuels que fait peser l'IA sur nos démocraties. En effet, le vrai danger serait plus à chercher du côté des conditions qui font qu'une société est démocratique ; en dénonçant le rôle de l'IA dans la qualité de l'information qui nous arrive, il nous montre comment la citoyenneté et la délibération collective, nécessaires à la création d'un espace démocratique, se retrouvent fortement impactées.

INFOX ET CITOYENNETÉ À L'HEURE DE NOS SOCIÉTÉS "SUPRACONNECTÉES"

Le Web, apparu et développé dans les années 2000, a littéralement transformé nos sociétés, la façon dont nous échangeons, communiquons, nous informons, etc. Pour qualifier cette transformation sociétale massive, J.G.Ganascia utilise, par une analogie scientifique, le néologisme de « supraconnectivité », traduisant l'idée de « supraconductivité », c'est à dire un phénomène physique caractérisé par une absence de résistance, insistant sur le caractère continu et ininterrompu de nos échanges et de notre activité en ligne. Aux débuts d'Internet régnait encore ce rêve utopique d'« agora numérique », une utopie d'horizontalité comme un espace public ouvert et accessible à tous où règnerait la transparence de l'information. Cette dernière n'est d'ailleurs pas nouvelle : la transparence serait en effet une idéologie qui aurait pris racine dans le mouvement libéral anglais dès la fin du XIX, matérialisée par des constructions architecturales comme le Palais de Cristal à Londres. 4 Loin de ccet idéal, la diffusion massive de « fake news » ou « infox », déstabilise complètement les institutions démocratiques. À la différence d'une conception exclusivement libérale de la citoyenneté et de la démocratie, notre auteur souligne en effet que ces deux notions résident avant tout dans la capacité des individus à prendre part à la vie politique. Or, dans un espace gangréné par la désinformation, comment retrouver les conditions d'une délibération collective ? On retrouve là l'idée chère à Arendt ou Habermas de la constitution d'un espace public, central pour échanger et délibérer, à l'origine d'une conception de la démocratie qui va au-delà de la simple protection des libertés fondamentales. A ce titre, J.G.Ganascia retient trois dimensions importantes dans la manière dont l'IA déstabilise nos démocraties.

INFORMATIONS ET IA: ZOOM SUR LES DANGERS DÉMOCRATIQUES DE…

→ La production et la diffusion massives d'infos : Nos sociétés, originellement structurées autour du concept de « surveillance » par les institutions de l'État moderne (architecture du panoptique, prisons, hôpitaux, asiles…) auraient basculé ou du moins seraient rattrapées par une nouvelle forme de « sous-veillance » qui viendrait faire contrepoids à l'autorité : le fait que n'importe qui puisse prendre une photo, une vidéo, la diffuser par les réseaux, contribuerait ainsi à une forme de démocratisation. Or cette habilitation est aussi source de grande déstabilisation dans la qualité de l'information diffusée, et parfois à l'origine de « fake news » ou d'infox. Si la fabrication de contenus fallacieux n'est pas nouvelle (notre intervenant cite entre autres les techniques de déstabilisation guerrière énoncées par Sun Tzu, les libelles et les canards lors du développement de la presse), ce phénomène prend une ampleur particulièrement effrayante aujourd'hui. En effet, il se trouve démultiplié car sa production et sa diffusion sont devenus un jeu d'enfant qui ne nécessite aucun (ou quasi) coût et est rendu accessible à tous. Contenu d'autant plus largement diffusé par les IA à l'heure où les algorithmes sont conçus pour relayer des vidéos de « buzz » qui engendrent le plus de visionnages possibles… Dans ce contexte, il apparaît de plus en plus difficile de distinguer le vrai du faux et d'avoir un accès égal à une information de qualité pour tous. → L'hypertrucage ou « deepfake » : En revenant sur le rêve culturel qui existait aux débuts d'Internet, celui d'une bibliothèque universelle, « mémoire collective de l'humanité » rendue accessible à tous à n'importe quel moment, Ganascia soulève un paradoxe : en plus de cette culture existante (qu'on ne retrouve en réalité souvent que partiellement, de manière tronquée), Internet abrite désormais aussi une culture alternative, basée sur la génération automatique d'images, de textes de sons, créant une mémoire encombrée et falsifiée. Les contenus générés par l'IA qui visent à reproduire la réalité de manière fallacieuse se nomment « deepfakes» (hypertrucages en français) inondent déjà les plateformes. 5 En reprenant la typologie sémiotique de C.Pierce introduite au XIXe siècle avec l'introduction de la photographie, Ganascia rapproche ces contenus des « indisignes» c'est-à dire des signes pointés sur des réalités qu'ils désigneraient. Ainsi l'infox donnerait l'impression qu'elle

nous pointe la réalité aujourd'hui, alors qu'elle est fabriquée de toute part, modifiant dangereusement notre perception du réel. → Le risque de profilage/ciblage : Enfin, l'auteur pointe le risque du profilage et du ciblage du public dans l'adressage des infox, creusant les dangers de la production d'une post vérité caractéristique de l'ère trumpienne. L'IA y joue un rôle central en effectuant un criblage de la population notamment sur les réseaux sociaux, en envoyant à chaque utilisateur du contenu susceptible de lui plaire ou de le faire réagir. Ainsi, ce dernier risque amènerait une fragmentation de la société avec des individus coincés dans des bulles algorithmiques, incapables de se confronter à différents points de vue, entérinant le dialogue social, marquant l'avènement d'une « civilisation des poissons rouges » (B.Patino). Le cas de la diffusion par le média X de vidéos ciblées durant la campagne américaine 2024 est parlant : sur certaines envoyées aux électeurs démocrates, la candidate Kamala Haris promouvait les positions de Benyamin Nétanyahou, alors que d'autres vidéos la montrant au côté des Palestiniens étaient adressées à des milieux juifs et conservateurs, en vue d'affaiblir le vote démocrate au profit du vote républicain. Certains réseaux sociaux apparaissent alors comme des espaces, favorisant des algorithmes idéologiquement biaisés, à l'image des idées de leurs détenteurs (le milliardaire libertarien Elon Musk pour X).

VERS UN USAGE PLUS ÉTHIQUE DE L'IA? PROPOSITIONS DE J.G. GANASCIA

Si Jean-Gabriel Ganascia met en garde contre les dangers d'une intelligence artificielle qui fragmente la société et compromet gravement l'existence d'un espace public démocratique fondé sur une délibération citoyenne libre et éclairée, il refuse néanmoins de céder au fatalisme. Il entrevoit, au contraire, des pistes pour faire de l'IA un outil bénéfique, à condition de l'utiliser avec discernement. Pour cela, plusieurs exigences s'imposent : - Sortir de la naïveté face aux géants de la tech et se méfier de leur « charité ensorcelée », pour reprendre l'expression rimbaldienne, qui dissimule souvent des intérêts économiques et politiques. Les discours de ceux qu'il appelle les « pompiers pyromanes qui brandissent les dangers fantasmés d'une IA qui dépasserait les hommes sont une manière de détourner notre attention des usages de l'IA qu'ils développent eux-mêmes, à l'image du projet d'interface cerveau-machine au coeur de Neuralink, la société d'Elon Musk. 6 - Comprendre les mécanismes d'une nouvelle « servitude numérique » afin d'imaginer les conditions d'une véritable émancipation citoyenne. Cela implique notamment de saisir le fonctionnement probabiliste des modèles de langage, comme les chatbots (type ChatGPT), et plus généralement les « orins » au cœur de nos vies (organismes informationnels) qui ne possèdent aucune compréhension réelle du monde mais calculent simplement la proximité d'emploi des mots (tokens) entre eux. Encourager un usage raisonné de l'IA, pourrait avec ces conditions, paradoxalement, redonner aux citoyens un certain pouvoir sur l'information. En exploitant l'IA pour identifier et confronter les différentes sources sur un sujet donné, il devient possible de reconstituer un espace de délibération collective, à l'image du travail journalistique qui repose sur la vérification des sources. Finalement, J.G. Ganascia appelle à l'élaboration de régulations plus claires et adaptées aux réalités des métiers, en évitant les recommandations trop générales qui se révèlent souvent inefficaces. À cette fin, la législation européenne ne devrait pas se limiter à la protection des libertés fondamentales (DSA, AI Act), mais aussi se charger de garantir les conditions d'exercice de la citoyenneté dans un espace public commun. Cela passerait notamment par une lutte efficace contre la désinformation, sans pour autant porter atteinte à la liberté d'expression.

IA ET ÉDUCATION (Florian Meyer)

Florian Meyer obtient d'abord une maîtrise en génie mathématique et informatique (Université de Franche-Comté) avant d'étudier à l'Université de Poitiers où il est diplômé en "Technologies audiovisuelles et informatiques pour l'éducation". Il continue ses études avec un doctorat en philosophie, technologie éducative et pédagogie jusqu'en 2010. C'est dans ce cadre qu'il réalise une thèse sur les "Effets d'un dispositif de formation exploitant des vidéos d'exemples de pratiques sur le développement d'une compétence professionnelle chez des enseignants du primaire"

Spécialiste du numérique éducatif, il donne également des cours en ligne portant sur l'intégration du numérique dans l'enseignement secondaire et supérieur ainsi que sur l'ingénierie pédagogique des formations à distance.

Ses travaux de recherche et ses collaborations internationales portent sur l'innovation pédago-numérique ainsi que sur l'usage des technologies éducatives dans les pratiques d'enseignement. Plus précisément, il travaille sur le développement des connaissances et compétences numériques des enseignants et des formateurs.

Pour cela il s'appuie sur la psychopédagogie, notamment à travers l'observation et l'analyse de vidéos de pratiques enseignantes.

Enfin, ses travaux de recherche sont récompensés par plusieurs distinctions, dont le Prix d'excellence en matière de diversité et d'inclusion de l'ARUCC et la Grande distinction en enseignement universitaire de l'Université de Sherbrooke.

Définition et applications générales de l'IA

L'IA est présente dans de nombreuses sphères et a un impact sur l'identité numérique tout en ayant la capacité d'enregistrer les usages.
Il existe plusieurs types d'IA, présentes dans nos vies sous différentes dimensions : de Netflix à Alexa en passant par les robots, Google Translate… Toutes sont appuyées par l'IA mais aucune n'a de réelle visée éducative (elles existent uniquement pour enrichir les capacités).
Ainsi, les IA génératives réfèrent à des systèmes informatiques capables d'utiliser des techniques avancées d'apprentissage profond, pour s'entraîner à partir d'une diversité de données (textes, images, musiques, etc.). Les IAG sont aptes à produire du contenu de manière autonome et en réponse aux requêtes d'une personne utilisatrice. (**Foster** et **Friston**, 2023). **Russell** et **Norvig** (2022) ont ainsi mis au point une typologie de base permettant de classifier divers outils alimentés par IAG selon la nature de leurs intrants et extrants.

Les IAG reposent sur le principe de décodage initial, c'est-à-dire l'interprétation des *tokens*, des symboles traités par la machine pour contextualiser l'ensemble du travail. Ces *tokens* sont organisés en segments, certains créés, d'autres existants, permettant d'identifier des associations

entre notions selon des probabilités calculées à partir de milliards de vecteurs structurés en réseaux de neurones. Ce processus aboutit à un réassemblage des informations, générant des réponses toujours probabilistes et jamais exactes, car une IA n'est pas intelligente en soi : elle dépend uniquement des données avec lesquelles elle a été entraînée. Plus une version d'un outil évolue, plus elle acquiert de compétences, mais son apprentissage reste une nourriture instable, continuellement alimentée et nécessitant une adaptation constante du réseau de neurones. Cette évolution perpétuelle peut néanmoins entraîner un effet d'aplanissement de l'information, voire un risque de tourner en rond.

L'IA appliquée à l'éducation

On voit pourtant se développer, à partir des années 80, des travaux de recherche sur le développement d'outils d'IA pour l'éducation et au service de l'apprentissage. Depuis une trentaine d'années, l'intelligence artificielle à visée éducative (« IAEd ») fait même l'objet de recherches académiques. Aujourd'hui, plusieurs systèmes exploitent ces technologies dans un but pédagogique :

- **Ali** et **ISA** : plateformes de suivi académique
- **Tableaux de bord intelligents** : capables d'analyser les difficultés d'un étudiant et de proposer des ressources adaptées
- **Systèmes intelligents adaptatifs** : capables de générer des exercices personnalisés en fonction des besoins de l'apprenant

Florian Meyer distingue trois grands types d'IA appliquées à l'enseignement :

- **L'IA centrée sur les étudiants** : adaptée aux difficultés d'apprentissage et aux besoins spécifiques (handicap, suivi pédagogique personnalisé).
- **L'IA centrée sur les enseignants** : sur la base d'outils d'aide à la préparation des cours, de correction automatique et de suivi des performances.
- **L'IA institutionnelle** : comme soutien à la gestion administrative et à la prise de décision stratégique.

Impact de l'IA sur les compétences et la pédagogie

1. *La notion de compétence*

Selon **Tardif** (2006), une compétence se définit comme :

"*Un savoir-agir complexe prenant appui sur la mobilisation et la combinaison efficaces d'une variété de ressources internes et externes à l'intérieur d'une famille de situations.*"

Une famille de situations étant un ensemble des situations de niveau de difficulté équivalent traduisant une même compétence. Les ressources internes correspondent aux connaissances et stratégies, c'est-à-dire aux capacités à agir dans une certaine situation. Les ressources externes mobilisent pour leur part les ressources humaines et tout outil susceptible d'aider ou d'assister.

On devient de plus en plus compétent dans ces situations en apprenant à résoudre des problèmes de plus en plus complexes. On distingue ainsi deux approches :

- **Être compétent** : agir et savoir agir dans une situation professionnelle, savoir mobiliser ses ressources « internes personnelles (connaissances, savoir-faire ou habileté, aptitudes, émotions, …) et externes (ressources de l'environnement) et en faisant appel à l'usage de fonctions de guidage » de manière adaptative dans un contexte donné.
- **Avoir des compétences** : posséder un ensemble de connaissances et de savoir-faire exploitables permettant ensuite d'agir dans des situations complexes. Cette notion renvoie à une ressource personnelle ou à une « combinaison de ressources personnelles nécessaires pour savoir agir en situation professionnelle ».

2. *Compétences traditionnelles et implication de l'IA*

Des compétences transversales peuvent être réparties en quatre sphères. Ces 4 sphères sont impactées par le développement de l'IA et son utilisation.

- **Compétences citoyennes** : responsabilité, éthique (ex : s'engager dans des manifestations, prendre en compte les différences de chacun, devenir un citoyen éclairé pour les sociétés)
- **Compétences apprenantes** : capacité à s'adapter, notamment aux outils numériques (ex : apprendre à gérer son stress, son organisation, son apprentissage)
- **Compétences disciplinaires** : acquisition de compétences relatives à son champ disciplinaire (ex : intégration de l'IA dans les pratiques académiques)
- **Compétences professionnelles** : adaptation au monde du travail et professionnel, (ex : rendre service, travailler en groupe, soutenir un projet, quel que soit le champ disciplinaire)

On peut aujourd'hui y ajouter une compétence supplémentaire : la **compétence numérique,** c'est-à-dire, à la fois la nécessité d'être un citoyen éthique, et d'avoir l'habileté technologique. Cette compétence numérique renvoie à une **littératie numérique** (**Fastrez** et **De Smedt**, 2012) telle que la capacité à lire, écrire, naviguer, ou encore organiser l'information, et plus largement à la compréhension du fonctionnement de cette technologie et à quel point celle-ci peut aussi affecter les autres utilisateurs dans leur pratique. La création d'une littératie de l'IA apparaît comme cruciale, incluant une éducation continue du personnel enseignant mais également administratif, pour mieux appréhender les outils de l'IA et leurs enjeux. Cette maîtrise leur permettrait ensuite de transmettre cet apprentissage de l'IA aux élèves dès le plus jeune âge, afin de développer des compétences solides et des connaissances dans des conditions optimales.
L'existence même de cette littératie numérique est encore étudiée, mais la recherche en dégage déjà six composantes :

1. **Recognize** : reconnaître l'IA
2. **Know and understanding** : comprendre comment elle fonctionne
3. **Use and apply** : être capable de l'appliquer à ses usages
4. **Evaluates** : être en capacité d'évaluer la qualité de ce qui est produit et les impacts que cela génère
5. **Create** : être capable de créer
6. **Navigate** : être en capacité de naviguer

L'UNESCO a également créé un référentiel de compétence en IA pour les apprenants témoignant de l'importance du sujet à l'échelle internationale[1].

3. L'agentivité

Selon **Jezegou**, l'agentivité constitue un « contrôle exercé sur leur propre fonctionnement conduite et l'environnement », c'est-à-dire « lorsqu'à un moment donné et au regard d'une situation ou d'un contexte spécifique, le sujet exerce une influence intentionnelle sur ses propres

conduites et modes de fonctionnement, sur ses actions, sur autrui ou encore sur les systèmes d'action collective, alors il fait preuve d'agentivité » (2019, p. 197). Pour **Bandura**, « être un agent signifie faire en sorte que les choses arrivent par son action propre et de manière intentionnelle » (2009, p. 17).

L'IA peut en effet constituer à certains égards une perte d'agentivité, une impression de perte de contrôle amenant à nous questionner : est-ce que c'est nous qui faisons réellement ? Est-ce que notre usage est éclairé, sécuritaire, éthique, déontologique ? Cet usage questionne sur la nécessité de se réapproprier l'IA pour s'assurer que c'est nous, utilisateurs, qui sommes en contrôle de l'innovation et qu'elle nous sert selon nos propres volontés. Dans l'univers de l'apprentissage, la menace serait le développement d'une forme de paresse cognitive des étudiants appelant à se questionner sur leurs capacités à faire les mêmes actions si la machine venait à disparaître, notamment concernant leurs ressources cognitives et leur dotation en connaissances, propres à leur statut d'apprenant.
Les professeurs développent de façon croissante une crainte de l'hégémonie de l'IA. Cela met en lumière la nécessité d'une **agentivité collective**, c'est-à-dire un questionnement collectif sur le recours aux usages de l'IA pour éviter la perte de contrôle humaine.

Usage des IA par les enseignants et étudiant.e.s

Résultats d'une étude menée à l'Université de Sherbrooke (2023-2024)

Une consultation menée à l'UdeS a permis d'analyser l'usage des IA génératives (IAg) par les étudiants et enseignants.

- **70 % des étudiants et enseignants** connaissent peu l'IA. Toutefois, l'utilisation reste plus courante chez les étudiants (**30 % l'utilisent régulièrement contre 20 % des enseignants**).
- **58 % des étudiants** et **53 % des enseignants** déclarent bien ou très bien connaître les IA génératives.
- **55 % des étudiants** et **74 % des enseignants** n'ont jamais utilisé ces outils dans leur parcours universitaire.
- L'outil le plus utilisé est **ChatGPT** (par **97 % des étudiants et 74 % des enseignants**), on note donc une adaptation rapide du corps enseignant.
- Les IA sont principalement employées par les enseignants pour s'inspirer (82 %), traduire du texte (61 %), et trouver des exemples (59 %).

Enjeux et inquiétudes

La **transformation des méthodes d'enseignement** impose l'intégration des outils d'IA dans la pédagogie, exigeant une adaptation des pratiques pour en exploiter le potentiel sans en subir les dérives. **L'impact sur l'évaluation** soulève des défis majeurs, notamment en matière de plagiat et d'équité entre étudiants, rendant nécessaire une réflexion sur les critères d'appréciation des compétences. Par ailleurs, la **dépendance à ces technologies** risque d'affaiblir la pensée critique et l'autonomie cognitive des apprenants, les rendant trop passifs face aux outils d'assistance. Ces évolutions s'accompagnent d'**enjeux éthiques** fondamentaux, tels que la transparence des algorithmes, la régulation de leur usage et la sensibilisation des acteurs éducatifs pour garantir une utilisation responsable et équilibrée.

Prévoir les attentes et impacts de l'IA sur la pédagogie

L'IA est souvent perçue comme un outil potentiellement bénéfique pour améliorer l'apprentissage et l'efficacité pédagogique. Toutefois, 34 % des enseignants déclarent ne pas

savoir si l'IA peut réellement soutenir l'apprentissage, tandis que la majorité estime qu'elle pourrait améliorer l'apprentissage dans leur discipline.

L'influence sociale joue un rôle déterminant : de nombreux enseignants ne savent pas si leurs collègues utilisent ces outils et ressentent parfois une réticence ou une impression négative quant à leur usage. Cette perception révèle un besoin de cadrage institutionnel et universitaire pour guider et normaliser son utilisation.

Toutefois un consensus semble s'affirmer dans la nécessité d'une adaptation pédagogique : enseignants et étudiants s'accordent sur l'idée que l'IA exige un changement des méthodes d'enseignement pour rester pertinente. Certains étudiants redoutent que l'IA ne devienne une béquille cognitive, risquant de diminuer l'autonomie et rendant impossible un retour en arrière. Du côté des enseignants, il faut noter un potentiel risque de perte d'agentivité, nécessitant un maintien de leur maîtrise des processus pédagogiques. De même, l'impact sur l'évaluation des compétences de leurs étudiants est réel : une machine ne peut juger une compétence, seulement une connaissance.

En somme, des recommandations communes peuvent être déduites. D'abord donc, former enseignants et étudiants à une utilisation critique de l'IA. Et puis dans un objectif d'efficacité, la nécessité de développer des ressources et outils adaptés aux besoins académiques spécifiques. Enfin, et ce dans l'intérêt commun de tout citoyen, encadrer son usage pour garantir un apprentissage éthique et efficace.

Recommandations pour une intégration éthique et efficace des IA dans l'apprentissage

Voici quelques IA utiles dans l'enseignement et à intégrer dans son apprentissage :

- **Search Rabbit, Consensus** : agrégation d'articles académiques, permet la connexion entre plusieurs ressources issues de revues universitaires
- **NotebookLM** : aide à la structuration des notes, à l'apprentissage, proposition de quizz…
- **Grammarly** : amélioration de la qualité linguistique.
- **Perplexity** : outil de recherche avancée.
- **SlidesAI** : création de présentations et diapositives automatiques.
- **Notion** : organisation du travail d'équipe et de projet de groupe
- **Site Web "There's an AI for that"** : permet de voir si une IA a déjà été créé pour le besoin rencontré
- **Comparia** : compare les réponses de 2 IA différentes ainsi que leurs empreintes environnementales
- **Article de Futurascience** : tutoriel expliquant aux enseignants comment créer un chabot personnalisé dédié à leurs cours et entraîné à partir de leurs propres contenus
- **Radar des incidences de l'IA générative de Gartner** : traque l'avancée de l'IA et montre à quel stade de compétences elle se situe

L'objectif est d'utiliser ces outils pour enrichir l'apprentissage, sans en devenir dépendants. Aujourd'hui, l'usage de l'IA doit être assumé et réfléchi, dans une logique d'accompagnement des processus pédagogiques.

[1] https://unesdoc.unesco.org/a

IA ET SON IMPACT ENVIRONNEMENTAL (Aurélie Bugeau)

Après avoir concentré ses recherches sur le traitement d'image, Aurélie Bugeau assiste à une conférence révélant les impacts du numérique sur l'environnement. Cette prise de conscience la pousse à changer de thématiques de recherche pour se concentrer sur cette question de l'impact environnemental de l'IA, en accord avec ses convictions.

Il existe un débat mouvant aujourd'hui autour de l'IA entre son impact négatif sur l'environnement et ce qu'elle peut au contraire lui apporter. Celle-ci est en effet capable de produire des modélisations climatiques, de permettre l'optimisation de la gestion des écosystèmes et des ressources ou de la consommation énergétique.

Nous faisons face aujourd'hui à un certain nombre de problématiques environnementales. Les Etats ont pu prendre des engagements face au changement climatique, notamment avec l'accord de Paris, pour essayer de limiter ce réchauffement à 1,5 degrés. En l'état des politiques mises en place, le réchauffement risque d'être supérieur à 3 degrés et impose une réduction drastique des émissions au niveau mondial pour se conformer aux objectifs.

La **perte de la biodiversité** a un impact sur les sociétés et la vie à travers ses effets sur l'alimentation, les ressources en eau, le logement, la santé (plantes, espèces utilisées.

En 2009, **neuf limites planétaires** ont été établies. Ce sont des seuils au-delà desquels les équilibres naturels terrestres pourraient être déstabilisés et les conditions de vie devenir défavorables à l'humanité. Six d'entre elles sont aujourd'hui déjà dépassées.

Les **matières premières et métaux rares** utilisés dans ces technologies que l'on doit extraire ailleurs posent des problématiques d'inégalité, de justice sociale, de colonialité. Cela nous pousse à interroger les **bénéfices-risques** de l'IA pour l'environnement.

Selon un rapport du parlement européen, l'IA peut être déployée pour un grand nombre d'applications pour promouvoir les objectifs du Green deal européen mais aussi empêcher d'atteindre ces objectifs.

Dans les discours des gouvernements et des grandes entreprises, l'IA est vue comme une promesse sur différents aspects. Cette semaine, est paru un **plan d'adaptation au réchauffement climatique** pour adapter la France et son budget à un réchauffement de 4 degrés. L'objectif « **utiliser IA pour l'adaptation** » en fait partie. En effet, l'IA permet de faire beaucoup de choses dans de nombreux domaines et de manière plus efficace qu'avant, surtout si elle est appliquée dans d'autres sciences. Elle peut par exemple être utilisée pour la détection précoce des **feux de forêt** par le biais de capteurs installés ou d'images satellites. Cela permettrait de **limiter ces catastrophes naturelles** qui émettent une grande quantité de gaz à effet de serre. En 2023, ils ont par exemple été responsables de l'émission de 2 giga tonnes (humains émettent 40). Une autre application est celle de la **détection de fuites de méthane** pour l'extraction d'hydrocarbures, une activité considérée comme une bombe climatique au vu de la quantité de méthane émise dans l'atmosphère, un gaz ayant une courte durée de vie dans l'atmosphère mais au pouvoir réchauffant élevé. Elle peut être utilisée en **appui à l'installation d'énergies décarbonées** comme le solaire ou le nucléaire qui requièrent un pilotage très fin. Les "smart grids" permettent une gestion intelligente de l'énergie en intégrant **des solutions techniques d'IA** pour surveiller et ajuster en temps réel la distribution d'électricité.

L'IA peut ainsi poursuivre un objectif environnemental dans de nombreux secteurs : les transports, l'industrie à travers des outils de prédictions environnementales, climatiques, en appuyant le développement de villes et de bâtiments intelligents, la gestion des sols, l'agriculture numérique pour produire mieux avec moins d'engrais, cibler les plans malades.

Personne n'envisage une absence d'IA en 2050. Des agences gouvernementales ou privées réalisent des prospectives pour guider les politiques d'aujourd'hui et dans les années à venir pour savoir comment atteindre la neutralité carbone.

Ainsi l'ADEME (Agence de la transition écologique) a créé **4 scénarios** qui comportent tous du numérique ou des sciences de l'IA. Parmi eux, un **scénario de sobriété** : grand défi humain, changement des modes de vies et à l'opposé un quatrième scénario reposant sur **l'ultra technologie** comme solution de croissance : pari humain vs pari technologique

En effet, l'IA peut être utilisée pour l'**adaptation sociétale** : la facilitation des changements individuels, la gestion des flux migratoires (répartition, outils démocratiques).

Les chiffres des agences internationales de l'énergie mettent en avant des progrès dans l'optimisation de l'utilisation d'énergie des centres de données. Toutefois, on note un manque de constance et une **explosion de la consommation.**

En 2026, on anticipe une **hausse de la consommation des data centers** due au développement de l'IA qui vient s'ajouter aux crypto-monnaies ayant créé une première augmentation dès 2022.

Au niveau régional, on note récemment, une importante **croissance des centres de données** en Europe, en Chine et aux US.

L'Irlande a permis l'installation de nombreux centres de données. Il est prévu en 2026 que ces centres représentent 30% de la demande électrique du pays (autant que le secteur du bâtiment). La nécessité de développer très vite de **nouveaux réseaux électriques** s'accompagne de nouveaux conflits.

En France, l'enjeu se situe dans la manière dont **alimenter les centres de données.**

Différentes tendances et scénarios sur la consommation électrique sont réalisés, situés entre une abondance sans limite et une impossibilité d'accès aux ressources.

Au niveau de l'entreprise elle-même : les chiffres sur la consommation énergétique et en eau pour les plus grosses entreprises du numérique dans les rapports environnementaux montrent une forte croissance. Ce constat est lié au déploiement rapide des centres de données pour l'IA. Ces chiffres incluent les centres de données mais aussi l'équipement du personnel, les voyages. Dans l'eau prélevée pour alimenter les bâtiments, une partie est rejetée dans les nappes ou les eaux usées.

La promesse pour 2030 d'entreprises comme Meta d'être "water positive" ou neutre en carbone est un objectif qui va être très compliqué à atteindre. **Une augmentation d'IA dans nos sociétés va entraîner une augmentation des besoins en électricité** et pourrait rendre incontournable la relance des centrales à charbon. L'impact de l'IA pour les digital companies est croissant.

D'où vient l'impact environnemental : du matériel et des infrastructures

En effet, le développement de l'IA nécessite **d'acquérir, traiter et stocker les données, entraîner des modèles, les tester et les déployer.** Toutes ces phases utilisent beaucoup d'équipement numérique tel que les serveurs de stockage et de calcul au sein des centres de données.

Les centres de données peuvent être comparés à de grandes armoires avec beaucoup de serveurs. Ils nécessitent des réservoirs de diesel pour prendre le relais en cas de coupures électriques. Ils sont composés de serveurs de calcul de stockage, système de réseaux et de refroidissement et de systèmes très performants pour maintenir le centre de données actif en cas de coupures d'électricité et enfin de capteurs pour acquérir des données.

La **phase d'entraînement d**e l'IA nécessite un centre de calcul et l'ordinateur du développeur. **La mise en réseau** nécessite aussi des équipements : des antennes 4G, 5G ainsi que des millions de kilomètres de fibre optique terrestre et sous-marine. Aujourd'hui, la taille des câbles est dimensionnée pour la vidéo à la demande. Ce qui consomme beaucoup dans les réseaux c'est l'allumage en permanence.

De plus, il est nécessaire de rajouter des équipements là où il n'y avait pas de réseau, notamment des antennes qui nécessitent un nombre conséquent de câbles d'alimentation.

La croissance de la consommation des centres de données est aussi principalement expliquée par les **cartes graphiques/GPU**, spécialisées pour l'affichage d'image ou sa performance de calcul pour faire fonctionner les machines.

La fabrication de câbles GPU (circuits imprimés) nécessite des usines spécialisées, un réseau électrique, de l'eau, de la recherche marketing, de la formation de personnes. Cela peut présenter des risques pour l'environnement.

L'utilisation de l'IA présente des **bénéfices et des risques**. Elle peut permettre une optimisation de la

consommation d'énergie tout en l'augmentant ailleurs selon un **effet rebond.** Plus les technologies permettent d'économiser de l'énergie, plus la consommation de cette énergie augmente car l'efficacité grandissante développe les usages et les marchés. Le bond d'efficacité promet d'être considérable, en concevant des objets et des processus beaucoup plus économes en temps et en énergie. Mais son développement élargit la demande et conduit pour l'instant à une **pollution supplémentaire.** L'IA modifie la façon de consommer et les flux financiers ce qui entraîne un impact direct sur l'environnement.

Pour évaluer cette consommation de l'IA, on regarde les **courbes de données (**Estimating energy consumption)

On prend en compte l'équipement sur lequel on fait tourner l'IA et on multiplie par le temps durant lequel on l'utilise. En pratique, c'est un procédé plus complexe car on ne sait pas toujours sur quoi tourne l'IA et combien d'autres personnes utilisent les cartes GPU en même temps. C'est pourquoi les chercheurs se basent sur des hypothèses formulées.

La **consommation** varie d'un modèle d'IA à l'autre. Les IA généralistes avec beaucoup de paramètres vont davantage consommer par rapport à des modèles plus petits. La consommation varie aussi en fonction du **nombre d'opérations effectuées chaque seconde.** Plus le modèle est grand, plus on a d'additions et de multiplications à faire.

Il existe d'autres phases dans la vie du matériel numérique tel que le **recyclage** ou la fin de vie en décharge. Ainsi, l'évaluation d'impact prend en compte le cycle de vie du matériel utilisé.

Au vu du nombre important de **métaux** utilisés dans le processus de fabrication, il est impossible de savoir d'où chacun provient. On choisit des hypothèses de simplification qui prennent en compte les **pollutions générées lors de la phase extraction, d'usage et la fin de vie.** Chacune de ces trois phases utilise des ressources naturelles : matières premières, produits chimiques, essence (mines), eau, électricité. Chaque phase génère également des émissions et déchets, pollution des sols. Tout cela crée un impact environnemental qui va au-delà des effets de serre.

La production de métaux pour le numérique se décompose en 4 étapes : on extrait la terre, on la broie, on sépare les différents métaux et ensuite on la purifie. Cette étape de **purification** est très importante pour le numérique. En effet, les calculs effectués nécessitent une grande pureté des métaux, semi-conducteurs afin que la transmission de données fonctionne très vite.

Les entreprises en mesure de fabriquer et graver de manière assez fine les cartes utilisées par l'IA sont très **concentrées à Taiwan**. Cette centralisation **amplifie la consommation d'eau** alors que Taiwan vit des sécheresses. Cet élément entraîne un **conflit d'usage** entre l'agriculture pour nourrir les populations et les entreprises de semi-conducteurs.

L'Europe essaie de développer ses propres entreprises mais ne peut pour l'instant se passer des importations de ces métaux.

Finalement, cette technologie utilise une grande quantité d'énergie, d'eau et de produits chimiques comme les solvants.

Le défi se trouve également dans la **gestion de la fin de vie des équipements numériques.** En effet, à ce niveau, la croissance très rapide de l'IA et la concurrence entraînent un **renouvellement fréquent des serveurs.** Les grands acteurs du numérique changent régulièrement leurs GPU alors que ces matériaux sont aujourd'hui **très peu recyclables** car les cartes sont gravées très finement et regroupent un nombre important de métaux. Ainsi sur l'ensemble des équipements électroniques, seulement 22% sont correctement collectés et génèrent sept kilos de déchets par personne dans le monde.

Exponential growth for notable AI models :

Calcul pour un modèle d'IA : Les **gros modèles s**ont très couramment utilisés. Le nombre de cartes GPU utilisées pour l'entraînement des modèles suit une croissance exponentielle. Ces modèles utilisent davantage de cartes, plus complexes car plus récentes, utilisant par conséquent plus de ressources tout en émettant plus de gaz à effet de serre.

Selon une étude de Berthelot 2025, les centres de données représentent une grande partie des émissions mais il ne faut pas négliger celle des **terminaux des utilisateurs,** c'est-à-dire les ordinateurs, smartphones et réseaux

qu'ils utilisent.

Conclusion

L'évaluation des impacts environnementaux de l'IA sont aujourd'hui compliqués à suivre, en partie car il existe un **nombre important d'outils** ne calculant pas exactement la même chose. Il est pour cela nécessaire d'être vigilant quant aux chiffres qui ne sont pas toujours comparables.

Les **évaluations sont souvent incomplètes** et concentrées uniquement sur les impacts directs des infrastructures.

Beaucoup d'évaluations des chiffres proviennent de **cabinets de conseil** dont la méthode de mesure n'est pas publique. Cette **opacité** pousse à se questionner sur leur fiabilité. L'accès aux données primaires est difficile.

Un travail important de recherche est encore nécessaire pour parvenir à mesurer tous les impacts et pallier les sous-évaluations.

Finalement, l'IA est une technologie pleine de promesses dont la croissance des impacts représente le plus grand enjeu. L'IA a aujourd'hui un rôle à jouer autour des questions de soutenabilité à travers le développement de technologies plus frugales qui utilisent moins de données, des modèles plus petits. Un renforcement de la régulation est également nécessaire, questionner et débattre ses usages, comment les prioriser, autour de sa généralisation et de l'influence des systèmes de recommandation.

Questions

1. Quel data center pollue le plus ?

La pollution des centres de données dépend fortement de leur l**ocalisation géographique,** notamment du mix énergétique du pays dans lequel ils sont implantés. Par exemple, un data center en France, où l'énergie est majoritairement décarbonée, émettra beaucoup moins de gaz à effet de serre qu'un centre situé aux États-Unis. Des indicateurs comme le **PUE** (Power Usage Effectiveness) ou le **PIE** permettent d'estimer l'efficacité énergétique des data centers. Il est aujourd'hui difficile, voire impossible, pour les citoyens de choisir sur quel centre de données leurs services numériques seront hébergés, notamment lorsqu'il s'agit de géants comme Google, Amazon ou Microsoft. Cela limite grandement l'autonomie des utilisateurs sur cet aspect écologique.

2. Quels usages prioriser pour réduire l'impact écologique ?

Plusieurs pistes ont été évoquées pour adopter des usages plus responsables :

- Adapter l'IA à ses besoins réels : utiliser une IA d'image uniquement si le besoin est visuel, et éviter les modèles généralistes trop lourds quand ce n'est pas nécessaire.
- Conserver les moteurs de recherche classiques : il est préférable de continuer à utiliser des outils comme Google pour les recherches web simples, plutôt que d'utiliser des IA comme ChatGPT qui sont beaucoup plus énergivores.
- Réduire le temps de réponse attendu : à l'avenir, il sera peut-être possible de choisir une réponse plus lente, mais moins consommatrice en énergie, ce qui constituerait une avancée positive.

3. IA et pollution : quelles différences selon les modèles ?

Les modèles d'intelligence artificielle diffèrent dans leur impact écologique :

- Les modèles récents comme ceux de Mistral ou DeepSeek tendent à n'utiliser que les paramètres nécessaires, ce qui permet de réduire les calculs et donc la consommation d'énergie.
- Pour ChatGPT, on ne dispose pas d'informations précises sur le fonctionnement interne, mais il est suggéré qu'à l'inférence (c'est-à-dire à l'utilisation), l'impact est moindre. Cependant, cette faible consommation unitaire est contrebalancée par une très large utilisation.

À noter que même si l'on changeait nos habitudes du jour au lendemain, l'impact resterait limité car les infrastructures (data centers) existent déjà. À l'échelle individuelle, un usage raisonné et occasionnel, avec des requêtes bien formulées, est recommandé.

4. Réutilisation de la chaleur des data centers : une fausse bonne idée ?

Une proposition d'EDF consistait à récupérer la chaleur dégagée par les centres de données pour d'autres usages. Selon l'intervenante, cette idée n'est pas réellement bénéfique sur le plan environnemental, car elle nécessite des infrastructures spécifiques lourdes à mettre en œuvre. Cela reste donc difficilement applicable dans la réalité.

5. Vers une IA disponible 24h/24 ?

Avec la multiplication des événements climatiques extrêmes, il n'est pas garanti que l'approvisionnement électrique soit toujours stable. Des coupures d'électricité pourraient rendre les IA inaccessibles à certains moments.
Cela soulève une question importante, notamment pour les étudiants : comment se préparer à un monde dans lequel l'IA ne serait pas disponible en permanence ?

LEXIQUE

Système de crédit social : système numérique sécurisé de surveillance, de saisie et des d'évaluation qui permet de classer et évaluer les individus, fonctionnaires, entreprises, organisations et associations. Permet un recensement de "bons" et "mauvais" comportements matérialisés dans un système de récompense/sanction. Ce système de notation est basé sur le « Projet de planification pour la construction d'un système de crédit social (2014-2020) » adopté par le Conseil d'État chinois le 14 juin 2014.

Libertés fondamentales : ici entendues comme l'ensemble des droits et libertés reconnus et recensés dans plusieurs textes internationaux, depuis la Déclaration des droits de l'homme et du citoyen jusqu'à la Charte des NU de 1958 ainsi que les textes au sein de l'Union européenne. Plus spécifiquement il est question ici avec les usages liberticides de l'IA d'atteintes aux libertés d'opinion, d'expression, de pensée, de religion, de réunion, de mouvement…

Techniques subliminales : terme introduit en 1957 par le publiciste américain, James Vicary, qui prétendait qu'en insérant, dans un film, tous les cinquantièmes de seconde, des images contenant des messages, ceux-ci influencent nos comportements sans que nous en soyons conscients. Théorie scientifiquement réfutée depuis.

Supraconnectivité : néologisme de l'auteur qui par analogie avec la science physique et le concept de « supraconductivité », décrit un phénomène caractérisé par l'absence de résistance dans le flux continu d'information et d'échanges en ligne dans nos sociétés connectées Sous-veillance: terme proposé par le Canadien Steve Mann pour décrire l'enregistrement d'une activité du point de vue d'une personne qui y est impliquée, souvent réalisée par un appareil enregistreur portable. Concept qui s'oppose à celui de surveillance qui caractériserait les sociétés modernes dans une conception foucaldienne ou benthamienne.

Infox: (de info[rmation] et intox[ication]) : information mensongère, délibérément biaisée ou tronquée, diffusée par un média ou un réseau social afin d'influencer l'opinion publique ; fausse information. Recommandation officielle pour la traduction de "fake news".

Deep fake (hypertrucage): enregistrement vidéo ou audio réalisé ou modifié grâce à l'intelligence artificielle. Ce terme fait référence non seulement au contenu ainsi créé, mais aussi aux technologies utilisées. Le mot "deepfake" est une abréviation de "Deep Learning" et "Fake", qui peut être traduit par "fausse profondeur".

Indisigne : l'une des catégories sémiotiques introduites par Charles Sanders Peirce (1839-1914) au tournant du XIXe et du XXe siècle, des signes pointés sur les réalités qu'ils désigneraient. Ils donnent l'impression d'en être la trace, au même titre que les photographies s'imprègnent de la trace lumineuse des choses. Les deux autres catégories sont le légisigne, qui est défini par une convention (par ex la lettre A), ou le symbole (défini lui par une relation d'analogie entre le signe et ce qu'il désigne, ex le dessin).

Profilage : Selon le RGPD (Registre de traitement des données personnelles), traitement automatisé de données à caractère personnel pour analyser ou prédire les intérêts, le comportement

et d'autres attributs d'une personne concernée. Le profilage algorithmique est central dans le fonctionnement d'offre de contenus sur les réseaux sociaux aujourd'hui.

Orins: 1. Filin qui relie un objet immergé à une bouée. 2. Contraction de termes "organismes" et "informationnels", repris de l'anglais "inforg" (L.Floridi): désigne un organisme incarné de manière informationnelle, une entité composée d'informations, qui existe dans l'infosphère. Par extension, ensemble constitué par les moteurs de recherche, réseaux sociaux, plateformes contemporaines, devenus indispensables pour nous relier au monde pour J.G. Ganascia.

Tokens: unité de base utilisée pour analyser et traiter le texte en intelligence artificielle et en traitement du langage naturel. Les tokens facilitent la compréhension et la manipulation des textes, en les divisant en unités plus simples et plus faciles à traiter

Agentivité : Capacité à exercer un contrôle sur son environnement et ses actions.

Littératie numérique : Ensemble des compétences permettant une utilisation critique et éclairée du numérique.

IA générative : IA capable de créer du contenu à partir de données existantes.

Paresse cognitive : Tendance à se reposer excessivement sur la technologie au détriment de la réflexion autonome.

Apprentissage adaptatif : Systèmes d'enseignement utilisant l'IA pour ajuster les parcours pédagogiques en fonction des besoins individuels des apprenants.

Éthique de l'IA : Étude des implications morales et sociétales des algorithmes et de l'intelligence

Biais algorithmique : Distorsion des résultats d'un modèle d'IA due à des biais présents dans les données d'entraînement.

Compétences transversales : Capacités mobilisables dans divers contextes professionnels et académiques, telles que la pensée critique et la collaboration.

Automatisation pédagogique : Utilisation de l'IA pour générer du contenu éducatif, évaluer les étudiants et assister les enseignants dans leur travail.

IA explicable : Développement d'algorithmes d'IA dont les décisions et les processus sont compréhensibles par les humains.