



Centre Catholique International
de Coopération avec l'UNESCO

Comment éduquer à un usage responsable des intelligences artificielles ?

Actes du cycle de tables-rondes 2024



INTRODUCTION

Depuis environ deux ans, « **Disruptif, Dystopie** », ces deux mots sont propulsés sur le devant de la scène. **Disruptif**, du latin : rompre, qualifie bien le fait qu'en étant livrée au grand public, du jour au lendemain, de façon gratuite, et avec une grande facilité d'utilisation, la publication par Open AI d'un outil d'intelligence artificielle générative généraliste, venait rompre avec l'évolution, certes rapide, des technologies mues par l'intelligence artificielle. En produisant des contenus dont on ne pouvait de prime abord discerner s'ils étaient issus d'une machine ou d'un humain, ChatGPT provoquait une véritable onde de choc qui dépassait largement le monde scientifique. Toute personne un peu curieuse se mit à le tester, lui posant 1000 sortes de questions, à toutes sortes de fins, et il n'échappa à personne qu'une révolution anthropologique s'amorçait, soulevant de nombreuses questions éthiques.

Génératives ou non, les intelligences artificielles évoluent à un rythme effréné, pour le meilleur en risquant le pire d'où ce deuxième mot de **dystopie**. Le fantasme récurrent de voir la machine dominer l'être humain, de Pygmalion à nos jours, voire même l'asservir, revient hanter nos esprits. Il y a bien matière à réflexion, et à réflexion Éthique.

En adoptant fin novembre 2021 la Recommandation sur l'éthique de l'[Intelligence artificielle](#), les 193 membres de l'Unesco saluaient le premier instrument normatif international sur l'éthique de l'intelligence artificielle.

Mettre l'humain au cœur du débat, l'Homme au cœur du monde, tel est le fil conducteur des réflexions, événements, contributions du Centre Catholique International de Coopération avec l'UNESCO (CCIC), en particulier dans son dernier forum international « Métamorphose du monde, jusqu'où l'Homme peut-il changer l'humain ? Quelle boussole pour l'éducation ? » sous le haut patronage de l'UNESCO. Il participait à la réflexion qui précédait l'élaboration de la recommandation de l'UNESCO, ainsi qu'à l'initiative UNESCO des « Futurs de l'éducation », et celle du Pape François du « Pacte éducatif mondial ». Les actes en ont été publiés, dont la version numérique est disponible sur le [site](http://site.CCIC-unesco.org) CCIC-unesco.org .

Le CCIC confirmant son désir de coopération souhaite par les travaux de ce nouveau projet s'associer à l'UNESCO dans le cadre global des Objectifs de Développement Durable 2030, notamment l'ODD4, celui d'une éducation de qualité pour tous. Ce projet propose une approche de la **mise en œuvre** de la recommandation de l'UNESCO, focalisée sur le Domaine Stratégique numéro 8, celui de l'éducation, ce en co-crédation avec la Commission Nationale Française pour l'UNESCO. Il rejoint donc de façon transversale le secteur des Sciences Humaines et Sociales et celui de l'Education, et s'attache à donner des pistes concrètes pour développer des attitudes et compétences autres que techniques pour l'usage des IA.

Le propos n'est donc pas celui de la gouvernance, nous nous intéressons à la personne humaine en tant qu'utilisatrice des outils de l'IA, au questionnement éthique lié à la personne humaine, et plus précisément à l'éthique de responsabilité, notion chère au philosophe Hans Jonas.

Riche de l'expérience de terrain des ONG qui le composent, le CCIC s'est appuyé sur les questions et les témoignages qu'elles ont recueillis auprès des personnes qu'elles accompagnent sur les cinq continents, et l'expérience d'experts d'horizons variés : ingénieurs, éducateurs, philosophes, neurologues, éthiciens. Le Père Thierry Magnin, vice-recteur de

l'Université Catholique de Lille et Monsieur Christian Byk membre de la Commission Nationale Française pour l'UNESCO ont accompagné le travail de réflexion.

L'objectif est de sensibiliser les jeunes et les moins jeunes à une utilisation des IA, éthique, pratique et spirituellement inspirée, par l'éducation. En mettant l'accent sur la sensibilisation et la formation des familles, des éducateurs et des enseignants, nous pouvons renforcer dans l'éducation la place de l'éthique de responsabilité dans l'utilisation des IA, en intégrant les dimensions de discernement et d'intériorité inspirées par la pensée de Teilhard de Chardin. Cette approche holistique favorisera un environnement éducatif où la technologie sera utilisée de manière responsable et éthique, contribuant au développement humain intégral des individus et de la société.

Cet ouvrage est la trace écrite du cycle de trois tables rondes qui s'est tenu en mode hybride dans trois tiers -lieux en lien avec une démarche éducative, les mardis 12 et 19 novembre et 3 décembre 2024

1^{ère} partie : « Ethique et IA, De quoi parle-t-on ? » Partant de l'historique de l'évolution du numérique, quelques aspects techniques sont abordés pour éclairer certains enjeux, et sont précisées plusieurs définitions dans le contexte du numérique, et celle de l'éthique, en particulier l'éthique de responsabilité

2^{ème} partie : « Comment éduquer à l'éveil de la conscience pour un usage éthique de l'IA ? » A l'heure du numérique, comment diriger notre vie ? Face à l'omniprésence de l'IA, il est important de chercher par l'éducation à développer et renforcer l'éveil de la conscience et de l'intériorité chez les jeunes comme les moins jeunes.

3^{ème} partie : « Repères et pistes pour éduquer » : Points de vigilance : rapport à la vérité, à l'exactitude scientifique, à l'identité, à l'estime de soi, aux aptitudes relationnelles, à la responsabilité individuelle et collective, au monde réel, à la liberté ...-pistes et exemples de formation à un usage éthique des IA, outils pédagogiques »

*« L'immense expansion de la technologie doit donc s'accompagner d'une **formation appropriée à la responsabilité** dans son développement... Nous avons donc le devoir d'élargir notre regard et d'orienter la recherche technico-scientifique vers la paix et le bien commun, pour le service du développement intégral de l'homme et de la communauté (...). Ce n'est pas la responsabilité d'un petit nombre, mais de toute la famille humaine ». Pape François, message pour la journée de la Paix, le [1er janvier 2024](#).*

Depuis 1947, le Centre Catholique International de Coopération avec l'UNESCO est une plateforme d'ONG internationales d'inspiration catholique. C'est une ONG partenaire officielle de l'UNESCO. Lieu de réflexion collective et force de proposition au sein de la société civile à l'UNESCO, elle se fait l'écho des missions et des expériences de terrain des ONG membres de la plateforme, et des millions de personnes qu'elles accompagnent dans plus de 140 pays.

Première partie :

Ethique et IA, de quoi parle-t-on ?

Isabelle Chaperon, Présidente du CCIC

Il me tient à cœur de remercier particulièrement nos hôtes pour cette première table-ronde, la communauté éducative Saint François-Xavier, qui crée et anime des écoles, collèges, et lycées et centre d'étudiants en France, en Côte d'Ivoire, au Tchad, au Togo et en Corée. Nous sommes très reconnaissants au CFRT- Jour du Seigneur, ainsi qu'à son équipe de techniciens, d'assurer la captation et la diffusion de ces trois tables-rondes. Elles sont l'aboutissement des travaux de toute une équipe, et des réseaux des ONG du CCIC. Qu'ils soient ici chaleureusement remerciés, ainsi que les intervenants qui nous font l'honneur de nous éclairer.

Eric Castelnau, Modérateur -Vice-Président du CCIC

Aujourd'hui, nous proposons d'explorer quelques bases :

-Comment fonctionnent les différents types d'IA ? : Certains mécanismes de l'intelligence artificielle (IA) et du cerveau humain nécessitent d'être connus et compris pour entrevoir les enjeux éthiques et anthropologiques.

-Quel est le sens de l'intelligence artificielle ?

-Y a-t-il un sens, et surtout, qu'est-ce que l'éthique de la responsabilité ?

Nous avons la chance d'accueillir ce soir des scientifiques, juristes et un entrepreneur, qui amènera un angle à la fois scientifique et entrepreneurial.

Et pour commencer, permettez-moi d'introduire notre première intervenante, Isabelle Chaperon. Vous l'avez peut-être vue dans son rôle de présidente, mais ici, elle interviendra en tant qu'ingénieure, formée à l'ENSTA à Paris, titulaire d'un Master of Sciences de Stanford.

1-L'IA : de quoi parle-t-on ? et comment ça marche ?

Isabelle Chaperon

L'objectif de cette introduction un peu technique est de faire émerger ce qui dans l'histoire et le fonctionnement des IA doit nous faire réfléchir à un usage responsable de ces outils.

Un rêve

Intelligence artificielle, ce terme apparaît en 1956, lors d'une conférence à Dartmouth, au NE des États-Unis. Elle rassemble quelques experts en informatique. L'IA naît d'avancées techniques, mais surtout du rêve exprimé par ces chercheurs, celui de remplacer l'intelligence humaine. L'idée de faire exécuter par une machine des tâches compliquées, fastidieuses, répétitives, existait déjà. Par exemple au 18ème siècle, Blaise Pascal -que je considère comme un génie eu égard à la diversité des domaines dans lesquels il excellait- avait dans sa jeunesse inventé une machine pour soulager son père dans la levée des impôts qui nécessitait des calculs complexes. Ce fut la Pascaline.

Je reviens au 20ème siècle, deuxième guerre mondiale : une avancée spectaculaire fut la conception et la réalisation par l'Anglais Alan Turing d'une énorme machine mécanique programmée pour tenter de décrypter le code Enigma qu'utilisaient les Allemands pour coder leurs messages. Le code changeait chaque jour. Les possibilités de chiffrement étaient innombrables ce qui rendait la tâche impossible en une nuit. In fine, l'idée géniale d'une femme de l'équipe fut de suggérer que les derniers mots des messages puissent être toujours les mêmes, comme une formule de politesse, et ceci permit d'économiser du temps de calcul. Grâce à cela, les messages secrets Allemands furent déchiffrés, à leur insu, facilitant un tournant décisif de cette guerre.

Convergence, progrès très rapides

L'informatique se développe rapidement, permet de programmer des calculs complexes et fastidieux, les chercheurs enthousiastes obtiennent de gros crédits, en proportion de leurs promesses. Les télécommunications se développent. Les ordinateurs peuvent communiquer entre eux et travailler en réseau, c'est l'arrivée d'internet. Les processeurs mécaniques au départ deviennent électromécaniques, puis électriques, électroniques. On a maintenant des micro-processeurs de plus en plus minuscules et l'on explore les possibilités de processeurs quantiques. Les sciences, la physique, les mathématiques, les neurosciences progressent aussi à vive allure, et c'est la convergence de toutes ces techniques qui a permis les progrès fulgurants du numérique, bien qu'il eût à subir deux « hivers », dûs à quelques espoirs déçus faute de puissance de calcul, voire de méthode.

Apparition progressive de plusieurs types d'IA

Nous n'avons pas choisi par hasard un pluriel pour les IA dans le titre de ce cycle, car il en existe de plusieurs natures. La plus récente, l'IA générative qui impressionne le plus le grand public est d'une nature très différente des précédentes et pose plus de questions.

L'IA symbolique : l'ordinateur exécute une suite d'ordres, de calculs, l'information est traitée de manière séquentielle au service d'un modèle numérique qui simule un phénomène physique. Le terme symbolique reflète le fait qu'une réalité concrète soit représentée par des chiffres, des symboles, ou des lois mises en équations.

Citons par exemple les modèles numériques de prévisions météorologiques : à partir de mesures, et des équations qui régissent les vents, températures, pressions atmosphériques (lois physiques de la thermodynamique), on peut prédire la météo du lendemain.

L'IA comportementale : Je rappelle que l'IA est née d'un rêve, celui de remplacer l'intelligence humaine, et cela a marqué l'inconscient collectif. Or sens, conscience, pensée, émotions humaines s'avèrent inaccessibles à l'ordinateur, un tournant s'opère alors pour simuler, ou imiter des comportements, pour remplacer l'intériorité mentale par le comportement du vivant. On peut par exemple programmer un robot muni de capteurs. Des caméras peuvent imiter le sens de la vue. Pour autant, l'imitation du vivant ne fonctionne pas pour tout : l'avion ne fonctionne pas en battant des ailes comme un oiseau !

L'IA statistique : C'est une autre approche, qui utilise des statistiques effectuées sur une base de données, dont on déduit des probabilités (par exemple ce que vous seriez intéressé d'acheter, de regarder, en fonction des statistiques d'achats de l'ensemble des clients, ou les vidéos qui devraient vous plaire sur YouTube), et on vous le propose, voire vous l'impose par le biais de

nudges, ces petits coups de pouce qui vous poussent à acheter, à décider. L'IA statistique fonctionne aussi déjà par apprentissage, mais sans commune mesure avec l'IA générative.

L'IA générative : L'événement disruptif fin 2022 : c'est ChatGPT, livré au grand public, du jour au lendemain, de façon gratuite et illimitée, sans nécessité d'équipement spécifique autre qu'un portable dont l'accès était déjà largement répandu, et avec une grande facilité d'utilisation. La publication par Open AI de cet outil d'intelligence artificielle générative vint rompre avec l'évolution, certes rapide, des technologies mues par l'intelligence artificielle. En produisant des contenus dont on ne pouvait de prime abord discerner s'ils étaient issus d'une machine ou d'un humain, ChatGPT provoqua une véritable onde de choc qui dépassa largement le monde scientifique. Alors qu'il avait fallu plus de deux ans à Netflix ou Twitter pour atteindre 1 million d'utilisateurs, ce fut fait en 5 jours par Chat GPT.

GPT : jusque-là, GPT signifiait *General Purpose Tool*, un outil à tout faire. Clin d'œil sans doute, car Chat-GPT est un agent conversationnel (Chat) qui donne accès à un *transformer* (T) pré-entraîné (P), génératif (G). En d'autres termes, il s'appuie sur une phase d'entraînement préalable où des myriades de paramètres ont été optimisés à partir d'encore plus de myriades de données trouvées sur internet, pour générer du texte. *Transformer* correspond au plongement lexical : les mots sont traités comme des vecteurs de nombres qui intègrent le contexte.

Les réseaux de neurones artificiels

Chat-GPT est l'outil qui a révélé au grand public l'IA Générative, il en existait d'autres mais peu connues du grand public. On l'a vu, les IA précédentes échouaient à traiter correctement le langage humain, les traductions, la reconnaissance d'images autres que préenregistrées. La convergence des divers domaines scientifiques permit à l'approche neuronale de faire une percée spectaculaire. Le fonctionnement des neurones stimulés par les synapses avait depuis des décennies inspiré les premiers perceptrons, mais la puissance de calcul des ordinateurs n'était alors pas au rendez-vous. Chat-GPT s'appuie sur un « grand modèle de langage » : *Large Language Model* ou LLM en anglais. D'autres modèles répondent à d'autres types de tâches, comme DALL-E, ou Midjourney pour générer des images, etc...

Fonctionnement et limites.

Mon objectif ici est de vous donner un aperçu de comment fonctionnent les IA génératives pour vous en faire entrevoir les limites. Elles nécessitent un entraînement ou apprentissage conséquent plus qu'une programmation logique, et s'appuient sur des réseaux de neurones artificiels.

Prenons l'exemple d'un modèle simple d'apprentissage profond- On dit qu'il est profond lorsqu'il comporte plusieurs couches cachées de neurones, ce qui est le cas par exemple dans le traitement automatique des langues ou la reconnaissance de la parole. Voici le cas simpliste de la reconnaissance du chiffre neuf manuscrit : sa photo est entrée dans l'ordinateur sous forme d'un vecteur de 784 variables qui correspondent à l'intensité de chacune des 28x28 pixels de sa photo.

Imaginons un modèle à deux couches cachées de 16 neurones chacune, le modèle est donc composé à l'entrée de 784 neurones artificiels porteurs des luminosités de chacune des 784 pixels de l'image à décrypter, qui transmettent chacun un « influx » vers chacun des 16 neurones suivants, qui eux-mêmes stimulent chacun plus ou moins chacun des 16 neurones de la 3ème couche, qui eux-mêmes influent chacun sur les 10 derniers neurones. Idéalement, on espère

qu'un seul de ces 10 neurone ait été activé pour donner le bon résultat, qui pour notre cerveau humain correspond évidemment au chiffre 9. La façon dont le signal se propage, à l'image des synapses du cerveau dépend de « poids » affectés à chaque intensité de transmission entre les neurones de chacune des couches. On calcule alors la fonction de « coût » qui est l'écart entre le résultat attendu et celui obtenu par la machine. Cette fonction dans notre exemple comporte comme variables les luminosités des pixels et comme coefficients les « poids ». C'est une simple fonction linéaire, mais elle a 784x16x16x10 coefficients. Là interviennent les mathématiciens : il s'agit de trouver quels poids assurent une valeur la plus faible possible de cette fonction d'erreur. Cela nécessite un entraînement sur encore bien plus de 13 000 échantillons, ou données d'entraînement. C'est ce qu'on appelle l'apprentissage, ou l'entraînement, qui fait partie de la mise au point de ce type d'IA, qui a recours à l'apprentissage profond. Les résultats obtenus sont impressionnants... mais non explicables.

Données numérisées : En bref, dans la phase d'apprentissage, on utilise des données qui peuvent être de natures diverses, qui ne peuvent être traitées que sous forme numérisée. Par exemple le texte, le son, l'image, le code sont convertis en paquets de 0 et de 1. Il est toutefois illusoire de vouloir traduire, c'est-à-dire numériser, la douleur, la saveur d'une pomme, le toucher du velours ou la beauté d'un nuage. J'entends par-là souligner que les données ne sont qu'une représentation, assez pauvre, d'une réalité. René Magritte l'exprimait très bien dans sa peinture surréaliste : il donnait à voir et à comprendre la différence entre une pomme et l'image d'une pomme. Dans l'éducation, il est important de faire prendre conscience que toute numérisation est réductrice. Ces données numérisées sont stockées dans des data centers, très gourmands en énergie, et en eau, et utilisées pour ajuster tous les « poids » dont on a parlé. C'est la base de l'apprentissage automatique, éventuellement supervisé (si les données sont étiquetées), ou renforcé (si une solution exacte est « récompensée »).

La phase opérationnelle : l'utilisation. Dans la phase opérationnelle de l'outil, l'utilisateur rédige une requête en langage naturel, *prompt* en anglais, fournit éventuellement des données d'autre nature (photo, vidéo, texte, son...) Tout ceci est analysé, et produit presque instantanément une réponse sous la forme choisie dans la requête : texte, image, son, vidéo, code, etc... (On entend souvent que l'IA peut fournir des données, en réalité elle fournit un produit d'IA qui peut servir de données mais dont la fiabilité n'est pas acquise). Cette rapidité, l'immédiateté de la réponse, met à mal chez les utilisateurs l'expérience du temps long, celui de la recherche, de la réflexion. Savoir rédiger une requête est tout un art auquel il convient de former les utilisateurs. Selon la qualité des *prompts*, l'IA peut générer des réponses insensées, absurdes, qu'on qualifie d'hallucinations, comme cette colombe à deux têtes et ses congénères à trois langues que j'ai obtenues après trois requêtes d'une image de colombes volant dans la même direction un jour où je concevais une carte de vœux avec l'IA intégrée dans un logiciel de graphisme. Dans l'exemple de Chat GPT, le texte produit est dans un langage qui fait illusion, il répond parfois des absurdités avec un aplomb déconcertant, et cette illusion n'est pas sans conséquences sur l'usage éthique qu'on peut en faire.

L'économie de l'attention, l'addiction

La performance des modèles d'IA Générative dépend très largement de la ressource en données d'entraînement, et du nombre de « poids » mis en œuvre. De GPT1 en 2018 à GPT4 en 2023, le nombre de « poids » ou coefficients est passé schématiquement de 10 puissance 8 à 10 puissance 14, soit un nombre à 15 chiffres. Ce qui importe pour mon propos est de mettre en évidence la quantité phénoménale de données nécessaires à de tels calculs, essentiels pour

l'élaboration du modèle. Et parmi les conséquences, la gourmandise, la voracité, l'impérieuse nécessité pour le concepteur d'amasser des données.

En particulier, plus l'utilisateur passe de temps sur une application, plus son fournisseur enregistre de données. La nécessité pour le fournisseur d'IA d'accéder à toujours plus de données le conduit à assortir ses produits d'une composante addictive, de la même façon que les cigarettiers ont toujours mêlé des produits addictifs au tabac des cigarettes. C'est ce qu'on appelle l'économie de l'attention, essentielle pour la publicité en particulier, qui est vitale pour leur modèle économique. Par exemple YouTube démarre des vidéos non sollicitées dès la fin de celle qu'on a choisi de visionner, on se laisse bercer, éventuellement pendant des heures. C'est l'origine d'addictions avérées, qui peuvent être gravissimes, à tout le moins chronophages, au détriment d'activités qui construiraient la personne dans toutes ses dimensions, de son insertion dans la société, et de sa réflexion propre.

L'usage responsable des IA passe par un questionnement éthique

Je vous livre en conclusion quelques-unes de ces questions :

- Quelle vérité peut résulter d'une machine qui ne fait que relier de façon probabiliste des mots sans en pénétrer le sens ?
- Quelles peuvent être les conséquences de l'illusion que donne l'IA générative d'une pensée et d'un travail humain ?
- Quelle est l'honnêteté intellectuelle d'une IA qui ne peut citer ses sources (la myriade de données d'entraînement), et de l'utilisateur qui ne mentionnerait pas que tout ou partie de son travail est le produit d'une IA ?
- La machine dispense-t-elle de penser par soi-même ?
- Ai-je conscience de « liker ma servitude » ? (Cette expression empruntée au titre d'un livre de Louis de Diesbach parle d'elle-même)
- Les mécanismes propres aux IA qui sous-tendent les réseaux sociaux n'amplifient-ils pas les discours de haine et autres fake news ?
- Réagir et partager sur les réseaux sociaux des informations non vérifiées est-ce responsable ?
- L'impact environnemental d'une IA générative n'inciterait-il pas à en faire un usage sobre ?
- Ne perdons-nous pas nos capacités à interagir, débattre, savoir accepter le dialogue et l'inattendu des autres à force d'abuser d'un tête-à-tête avec des IA ?

Et plus largement :

- Mon être se réduit-il à mes comportements ?
- Qui décide de mes comportements, mes choix, ma pensée ?
- Comment préparer un jeune à être le pilote de sa propre vie, à être responsable de ses actes, des conséquences de ses actes ?

Autant de questions qui restent ouvertes, que je laisse à votre réflexion...et à la deuxième table-ronde de ce cycle

2- L'IA, les neurosciences et l'éthique

E. Castelnau : Avant de plonger dans les détails éthiques, explorons d'abord la notion de responsabilité face à ces technologies.

Laure, vous êtes éthicienne, neuroscientifique, PhD, chercheuse à Paris -Saclay en neurosciences, cheffe de projet en biomédecine et déontologue au laboratoire de l'INSERM, et professeure à l'EDHEC.

Laure, expliquez-nous ce qu'est l'éthique, et de votre point de vue d'éthicienne, jusqu'où peut-on aller ? Peut-on se limiter à une IA sobre, comme l'évoquait Isabelle, faut-il en limiter l'usage, le peut-on seulement, et quel est le nouveau rapport entre l'élève et le maître- Vous avez fait un TedX sur le sujet- pouvez-vous nous expliquer votre vision de chercheuse et de professeure ?

Laure Tabouy

Bonjour à tous, merci de cette opportunité de vous partager mon expérience, mes ressentis, mon point de vue par rapport à ces questions

Je vais me focaliser comme en a parlé Isabelle sur la convergence neurosciences, informatique, ingénierie car c'est sur ces thématiques que je travaille particulièrement en ce moment. En effet je suis neuroscientifique, donc je rentre par la porte des neurosciences, et de mon point de vue, celui d'une personne qui a passé 17 ans sur les paillasse à triturer des cerveaux. J'ai cet axe scientifique. Je vais aborder l'axe éthique, et je vais laisser Monsieur Byk parler de l'aspect juridique.

Mon point de vue d'éthicienne aujourd'hui, est que l'on a plein de choses dans la tête : lire dans les pensées, décrypter les rêves, communiquer, écrire par les pensées, télécharger le contenu du cerveau dans une machine, et on a une frontière entre un monde fictionnel, de nos œuvres fictionnelles et un monde réel, la réalité, et ces deux mondes se frictionnent. La frontière entre fiction et réalité est de plus en plus poreuse, cela pourrait éventuellement un jour disparaître- mais c'est au conditionnel.

La question qui me taraude c'est qu'est-ce que nous construisons ? Quel monde voulons-nous ? Je ne vais pas vous redire l'histoire de la convergence, Isabelle en a déjà parlé, mais insister sur la convergence neurosciences-numérique depuis les années 50, avec une croissance exponentielle, vertigineuse. Tout ce qu'on voit aujourd'hui, c'est parce qu'on y travaille depuis les années 40, et même avant. Les neurosciences en tant que discipline académique sont nées entre les années 50 et 60 vraiment aux confins de la cybernétique, née dans les années 40-50 avec l'IA.

L'idée de décoder le cerveau, la pensée dans les neurosciences, va de pair avec les avancées des neuro-technologies, qui permettent d'explorer le cerveau de manière nanométrique, de manière beaucoup plus précise, plus profonde.

L'IA, le *machine learning*, le *deep learning*, c'est pour moi une ouverture sur la façon dont le cerveau fonctionne. Au départ elle n'est pas née de la volonté de simuler le cerveau, mais plutôt d'une volonté de comprendre comment les synapses et les neurones fonctionnaient, et imiter comment l'information allait de neurone en neurone. Le cerveau, en soi, c'est une boîte noire, un circuit électrique fermé, le plus performant du monde avec 86 milliards de neurones et autant de cellules gliales, c'est mille milliards de synapses à la seconde, un circuit électrique hors norme.

Cette boîte noire est très mystérieuse, parce que, en rapport avec l'histoire de la philosophie, c'est une porte vers ce que l'on est, l'humanité, l'Homme, la pensée, la conscience. En tant que chrétien c'est aussi une porte vers la foi, vers une relation autre, la place que l'on donne à Dieu. C'est un organe du corps humain, le plus mystérieux. Il est vrai que la course à l'accès à la pensée, au comportement, à la voix, à la vision, c'est un tout, qu'il faut prendre de façon beaucoup plus large. Ce qui est source d'inquiétude, c'est l'ensemble, c'est l'accélération, pas seulement l'IA générative mais cet ensemble.

La perspective 2024-2025 est assez excitante et en totale ébullition, Monsieur Byk en parlera sans doute.

Depuis 2019, 2021 on a des recommandations de l'OCDE sur l'IA et les neuro-technologies sont tombées ; en France, on a une charte des neuro-technologies responsables inaugurée en 2022, c'est la première au monde ; et à l'UNESCO on a différentes recommandations en 2021, 23, sur l'éthique de l'IA, bientôt sur l'éthique des neuro-technologies. En 2025 sonne aujourd'hui un travail collectif mondial sur les textes sur les neuro-technologies, la relation neuro-sciences-IA et IA.

Les données numériques, sont la problématique d'aujourd'hui. On a certes du fait de cette convergence IA et neurosciences, des données personnelles de toutes sortes, mais se présente maintenant le sujet des données cérébrales, elles sont en cours de discussion, de qualification. Elles sont peu protégées, sauf dans le domaine médical, et sont un véritable enjeu actuel sur tout ce que l'IA capte par le biais de ces neuro-technologies.

Ça va de pair avec la manière dont on appréhende toutes ces données et dont on les fusionne. Elles peuvent être personnelles, génétiques, médicales, et la manière dont on fonctionne, on réagit, on les préserve, est important car l'enjeu aujourd'hui et demain c'est la fusion de toutes ces données, pour encore plus d'accélération. D'où l'importance de se poser la question de ce que l'on fait de toutes ces données-là.

L'éthique et le digital interfèrent avec le droit. Un nouveau domaine émerge : le droit du cerveau, les neuro-droits, mélange de réflexion entre l'IA et les neurosciences, cette convergence. Quatre nouveaux droits arrivent en discussion aujourd'hui : le droit à la liberté cognitive, le droit à la vie privée mentale, le droit à l'intégrité mentale, le droit à la continuité psychologique. Ce sont des discussions actuelles.

Dans le cadre de l'éthique en France, on a légiféré par le biais de la loi relative à la bioéthique, du 2 août 2021 qui commence à ouvrir la réflexion sur les neurosciences, l'IA, les algorithmes, les neuro-technologies, les big data avec ses articles 17, 18 et 19, qui ouvrent le champ de nouveaux droits, de nouvelle éthique de données cérébrales, de fusion de données de big data/ IA.

En tant qu'éthicienne c'est autour de cette convergence des neurosciences, de l'informatique, de l'ingénierie et des technologies qu'il est intéressant de s'interroger.

L'enjeu est de savoir où se placer dans cet équilibre et de s'interroger : quelles sont les opportunités, les futurs souhaitables, les dangers et les risques ; plein de promesses, un équilibre entre bénéfice et risques, qu'il faut gérer. C'est aussi une invitation à aller plus loin que la simple utilisation pour se focaliser davantage sur le comprendre, comment ça marche, d'où ça vient, quelle est son histoire, pourquoi on en est là aujourd'hui, plutôt que le savoir instantané, sur simplement ce que l'on a dans les mains, et ce que l'on peut regarder sur internet sans critique.

Et pour moi, c'est m'appuyer sur l'éthique, sur cette boussole : c'est ce qui est important pour moi, qui me permet de rester chef d'orchestre de ces innovations.

Pour répondre à la question d'Éric : qu'est-ce que l'éthique pour moi ? Pour moi c'est une boussole.

Je m'appuie beaucoup sur une phrase de Hans Jonas de 1978 du principe responsabilité : « Agis de telle sorte que les effets de tes actions soient compatibles avec la pertinence d'une vie authentiquement humaine sur terre », et sur la phrase de Paul Ricoeur : « L'éthique est le mouvement même de la liberté qui cherche une vie bonne, dans la sollicitude envers autrui et dans un juste usage des institutions sociales ».

Pour moi c'est vraiment l'homme qui est responsable de ses actes et décisions, d'orienter ses actions vers le bien commun, le bien agir. L'éthique, ce n'est pas un ensemble d'interdits, d'obligations, de bien, de mal, le devoir de chacun de considérer l'impact de ses actions dans le futur, une invitation à la réflexion, à prendre de la hauteur, à voir les perspectives, les tensions que ça implique, les dilemmes... ça va beaucoup plus loin que ce que l'on voit de prime abord comme bien ou mal. Ce questionnement va utiliser la philosophie, l'histoire des sciences, de la médecine, la bioéthique, l'anthropologie, la sociologie, pour alimenter la réflexion, clarifier.

Il faut dire aussi que toutes ces questions de neurosciences, de l'IA aujourd'hui dépendent des cultures, dépendent aussi du pays d'où l'on vient. Chacun, chaque culture a sa définition du cerveau, chaque culture a sa définition de l'IA, chaque culture a aussi sa perception des choses et c'est très important de mesurer à quel point les cultures sont importantes dans la réflexion et dans ce questionnement. On peut ne pas être d'accord, mais pour moi l'éthique ne donne pas de réponses, on n'est pas là pour donner une réponse d'éthique, on est là pour vraiment ouvrir une voie de réflexion, mettre en dialogue, mettre en tension et permettre d'identifier les questions importantes, pour pouvoir agir, puis basculer peut-être vers la morale, le droit, et légiférer ensuite.

Donc l'éthique pour moi est un pivot réflexif entre une démarche normative et une démarche plus politique, c'est vraiment le pivot réflexif pour articuler les deux.

Et dans tout cela, la question de l'IA et des neurosciences est très importante. Depuis 2002 on a une nouvelle discipline, jeune, la neuro-éthique, née en 2002 et qui va ouvrir la réflexion des neuroscientifiques de manière plus spécifique. C'est par ce biais-là, à travers le questionnement sur la responsabilité, la vulnérabilité, la vie privée mentale, l'accessibilité, la justice sociale, qui couvrent effectivement toutes ces questions de neurosciences et l'IA, et de cette convergence, qu'aujourd'hui on ouvre toutes ces questions de l'UNESCO et de l'OCDE. Et c'est très intéressant.

Donc finalement cette convergence, les progrès depuis 1950, seront un outil puissant au service de l'humain, du citoyen, de son bien-être, et permettent de mieux percevoir dans le cas des neurosciences le cerveau. Ils génèrent de grosses bases de données mais génèrent aussi des inquiétudes dans la question du cerveau. Va-t-on découvrir la signature neuronale de la conscience ? C'est une vraie question, qui donne du poids à cette convergence. L'IA questionne aussi l'approche philosophique très importante de la conscience, mais on voit que ça ouvre de nouveaux dilemmes sur les droits de l'homme, sur les droits fondamentaux. Je vais passer la parole à Christian Byk sur ce sujet, cela remet aussi en question le principe juridique du libre arbitre, donc le fondement de la responsabilité juridique mais aussi de la responsabilité éthique. Cela nécessite aussi d'ouvrir à la réflexion sur ces enjeux : réflexion citoyenne, réflexion dans la recherche, car il y a là une réelle nécessité, et puis de faire le lien entre la recherche et le citoyen,

une conception aussi d'un nouveau système interdisciplinaire et de mise en place de cadres de gouvernance, c'est ce qui se passe avec l'UNESCO, avec l'OCDE et la charte française des neuro-technologies responsables. Le fait est qu'on travaille déjà vraiment sur ces cadres de gouvernance. On navigue entre espoir et crainte, il faut vraiment se dire que c'est porteur de promesses, donc il ne faut pas tout jeter. Il est important de savoir aussi se réjouir de ce qui se passe aujourd'hui, et en même temps, peut-être qu'en tant que chrétiens, nous pouvons aussi nous poser la question de savoir comment nous avons une vision chrétienne de l'Homme. Comment ma foi- je fais référence à une parabole- comment la foi peut être aussi le levain dans la pâte, comment elle peut aussi donner de la consistance à tout ce que nous voyons, à tout ce que nous avons aujourd'hui entre les mains et considérer qu'à travers cette réflexion chrétienne on peut aussi basculer et voir toutes ces opportunités, ces promesses, ces limites et ces dangers, comment je peux avertir, et aussi comment en parler.

Voilà, pour moi, le mot de la fin, c'est une nécessité de définir et d'anticiper la responsabilité de chaque acteur et nous sommes tous responsables. Nous avons une responsabilité individuelle face à toute cette technologie, à l'IA, aux neurosciences, et je pense qu'il est très important de se dire aussi que nous sommes tous responsables à notre niveau, à notre manière, chacun avec notre propre façon d'agir, là où nous en sommes, et d'évaluer l'impact et l'influence de chacune de nos actions et d'évaluer cet équilibre bénéfice/risque.

3-Promouvoir les savoirs et l'éducation, au cœur de la mission de l'UNESCO

E. Castelnau

Merci Monsieur Byk d'être parmi nous. Je ne peux pas lister les nombreux postes que vous occupez ou avez occupés. Vous êtes magistrat honoraire donc vous allez nous parler du droit. Vous êtes membre de la Commission Nationale Française pour l'UNESCO, et c'est en co-crédation avec elle que nous avons élaboré ces différentes tables- rondes. Je rappelle que vous avez présidé le CIGB, le Comité Intergouvernemental de Bioéthique de l'UNESCO, dans lequel Laure intervient également, et que vous êtes Secrétaire Général de l'Association Internationale Droit, Ethique et Science. Je vois deux sujets qui nous intéressent : ce que fait l'UNESCO en matière d'éducation et d'IA responsable - et vous avez beaucoup travaillé sur ce sujet- et deuxièmement, quels sont les dangers de ces technologies dont le droit peut et doit sans doute nous protéger. Par exemple comment préserver plus largement le droit à l'image, de la propriété intellectuelle... C'est à vous...

Christian Byk

Merci pour cet échange, qui est le fruit de trois années de coopération souhaitée par la Commission Nationale Française pour l'UNESCO avec l'équipe du CCIC, que je remercie pour cet engagement et sa perspicacité.

1-J'aimerais commencer ma réflexion en rappelant les trois exemples qui ont été donnés : la Pascaline, merveilleuse machine destinée à mieux faire rentrer l'impôt, l' « ordinateur » mis au point par le mathématicien Alan Turing pour décoder les messages que les Allemands, au cours de la Seconde Guerre mondiale, envoyaient grâce à Enigma, et enfin tout ce qui sert à nous faire consommer plus. Telle est, en effet, la réalité que l'histoire récente et fulgurante de la technoscience a construite. Elle n'est pas une fiction mais l'emballement d'un monde qui vit au rythme de la devise olympique « plus vite, plus haut, plus fort », dont on rappellera qu'elle est

l'œuvre d'un prêtre qui lui donnait un sens chrétien. Emballés dans une course de vitesse à produire et consommer, nous sommes les sujets d'un processus d'innovation à grande échelle, qui influe sur nos désirs pour nous vendre de nouveaux produits et services. C'est ainsi que se présente aujourd'hui la société techno-industrielle et, en particulier l'IA.

2-Pour savoir à quoi sert le droit dans ce « tourbillon de la vie », il convient préalablement de se demander si la manière dont nous le produisons et les usages que nous en faisons gardent la même pertinence à travers l'histoire. Des questions plus ciblées, s'agissant de l'UNESCO seraient : pourquoi a-t-on besoin d'un droit international et quelle est la légitimité de ce « grand machin » qu'est l'UNESCO à en être maître d'œuvre ? A quoi servent donc de telles organisations, qui « prétendent » pouvoir nous aider à faire face à la mondialisation ? N'étions-nous pas déjà avec les Romains dans un premier type de mondialisation, qui a profondément marqué, entre autres, le droit et les routes commerciales ? Et sur les débris de cet empire, dont elle a hérité, une institution universelle, l'Église catholique, n'a-t-elle pas proposé dès son expansion une vision globale du Monde, de la place de l'homme et des principes qui doivent s'y appliquer ? Est-ce qu'aujourd'hui l'UNESCO met, elle- aussi, en avant une vision du Monde, et laquelle ? N'y a-t-il pourtant pas d'autres visions et comment interviennent-elles dans cette synergie d'attentes et de craintes, d'enjeux économiques et politiques autour de l'intelligence artificielle ?

3-A cette fin, il convient de s'interroger pour resituer ici et maintenant notre rapport avec nos valeurs, nos principes, nos modes de vie et déterminer si leur utilité reste pérenne, garde son sens ou, si les circonstances ne nous obligent pas, du fait de leur gravité ou des injustices qu'elles laissent subsister, à réviser telle ou telle de ces règles et façons d'agir. Autrement dit, ne devons-nous pas regarder le Monde tel qu'il est avant de le vouloir tel qu'on le souhaite ?

Dans le cadre du Forum franco-latinoaméricain qui s'est déroulé en octobre 2024 sur le thème de l'intelligence artificielle dans le domaine de la santé, le philosophe Thierry Ménissier nous a rappelé que Nicolas Machiavel était essentiellement quelqu'un qui regardait la réalité telle qu'elle est. Alors comme juriste, ce que je vous invite à faire, c'est de regarder cette réalité telle qu'elle est aujourd'hui. C'est de savoir pourquoi l'intelligence artificielle, comme d'autres technologies, gouverne, d'une certaine manière, le Monde et pose tant de questions, avec un certain succès médiatique. Mais, dites-vous que la première étape, celle de la prise de conscience, doit dépasser le simple traitement médiatique : il ne s'agit pas de diffuser la peur et l'anxiété mais il s'agit de savoir faire rationnellement des analyses et y inclure, par la connaissance des comportements humains, le jeu des émotions et des sentiments.

4- Vous m'avez invité à parler de droit mais je parlerai aussi de politique. Le juriste, quand je dis qu'il a « les pieds sur terre et la tête dans les nuages », cela signifie qu'ayant la tâche de réguler la société, il est au cœur de ce qui est politique, même si le politique exige que le droit s'implique à analyser la complexification du Monde par des messages clairs non seulement pour en décrire exactement l'état mais aussi pour indiquer les quelques chemins qu'il serait possible d'emprunter avec les avantages et inconvénients que chaque route présente.

Karl Marx, dénonçant la démarche philosophique, disait que « Les philosophes n'ont fait qu'interpréter le Monde de différentes manières, ce qui importe c'est de le transformer. » (conclusion des Thèses sur Feuerbach, 1845). Il est vrai que l'interprétation n'est pas l'action mais le juriste a cette particularité sur le philosophe que ce qu'il interprète peut ouvrir la voie à une reconfiguration du droit non en théorie mais de façon réelle et concrète, du moins dans les sociétés où la culture du droit permet au législateur et au juge de le faire évoluer au gré des transformations sociales.

5-Mais, convenons que les choses ne sont pas si simples. Il existe effectivement un mouvement universaliste qui tente de répondre à la mondialisation sauvage par une extension du droit des Droits de l'Homme. Il se construit cependant sur des racines socio-culturelles extrêmement différentes. A cet égard, je me souviens d'un débat que j'avais organisé en 2000 au Japon. J'avais sollicité deux philosophes, le président de la Société japonaise de philosophie, H. Sakamoto, et le Français Michel Serres pour dialoguer sur la perception de la nature dans nos deux pays et la manière dont cela influence l'organisation et la vie de nos deux sociétés. L'idée même d'un dialogue a pris fin dès le moment où l'orateur japonais, qui était aussi président de la Société japonaise de Bioéthique, a affirmé que la culture de son pays n'avait nul besoin de la Révolution des Lumières et du dogme des Droits de l'Homme pour aborder les questions de bioéthique et comprendre les relations de l'homme avec l'environnement.

Il nous a signifié, ce que M. Serres a dû ressentir comme une déclaration de guerre, qu'il y avait en Asie toute une philosophie qui préexistait à notre philosophie des Droits de l'Homme et qui avait permis à l'homme de vivre en harmonie avec la Nature. C'est le souvenir du film japonais, « la Balade de Narayama » (Keisuke Kinoshita, 1958), qui m'est revenu alors à l'esprit. On y voit que les Japonais ont une conception de l'Homme qui ne fait qu'un avec la Nature, dont il est un élément parmi d'autres. Ce n'est pas forcément notre façon de voir mais, dans ce Monde qui est « globalisé » plus qu'universel, se perpétuent, voire s'affrontent, des visions qui correspondent à la façon dont, à travers nos cultures, nous l'avons construit et pensé. Certes, il existe des similitudes entre les religions, les grands mythes. Ainsi, le temps perpétuel s'organise suivant un cycle annuel plus ou moins commun : le Printemps est toujours une renaissance des choses.

6-Il y a ainsi dans ces rythmes anthropologiques, qui se ressemblent et s'opposent, le reflet des récits que différentes cultures font de notre Monde. Cela transparait aussi dans l'organisation politique de nos sociétés. Tout en se déchirant entre Réforme et Contre-Réforme, l'Europe chrétienne, dans sa quête d'universel, est partie à la conquête du Monde, de ses richesses, de ses hommes et de ses terres pour construire des empires coloniaux. Puis, au début du XXème siècle, dans la ruine d'anciens empires, de nouveaux sont apparus à l'Est tandis que l'Occident semblait dans le fascisme et le nazisme.

L'Europe dans laquelle nous vivons, avec sa mentalité, sa philosophie, sait très bien mettre en avant le rôle de l'individu dans l'exercice des Droits de l'Homme, quitte à prendre le risque aujourd'hui, avec certains acteurs majeurs de la promotion de l'intelligence artificielle, de se trouver face à un individualisme exacerbé, mêlé de nationalisme, et visant à installer sur le plan politique un libertarianisme appliqué sans règles autres que celle de la « dynamique de l'innovation » !

Tout ceci vise à expliquer ce que sont des institutions comme l'UNESCO et à en justifier les missions. Elles ne sont pas là pour porter un angélisme naïf mais elles ont précisément vocation de n'ignorer aucun des conflits que notre Monde traverse afin d'être les lieux où ceux-ci se traitent avec des procédures et des objectifs, qui mettent des intérêts opposés en situation de trouver des accords.

7-L'UNESCO est l'organisation du système des Nations-Unies spécialisée dans les domaines de l'éducation, de la science, de la culture et de la communication, qu'elle aborde avec une vision transversale. C'est le sens de la mission de coordination inter-institutions internationales que lui a confié le Secrétaire général des Nations-Unies dans le domaine de la bioéthique et de l'éthique des sciences.

Elle développe ainsi depuis les années 1990 une stratégie, qui est celle d'une communauté d'Etats, 194 fin 2024. Durant cette période, l'UNESCO a adopté trois déclarations dans le domaine de la bioéthique (Déclaration universelle sur le génome humain et les droits de l'homme, 1997 ; Déclaration internationale sur la protection des données génétiques humaines, 2003 ; Déclaration universelle sur la bioéthique et les droits de l'homme, 2005) et trois textes dans le domaine de l'éthique des sciences (Déclaration de principes éthiques en rapport avec le changement climatique, 2017 ; Recommandation concernant la science et les chercheurs scientifiques, 2017 ; Recommandation sur l'éthique de l'intelligence artificielle, 2021).

Il est intéressant de noter qu'avec cette dernière recommandation, l'UNESCO a intégré à ce type de document normatif une partie qui va au-delà des valeurs et principes puisqu'elle concerne leur mise en œuvre dans les pratiques d'une douzaine de secteurs d'activités, qui inclut notamment la protection des données, l'environnement, l'éducation et la recherche ainsi que la santé, et impliquent tous les acteurs, y compris la société civile, la communauté des chercheurs et les acteurs industriels et économiques.

Cela signifie que l'UNESCO a choisi de façon pertinente, s'agissant d'applications technologiques qui bouleversent une diversité d'activités humaines par de nouvelles pratiques, d'inviter leurs acteurs à incorporer dans celles-ci une dimension éthique concrète et adaptée.

Ce n'est pas sans lien avec le Consensus de Beijing sur l'intelligence artificielle (IA) et l'éducation, adopté en 2019, qui est le premier document proposant des conseils et des recommandations sur les meilleures façons d'exploiter les technologies de l'IA pour la réalisation de l'agenda pour l'éducation 2030.

Et il n'est pas anodin de remarquer que l'UNESCO a eu le souci de demander dans un bref délai aux Etats membres de faire connaître en 2025 ce qu'ils ont entrepris pour mettre en œuvre la Recommandation adoptée en 2021. Pour notre part, les table-rondes organisées avec le CCIC, sont avec les deux colloques du Forum franco-latinoaméricain de bioéthique d'octobre 2024 (IA et santé, Montevideo et IA et système judiciaire, Buenos-Aires) des exemples de la contribution de la Commission française pour l'UNESCO au rapport déposé par la France.

8-Vous m'avez posé des questions sur l'IA et la propriété intellectuelle, qui touche autant le domaine économique que celui de la culture. En l'état, pour nombre de droits nationaux, dont la France, le droit d'auteur ne peut être accordé qu'à des personnes physiques car cela implique une œuvre de l'esprit, donc une création. Toutefois, l'IA conduit à s'interroger sur ce qu'est une œuvre de création, débat que la photographie et le cinéma ont déjà connu en d'autres temps. On peut penser à élaborer des droits spécifiques, des droits voisins, voire des droits hors du domaine de la propriété intellectuelle. Il y a déjà des propositions qui visent soit à modifier le droit existant, soit à ne pas le changer en laissant le juge faire preuve d'imagination. Certains pays ont déjà répondu en faisant évoluer certains droits ou, au contraire, en estimant qu'on devait en rester au statu quo. En tout état de cause, il paraît probable d'envisager de rémunérer un certain nombre de personnes, qui participent au processus de mise en œuvre de l'IA, sans qu'il soit nécessaire pour autant de donner une personnalité morale à l'intelligence artificielle elle-même. On pourrait faire en sorte de reconnaître, comme dans les droits d'auteur collectifs, que la personne morale qui a contribué pour partie au processus utilisant l'IA puisse en bénéficier.

4-Des technologies au service des *soft powers*

E. Castelnau

Avec Timothée Hunter on va changer un peu d'angle de vision. J'ai tenu à ce qu'il y ait une part de jeunesse, non que nous en soyons totalement exempts, et vous avez vu que chacun des intervenants avait sa capacité à comprendre l'intelligence artificielle. Mais là, nous avons affaire à Timothée, polytechnicien, qui est parti de l'autre côté de l'Atlantique où il a fait un PhD en intelligence artificielle. Il y a travaillé avec de grands chercheurs en IA, puis dans une entreprise dans laquelle il a commencé avec 5 personnes, il y en a 5000 aujourd'hui. La question que j'ai envie de te poser est la suivante : à l'aune de ton expérience, de ce que tu as vécu, de ce que tu as appris, comment vois-tu les choses, quelle est ta vision de l'IA générative, comment la traitons-nous en Europe, comment la traite-t-on aux Etats-Unis ? Quelle est l'éthique dans ces contextes, et l'éthique, est-ce que ça existe, ou pas du tout ? et par ailleurs quelles sont les conséquences géopolitiques ?

Timothée Hunter

Merci Eric. C'est une longue question qui pourrait nous occuper une bonne partie de la soirée. Je vais simplement essayer d'aborder quelques points qui me semblent particulièrement importants dans ce débat, qu'on pourra ensuite approfondir dans les questions. Et comme vous évoquez l'intelligence artificielle en Californie, j'aimerais commencer par une petite anecdote. Quand j'ai commencé mon Master avec la personne qui est maintenant devenue une des stars de ce domaine, il m'a dit « Timothée va dans un magasin de jeux vidéo pour obtenir toutes les cartes de jeux vidéo que tu trouveras, et c'est ainsi que l'intelligence artificielle moderne a commencé, c'est avec du matériel de jeux vidéo ». Pour vous donner une idée de comment ce type de système est construit en pratique, car mon point de vue est vraiment celui du créateur, du concepteur de ces intelligences artificielles, comme l'a rappelé Isabelle, on commence par récupérer un ensemble de textes. Dans le cas des intelligences artificielles modernes c'est environ 15 000 milliards de symboles, et pour donner un ordre de grandeur, si vous vous souvenez de vos dictionnaires Larousse, ça représente 1 million de dictionnaires Larousse mis bout à bout. Ensuite on met tout ça dans un système qui va apprendre à reconnaître les textes, puis à prédire les textes suivants. Ceci prend entre 10 et 300 ingénieurs et coûte entre 5 et 200 millions de dollars. Donc une IA moderne coûte quand même assez cher, et au final le résultat semble assez mince. Ça représente 400 milliards de paramètres, qui ne sont en réalité que des chiffres que l'on peut mettre sur une simple clé USB. En somme, 200 milliards de dollars plus loin, on a une clé USB bien remplie.

Avant d'expliquer comment on utilise ces IA, j'aimerais insister sur une nuance qui me paraît assez fondamentale. Ici on a parlé à la fois de Chat-GPT et de clé USB, et il me semble important quand on parle d'éthique de savoir si on parle de l'IA en tant que technologie ou en tant qu'application. Et en particulier, est-ce qu'on parle des diffuseurs d'IA ou bien des concepteurs, des créateurs d'IA, parce que l'IA est entrée dans notre langage dans les deux domaines : c'est à la fois une technologie dans le sens où les scientifiques écrivent des articles- on peut expliciter ce que c'est, on peut le mettre sur une clé USB, le copier, le distribuer- mais c'est

aussi un outil et une application. Quand on parle de Chat-GPT, on parle d'une interface que l'on peut utiliser, avec laquelle on peut dialoguer, et pour laquelle une entreprise fournit ce service. Pour prendre une analogie qui a été utilisée par un chercheur de l'Université de Berkeley, le moteur électrique est aussi une technologie. On peut mettre un moteur électrique dans un aspirateur et personne ne dira rien, mais on peut aussi le mettre dans un drone de combat. Et pourtant parle-t-on d'éthique du moteur électrique ?

Donc quand on parle d'éthique de l'IA, pourquoi, si c'est une technologie, devrait-on même aborder ce sujet ? C'est pour cela que je vais développer ici un point de vue qui sera sans doute assez pragmatique. L'IA comme je l'ai dit peut être vue comme une technologie assez malléable. Ce que j'entends par là, c'est que l'on peut lui faire véhiculer certaines valeurs. Et justement c'est là que le contexte géopolitique, ou l'idée d'affrontement entre différentes visions du monde, est véritablement au cœur de la création des IA. Par exemple si l'on considère l'IA faite en Chine, les lois chinoises stipulent explicitement que l'IA chinoise doit adhérer à un leadership politique correct et ne pas perturber l'ordre économique et social. Je vous laisse donc imaginer ce que cela signifie concrètement. Une chose est claire : si vous demandez à une IA chinoise ce qui s'est passé place Tian'anmen en 1989, elle ne vous répondra certainement pas, et toutes les données qui auront servi à l'entraîner auront soigneusement exclu ce genre de référence. Une IA ne pourra jamais compléter ce qu'elle ne sait pas, à la différence d'un philosophe, un mathématicien ou une personne. D'ailleurs, là j'évoque un exemple chinois mais tous les créateurs d'IA en Occident, en Amérique comme en Europe essaient aussi d'ajouter certaines barrières ou garde-fous.

Quand on parle d'appliquer des règles d'éthique en IA, il nous faut à mon avis distinguer deux acteurs très différents : d'une part les fournisseurs d'IA génératives, donc Chat-GPT, Google... qui devront suivre des règles de droits fondamentaux, des règles de liberté d'expression, et qui sont réellement en charge d'être les médiateurs de l'IA avec les utilisateurs que nous sommes, et d'autre part les concepteurs d'IA génératives, les créateurs- je prends l'exemple de la société française Mistral, ou des groupes de chercheurs comme l'Allen Institute of AI à Washington, ou même Meta- qui créent juste des intelligences artificielles et les mettent à la disposition d'autres fournisseurs qui vont ensuite les utiliser. Pour ces concepteurs, il est important à mon sens que leurs tâches se concentrent plutôt sur le fait de comprendre les risques liés à l'IA générative et de faire preuve de transparence quant à l'utilisation de leurs sources et à la capacité de leur IA à ne pas générer contenus illégaux. Par exemple les IA chinoises n'ont aucune obligation de dire qu'elles n'incluent pas certaines références historiques dans leurs sources, comme je l'ai mentionné. Donc quand on voit l'IA en tant que technologie cela implique que c'est un choix social et humain et non un choix technologique de mettre et véhiculer certaines valeurs dans une IA. De mon point de vue de chercheur Outre-Atlantique, il y a un point qui me semble important de mentionner qui sera au cœur de ce débat : il n'y a pas de consensus au sein de la communauté scientifique quant à la meilleure approche pour créer et améliorer ces IA. Vaut-il mieux un cadre légal strict- c'est l'approche européenne- dans lequel on s'intéresse au risque potentiel lié à l'utilisation de l'IA, ou vaut-il

mieux un cadre plus souple- ce qui est le consensus américain- qui se base davantage sur les risques observés en pratique quand on commence à déployer ces systèmes. A mon avis il est préférable de concentrer les aspects légaux sur les applications et non sur la création de l'IA, et je vais vous expliquer d'où je tiens ce point de vue. Si l'on regarde l'IA utilisée dans les entreprises, une chose est évidente, c'est qu'une entreprise a un objectif commercial, donc l'IA dans une entreprise- je ne suis pas juriste, mais c'est en tout cas ainsi que je le verrais- l'IA dans une entreprise obéit à une certaine finalité, donc à un souci d'efficacité, de performance, de productivité et en particulier elle est utilisée pour prendre des décisions. Je vois un exemple à côté de chez moi, le port de Rotterdam, un port entièrement automatisé qui fonctionne 24 heures sur 24 et qui utilise des dizaines de milliers d'agents que l'on appelle désormais agents d'intelligence artificielle pour manipuler des conteneurs. Ils utilisent aussi d'autres types d'applications qui consistent par exemple à réaliser des recommandations automatiques de CV pour les candidats à certains postes. Ce sont deux exemples d'IA complètement différents quant à leur finalité mais qui répondent tous deux au cadre de ce qu'on appelle l'IA. Alors quand on parle des IA génératives comme Chat-GPT ce qu'il faut voir, c'est qu'en entreprise elles sont rarement utilisées directement. En pratique, au lieu d'avoir une IA très générale et lui poser une question complexe comme par exemple générer des contrats d'assurance, il est beaucoup plus efficace de prendre cette IA et de la focaliser sur une tâche spécifique comme « quel est le nom d'une personne, quel est le nom d'une entreprise, quel est le nom d'une compagnie d'assurance ? » et ainsi de suite. Pour prendre un exemple concret, une compagnie d'assurance américaine a repris tout son processus de création de contrats pour la souscription d'assurance, processus qui lui prenait 2 semaines avant l'introduction de l'IA, et a décidé de l'automatiser avec 78 micro IA, toutes focalisées sur une tâche spécifique. Cela a permis à cette compagnie de passer d'un délai de souscription de 2 semaines à 3 heures, et répond concrètement à un souci d'automatisation et d'efficacité.

Considérant les IA concentrées sur des tâches spécifiques, comment s'inscrit l'éthique et l'idée de responsabilité dans ce domaine ? Quand je vois les textes juridiques sur ce sujet, en tant qu'utilisateur et créateur d'IA, j'y vois principalement un cadre conceptuel qui m'éclaire quant à l'orientation générale et qui répond tout à fait sur le plan personnel à mes valeurs, à mes valeurs chrétiennes, mais après cela, j'ai du mal à connecter ça à la réalité de la création d'une IA. Je vais prendre un exemple très concret : une IA ne doit pas présenter de caractère discriminatoire. Comment définir un caractère discriminatoire en pratique ? Or cela à mon avis n'est pas possible. L'IA est une technologie basée sur les mathématiques et la statistique. Si nous parvenions à avoir une IA entièrement éthique, cela voudrait dire que nous serions parvenus à prendre toutes nos valeurs, tous nos choix de société, et à les représenter dans un langage mathématique. Et c'est impossible, donc parler d'IA éthique de façon fondamentale ne sera à mon avis pas possible, du moins dans le cadre d'une IA statistique comme nous la connaissons maintenant. Je reprends l'exemple de notre application de sélection de CV, donc choix binaire : est-ce que je rejette ou est-ce que je garde ce CV ? Pour définir l'équité et la discrimination dans un cadre assez simple, des chercheurs ont dénombré 14 définitions

différentes de l'équité et de la discrimination, définitions mathématiques qui sont toutes contradictoires. Ce que je veux dire par là, c'est que si j'en choisis une et construis une IA équitable par rapport à cette définition, elle sera automatiquement inéquitable au regard d'une autre définition ; et pire encore, aucune de ces définitions ne répond pleinement à des critères généraux tels que l'indépendance d'informations masquables, la séparation, la suffisance... ce sont des détails mathématiques.

C'est vraiment très difficile d'arriver à définir comment avoir une conception éthique d'une IA. Revenons à l'exemple des CV. J'ai vu, dans les entreprises qui essayent d'utiliser l'IA pour la sélection d'un poste à pourvoir dans le domaine informatique, arriver neuf fois plus de candidats masculins que féminins. Si un chercheur utilise la notion de parité statistique, une des définitions, il obtiendra neuf fois plus d'hommes que de femmes, alors que s'il utilise un critère de parité démographique, il recevra autant de CV d'hommes que de femmes... mais à ce moment-là, les hommes pourront potentiellement dire qu'ils ont été victimes de discrimination ; s'il utilise une notion d'égalité des chances, à ce moment-là on peut montrer que statistiquement il y aura plus d'hommes que de femmes, et ainsi de suite. Dans tous les cas une des classes y sera discriminée. Cet exemple ne met en jeu que le critère hommes-femmes. Imaginez ce qui se passera si je commence à inclure l'âge, ou l'origine, et ainsi de suite. Donc en aucun cas il n'y aura de bon choix, mais une chose est sûre, une entreprise qui utilise un système automatique utilisera implicitement ou explicitement l'une de ces définitions. Dans la pratique, ce que je veux dire par là, c'est qu'il y aura des entreprises qui fermeront les yeux et feront confiance à leurs ingénieurs, ou des entreprises qui bricoleront une idée de comment intégrer l'éthique dans un cadre technique. IBM avait demandé il y a 2 ans, parmi ses clients qui utilisaient l'intelligence artificielle, lesquels avaient des concepts éthiques, et il y a 2 ans, les 3/4 n'avaient aucune considération éthique concernant l'utilisation de l'intelligence artificielle. C'était au niveau mondial. J'imagine qu'au niveau européen un nombre beaucoup plus élevé d'entreprises avaient des considérations éthiques.

Donc pour moi, la question en tant qu'utilisateur, en tant que concepteur d'intelligence artificielle, est de trouver un juste milieu entre transparence et secret industriel : avons-nous besoin de savoir comment fonctionnent les 78 petites unités autonomes d'une compagnie d'assurance ? mais aussi de décider en tant que société, tous ensemble, en tant qu'état, quels sont les critères autorisés pour définir l'éthique dans la pratique. Par exemple parmi les 14 définitions, y en a-t-il une que nous préférons ? et ainsi de suite. On a vu ce débat aux États-Unis pour un logiciel appelé Compass utilisé dans le contexte de la justice prédictive. Il ne dit pas qui est un criminel, qui est coupable, il indique juste un score de probabilité de récidive. L'entreprise a été attaquée par un organisme journalistique disant que ce logiciel était biaisé, n'était pas équitable envers certaines populations plus défavorisées comme les Afro-Américains, ce en quoi elle avait raison dans sa propre définition de l'équité. L'entreprise derrière Compass a répondu qu'elle utilisait des critères démographiques qui n'avaient pas été pris en compte par cette entreprise journalistique, ce en quoi elle n'avait pas non plus tort selon sa propre définition de l'équité. Alors, comment allons-nous parvenir à résoudre ces

débats sur ce qu'est précisément l'équité en pratique, quand on va utiliser des systèmes de décision automatiques ? D'ailleurs je voudrais faire ici une remarque sur la transparence : pour déterminer si un outil d'IA a un caractère discriminatoire, il faut le mesurer, il faut savoir mesurer cette discrimination. Et pour le moment, c'est une spéculation, mais je ne serais pas surpris si dans un futur relativement proche on arrivait à une situation où pour postuler à un poste par exemple, on demande au candidat de fournir non seulement son âge mais aussi son origine, potentiellement des informations raciales... un nombre d'informations qui seraient maintenant considérées comme sensibles, mais qui permettraient de détecter les caractères discriminatoires de ces IA. Pour l'instant cela reste une spéculation. Dans ce contexte, l'utilisation de l'IA pour les entreprises européennes peut être assez complexe, pour ne pas dire délicate, et on pourrait être tenté en raison de ces questions complexes d'équité, d'utiliser des IA américaines

Cela pose des enjeux géopolitiques et géostratégiques majeurs, et pour expliquer pourquoi, je vais parler ici des grands modèles d'IA - car dans l'IA il y a des grands modèles et des petits modèles. Les grands modèles sont ceux qui font la une des journaux. Ce sont les petits modèles qui décident si vous recevez une publicité dans votre boîte mail. Quant aux gros modèles, si on regarde où sont faits les investissements, pendant l'année 2023, les Américains ont construit, conçu et déployé 95 grands modèles d'IA, ce qui inclut Chat-GPT, tous les modèles de création d'images et ainsi de suite, la Chine en a construit 65, l'Europe 20, dont 9 en France. On pourrait se féliciter d'avoir construit près de la moitié de tous les IA en Europe l'année dernière, mais il faut considérer aussi que ça représente 5% des grosses IA déployées à travers le monde. C'est pas mal mais c'est loin d'être suffisant, et on pourrait se demander quel est le problème. Si l'on considère des modèles comme ceux fabriqués par Meta pour lesquels on a accès à des statistiques, 80 à 90% des données utilisées sont en anglais, cela signifie que toutes les autres langues qui sont incluses dans ces modèles ne sont pas aussi bien représentées que l'anglais. Cela signifie aussi que toutes les associations conceptuelles qui sont derrière sont également dominées par des idées anglosaxonnes. Pour tester cela j'ai fait une expérience ce week-end : j'ai demandé à Chat-GPT, et à bien d'autres modèles d'ailleurs, si je devais travailler ce week-end pour terminer ma présentation de ce soir, et la réponse a été particulièrement fascinante. Toutes les IA que j'ai testées, dont les cinq versions de Chat-GPT, testées également en langues française, anglaise, néerlandaise, allemande, toutes des langues pour lesquelles j'ai su écrire ma question, puis comprendre la réponse, toutes m'indiquent qu'il me fallait travailler ce week-end pour impressionner mon patron, donnant de nombreuses raisons.

Non seulement c'est bien cela leur réponse - et j'ai effectivement travaillé ce week-end- mais aucune d'entre elles n'a même demandé ou remis en question le but ou l'importance de cette présentation, aucune, dans aucune langue. Je suppose que si c'était une IA conçue en France, le résultat aurait été un peu différent. J'ai même continué à tester en me disant syndicaliste. La réponse a été que je dois démontrer au reste de mon entreprise que j'ai une forte éthique de travail. Je ne sais pas si c'est une conception particulièrement française...

Je vais tenter de répondre à votre question sur les conséquences du peu de liberté laissée au système européen. C'est quelque chose qui à mon avis en tant que créateur est particulièrement important. Le secteur de l'IA évolue extrêmement vite, et si l'on regarde comment la loi européenne sur l'intelligence artificielle a évolué, Chat-GPT fin novembre 2022 est apparu extrêmement tard dans la course et a été ajouté quasiment à la dernière minute dans un projet de loi. Il me paraît particulièrement important d'avoir une interprétation souple et évolutive de nos cadres juridiques pour assurer qu'on se concentre sur les aspects importants. En effet, d'expérience, tous les modèles qu'ils soient open-source réalisés par des chercheurs ou bien réalisés sous le secret commercial par des entreprises, tous les modèles un tant soit peu intéressants seront classifiés « à haut risque » par la loi européenne. Par ailleurs, lorsqu'un modèle est dit à haut risque, les contraintes sont fortes sur les créateurs d'intelligence artificielle en matière de traçabilité et de transparence, non seulement les chercheurs mais aussi Mario Draghi, comparent cela avec le cadre juridique des entreprises pharmaceutiques où les coûts documentaires pour expliquer quelles décisions sont prises à quel moment sont tels qu'en fin de compte, les seules entreprises capables de gérer ces contraintes sont les gros groupes tels Google et compagnie, tandis que les PME seront incapables de créer une IA dans un système au sein d'un cadre juridique européen, d'autant que le coût de l'erreur est élevé : il atteint jusqu'à 7% des revenus annuels mondiaux, ça peut aller très vite. Si on se place de mon point de vue américain, un contre-exemple est le projet de loi américaine qui visait justement à encadrer certaines IA sur le modèle européen. Le projet de loi californien visait à réguler les systèmes comme Chat-GPT, il a été rejeté en octobre il y a moins d'un mois. La raison avancée par le gouverneur est précisément que ce système était plus focalisé sur la création que sur l'application des IA. Il y a un risque que nous nous retrouvions dans un monde où les Américains continueront à créer, tandis qu'en Europe nous passerions plus de temps à documenter ce que nous faisons qu'à créer de l'IA. Il est possible d'explorer aussi d'autres directions, mais je dois laisser la parole...

5-Débat

Eric Castelnau : J'ai une première question pour Isabelle : Vous avez terminé par la nécessité d'un questionnement, par une batterie de questions. Celle-ci retient particulièrement mon attention : -Les mécanismes propres aux IA qui sous-tendent les réseaux sociaux n'amplifient-ils pas les discours de haine et autres fake news ? -Réagir et partager sur les réseaux sociaux des informations non vérifiées est-ce responsable ? Pouvez-vous nous dire ce à quoi vous faites allusion ?

Isabelle Chaperon : Cela concerne les réseaux sociaux et c'est une question qui me semble de la première importance, car beaucoup de gens imaginent que les réseaux sociaux ne sont pas de l'IA. En fait elles comportent des petits programmes d'IA, petites IA au sens évoqué par Timothée. Et si je souhaite parler des réseaux sociaux en particulier dans ce débat sur les IA, et où l'on réfléchit aux conséquences de nos actes, c'est que cela concerne un très large public

y compris dans beaucoup de pays ou dans beaucoup d'environnements où l'on n'a pas accès à tous ces très grands modèles d'IA. C'est une enseignante d'un collège ivoirien qui a attiré mon attention sur cet usage. Les jeunes, même parfois beaucoup trop jeunes, ont leur téléphone portable et usent et abusent des réseaux sociaux. Comment les réseaux sociaux amplifient l'impact des fake news et discours de haine ? Le modèle économique de l'entreprise Facebook, ou TikTok par exemple, est de gagner de l'argent en plaçant un maximum de publicité, donc maximiser l'audience et le temps passé par les utilisateurs. Le concepteur demande à une IA de maximiser ce qu'on appelle « *l'engagement* ». Comment une machine qui est encore très stupide au départ mesure-t-elle l'engagement ? Elle enregistre le temps passé à lire l'article posté, les réactions comme par exemple les émoticônes de colère (le petit bonhomme tout rouge), et le nombre de commentaires et republications dont l'article fait l'objet. Il suffit d'une petite IA qui sélectionne les posts ayant suscité un bon « *engagement* » pour les mettre en avant dans le fil d'actualité, c'est-à-dire ce qui vous impose des articles que vous n'avez pas sollicités. Le résultat est que ça amplifie considérablement la diffusion, donc l'impact, en particulier des fake news, discours complotistes ou de haine. Vous comprenez ainsi pourquoi *liker*, commenter, republier n'est pas sans conséquence, pensez responsabilité. Rien ne parle mieux qu'un bon exemple. Celui-ci est terrifiant, je l'ai trouvé dans le dernier livre de Yuval Harari sorti fin septembre. Il cite des études très documentées sur le rôle qui s'est avéré extrêmement important des réseaux sociaux dans le massacre des Rohingyas au Myanmar... Une petite partie musulmane de la population birmane était en conflit avec tout le reste de la population essentiellement bouddhiste. Des référents bouddhistes avaient commencé à poster sur Facebook des propos très hostiles aux musulmans, qui ont vite fait le *buzz*, ces publications ont été surmultipliées et en ont appelé d'autres, par le processus que je viens d'évoquer, ce qui signifie qu'il est devenu dans l'esprit commun des bouddhistes de base qu'un Musulman est un sous-humain, et qu'on peut tuer un Musulman. Si on ne peut pas accuser Facebook d'avoir créé ce massacre, le rôle de Facebook a cependant été déterminant. L'autre problème des réseaux sociaux est qu'il est avéré, et bien documenté, que sur Twitter -donc sur X maintenant- 3% des comptes Twitter sont en fait générés par des IA, surnommés Trolls. 3% ce n'est pas le bout du monde, seulement on observe que 20% du trafic donc 20% ce qu'on voit sur les réseaux, sur Twitter, a été produit par ces 3%, en raison encore de ce processus qui amplifie l'impact des articles problématiques, et le risque est maintenant avéré de manipulation. J'ajouterai pour finir qu'une étude réalisée par un organisme français montre que les journalistes utilisent beaucoup Twitter, X, comme source d'information, ce qui pose question quand le même organisme démontre que les fake news circulent six fois plus vite que les posts ordinaires. D'où ma recommandation d'un usage sobre, si *liker* vous donne l'impression d'exister dans le débat, c'est une pure illusion, et dont il faut mesurer les conséquences.

Question dans la salle : Concernant les biais de construction des algorithmes, quelle différence faites-vous entre transparence et lisibilité ?

Christian Byk Pour une mise en perspective, faut-il prendre cette question comme une question technique ? La transparence est quelque chose de technique, la lisibilité est quelque chose de technique. Dans le discours de Timothée, j'ai le sentiment qu'il y avait le souhait de séparer le bon grain de l'ivraie. Le bon grain c'est la technique, l'ivraie c'est éventuellement l'éthique quand elle se mêle de ce qui ne la regarde pas. Non, je pense que c'est une erreur fondamentale, et l'exemple que vous venez de donner est un bon exemple. L'appropriation que tout un chacun se fait aujourd'hui de la technologie signifie que nous ne pouvons pas séparer la technologie des applications, de l'éthique, du droit et du politique. Je vais retourner à l'exemple de Machiavel. Machiavel veut imposer à Florence la conscription militaire volontaire. Pourquoi ? Parce que qu'il a compris que ce qui ruine Florence c'est de devoir payer des condottieri pour lever des armées - on est en pleine guerre avec les Français. Mais qui y a intérêt ? Les grandes familles oligarchiques qui prêtent de l'argent à Florence pour payer les militaires, donc les militaires n'ont jamais intérêt évidemment à ce que ça s'arrête, et les financiers non plus. Il va alors procéder à une réforme du système politique de Florence, avec une grande assemblée de 1300 citoyens et réduire la part de l'oligarchie. Ce que je veux dire, c'est que quand on a un système, c'est un tout. Ce que la technologie et surtout nos sociétés techno-scientifiques qui ne sont plus simplement une technique purement cognitive, apportent, se construisent là-dessus, c'est un tout indissociable.

Timothée Hunter : Je vais rapidement rebondir, je dois être 100% clair. Je souhaite de tous mes vœux que nous ayons un cadre éthique clair pour l'utilisation de l'intelligence artificielle, cela me paraît évident. Comment allons-nous le définir et le mettre en œuvre ? Donc la question se pose quand il s'agit d'entrer dans les détails, c'est là que ça peut devenir plus compliqué. Pour répondre à la question sur la transparence et la lisibilité, je vois deux points qui doivent absolument être mis en place pour tous les systèmes d'intelligence artificielle, au minimum ceux dits à haut risque. La façon dont je le conçois, c'est que la transparence doit servir à fournir toutes les informations nécessaires pour comprendre le fonctionnement et les biais des systèmes d'intelligence artificielle. La lisibilité est là pour apporter un éclairage sur comment les décisions sont prises. Donc transparence ça apporte des statistiques disant que ce système est fiable ou pas fiable pour telle ou telle raison, la lisibilité indique ce qui a permis d'arriver à cette décision plutôt qu'une autre.

Laure Tabouy : Du coup, je vais trancher entre les deux ! Peut-être que la solution est justement pour concilier tout ce qui est vraiment créateur sur le terrain et puis et puis réflexion, droit, etc... C'est ce qu'on peut appeler une éthique by design ou une éthique par conception. L'éthique par conception est en cours de discussion, en cours d'élaboration. C'est aussi philosophiquement complexe car d'un côté on a une éthique orientée vers l'objet, et de l'autre une éthique orientée vers la conséquence, l'aspect déontologique d'un côté et conséquentialiste de l'autre, ce qui est plutôt philosophique mais du coup l'éthique est vraiment coincée entre les deux. Où placer l'éthique et où placer le droit, c'est une question vraiment actuelle. Comment pouvons-nous mettre en œuvre l'éthique, le droit depuis le début de la conception jusqu'à la fin pour laquelle il est construit. C'est un débat actuel qu'il faut

construire et qui n'est pas si simple car coincé, encore une fois, entre des aspects philosophiques plutôt conséquentialistes d'un côté et déontologistes de l'autre. Peut-être y a-t-il une troisième voie qui concilierait les deux, ça nécessite de travailler sur du *by design*, qui est à mon avis une bonne voie.

Question dans la salle : Pour illustrer un peu peut-être les limites de l'IA et en particulier dans son rapport avec l'éthique, je voulais poser la question suivante un peu iconoclaste : Avez-vous posé à Chat-GPT la question « Est-ce que Dieu existe ? »

Timothée : Je ne l'ai pas fait mais la réponse est simple, c'est ce qui s'appelle l'alignement des valeurs. Chat-GPT, comme tous les modèles d'ailleurs, a un certain nombre de questions prédéterminées auxquelles il refuse de répondre. De la même façon, si on demande comment fabriquer une bombe, il dira seulement que ce n'est pas légal. A la question de l'existence de Dieu, il répondra que c'est une question à laquelle il ne peut pas forcément répondre et ça fait partie des valeurs qui sont intégrés dans Chat-GPT. Il y a d'autres moyens de lui demander vraiment ce qu'il pense, mais c'est un autre débat.

Question de la salle : Ce qui m'intéresse vraiment c'est de savoir, peut-être à court, moyen, long terme, quel sera le contrôle démocratique de l'application de l'IA, parce que moins que de savoir si la technologie aura des biais dans la façon dont elle est conçue, c'est plus mon expérience personnelle qui me préoccupe. Je suis une jeune femme racisée. Quand je vais demander un prêt à la banque, est-ce que j'aurai moins de chance qu'un collègue blanc et masculin d'avoir accès au crédit ? Vous avez mentionné Elon Musk... Nous sommes les gens du peuple, comment pouvons-nous donner notre avis là-dessus, j'entends que c'est aussi vraiment ce problème de ne pas être laissés à la traîne par les Américains...

Eric Castelnau : qui souhaite répondre ? ...Monsieur Byk ?

Christian Byk : Sur les deux questions concrètes, on sait déjà comment est l'IA quand il n'y a pas de démocratie. Si l'on parle de Chine, pas seulement les Chinois mais les GAFAM s'en sortent très bien. Pour ne pas rater le marché chinois, ils négocient avec le gouvernement chinois, ils acceptent de couper les canaux quand on le leur demande, pour garder le marché. Donc les GAFAM, ces grosses entreprises américaines - comme aujourd'hui, même les chinois - on ne peut plus s'en passer - s'arrangent très bien ensemble. Et de la même façon, les discriminations, on sait très bien ce que c'est : allez devant les tribunaux, vous avez des décisions en France, à l'étranger. La loi, la Convention Européenne des Droits de l'Homme, la Convention Européenne sur la Biomédecine, les protocoles additionnels dans le domaine des tests génétiques, les déclarations de l'UNESCO, vous donnent les catégories de discrimination, elles sont là. Et dans la réalité, je suis d'accord sur le fait qu'il y a plusieurs définitions, mais c'est celle de la loi qui compte, la loi telle qu'interprétée par le système judiciaire.

Question : En tant que réalisatrice, qu'utilisatrice de l'IA en image : on a parlé d'intelligences artificielles qui pourraient varier selon les cultures, mais je vois que les visuels sont majoritairement inspirés par l'imaginaire américain, cela pourrait-il changer ?

Laure Tabouy : La question de la culture est fondamentale. En fait on remarque une chose, et Timothée l'a bien dit, c'est que la façon dont on construit effectivement les algorithmes des IA, porte vraiment l'empreinte des valeurs que l'on souhaite, et en même temps c'est aussi la main de l'homme, donc il est peut-être nécessaire que nous réfléchissions à une autre façon de construire et voyions comment avec notre diversité culturelle, nous pourrions mettre cela en œuvre autrement. Je ne suis pas geek et ne suis pas informaticienne...

Timothée Hunter : petite anecdote : le premier jeu vraiment complet d'images utilisé en intelligence artificielle surtout pour les visages, *imagenet*, inclut plus de 50% d'images anglaises et américaines. Presque tous les jeux de données ont été créés à partir de celui-là, d'où la prépondérance des idées anglo-saxonnes. Quant à quelque chose de propre à une culture, pour moi la réponse est très simple, oui, c'est possible, et la plupart des pays y travaillent. D'ailleurs nous l'avons montré sur le plan juridique avec des lois comme le RGPD on peut avoir un contrôle des données au niveau régional ou au niveau national, donc c'est tout à fait possible, et je ne serais pas surpris si à long terme par exemple, la France ou même l'Europe imposait des règles comme le fait que la moitié des données utilisées pour créer un modèle utilisé en France soient des données provenant de France. C'est possible et ce pourrait être une conséquence assez simple à mettre en œuvre des lois que nous avons déjà au niveau européen.

Laure Tabouy : j'ai juste une question : nous avons parlé de la Chine, de l'Amérique du Nord, de l'Amérique du Sud, mais qu'en est-il de l'Afrique ? Il y a là un grand vide, et pourtant beaucoup de pays africains sont en train de monter en puissance, de nombreux pays africains travaillent vraiment sur l'IA, aussi bien que les neurosciences, que l'éthique, que la technologie et l'algorithmique. Dans cette question de culture on se rend compte qu'il y a un là un biais et qu'on a un manque réel de considération d'une partie du monde, ce qui est à mon avis discriminatoire.

Isabelle Chaperon : Le fonctionnement lui-même d'une IA, en particulier l'entraînement d'une IA générative, nécessite des myriades de données. Si les données venant d'Afrique sont sous-représentées, c'est essentiellement parce que l'Afrique produit peu de choses sur internet en comparaison de l'Amérique du Nord et l'Europe en général. Dans l'exemple d'une IA de reconnaissance faciale, la faible proportion de données d'entraînement venant d'Afrique plombe sa performance. Ce n'est pas une discrimination intentionnelle, mais un manque de données. Pour revenir à l'UNESCO, il existe une chaire UNESCO sur l'éthique en Côte d'Ivoire par exemple, dont le Professeur Poamé est titulaire. J'ai trouvé fort intéressante à cet égard son initiative d'un concours de traduction la Recommandation sur l'éthique de l'intelligence artificielle de l'UNESCO en plusieurs langues locales africaines : Swahili, Bantou, etc... Outre la

diffusion de la recommandation, cela aura le mérite de fournir des textes dans plusieurs de ces langues, qui pourront être utilisés pour l'entraînement de modèles langagiers.

Eric Castelnau : Merci à chacun de nos intervenants. Je crois que c'était très instructif. Vous avez compris comment ça marche, quel est le sens de l'IA, de l'éthique et les limites de l'éthique, mais aussi tous les progrès que l'on peut faire en matière d'éthique, d'éthique by design, et troisième chose, c'est en effet tout ce que fait l'UNESCO, et tout ce qui se fait partout dans le monde pour enseigner, prévoir, imaginer, alerter, en impliquant toutes les parties prenantes ; et puis on a eu le point de vue des américains, tempéré – merci Timothée d'avoir parlé français car c'est un exercice pour toi après tant d'années Outre-Atlantique parmi les startupper créateurs d'IA.

Deuxième partie

Eduquer à l'éveil de la conscience pour un usage responsable des IA

Isabelle Chaperon

Aujourd'hui, pour la table ronde centrale de ce cycle, nous sommes accueillis aux Facultés Loyola, nouveau nom du Centre Sèvres, que nous tenons à remercier, tout comme le CFRT/Jour du Seigneur qui assure gracieusement la captation de cette table-ronde. Je suis reconnaissante à toute l'équipe qui l'a préparée au sein du CCIC, en particulier Roseline Moreau, Monique Grandjean, Christine Roche, Françoise Meauzé et Carol Jardin, ainsi qu'à Didier Coulomb, qui en sera l'animateur. Un [essai](#) en reprend le travail.

Merci chers orateurs d'avoir accepté notre invitation à cette deuxième table-ronde.

« *Que sert à l'Homme de gagner l'univers s'il vient à perdre son âme ?* » (Marc, 8, 36). Cette question est plus actuelle que jamais...

Didier Coulomb

Quatre intervenants vont éclairer cette question : Monique Grandjean, Louis Lourme, Thierry Magnin et Alexandra Maurice.

Notre première intervenante, Monique Grandjean est sociétaire de l'Académie Catholique de France, membre fondatrice de l'Association Européenne François Mauriac ; elle est philosophe et auteure notamment avec Roseline Moreau du livre du CCIC « Et mon âme dit le robot » (<https://ccic-unesco.org/et-mon-ame-dit-le-robot/>) La question posée est : comment éduquer à l'éveil de la conscience pour une utilisation éthique de l'IA ? Monique je vous laisse la parole pour l'expliquer.

Monique Grandjean

Bonsoir chers amis. Nous nous interrogeons sur notre façon de mener nos vies en donnant du sens à l'aventure humaine. A cet effet, nous avons besoin d'un permis de conduire, d'un état des routes pour étudier des cartes mais surtout d'avoir une boussole d'intériorité parce que nous ne prenons plus le temps du silence intérieur. Nous vivons trop à la surface de nous-même, ce qui n'est d'ailleurs pas nouveau puisqu'au début du siècle dernier, Bernanos s'en plaignait déjà et le craignait pour nous. Aujourd'hui, en effet, allons-nous laisser notre vie pilotée par des algorithmes ou aurons-nous la force de les laisser à leur place d'instrument destiné à aider les humains ? En fait qu'est-ce que le progrès si ce n'est un outil pour faire progresser l'homme, ce qui nous amène à penser comme le philosophe Ellul que le véritable

défi de la technique se situerait dans l'homme lui-même ? Ces questions sont bien à leur place dans ce bel espace où souffle l'esprit et principalement celui de Saint Ignace de Loyola.

Les jésuites ont fondé le Centre Sèvres en 1974 qui est devenu la faculté Loyola Paris en 2024. Cette faculté propose une pédagogie inspirée de la tradition des exercices spirituels d'Ignace de Loyola dont les maîtres mots étaient la vigilance et le discernement, deux attitudes qui sont bien les premiers articles du code de notre permis de conduire, notre vie intérieure.

Mais cette recherche prend un certain temps et cela fait déjà 7 ans, depuis 2017, que le Centre Catholique International de Coopération avec l'UNESCO, le CCIC, réfléchit, en partenariat avec l'UNESCO, l'Académie catholique de France et les ONG de sa plateforme, sur les transformations numériques. Il a organisé des conférences, des forums, des journées de réflexion sur le thème de la place de l'homme au cœur du monde, au milieu de toutes ces technologies. Les thèmes étaient l'homme augmenté, le transhumanisme, le pouvoir de décision laissé aux puissances technologiques. En 2019, malgré le Covid, il y a donc eu trois webinaires puis un livret que vous pourrez acheter plus tard « Métamorphose du monde, quelle boussole pour l'éducation ? », soit les Actes du forum.

Ce livre a reçu les félicitations d'un certain nombre de personnes importantes et l'intérêt de la Commission Nationale Française pour l'UNESCO (CNFU) qui nous a demandé de poursuivre le travail sur la question suivante : quelle éthique de l'IA dans l'éducation de la jeunesse ? Après deux ans de travail d'enquête avec les ONG de la plateforme du CCIC, nous avons rédigé ce document que nous sommes heureux de vous présenter et surtout en espérant l'enrichir par le dialogue avec vous.

Car, en fait, nous sommes à une époque où le conceptuel n'est pas une fin en soi. Il appelle un débat suivi d'une action, une mise en mots car savoir débattre est devenu aujourd'hui, surtout dans les nouvelles technologies, une compétence clé de l'éducation civique : elle développe l'autonomie morale, aide à devenir citoyen. C'est aussi l'apprentissage de la démocratie, c'est un art de vivre ensemble tout au long de nos vies, de l'école à l'université, la vie professionnelle et même dans la vie privée. Sans la science du débat, nous tombons dans la violence physique, les insultes verbales, le divorce et la guerre. Mots pour le dire : c'est déjà le titre d'un livre sorti en 1960 et dans la forêt des technosciences, des nanotechnologies, des biotechnologies, des sciences de l'information, du numérique etc, nous avons toutes les raisons de nous perdre ou de céder à tout ce qui nous est offert.

C'est pourquoi nous avons construit une boussole d'intériorité dont nous avons emprunté l'idée et le design à la juriste récemment décédée Mireille Delmas -Marty. Elle pensait à la gouvernance mondiale, nous avons pensé à la gouvernance individuelle : comment enseigner l'ère numérique à l'avenir ? C'est une boussole très spéciale car elle a comme seul pôle magnétique dans son centre d'attraction l'éveil de la conscience qui irrigue quatre flèches : honorer l'esprit critique et le discernement, comment avoir un lien avec l'altérité et avec le monde et, en Occident, le sens de la transcendance.

Je n'en parlerai pas davantage car c'est mon amie Alexandra Maurice qui vous en parlera plus tard, elle qui est formatrice et professeure et dont le rôle est très important puisque récemment le pape François a mis entre les mains des professeurs l'avenir de la civilisation, sa grandeur ou sa décadence. Mon homme appelé le robot n'imagine pas dans cet essai les angoisses métaphysiques. Un petit bonhomme de silicium même s'il est équipé d'outils super sophistiqués, n'a rien de commun avec l'être humain, dont nous donnerons la définition déjà donnée par saint Irénée au premier siècle après Jésus-Christ mais qu'aujourd'hui François Cheng a reprise dans toute son œuvre : la triade indissociable corps, âme, esprit à l'ère des algorithmes et de l'homme contemporain. Le moi, dit François Cheng, a d'autant plus besoin de l'âme des poètes et des artistes.

Cette question vient du livre qu'avait écrit en 1987 le professeur Jean Bernard, hématologue dont les amis étaient Jacques Monod, François Jacob, François Gros. Ils s'intéressaient aux neurosciences et se posaient la question de savoir si la pensée, la morale, les passions pouvaient s'expliquer par les progrès de la science. Tous ces grands professeurs s'intéressaient à l'éducation et allaient dans les écoles pour parler de leurs livres à leurs étudiants. Et un jour une jeune Brigitte de 10 ans a dit au professeur : peut-on regarder l'âme derrière ton microscope ? Et si on fait une prise de sang est-ce qu'on va pouvoir la voir ?

Ce livre a eu un fort impact sur moi parce qu'il posait les bonnes questions. Ce sont les mêmes questions qui traversent toutes les recherches aujourd'hui dont nous, savants, philosophes, théologiens, artistes, dans cette période si riche en nouvelles technologies, tout le monde, nous inquiétons de l'avenir. Même le prix Nobel de physique, en cette fin d'année 2024, se posait des questions et se demandait comment les humains pouvaient rester non seulement fidèles à eux-mêmes, mais toujours à l'avant-garde de toutes les nouvelles technologies, comment ils pouvaient garder leur intériorité psychologique, leur intériorité morale, leur transcendante intériorité et pour ce faire je vais donner la parole à notre ami Thierry Magnin qui nous a accompagnés tout au long de notre travail et je lui passerai la boussole qui est vraiment devenue mon compagnon de vie.

Didier Coulomb

Thierry Magnin est prêtre mais aussi physicien et théologien, donc il a toutes les qualités pour commenter ces problèmes de technologie que Monique vient d'expliquer. Recteur adjoint de l'Université Catholique de Lille après avoir dirigé celle de Lyon, il est responsable de la chaire des sciences, des technosciences et de la foi à l'heure de l'écologie intégrale. Je profite de l'occasion pour signaler un de ses livres qui est aussi en vente en ce moment : « Vers une civilisation de l'algorithme ; un regard chrétien sur un défi éthique ».

Thierry Magnin

Bonsoir à tous, merci de m'avoir fait signe. Alors comment mettre l'IA au service de l'éducation sans perdre la boussole intérieure qui vient de nous être offerte ? L'IA présente des opportunités tout aussi nombreuses que ses risques. Cela me semble être le dilemme de l'IA,

d'ailleurs bien illustré par les deux récents Prix Nobel 2024 de physique qui ont tous deux été récompensés pour leurs travaux sur les réseaux de neurones artificiels qui sont à la base de ce qu'on appelle le *deep learning*, qui est une méthode essentielle aujourd'hui en intelligence artificielle. Ils ont en effet en même temps lancé un véritable cri d'alarme sur une possible perte de contrôle par l'homme. On pourrait dire qu'il y a « attention danger ».

Et puis les lauréats du prix Nobel de chimie attribué quelques jours avant ont utilisé l'IA pour inventer des protéines à des fins thérapeutiques et là, ils ont crié bravo l'IA. Devant ces deux attitudes auxquelles nous sommes conviés, il semble nécessaire de faire preuve d'une profonde vigilance mais aussi, peut-être, de regarder l'IA comme nous le suggère le pape François dans sa lettre du 1er janvier 2024 à l'occasion de la Journée mondiale de la paix. Dans son texte intitulé « l'IA et la paix », où il reprend notamment une phrase d'Isaïe au chapitre 2 : transformer nos épées en socs de charrue, il s'agit peut-être de transformer nos épées technologiques en socs de charrue pour cultiver et garder la terre. Comme exemple en France, vous connaissez sans doute mon Espace Santé qui est cet espace où l'on peut gérer ses données de santé et bénéficier des 14 millions de personnes qui ont mutualisé leurs données de santé pour créer un élément de santé publique. Toutes ces données sont pilotées par l'IA avec des systèmes de sécurité qu'il faut toujours vérifier mais c'est quelque chose au service du bien commun.

Comment le rendre de plus en plus accessible à tous et de plus en plus sûr avec une sécurité réelle voilà un type d'exemple très concret. En se focalisant un peu plus sur l'éducation, on sait que l'IA générative se diffuse très vite comme le montre une étude publiée par l'Académie de Paris en 2023 auprès de 5600 enseignants et étudiants de l'enseignement supérieur en France. Cette étude montre que 35% des enseignants et 55% des élèves utilisaient régulièrement l'IA et les élèves l'utilisaient beaucoup plus pour l'aide à la rédaction que pour les sources documentaires, ce qui est une bonne chose, sans tenir compte du problème du plagiat, rappelant que l'IA générative est toujours susceptible d'hallucinations, l'IA fournissant aux élèves des références plausibles mais pas toujours correctes.

On se souvient aussi que Chat GPT a brillamment réussi l'examen du barreau de New York, qui cependant est réputé pour être très difficile : c'est un exploit qui est, d'une certaine manière, stupéfiant, comme disent les jeunes. Mais nous savons que l'éducation ne se limite pas à l'accumulation de connaissances. Nous pensons aussi aux images générées par l'IA avec toutes les possibilités de manipulation et de fausses nouvelles. Comme nous recherchons et nous travaillons pour cela, nous pouvons espérer qu'elle offre une solution d'aide personnalisée pour chaque étudiant.

Nous le voyons déjà avec l'expérience de l'apprentissage des langues pour les étudiants qui, par exemple, arrivent en France en voulant apprendre le français, en provenance de nombreuses régions du monde avec des accents qui sont évidemment très différents. Mais la machine à intelligence artificielle est capable de s'adapter à tous les accents et de traduire le français dans de nombreuses langues et inversement. Il n'est pas impossible que les nouvelles

technologies éliminent les barrières linguistiques entre les êtres humains et c'est en cours et sera d'autant mieux dans un futur plus ou moins lointain. Nous pensons que l'apprentissage des langues restera une activité centrale de l'éducation au-delà de toute technique, y compris avec l'aide de la technologie, car comme nous le savons, il s'agit de donner vie aux gens et de les rassembler, bien que de cultures différentes. De plus la linguistique appliquée a démontré que l'apprentissage des langues, pas simplement avec une machine, améliore nos capacités analytiques, en particulier notre aptitude à extraire des règles des usages, ainsi que nos capacités d'abstraction et de généralisation.

Oui, l'utilisation critique de l'IA générative peut offrir de nombreuses possibilités pour soutenir l'apprentissage dans de nombreux domaines mais pour cela nécessite un bon nombre de comportements éthiques et d'apprentissages. Par exemple, faire un bon *prompt*, ce qu'on peut appeler l'art de l'invite, ciblée, efficace, cela s'apprend. Il faut être précis dans les requêtes que l'on fait à la machine intelligente comme Chat GPT, pour bien les situer dans un contexte. Il s'agit aussi de gérer de manière critique les réponses de l'IA générative, par exemple en posant la même question avec des mots différents et donc nécessairement en voyant des réponses différentes et en analysant les différences. Relancer la demande pour spécifier une première réponse de la machine, reformuler les réponses pour être sûr de les comprendre et être sûr de leurs performances, des biais, hallucinations, bref, travailler de manière itérative.

Pourquoi ne pas utiliser la machine Intelligence artificielle par exemple pour produire un premier jet sur un sujet donné, pourquoi pas, à condition qu'on s'interroge, que l'on mette notre marque dessus à partir de la toile donnée par la machine et que l'on sache rendre compte intelligemment de ce que la machine a pu produire et de ce que l'on a pu en faire, l'essentiel étant toujours d'avoir ce sens critique, ce sens de la curiosité et l'art de l'argumentation. Selon moi, c'est dans les pédagogies actives aujourd'hui, de construire un texte avec plusieurs humains, un peu sur le modèle de Wikipédia, et ensuite de le passer au moulin de la machine intelligente pour avoir les réactions de la machine afin de compléter le premier texte à partir d'idées données par la machine, évidemment en faisant preuve d'esprit critique.

Bien sûr en éducation il faut plus que trouver une bonne opérationnalité des machines : parmi les points durs des mentalités engendrées par les machines intelligentes, on retrouve la définition matérialiste de la pensée humaine que l'on appelle computationnelle. Elle serait réduite à un traitement de l'information, comme l'indiquent certaines tendances de ce qu'on appelle aujourd'hui la philosophie de l'esprit, pour laquelle la machine devient de plus en plus analogue au cerveau humain. On voit des discours et articles dans quelques magazines selon lesquels si on enlève le couvercle d'un ordinateur on voit un réseau de câbles et de transistors et si on enlève le couvercle de notre cerveau, notre crâne, on verrait une masse de veines et de nerfs : nous serions des machines biologiques, et des machines artificielles seraient déjà capables de résoudre des problèmes de la même manière que les humains. Nous entendons parfois ce discours. Il est essentiel de ne pas rester trop réductionniste.

Personnellement j'ai montré de grandes différences entre l'intelligence des machines et l'intelligence humaine, notamment en termes d'importance du corps, de l'incarnation dans la manière d'éprouver des émotions et non de les simuler, et avec des neuroscientifiques comme Damasio rendant les liens entre émotion et raison avec les mécanismes d'accès à la conscience beaucoup plus complexes que ce produit sur silicium. Oui, la machine intelligente peut avoir une forme d'intelligence et de créativité que les statistiques permettent, comme c'est le cas pour les protéines des lauréats du prix Nobel de chimie que j'évoquais précédemment : la machine leur a, en fait, permis de trouver des combinaisons d'atomes et de molécules que leur pensée personnelle ne permettait pas de trouver, qu'ils n'avaient pas imaginées et du coup de les tester et évidemment de n'utiliser que celles qu'ils ont su retenir comme efficaces. Mais c'est le jeu des statistiques sur de grandes masses de données.

La pensée humaine n'est pas réduite à un jeu de statistiques même si elle a aussi un aspect calculateur : elle n'est pas réduite au quantifiable, au mesurable. Certes, les vies humaines sont en partie dans ce mesurable mais aussi bien au-delà, dans une singularité qui ne se réduit pas à des statistiques, dans le non-quantifiable qu'on appelle la créativité humaine, la capacité de sortir de la logique connue, ce que les poètes, les artistes et les spirituels nous montrent si souvent : le cœur de l'homme n'est pas un réseau de neurones artificiels, même sophistiqué. D'ailleurs c'est souvent de notre vulnérabilité justement ou même de nos fissures que la créativité prend son ampleur chez l'homme. Eradiquer notre déterminisme biologique et notre vulnérabilité transforme l'homme en cyborg.

Le cyborg est la fusion de l'homme et de la machine invulnérable et c'est justement ce qui nous fait sortir de l'humain : la machine reste un objet qui n'a pas de sensibilité propre, ni de cœur au sens de conscience profonde, ni de volonté, ni d'intention propre au sens anthropologique, C'est pourquoi il me semble essentiel dans l'éducation de bien différencier le rapport à une machine et les rapports humains ; le rapport à une machine qui est un objet, qui reste un objet même si c'est un objet parlant et certes aussi efficace mais qui reste du silicium avec les rapports entre les humains et les rapports de l'homme avec l'ensemble du monde vivant à l'heure de l'écologie intégrale. Ces différences entre rapport aux choses et les relations humaines sont particulièrement essentielles à distinguer même si je fais partie de ceux qui pensent que le terme d'intelligence n'est pas complètement usurpé pour une machine du point de vue de la capacité d'une machine à s'adapter à une situation nouvelle mais sans conscience, sans volonté.

J'ajoute tout de suite que cette intelligence est précisément artificielle, différente de l'intelligence humaine avec ses compétences très développées aujourd'hui dans la science de l'éducation, intelligences notamment émotionnelles, existentielles qui viennent croiser l'hypothético-déductive. L'intelligence, certes, de Chat GPT est de plus en plus impressionnante quant aux réponses qu'elle donne aux questions les plus subtiles mais il n'y a pas, il n'y a rien de magique ici, c'est le jeu des statistiques à grande échelle, la probabilité qu'à un mot en succède un autre dans la réponse ainsi construite et c'est justement, il me

semble, en apprenant à différencier ces différents types d'intelligence, que l'homme sera d'autant plus libre d'utiliser la fameuse boussole pour vivre et vivre notamment avec les machines dites intelligentes sans confusion et en essayant de les mettre au service du bien commun comme le pape François nous invite à le faire dans la lettre que j'ai initialement citée du 1er janvier 2024. Mettre l'IA au service de l'éducation pour la paix et non pas le remplacement de l'humain par des machines, c'est tout un programme.

Didier Coulomb

En effet, il y a beaucoup à faire... et en particulier dans l'éducation. Nous allons écouter Alexandra Maurice qui est en charge de l'innovation et du numérique à l'Institut Supérieur de Formation de l'Enseignement Catholique de Nouvelle-Aquitaine.

Alexandra Maurice

Quelques petits éléments de contexte. J'ai voulu à mon tour aborder ce thème et compléter les remarques sur les usages de l'IA dans l'éducation : on va faire un petit détour par l'émerveillement. Le monde dans lequel nous vivons aujourd'hui est un monde de données. On en produit, on le fait en *likant*, en le consommant, en allant sur Internet, finalement c'est l'homme qui a fait ce monde. Il ne serait pas un peu responsable de ce qui lui arrive aujourd'hui ? Depuis 2022 et le lancement de Chat GPT pour le grand public, tout le monde peut bénéficier d'une IA générative partout, tout le temps, à n'importe quel moment et sur n'importe quel appareil. L'IA est et va être omniprésente. Il va falloir faire face à cela dans un monde où l'homme sera assisté, ou pourra l'être, s'il le souhaite, par des machines.

La cobotique, c'est-à-dire la collaboration homme-machine, sera de plus en plus fréquente. Certains y verront une mode, d'autres une vraie révolution. Il y a 2 ans, on parlait d'un tsunami, une catastrophe pour de nombreux enseignants. En tant que responsable de l'innovation pédagogique, c'est mon rôle de voir venir la vague, de détecter les signaux faibles. D'autres l'ignorent mais tous en entendent parler aujourd'hui. Il faut donc former des enseignants qui pourront alors sensibiliser leurs élèves et avoir une vraie éducation à l'IA. Mon rôle de formatrice et vulgarisatrice techno-pédagogique, c'est de leur faire découvrir l'IA et comprendre qu'ils peuvent en tirer le meilleur parti tout en étant conscients des mécanismes commerciaux qui vont de la commercialisation de leurs données à la supervision de l'IA, que les IA sont omniprésentes et comment elles fonctionnent.

La formation que nous donnons à l'ISFEC François d'Assise sert à démystifier pour rassurer mais aussi informer. Expérimenter et débattre parce que les enseignants, comme de nombreux citoyens, partagent des croyances et peuvent avoir peur de ce qui est nouveau et de ce qu'ils ne maîtrisent pas. Ils ont besoin d'être guidés et soutenus. La question que je vous pose ce soir à vous aussi : vous diriez que l'utilisation des intelligences artificielles et des intelligences artificielles génératives dans un contexte professionnel est fabuleuse et magique, menaçante et angoissante ou un peu des deux ? Votre réponse est surtout un peu des deux. Donc la

formation devient indispensable pour faire découvrir des éléments importants et démystifier certaines choses que je partage avec vous ici.

Quelques exemples de découvertes faites par nos apprenants lors de notre formation découverte :

- L'intelligence artificielle générative est une toute petite partie de l'intelligence artificielle ; la sensibilisation permet effectivement d'éviter les raccourcis.
- L'IA aura toujours besoin de l'homme pour la créer, pour la faire fonctionner, l'entretenir et la réguler.
- Chat GPT n'est pas seul : chaque outil a sa spécificité liée à son origine culturelle et géographique : il a des fonctionnalités spécifiques, il a son algorithme initial, fait de données sélectionnées précisément par l'homme.
- L'IA n'est pas un humain, on a trop tendance à l'anthropomorphisme. Voici un exemple : un humain demande à une IA qui modélise une araignée d'optimiser ses mouvements pour toujours franchir une certaine hauteur : l'algorithme répond à la commande en faisant « la roue ». L'humain pensait en fait que l'araignée allait sauter. Ce n'était pas le cas et il ne s'attendait pas à ce qu'elle fasse une roue, ce qui lui a permis du moins ce qui a permis au modèle, de franchir une hauteur supérieure à ce que pouvait anticiper l'humain. C'est juste pour illustrer qu'il est très difficile aujourd'hui de faire comprendre ce que l'on veut vraiment, ce que l'on a dans le crâne, comme l'a dit M. Magnin tout à l'heure, d'où l'importance de la formation à l'écriture des invites, par exemple, et de l'itération avec ces agents conversationnels.

On a souvent tendance à l'anthropomorphisme : on parle à Chat GPT. Si on échangeait avec lui en anglais ou en allemand on n'aurait pas de problème car le neutre existe mais pas en français. Il y a même une intelligence artificielle génératrice de texte qui s'appelle Claude, pour anthropomorphiser la machine encore plus.

J'estime qu'une formation est réussie quand les enseignants ont pu échanger sur leurs croyances et les faire évoluer, qu'ils ont pu expérimenter par eux-mêmes, débattre des enjeux et qu'e dorénavant ils seront prudents. Comme toute nouvelle technologie, il faut savoir écouter les peurs bien réelles, la désinformation intellectuelle, les biais et les hallucinations des IA génératives et le manque de créativité que cela pourrait susciter, mais surtout il faut éviter le sensationnalisme. Il faut rester vigilant donc il faut un guide, une boussole pour nous orienter dans cette évolution. C'est pourquoi la boussole que nous proposons avec ses quatre axes est un excellent outil pour guider dans cette éducation, pour éveiller, générer le débat et informer les communautés éducatives.

On va commencer par le Nord : si on regarde la flèche du Nord de cette boussole, elle nous indique que l'esprit critique doit être central dans l'apprentissage. Nous devons tous être capables de questionner les résultats produits par les algorithmes, pour évaluer la fiabilité de l'information et pour identifier les biais potentiels pour les jeunes. Cela inclut la préparation des jeunes à l'enseignement supérieur, leur orientation mais aussi leur rôle de citoyen dans un monde de plus en plus automatisé.

Selon la flèche du Sud, nous devons cultiver la curiosité face aux avancées technologiques. Plutôt que de les craindre, il faut encourager une ouverture d'esprit face aux innovations qu'elles peuvent apporter, rester informé, rester vigilant sur le sujet. Cela implique aussi une participation active au débat sur les implications sociétales et éthiques de l'IA. Par exemple, j'ai créé un jeu de simulation où les apprenants incarnent des invités qui représentent divers secteurs de la société mondiale et leurs opinions lors d'une conférence internationale sur le sujet : s'agit-il d'une menace pour l'humanité ou d'un avenir prometteur ?

La flèche suivante est l'attitude responsable. L'utilisation doit toujours être accompagnée d'une réflexion morale. Cela signifie apprendre à utiliser cette technologie dans des cadres éthiques et dans le respect des valeurs fondamentales telles que la confidentialité, la justice et l'équité. Les usages que nous proposons doivent être pensés et encadrés par des principes pour faire en sorte que l'IA soit un outil au service de l'humain et non l'inverse. Par exemple ce fameux rapport 101 thématique 3 de la délégation pour l'éducation prospective du Sénat en date du 30 octobre et bien sûr toutes les recommandations récentes de l'UNESCO sur le sujet.

Pour la flèche de l'Ouest peut-être la plus difficile à transmettre mais néanmoins la plus fondamentale qui guide nos actions de formation car il est crucial de cultiver les aspects de l'humain que la machine ne peut pas imiter comme la créativité, l'empathie, l'émerveillement. Je vais vous donner un exemple très concret : un de mes étudiants lors d'un cours sur l'IA m'a demandé la permission de photographier le coucher de soleil. Je l'y autorise et je le remercie de m'avoir montré à quel point un humain a la capacité de s'émerveiller de ce spectacle.

L'IA ne doit pas être considérée comme une force remplaçant l'humain mais comme une opportunité de renforcer notre singularité. Nous devons en même temps encourager la réflexion spirituelle et philosophique sur la place de la technologie dans nos vies. Cela place aussi la réflexion sur notre capacité à nous émerveiller, à croire, c'est-à-dire à nous laisser habiter par quelque chose qui nous dépasse, à chercher du sens dans le monde, à aller à la rencontre de la beauté et de la sensorialité, imaginer ce que l'on ressent. Lorsque vous êtes près de la mer ou de l'océan, tous vos sens sont en éveil simultanément : l'ouïe par le bruit des vagues, l'odeur par le parfum des embruns, le toucher avec le vent qui caresse votre peau, la vue par le spectacle des vagues ou l'immensité, le goût de l'iode qui se dépose sur votre peau et tous les souvenirs associés à cette expérience, c'est ce que l'IA ne sera jamais capable de sentir et ne pourra jamais retranscrire.

Toutes les orientations de la boussole sont bien évidemment dépendantes et complémentaires les unes des autres, il faut considérer l'approche comme étant complètement systématique.

Pour finir, l'éducation à l'IA et avec l'IA est un défi qui nous concerne tous et c'est en unissant nos efforts et nos pensées que nous pouvons créer un avenir où la technologie amplifie notre humanité plutôt que de la diminuer, c'est aussi admettre que les plus beaux chapitres de l'histoire humaine n'ont pas été écrits par des machines mais plutôt avec des rêves, des sacrifices et l'ambition de laisser à tout prix un monde meilleur à nos enfants.

Didier Coulomb

Merci Alexandra. Je passe maintenant la parole à Louis Lourme. Vous êtes Recteur des Facultés Loyola et Docteur en philosophie de l'université de Bordeaux. Vous avez écrit des livres dont je recommande la lecture. Louis, vous êtes le dernier orateur de cette table-ronde. Pouvez-vous conclure comment l'IA peut changer notre vie et comment il faut organiser l'enseignement de l'intelligence artificielle ?

Louis Lourme

Merci pour cette invitation à cette table ronde et pour la manière dont la table ronde est organisée : cela nous permet d'avoir des interventions très variées sur la question qui nous concerne. La mienne clôt cette première prise de parole, donc je vais aller à l'essentiel et je vais m'appuyer sur des penseurs de l'objet technique en me disant que cela pourrait peut-être intéresser certains pour approfondir les pistes évoquées. J'ajoute pour finir cette petite introduction que je suis moi-même un utilisateur très régulier de ce qu'on appelle l'intelligence artificielle. Pourquoi je le précise ? Parce que, comme je m'apprête à développer un point qui nuance notre relation avec l'intelligence artificielle et les questions éthiques qui y sont associées, je ne veux pas donner l'impression de tout rejeter en bloc, ce n'est pas du tout ma position.

Mon point de départ pour cette brève prise de parole, c'est un étonnement très personnel devant l'émergence de l'intelligence artificielle et les réactions que cette émergence a suscitées dans tous les groupes que j'ai pu rencontrer et peut-être en particulier dans le domaine éducatif qui était celui dans lequel je travaillais au moment de cette émergence. Je crois pouvoir dire que cette émergence a été quelque chose comme un choc, cette émergence a été très soudaine. Tout d'un coup effectivement il fallait en parler, il fallait réagir, avoir un avis et on a parlé de cette étrange temporalité tout d'un coup même si ces technologies sont le fruit d'une longue période de développement et de recherche.

Il y a une accélération à laquelle on a été soudainement appelé à réagir et ce qui m'intéresse, c'est avant tout ce que ce choc a provoqué, c'est-à-dire les réactions chez les personnes qui se sont retrouvées brutalement confrontées à cela. C'est peut-être un peu caricatural de le dire ainsi, mais je pense qu'on peut dire que ces réactions étaient généralement des réactions d'inquiétude à propos de ce nouvel objet, je ne dis pas qu'il ne s'agissait que de cela, mais tout de même, dans le monde des enseignants en particulier, la première réaction a été celle de l'inquiétude, sans parler par exemple de ceux pour qui c'était vraiment une catastrophe, et je pèse mes mots.

Maintenant que le temps a passé par rapport au choc dont je parlais tout à l'heure, je pense qu'on peut essayer de distinguer différentes raisons à cette préoccupation, c'est ce que je veux essayer de faire rapidement. Il y a tout d'abord, je crois, le sentiment de danger, dans cette inquiétude : nous avons déjà parlé de ce qui est mis en danger dans nos habitudes ou nos pratiques par l'intelligence artificielle. Nous en avons parlé même dans le cadre de la précédente table ronde, je crois. Mais peut-être pouvons-nous simplement observer ici que

notre attitude envers l'intelligence artificielle est une attitude au moins en partie contradictoire à partir de ce point de vue qui retrouve un peu des intuitions de Simondon dans son travail de 58 sur le mode d'existence des objets techniques.

De ce travail, tout le monde peut lire l'intro avec beaucoup de bienfaits : ce qu'il dit des objets techniques décrit assez bien notre rapport à l'IA aujourd'hui. Il dit que nous les traitons d'une part comme de purs assemblages de matière dépourvus de sens et qui présentent simplement une utilité pour nous tout en supposant par contre qu'il y a aussi un danger, presque une intentionnalité hostile, ce sont ses mots. Eh bien, il me semble que cela en dit un peu sur les premiers instants de réaction, sur le sentiment de danger et quand on essaie de comprendre cette inquiétude dont je parlais plus tôt, il y a probablement quelque chose comme le sentiment de perte ou de blessure comme si on nous emportait une caractéristique de notre humanité.

Je parle de blessure pour faire référence à la notion de blessure narcissique que l'on retrouve chez Freud. Vous savez que l'humanité a vécu trois grandes blessures narcissiques, c'est-à-dire des blessures infligées à l'idée que l'homme a de lui-même : avec Copernic tout d'abord, l'homme n'est plus le centre de l'univers ; avec Darwin, l'homme n'est plus le centre de la création ; avec Freud, c'est Freud qui parle, c'est sa formule que l'on connaît : je ne suis plus maître dans ma propre maison. Ce sont les trois blessures narcissiques et très probablement la crainte que le développement de l'intelligence artificielle fasse partie de cet ordre, une forme de 4ème blessure narcissique que l'humanité s'impose à elle-même : le sentiment d'être en partie dépouillé non seulement de la maîtrise de notre pensée consciente, ce qui a été le cas de la troisième blessure narcissique, mais de notre exclusivité dans la pensée complexe elle-même.

Donc quoi faire et en particulier, qu'est-ce que cela signifie de dire, comme nous le prétendons penser ensemble ici, que nous devons éduquer à une utilisation responsable ou éthique de l'IA ? La question n'est pas simplement de savoir quoi faire avec l'intelligence artificielle mais que faire de l'injonction à l'utilisation éthique de l'intelligence artificielle. Je voudrais justement sur ce point, proposer deux pistes critiques dans les quelques minutes qu'il me reste. Ma première remarque concerne la décorrélation qui s'opère entre la technologie et l'éthique. On s'y habitue. C'est une décorrélation qui me semble périlleuse. Je crois qu'il faut se méfier de ce qu'on pourrait appeler une conception du pardon de la technique, ou pour le dire plus simplement, je crois qu'il est bon de se méfier des discours qui font de l'objet technique un objet qui serait parfaitement neutre sur le plan éthique même si c'est la première intuition.

Est-ce vraiment le cas ? Une question importante chez les philosophes de la technique, c'est par exemple ce que dit Jacques Ellul : la possibilité d'une utilisation éthique d'un objet technique se réfère à un présupposé qui n'est pas anodin et qui consiste à dire que l'objet technique serait neutre en lui-même sur le plan éthique et que ce serait notre usage seul qui pourrait être non éthique. Disons que cette éthique d'utilisation est basée sur l'idée que l'on

pourrait utiliser un outil, une machine, bien ou mal, c'est certain bien sûr personne n'en doute mais on associe deux points à cela : d'abord que l'outil ou la machine n'a aucune valeur morale en soi et ensuite que c'est l'utilisateur, vous et moi qui avons tout le poids éthique sur les épaules.

Que l'utilisateur porte ce poids n'est pas une évidence. L'objet technique ne vient pas de nulle part, Chat GPT ne sort pas de nulle part. Il a nécessité des conditions de production qui ne sont pas éthiquement neutres et il répond à une intention qui ne l'est pas non plus. C'est tout pour le premier point. Ensuite, l'environnement technique que nous nous sommes construit contribue en partie à la détermination de nos repères éthiques contemporains. Pourquoi dire cela ? Certainement pas pour dire que tout cela est très mauvais mais plutôt pour remettre notre responsabilité éthique à sa juste place car c'est très malin de dire à un technicien, construisez un objet neutre et c'est vous qui vous occuperez de son utilisation éthique ou non. Voilà donc ma première remarque, l'éthique ne vient pas seulement en second lieu ou après coup, elle est présente dans la constitution de l'objet technique.

Ma deuxième remarque sera aussi mon dernier point. Ce sera une évocation seulement. Elle consiste à prendre au sérieux la remarque qu'Hans Jonas a développée sur ce que l'on pourrait appeler la temporalité de l'éthique dans une conférence remarquable de 72 dont l'énoncé du titre suffira à montrer son intérêt pour nous, je vous le donne : « Technologie et responsabilité des points, réflexion sur les nouvelles tâches de l'éthique, technologie et responsabilité ». Il a cette très belle formule : l'univers éthique est composé de contemporains. Ce qui le préoccupe, c'est qu'en éthique on procède souvent par vieux réflexe, on reproduit des habitudes éthiques, mais comme il le dit, pour une grande partie les leçons de l'expérience sont impuissantes, et son travail de philosophe a consisté à montrer qu'une nouvelle éthique était nécessaire pour un monde post-moderne.

Le principe éthique qu'il construit s'appelle le Principe de Responsabilité. C'est son travail de 79 qui a été très significatif. Je crois que si l'on veut réfléchir sérieusement à une éducation à une utilisation éthique de l'intelligence artificielle, il peut être pertinent de rappeler la manière dont il a dénoncé une dérive utopique qui contribue à créer, selon lui, un nouveau besoin de sagesse, oui, mais une sagesse d'un type nouveau et d'une ampleur nouvelle, ce qui a inspiré cette très belle observation que je vous donne à finir comme des phrases à méditer pour nous : « Nous avons le plus besoin de sagesse quand on y croit le moins. »

Didier Coulomb

Je vous remercie, je remercie en vous le philosophe comme l'éducateur, et propose maintenant que les participants dans la salle et sur internet posent des questions à nos orateurs.

Débat

Eric Castelnaud

Je vais peut-être commencer par une question qui nous vient des internautes : vous avez dit, mon Père, que vous n'êtes pas contre le fait que l'IA ait une forme d'intelligence humaine mais sans conscience. Pourtant certains disent qu'il y a une conscience.

Thierry Magnin

En fait l'intelligence artificielle aujourd'hui, en tout cas les machines d'intelligence artificielle, n'ont pas de conscience et ne comprennent pas ce qu'elles disent. C'est quelque chose de très important. Elles répondent par un calcul statistique. Donc, en ce sens, elles n'ont pas de conscience. Et ce que j'ai essayé de montrer à plusieurs reprises, c'est que les neuroscientifiques d'aujourd'hui nous disent que pour qu'il y ait une conscience qui ressemble à une conscience humaine, il doit y avoir un corps. Il n'y a pas de corps sans conscience, il n'y a pas de conscience sans corps. J'appelle souvent l'intelligence artificielle, une intelligence sans corps et qui est capable de simuler et de simuler remarquablement bien d'ailleurs, qui est capable de produire des choses que nous n'avions pas imaginées nous-mêmes, mais sans émotions, sans raison et sans conscience : la place du corps est primordiale dans l'intelligence de type humain.

Un participant : Je profite de ma retraite pour offrir des cours particuliers du soir à mes petits-enfants. Ma question porte justement sur l'éducation parce que ce qui a été dit dans cette conférence où l'on met en avant le corps et le dialogue, revient constamment dans les commentaires des intervenants mais l'éducation nationale de notre pays aujourd'hui, malheureusement, ne met pas en œuvre le corps et le dialogue dans son système éducatif. Les enfants doivent être assis, ne pas bouger, ne pas dialoguer et accepter qu'un professeur leur dise ce qu'ils ont besoin de savoir. Je me demande si l'UNESCO peut avoir un peu d'influence sur le système éducatif français ?

Didier Coulomb : Cela lui donne peut-être beaucoup d'ambition mais normalement l'UNESCO a une vocation universelle donc elle peut peut-être donner quelques inspirations. Parmi les éducateurs : Louis Lourme, avez-vous des commentaires à faire sur cette exigence d'éducation sur le statut du corps dans le système éducatif ?

Louis Lourme : Écoutez, je vais vous donner une réponse. Je ne partage pas votre impression, c'est-à-dire, je ne crois pas que le corps soit écarté par l'Éducation Nationale. Je crois aussi qu'un corps qui ne bouge pas et un esprit qui ne dialogue pas avec l'enseignant en permanence ne nous ont pas empêchés dans la construction des gens que nous sommes devenus. Je ne suis pas sûr que ce soit quelque chose qui repose encore sur les épaules de l'école et je ne suis même pas sûr qu'on puisse vraiment reprocher à l'école d'aujourd'hui de ne pas prendre en compte le corps de l'enfant. Je crois vraiment que c'est une nouvelle préoccupation et quelque chose qui émerge par rapport à ce que nous avons connu, il me semble.

Une participante : Je voudrais m'adresser au dernier intervenant car vous avez beaucoup piqué ma curiosité en disant que vous utilisez énormément l'intelligence artificielle. Pourriez-vous nous donner des exemples ?

Louis Lourme : Je l'ai fait exprès pour aiguïser la curiosité mais vraiment je l'utilise beaucoup dans des situations très différentes et avec, je crois, un peu de conscience des limites qui y sont associées. Je vais vous donner deux ou trois exemples très concrets. Dans la dernière version pro de Chat GPT par exemple vous pouvez forcer Chat GPT à travailler sur un corpus que vous lui donnez : vous ouvrez une porte, vous mettez les documents que vous voulez, vous fermez et puis quand vous échangez avec lui, il s'appuie sur les textes que vous lui avez donnés. Je suis allé par exemple prendre tous les documents du concile Vatican 2, je les ai mis, j'ai fermé, je l'ai appelé GPT Vatican 2 et si je lui dis : fais-moi un lexique des 20 termes clés du concile Vatican 2, eh bien ça me produit un lexique de 20 termes clés de Vatican 2. Cela ne veut pas dire que je ne priorise pas comme je le veux.

Je vous donne d'autres exemples sur la doctrine sociale de l'Église. Je suis allé sur le site du Vatican. J'ai pris les différentes encycliques qui composent le corpus classique de la doctrine sociale de l'Église, et je les mets au format pdf. Je les lui ai donnés (et on peut le prendre dans les langues qu'on veut bien sûr, vu que cela l'indiffère). Je les ai mis dans un dossier que j'ai fermé et je peux alors lui expliquer que je ne connais pas la notion de bien commun. Je demande : « Comment structureriez-vous une réflexion sur le bien commun à partir de ces documents ? »

Je l'ai fait aussi pour la pédagogie ignacienne : j'ai pris tous les textes de notre tradition pédagogique depuis 60 ans environ, les grands textes des supérieurs généraux, les textes produits par les responsables de formation, je les ai mis dans un GPT, j'ai clôturé. Les résultats sont bluffants. Si je lui dis : « Prépare-moi une séance de 20 minutes pour expliquer à un élève de CM2 ce que c'est que le tribunal personnalisé, sachant que cet enfant n'a pas de culture chrétienne, vous commencez par un exemple, vous mettez quelques blagues, mais je veux qu'on arrive à quelque chose de sérieux ». Si je lui dis ça, il fait un boulot pour moi, ce qui est très bien, je ne prendrai pas forcément tel quel cet exemple. Il répond : « Tu sais chez tes parents il y a plusieurs plantes vertes, certaines ont besoin d'eau et moins de soleil, d'autres ne peuvent pas être en lumière directe, etc. Si vous donnez à toutes les plantes la même quantité d'eau et de soleil, eh bien il y en a qui vont périr. »

C'est remarquable comme image et c'est une aide à la réflexion, en gros c'est une bonne synthèse ; j'y vois un intérêt. Si je lis les textes de la doctrine sociale de l'Église, quand je lis *Rerum Novarum*, évidemment j'ai un parti-pris. Je le lis en sachant ce qu'il représente et je ne vais pas lire tous les textes de la même manière. C'est ce qui rend ma propre lecture intéressante. Je le lis en étant chargé de ce que je suis et on peut considérer que c'est aussi intéressant qu'il puisse être lu par un autre biais, un parti pris (qui n'a pas de parti pris ?). Je ne dis pas c'est mieux, je dis en tout cas que si je lui demande de me faire ce que je lui ai demandé, de me faire une recherche triviale sur la doctrine sociale de l'Église, eh bien, il me

le fait à partir d'une lecture d'une suite de 0 et 1 où tout est de niveau. Ce n'est pas mieux mais ça apporte quelque chose.

Deux dernières remarques. La première c'est l'intérêt pour les langues : j'ai pu lui donner des textes polonais, des textes allemands, c'est pour moi utile de le nourrir avec toutes ces langues et que son travail dépasse cette limite. Je ne dis pas qu'il n'y a pas d'intérêt à rencontrer la langue de l'autre, mais c'est aussi intéressant de pouvoir l'utiliser comme ça. Et puis il y a autre chose c'est le calcul, je le prends comme une calculatrice. Si vous voyez les réactions des collègues dans la salle des enseignants, ils sont terrifiés, pensant que le monde s'effondre, etc. Je dois avouer que c'est une langue que je ne comprends pas bien, il ne viendrait pas à l'idée d'un prof de physique ou de maths de se dire : je sais que vous avez une calculatrice mais faites-le sans calculatrice, ça change juste notre rapport ; en tout cas on pourrait en débattre. J'ajoute que c'est une puissance de calcul tout de même !

Si je dois passer un accord entre mon établissement et un autre établissement, ceci m'est arrivé il y a 9 mois quand j'étais chef d'établissement à Bordeaux, je lui ai dit : « Voilà je suis tel établissement. Je veux passer un accord avec tel ou tel établissement. Voici nos caractéristiques, voici les leurs, voici ce que je voudrais, pourriez-vous me faire un modèle de convention ? ». Il me fait un modèle de convention. Vous allez me dire mais il y a plein de gens qui savent faire ça. Oui, mais il l'a fait pour moi en 15 secondes. On l'a fait regarder par nos avocats. C'est fait mais c'est quand même formidable. De la même manière, j'étais en train de passer un autre accord que je lui ai donné lui disant que dans cet accord je suis tel et tel établissement, quel problème voyez-vous pour ma structure ? Cela ne veut pas dire que je ne l'ai pas donné à mon RH, que je ne l'ai pas donné à mon DAF et à mon avocat, mais c'est quand même assez amusant et assez stimulant d'avoir ce type d'aide. Mais ce n'est que de l'aide pour tous les exemples que je donne.

C'est seulement une aide que j'ai en une seconde ou pas. J'ai fait un GPT personnel privé mais tous ceux dont je vous ai parlé sont publics, sont accessibles. Je travaille à Abidjan sur la doctrine sociale de l'Église : si je vais sur GPT, je peux trouver le GPT qui s'appelle doctrine sociale de l'Église où il y a la liste de tous les textes que j'ai intégrés. Pour que ce soit transparent je l'ai fait aussi en termes privés pour mon propre travail. J'ai pris ma thèse, tous mes articles, mes livres, je les mets à un endroit, je l'ai fermé, j'ai appelé ça une recherche personnelle et ça veut dire que j'ai accès à ce que j'ai écrit il y a 15 ans, même si je ne l'ai pas parfaitement présent en tête tout le temps. Je ne dis pas qu'il n'y a pas d'intérêt à ne pas l'avoir présent à l'esprit tout le temps, mais ce que dit Nietzsche en particulier, c'est qu'il y a des vertus d'oublier.

Nous sommes des êtres humains parce que nous devons oublier. Ça a quand même un intérêt, ce GPT de recherche personnelle, dites-lui : « Donnez-moi une mise à jour sur le concept de Parlement mondial, vous gardez les trois idées clés de mon travail ». Il me les fait ressortir. C'est une sorte de réactivation, c'est un aide-mémoire, c'est très riche et très pratique.

Didier Coulomb

Est-ce que d'autres personnes de la table ronde, en ligne ou pas, ont des commentaires ?

Monique Grandjean

Vous avez dû vous rendre compte tout à l'heure que j'étais tout à fait d'accord avec Ellul et d'ailleurs avec Paul Ricoeur aussi pour dire que c'était cet usage qui était le plus important, la chose la plus importante par rapport à l'objet qui est devant nous. Il y a une chose que vous comprendrez très bien, j'ai un piano en face de moi, si celui qui est derrière le piano est un artiste accompli, vous aurez du Schubert et du Chopin, ce sera magnifique. Si vous avez un enfant de 3-4 ans ou même si vous avez quelqu'un qui n'a aucune technique ou aucune créativité du tout, vous n'obtiendrez aucun résultat du piano.

Pourquoi avez-vous dit plus tôt, et dans quelle mesure pensez-vous que c'est totalement différent quand vous êtes en présence des nouvelles technologies, qu'à ce moment-là ce n'est plus l'usage ou la manière d'utiliser ou la neutralité ou pas la neutralité de la personne qui est derrière l'objet en question et qu'une nouvelle éthique serait nécessaire devant un nouvel objet ?

Louis Lourme

Je pense que je suis allé trop vite parce que ce n'est pas exactement ce que je voulais dire donc c'est très bien que vous me donniez l'occasion de clarifier.

Ellul ne dirait pas exactement que le marteau est neutre et que je peux l'utiliser pour construire des maisons pour les sans-abris ou pour attaquer quelqu'un. Si j'ai une conscience ou si je suis quelqu'un qui sait choisir entre le Bien et le Mal, je vais utiliser le marteau de telle ou telle manière. Oui, cela va de soi et je ne dis pas : sortons nos marteaux et attaquons-les. Bien sûr je suis tout à fait d'accord et Ellul dit bien sûr qu'il y a une utilisation éthique ou non et que notre travail en tant qu'humains est de réfléchir à ce que sera le meilleur usage que nous puissions faire de ces outils techniques, mais il dit qu'il y a un système technique, c'est-à-dire qu'il y a une façon de développer la technologie qui enlève progressivement la question même de l'éthique. Il existe une façon de développer la technologie qui élimine progressivement la question du comment de notre environnement ; ce n'est pas que le marteau lui-même serait bon ou mauvais, c'est juste que nous construisons un système, nous construisons un environnement.

Nous sommes de plus en plus dépendants du comment, c'est plutôt une façon de dire que nous avons besoin de plus en plus de la question du pourquoi, mais ce n'est pas seulement une question entre l'objet et moi, cela surgit entre l'objet et moi : quel usage je vais faire de cet objet technique. Mais se pose aussi, plus largement, la question éthique et il ne faut pas

croire qu'elle repose uniquement sur l'épaule de l'utilisateur, c'est le premier point. Le deuxième point, c'est cette nouvelle éthique dont parle Hans Jonas, c'est une éthique qui correspond pour lui à un moment nouveau dans notre rapport à la planète, c'est un moment où le développement technique est tel qu'il met en danger l'avenir et donc on ne peut plus penser à cet état de fait armé seulement des outils éthiques qui nous précèdent. C'est pourquoi il produit le Principe de Responsabilité, la formulation du Principe de Responsabilité, très simple, en date de 79. Maintenant, si nous voulons faire une formulation à la manière de Kant, agissez de telle manière que votre action actuelle soit compatible avec la permanence d'une vie humaine authentique sur terre, voici le principe de précaution aujourd'hui.

Troisième Partie

Repères et pistes pour éduquer

Isabelle Chaperon, présidente du CCIC

Bienvenue ce soir chez « les Apprentis d'Auteuil » que nous remercions chaleureusement de leur accueil. C'est le troisième des « tiers-lieux » emblématiques de lieux d'éducation sous des formes très diverses. L'institution met en œuvre depuis de nombreuses années l'accueil, l'accompagnement et la formation professionnelle d'un grand nombre de jeunes dont la plupart vivent des situations familiales difficiles. Or, les familles, nous ne nous lasserons pas de le répéter, sont les premiers éducateurs de leurs enfants,

Nous vous avons invités à un questionnement : « Comment éduquer à un usage responsable des intelligences artificielles ? », avec l'objectif de contribuer à une mise en œuvre concrète du volet éducation de la recommandation de l'Unesco sur l'éthique de l'intelligence artificielle.

Les regards croisés des panélistes de ces trois tables-rondes abordent la question sous des angles différents et complémentaires, ils appellent les usagers et les formateurs à des points essentiels de vigilance, et nous ouvrent des pistes pédagogiques nécessaires dans le paradigme nouveau de l'omniprésence des IA

Telles que, entre autres :

- La nécessité d'une boussole éthique, de l'éveil de la conscience des jeunes et de leurs responsabilités,

- La connaissance de soi, l'estime de soi en tant qu'être humain singulier, le rapport aux autres, à l'altérité, le sens du débat, le sens du concret, tout cela doit être développé ou renforcé.

Ces approches à la fois philosophique, psychologique, scientifique, il serait important des intégrer en tant que *soft skills* à l'enseignement de ce qu'on appelle la littératie numérique, et plus largement dans des modules pédagogiques transversaux.

L'objectif est que l'humain grandisse en humanité et qu'il sache mettre ses créations technologiques à son service et à celui du bien commun dans le respect de son environnement.

Chers panélistes, un immense merci de venir partager ce soir votre expérience et votre vision.

Je tiens à associer à ces remerciements l'équipe du CCIC qui a pensé et réalisé ce cycle de tables-rondes, en particulier le vice-président et chef de ce projet Eric Castelnau. Nous sommes très reconnaissants au CFRT/Jour du Seigneur pour la mise en œuvre des moyens techniques de diffusion de cette soirée.

La parole est à vous, chers orateurs, et à notre modérateur Marc Deluzet, Délégué général du CCIC.

Marc Deluzet

Merci Isabelle. Cette troisième table ronde est effectivement davantage orientée vers les solutions, les pistes et les repères en direction des éducateurs au sens large, familles et enseignants bien sûr, mais aussi accompagnants, animateurs.

Je vais tout d'abord donner la parole à Christophe Salomon, qui est physicien et membre de l'Académie des Sciences. Il dirige un laboratoire de physique quantique à l'Ecole Normale Supérieure. La physique quantique, c'est un domaine où règne le principe d'incertitude d'Heisenberg : nous ne savons pas exactement à l'instant t où se trouvent les électrons. Et cela présente des analogies avec l'intelligence artificielle, puisque nous sommes là aussi dans un monde régi par les probabilités, et non dans le domaine des sciences exactes comme la physique classique de Newton. Christophe, je vous laisse la parole.

Christophe Salomon

Merci Marc. Je vais vous exposer le point de vue du physicien et je serai ravi de répondre aux questions et aux points qui seront soulevés avec mes collègues ici présents.

Je voudrais juste commencer par une remarque personnelle. Je n'aime pas vraiment l'expression "intelligence artificielle", car l'intelligence est quelque chose qui s'applique essentiellement aux humains. On peut bien sûr parler de l'intelligence des singes, par exemple, mais le mot "artificiel" sous-entend que l'on va remplacer l'intelligence humaine par quelque chose d'autre, et que ce quelque chose est une machine.

Et là, je voudrais vous rassurer tout de suite : les machines sont encore très, très loin d'être capables de faire ce que nous faisons avec notre cerveau aujourd'hui. Nous avons une capacité incroyable à acquérir des signaux visuels, auditifs et à prendre rapidement des décisions qui sont fondées sur notre histoire, notre savoir-faire, notre mémoire. Nous ne devons donc pas nous effrayer de cela : les machines font certaines choses mieux que l'homme, oui, mais pas tout.

C'est le premier point sur lequel je voulais insister, la machine peut égaler l'homme, nous avons vu plein d'exemples, elles jouent mieux aux échecs que les humains maintenant. Mais il ne faut pas penser qu'elles peuvent tout faire.

Mon deuxième point est que l'année 2024 est véritablement l'année de l'intelligence artificielle. D'ailleurs, Pierre-Yves, qui interviendra plus tard, nous le rappellera avec son livre.

A tout seigneur tout honneur, je commence par rendre hommage aux physiciens : ce sont eux qui ont remporté le Prix Nobel de physique cette année en 2024. Les lauréats sont Hopfield et Hinton, leur approche est très intéressante. Ils se sont dit : "Le cerveau est complexe, mais nous savons qu'il est composé de nœuds appelés neurones et de synapses qui relient ces neurones. Ne pouvons-nous pas créer des petits modèles, mais ultrasimples, pour essayer de comprendre comment on peut créer une mémoire et des processus cognitifs ?" Finalement, ces modèles ultrasimples se sont fortement développés - cela a pris cependant des années, leur premier modèle date des années 1970 et Hopfield a franchi un cap majeur en 1984 – et aujourd'hui, grâce aux progrès des ordinateurs et à leur puissance croissante, nous avons un développement dans un domaine tout à fait nouveau pour résoudre des problèmes très complexes. Mais rassurez-vous, je ne vais pas vous décrire tous les types de problèmes complexes que nous sommes capables de résoudre aujourd'hui. Cela prendrait toute la soirée, et ce n'est pas l'objectif de notre discussion.

Ce qui est important, c'est que ces chercheurs, en cherchant à comprendre le cerveau humain, ont conçu de petits modèles basés sur des nœuds, des intersections de points et des liens avec un poids entre ces points, en cherchant à optimiser une certaine fonction pour créer quelque chose qui va servir de mémoire. En termes plus scientifiques, nous parlons de rechercher l'état fondamental d'un système. Il y a donc une explosion, due à la créativité de départ de ces physiciens, qui s'est développée de manière fulgurante dans énormément de domaines.

Le second point frappant de l'année 2024 est le prix Nobel de chimie qui a été attribué à John Jumper et ses collègues de Google, qui ont utilisé l'IA pour déterminer la structure des protéines dans le corps humain. Les protéines sont des molécules complexes à l'origine de la vie. Elles permettent le bon fonctionnement des réactions chimiques.

Mais leur structure est très difficile à déterminer car ce sont des chaînes d'une vingtaine d'acides aminés, nombreuses, complexes, enroulées comme des spaghettis, bien organisées. Avant, il fallait des années de travail pour analyser leur structure. Aujourd'hui, ces chercheurs de Google DeepMind ont entraîné leur modèle d'intelligence artificielle avec leurs réseaux de neurones sur une base de données d'une centaine de milliers de structures connues de protéines et ont réussi à déterminer la structure de 200 millions de protéines, soit pratiquement toutes les protéines connues ! C'est une avancée absolument majeure, qui va avoir des applications considérables en biologie, car on commence à comprendre où vont d'attacher les bonnes cellules, ce qui marche bien, ce qui se passe dans telle ou telle pathologie.

Une autre classe d'applications dans laquelle l'intelligence artificielle est très performante est le langage. Elle excelle dans l'association de mots et permet désormais des traductions automatiques de très bonne qualité. Moi-même, j'utilise des traducteurs automatiques pour rédiger des textes en anglais, car mon français est meilleur que mon anglais. Cela me fait gagner du temps et produit des lettres de recommandation magnifiques. Mais il faut toujours relire et garder un esprit critique car c'est nous qui signons, nous avons une responsabilité.

L'IA apporte aussi de la vitesse dans l'analyse des images à partir d'images entraînées, avec toute sortes d'applications, certaines bénéfiques, d'autres problématiques. Ce qui pose la question du type d'image sur lequel s'opère l'entraînement : sur quelles bases de données ? À qui appartiennent ces bases de données ? Quels sont les biais présents dans ces images générées par l'IA ?

Si j'en arrive au point éducatif, mon premier conseil est qu'il faut jouer avec l'IA ! N'hésitez pas à essayer ChatGPT ou d'autres IA disponibles gratuitement. Posez-lui des questions, même des questions toutes bêtes. Vous verrez alors assez vite deux ou trois choses : cela va vous aider à réaliser des tâches, produire des textes et des mails propres, cela va vous donner deux ou trois idées ; mais cela va aussi vous donner des réponses fausses.

C'est une question fondamentale, car ce n'est pas un système vraiment intelligent, un système, par exemple, comme les recherches dans des bases de données Google, où il y a des traces, jusqu'à des articles scientifiquement prouvés. Une IA associe des mots, et parfois, si elle associe des bons mots et des mauvais mots, elle va conclure que 1 égale 0, ou délivrer des affirmations fausses. On ne sait pas pourquoi. Pour certains algorithmes, on constate qu'ils marchent très bien et c'est d'ailleurs un sujet de recherche actuel que de comprendre pourquoi ça marche et pourquoi ça produit des erreurs. C'est pour cela qu'il faut être prudent.

En termes de conseil éducatif, il faut donc à la fois jouer avec les IAs et en même temps garder son sens critique : interrogez-vous : « ce qui est écrit là ou bien la réponse que me donne la machine, cela a-t-il un sens ou non ? Et pour cela les parents, l'environnement éducatif ont un rôle à jouer. Cela va dépendre de l'âge mais il ne faut pas avoir de peur démesurée. Les professeurs, par exemple, ont eu très peur à une époque, en se disant que leurs élèves allaient tous faire leurs devoirs à la maison avec ChatGPT et qu'ils auraient toujours la même copie à corriger. Il y a bien sûr des alertes à poser et des corrections que les professeurs sont tout à fait capables d'apporter, mais jouer avec et se rendre compte par soi-même avec sa conscience et ses valeurs, est essentiel.

Marc Deluzet

Il est donc important de permettre à des jeunes (et à des moins jeunes) d'avoir du recul, du sens critique en sachant comment fonctionne la machine.

Nous allons poursuivre avec Joséphine Corcoral qui est la Secrétaire Générale Adjointe du Conseil National du Numérique. Elle va nous présenter cet organisme qui a développé une soixantaine de « cafés IA » à travers la France, avec tous publics, des jeunes et des moins jeunes, pour mieux appréhender l'IA, échanger et débattre sur son utilisation.

Joséphine Corcoral

Merci beaucoup Marc. Avant de vous présenter plus en détail ce que nous faisons avec les « cafés IA », je vais vous expliquer ce que fait le Conseil National du Numérique, car tout le monde ne connaît pas nécessairement notre institution.

Le Conseil National du Numérique existe depuis 2011. C'est une commission consultative indépendante. À l'origine, il a été créé pour accompagner le développement de l'écosystème entrepreneurial dans le domaine du numérique, ce que l'on appelle aujourd'hui la French Tech. Mais au fil des années, nous avons réalisé que le tissu économique du secteur commençait à se densifier et que les enjeux du numérique allaient bien au-delà du simple développement économique. Ils posaient de nombreuses questions de régulation et il y avait un besoin croissant de conseils en matière de politiques publiques.

Nous sommes donc entrés dans un deuxième âge du Conseil, où notre rôle a été d'aider à réguler le numérique et à orienter les politiques publiques. Puis, à partir de 2020-2021, la régulation du numérique s'est déplacée au niveau européen, avec l'adoption de textes de lois de plus en plus nombreux et en même temps de plus en plus de débats publics clivants sur ces questions. Je pense par exemple aux controverses qui ont entouré le confinement et le passage au tout-numérique, aux polémiques autour du compteur Linky ou bien de la mise en place de la 5G.

Face à ces sujets sensibles, nous sommes entrés dans un troisième âge du Conseil, dans un rôle d'animation du débat public, de réflexion et de prise de recul sur notre rapport aux outils numériques, pour en discuter plutôt que de s'y opposer. Nous n'avons ni une vision techno critique, ni une vision technophile naïve. Notre objectif est d'écouter toutes les opinions et de mettre tout le monde autour de la table pour se parler du numérique.

Depuis 2021, nous avons publié de nombreux travaux librement accessibles sur notre site web. Nous avons exploré différentes facettes de nos relations avec le numérique, par exemple la circulation des fausses informations en ligne, l'économie de l'attention, les interactions homme-machine au travail.

À chaque fois, nous faisons un état de l'art, de la problématique, puis nous proposons des leviers d'action pour la puissance publique et tous les acteurs concernés. Nous nous sommes aperçus que quel que soit le sujet traité, nous arrivions toujours à la même proposition commune : la nécessité de diffuser une culture numérique partagée. Qu'est-ce que cela signifie ? Je prends souvent la métaphore du permis de conduire. Quand on a le permis, on ne sait pas forcément comment le moteur de la voiture fonctionne, mais on connaît tous le Code de la route. Pour le numérique, c'est exactement la même chose, nous n'avons pas besoin de savoir coder précisément un algorithme pour néanmoins comprendre comment cet outil fonctionne, ce qu'il peut faire ou ne pas faire et l'utiliser ou de ne pas l'utiliser en connaissance de cause.

A ce moment-là, nous avons donc voulu joindre le geste à la parole et avons lancé en 2022 une initiative appelée Itinéraire Numérique. Nous avons parcouru la France en réalisant plus de 120 déplacements, à la rencontre de celles et ceux qui organisent des temps d'échange avec les citoyens pour diffuser cette culture numérique partagée. Nous avons rencontré tous types de publics : salariés en entreprise, écoliers et enseignants, personnes âgées en EHPAD, habitants de quartiers populaires. Et nous en avons tiré une restitution qui appelait à poursuivre et à généraliser cette démarche à l'échelle nationale, avec une plus grande ampleur que ce que la dizaine de personnes que nous étions, avaient développé à la fin 2023.

En 2024, une Commission sur l'Intelligence Artificielle a été mise en place par la Première Ministre, Élisabeth Borne à ce moment. Cette commission a rendu son rapport en mars 2024. Sa première recommandation était la diffusion d'une culture numérique partagée sur l'IA. En mai 2024, le Président de la République a confié au Conseil National du Numérique la mission de coordonner cette démarche qui a pris un nouveau nom.

Nous l'avons appelée Café IA. L'idée est simple : se réunir pour parler d'IA autour d'un café. L'objectif est de soutenir, promouvoir, fédérer toutes les initiatives - elles sont extrêmement nombreuses - d'échanges, de débats, d'expérimentation autour de l'IA. Un Café IA est un moment où on se parle d'IA, je le dis de manière très large et c'est à dessein, puisque cela peut toucher des publics très différents et être conduit de façons très diverses ; cela peut être très simplement des chaises en cercle avec des post-it pour noter des appréhensions, des attentes vis-à-vis de l'IA qu'on met ensuite en commun anonymement ; et cela peut aller jusqu'à des formats plus poussés de jeux, d'expérimentation d'outils, de conception, par exemple, de son assistant personnalisé grâce à l'IA.

Il est possible d'imaginer plein de formats différents, le but étant d'avoir toujours quatre piliers : la montée en compétences (comprendre comment cela marche, qu'est-ce que cela fait et ne fait pas), le débat (toujours dans une posture d'écoute et de dialogue), l'expérimentation (nous sommes convaincus que l'on débat mieux d'un sujet quand on l'expérimente concrètement) et la mise en action (à l'issue d'une Café IA toujours se demander ce que l'on fait pour que cela ne reste pas lettre morte).

Nous pouvons imaginer plein de débouchés : si dans une école apparaît l'envie d'organiser un Café IA parce qu'un outil pédagogique à base d'IA est proposé à des élèves pour permettre une pédagogie plus personnalisée, il est possible à l'issue de ce Café IA, de dresser une charte du bon usage pédagogique de l'intelligence artificielle. De la même façon en entreprise, si un système informatique propose un nouvel outil, il est possible d'organiser un Café IA pour s'en parler, comprendre à quoi cela sert, faire part de ses besoins en tant que travailleur, en tant que manager, et à l'issue de ce Café IA, amorcer un cycle de formation pour monter en

compétences sur cet outil ou plus généralement sur l'IA. Voilà un panorama assez large pour vous montrer à la fois la diversité des situations dans lesquelles il peut être pertinent d'organiser un Café IA et ce qui peut en ressortir.

Nous sommes sur ce projet depuis début mai et nous avons organisé une soixantaine de Café IA, de nombreux autres se sont organisés sans nous, ou juste avec notre soutien à distance. Notre objectif est d'en organiser 100 d'ici février 2025, date à laquelle la France accueillera le sommet concernant l'Action sur l'Intelligence Artificielle. En marge de cette manifestation, nous organiserons une journée citoyenne, le 7 février, dont l'objectif sera de faire un premier point d'étape de ces Café IA afin de faire remonter les messages qui ont été partagés sur le terrain. Nous tiendrons toute une série d'ateliers pour porter la voix des citoyens en regard de ce sommet qui réunira essentiellement des dirigeants et des chefs d'entreprise.

Si vous souhaitez en savoir davantage ou organiser un Café IA, vous pouvez nous contacter sur le site internet « cafeia.org », tout neuf d'aujourd'hui, lire la lettre d'information hebdomadaire Cénum, nous écrire sur l'adresse électronique dédiée bonjour@cafeia.org.

Marc Deluzet

C'est un peu comme si le CCIC faisait ce soir son troisième Café IA ! Puis-je vous demander de dire en quelques minutes ce que vous tirez aujourd'hui comme principes, leçons d'expérience vis-à-vis de l'appropriation de l'IA dans la population, par des publics très diversifiés, notamment chez les éducateurs, dans l'enseignement et les familles.

Joséphine Corcoral

Très rapidement car il y a beaucoup à dire et je crains d'oublier des points importants. La première chose qui me frappe actuellement, c'est à quel point le sujet est connu de toutes et tous, car il est omniprésent dans le débat public. Je pointe une très grande confusion entre IA et IA générative, et donc l'enjeu de préciser de quoi l'on parle et quelles sont les problématiques en fonction des sujets abordés.

Deuxième point, des questionnements sont récurrents, quels que soient les publics, quels que soient les âges, quels que soient les contextes. Il y en a deux principaux : la peur du remplacement par la machine, notamment au travail, et la fin du travail humain ; le second est l'impact environnemental de l'IA. Ce sont des questions omniprésentes quel que soit le public que nous rencontrons.

C'est très éloquent d'un débat très anxiogène sur l'IA qui se concentre beaucoup sur l'IA générative et d'une projection dans le temps long avec ce que pourrait faire hypothétiquement la machine, alors qu'à date c'est plutôt un compagnon comme cela a été très bien dit précédemment. Débat très anxiogène également, avec cette angoisse beaucoup plus immédiate de l'impact environnemental, sujet qui est, pour le coup, beaucoup moins traité et que les citoyens ont envie de mettre sur le devant de la scène pour en parler davantage, d'autant que les premiers chiffres qui nous parviennent sur l'IA générative sont particulièrement préoccupants.

Marc Deluzet

Je vais donner la parole à Pierre-Yves Oudeyer qui est directeur de recherche à l'INRIA et qui est l'auteur avec Didier Roy de ce livre passionnant : « C'est pas moi, c'est l'IA ! » pour les jeunes ados, et aussi pour leurs parents. Il peut le dédicacer ce soir et c'est un excellent cadeau

de Noël. Il va poursuivre sur le même registre : quelles solutions pour mieux s'approprier l'IA, pour devenir compétent et pouvoir jouer avec, quels enseignements et point de repères il en tire.

Pierre-Yves Oudeyer

Merci beaucoup, je suis ravi d'être ici avec vous ce soir. Je me présente en deux mots avant d'entrer dans le vif du sujet. Je dirige un laboratoire de recherche à l'INRIA (Institut National de Recherche en Sciences et Technologies du Numérique), qui est un peu comme le CNRS, mais spécialisé dans l'informatique. C'est la plus grande structure de recherche en informatique d'Europe. Mon laboratoire travaille à la frontière entre l'IA et les sciences cognitives. Nous nous intéressons à la modélisation des mécanismes de la curiosité et à leur rôle dans l'apprentissage chez les enfants. Cela nous amène à beaucoup d'interactions avec les questions éducatives, et c'est justement de ces questions éducatives dont je vais vous parler ce soir.

Je ne vais pas vous parler d'IA en général mais plutôt de l'IA générative. L'IA générative, de quoi parle-t-on exactement ? Nous parlons de ces logiciels comme ChatGPT, qui sont capables de produire du texte qui ressemble à celui écrit par des humains. Nous parlons de logiciels comme Midjourney, qui est capable de produire des images similaires à celles produites par des humains. Et nous parlons de logiciels qui génèrent des vidéos qui ressemblent aux vidéos qui peuvent être produites par des humains, à tel point que dans quelques mois, vous verrez apparaître des films entiers, avec un scénario cohérent et des images qui ressembleront à des images et des vidéos totalement réelles. Dans très peu de temps, nous allons vivre dans un monde où il sera techniquement impossible de savoir si une vidéo a été réellement filmée ou si elle a été générée par un logiciel d'IA.

C'est une révolution extrêmement récente. Rappelez-vous que ChatGPT est apparu il y a juste deux ans. En deux mois, il avait déjà 100 millions d'utilisateurs. Aujourd'hui, il en compte plus de 180 millions, et encore davantage si on additionne tous les logiciels d'IA similaires. L'usage est devenu absolument massif. Il y a beaucoup de modèles de logiciels similaires à ChatGPT : comme par exemple « Hugging Face », qui est un site web qui agrège plein de modèles *open source* que l'on peut télécharger et faire tourner sur son ordinateur, il y a en a plus de 400.000 aujourd'hui.

En France, les plus de trente-cinq ans n'utilisent pas encore beaucoup ces logiciels, seulement 22%, mais chez les 25-34 ans, la proportion est autour de 47% et chez les 18-24 ans, elle est environ de 70%. Nous allons dans les collèges et les lycées et nous nous apercevons qu'aujourd'hui plus de 80% des collégiens et lycéens utilisent ces outils-là régulièrement.

On observe aussi que, pour de nombreuses tâches, 44% de tous ces utilisateurs reprennent les résultats sans les modifier : ils font du copier-coller. Chose encore plus importante, quand les adolescents commencent à utiliser ChatGPT pour s'aider à faire leurs devoirs, ils ont ensuite une forte tendance, face à n'importe quelle question sur n'importe quel sujet, à ne plus passer par les moteurs de recherche et à directement utiliser ChatGPT.

C'est extrêmement important d'avoir cela en tête, car il y a plusieurs façons d'appréhender ces IA génératives. L'une des plus intéressante, consiste à les envisager comme des outils de transmission culturelle. Qu'est-ce que cela veut dire ?

Ce sont des logiciels qui ont été fabriqués sur la base de milliards de textes, d'images, de vidéos, à partir desquels ils vont encoder les valeurs, les connaissances, les croyances, les préférences, les biais, correspondant au corpus de textes et d'images sur lesquels ils ont été entraînés. Autrement dit, ils véhiculent une certaine vision du monde, et celle-ci n'est pas neutre. C'est une question fondamentale quand on réalise que de plus en plus de personnes utilisent ces outils pour accéder à la connaissance et à l'information.

Il est essentiel de comprendre qu'une grande partie du contenu d'Internet lui-même sera de plus en plus généré par ces IA. Cela signifie que même ceux qui n'utilisent pas ces logiciels, verront des images et liront des textes qui auront été générées par ces outils.

Sur les réseaux sociaux, il y a aujourd'hui en permanence des millions de logiciels qui simulent des personnes réelles et qui sont lancés à grande échelle par des organisations, parfois petites, parfois grandes, pour influencer l'opinion des populations. Aujourd'hui la plupart de ces grands modèles sont fabriqués par des entreprises américaines ou chinoises, qui véhiculent donc leur propre vision du monde tout à fait particulière.

On peut repérer ces modèles culturels quand on regarde les biais ou les stéréotypes qu'ils reproduisent. Par exemple, si on demande à ChatGPT de donner le féminin d'un métier : "Il est docteur, elle est...", en lui demandant de compléter, il va dire "infirmière". On essaie alors de tester s'il arrive à se corriger dans l'autre sens : "Elle est docteur, il est..." et il répond "ingénieur". S'il voulait être infirmier, il n'a pas le droit.

De la même façon, si on demande à une IA de produire l'image d'une personne travaillant dans le social, on obtiendra une femme et si on demande une personne en costume dans un bureau, on obtiendra un homme blanc.

Il y a un certain nombre de personnes qui vont utiliser volontairement la capacité de produire des images fausses pour manipuler les gens, par exemple, pour des arnaques commerciales ou dans des contextes de guerre. Ici, vous voyez une vidéo du président ukrainien à qui on faisait dire au début de la guerre que sa population devait rendre les armes. Il y a donc des enjeux très grands pour la démocratie.

Mais à côté de ces enjeux-là, comme on l'a vu auparavant, il y a des opportunités immenses avec ces outils. Je ne vais pas rentrer dans le détail. Mais les prix Nobel de physique et de chimie qui viennent d'être décernés à des chercheurs en intelligence artificielle, montrent vraiment le rôle grandissant de ces outils qui vont littéralement bouleverser non seulement la science, mais aussi la manière dont on fait la science, dont on apprend à en faire, et pas simplement chez les débutants, chez les étudiants à l'université, mais surtout chez les plus grands scientifiques. Il y a de plus en plus de mathématiciens, des médailles Fields, qui utilisent ces outils pour les aider avec leur équipe à trouver des solutions à des problèmes extrêmement difficiles.

A côté de ces opportunités et de ces défis, il ne faut jamais oublier que cet ensemble de technologies possède une empreinte environnementale extrêmement forte. Dans le détail, il est encore difficile de la quantifier, mais on sait qu'elle est forte et il faut toujours avoir cela en tête.

Je vais maintenant revenir à des questions vraiment éducatives. On vient de voir que ces outils sont en train de changer le monde. Il est donc important que tous les citoyens, les jeunes et les moins jeunes soient vraiment éduqués pour comprendre ce qui se passe. Si l'on regarde

plus précisément ce qui se produit à l'école, il y a une déferlante absolument gigantesque qui est en train d'arriver. D'abord un certain nombre d'usages commencent à être massifs, qui ne sont pas extraordinairement positifs d'un point de vue pédagogique. Par exemple, beaucoup d'élèves se disent « je vais utiliser ChatGPT pour faire plus vite mes devoirs et en plus avoir de meilleures notes ». Cela évidemment n'est pas une très bonne idée car ils ne vont pas apprendre grand-chose et faire perdre du temps à leurs enseignants. Il y a d'autres idées qui sont meilleures. Par exemple, il y a des élèves qui se disent : « Je prends une photo de mon cours d'histoire, je l'envoie à ChatGPT et je lui demande de me poser des questions pour vérifier si j'ai bien compris ou si je sais ma leçon. »

D'un point de vue scientifique, on peut se dire aujourd'hui à peu sûrement qu'on ne comprend rien car cela est arrivé tellement récemment et tellement vite qu'en fait, il n'a pas été possible de faire des expérimentations scientifiquement contrôlées. On commence cependant à avoir de petits indices. Par exemple, j'ai des collègues de l'Université de Stanford qui ont fait à assez grande échelle une expérience avec des cours en ligne pour lesquels ils ont donné à une partie des étudiants du monde entier qui se connectaient un accès gratuit à ChatGPT. Celui-ci a été paramétré de manière à les aider à mieux comprendre le cours. Et ils ont évalué l'impact qui en résultait sur l'engagement des élèves dans les cours et la réussite aux examens. Ils ont observé des choses étonnantes : dans les pays à haut indice de développement (Amérique du Nord, Europe), l'accès à ChatGPT avait tendance à diminuer l'engagement des élèves et leurs résultats aux examens, tandis que dans les pays à bas indice de développement humain (Pakistan, Soudan par exemple), les élèves qui avaient accès à ces outils avaient un engagement et des résultats bien meilleurs. On ne comprend pas encore totalement pourquoi mais on voit tout de suite qu'il doit y avoir des opportunités d'éducation extraordinaires pour tout un tas de pays qui n'ont pas un système éducatif aussi développé que le nôtre.

Maintenant faisons un petit zoom sur ce qui se passe dans les classes aujourd'hui en France. Nous avons fait une expérimentation au mois de mai-juin dernier dans les collèges de Nouvelle Aquitaine, en nous posant la question : comment les élèves parviennent un tout petit peu à utiliser ChatGPT, par exemple quand on leur donne un petit exercice de science et de recherche. Ainsi, nous leur avons montré une image imprimée sur un papier en leur disant : « Anna veut comprendre les propriétés de la lumière, elle a commencé par un test simple en laissant tomber son crayon dans un verre d'eau, voici ce qu'elle a vu. Ton but est de comprendre pourquoi l'apparence du crayon est différente ». Et à partir de là, nous leur donnions accès à ChatGPT, avec des questions qu'ils pouvaient potentiellement poser à ChatGPT. Volontairement certaines questions étaient bien formulées, d'autres étaient moins bien formulées. Exemple de question bien formulée : qu'est-ce qui fait qu'un crayon change de direction quand il est mis dans un verre d'eau ? Exemple de question moins bien formulée : pourquoi un crayon devient-il différent ? Nous leur demandions à chaque fois : « est ce que vous pensez que c'est une bonne question ? une mauvaise question ? Si vous voulez la poser, vous pouvez, si vous pensez qu'elle est mauvaise, vous pouvez poser la question que vous souhaitez. Ensuite, ils avaient la réponse de ChatGPT et ils avaient la possibilité de reformuler la question. Figurez-vous que ces exercices qui n'avaient rien de sorcier, sur la population d'élèves que nous avons étudiée, la moyenne était de 10/20. Ce n'est pas très glorieux. Il y avait quelques élèves qui avaient d'excellentes notes mais la plupart avaient autour 10/20. On s'est aperçu que pour la plupart des élèves, ils étaient incapables de faire la différence entre une bonne question et une mauvaise question posée à l'outil. Par ailleurs, la plupart d'entre

eux (plus de 90%), quand ils obtenaient une réponse qui n'était pas adaptée, n'avaient absolument pas le réflexe de reformuler leur question et d'approfondir. Ils prenaient la réponse telle quelle, comme argent comptant, « je donne ça comme réponse ». Nous nous sommes aperçus, en analysant plus finement, que ce n'était pas exactement tous les élèves. Nous avons observé le rôle de leur métacognition, c'est-à-dire leur capacité à savoir ce qu'on sait ou ne sait pas, à définir ce qu'on comprend et qu'on ne comprend pas. Les élèves qui avaient une métacognition un petit peu plus développée que les autres s'en sortaient mieux. En revanche, chez les autres, nous nous sommes aperçus d'une chose particulièrement préoccupante. Nous avons aussi demandé à tous les élèves dans quelle mesure ils avaient déjà utilisé ChatGPT : se sentaient-ils familiers, se sentaient-ils comprendre l'outil. Plus les élèves disaient qu'ils avaient déjà essayé ChatGPT et comprendre ce que c'était, moins bons étaient leurs résultats. Cela veut dire qu'en fait, en utilisant de manière autonome et en explorant de manière autonome cet outil-là, ils se sont fait des images mentales de plus en plus fausses qui les amenaient à utiliser le logiciel de plus en plus mal.

Cela nous amène à deux défis éducatifs extrêmement importants : le premier ne concerne pas seulement l'IA : il est essentiel à l'école d'entraîner la métacognition, il y a tout un tas d'initiatives en cours dans cette direction ; la deuxième chose est qu'il est très important d'accompagner l'éducation à l'IA pour que les enfants se construisent les bonnes métaphores.

Je conclus en soulignant que pour l'éducation à l'IA générative, des outils sont en cours de développement. Par exemple, avec mon équipe, nous avons développé une série de 8 vidéos extrêmement courtes, qui s'appellent « ChatGPT expliqué en 5 minutes » et s'adressent aux adolescents, aux enseignants, aux parents, à tout le monde. Elles abordent les mécanismes, les applications, les défis, les opportunités, les enjeux sociétaux de l'IA générative. Nous avons volontairement partagé cette série de vidéos selon une licence « *créative commons* » qui permet à tout le monde de les utiliser librement sans nous demander l'autorisation, pour des usages gratuits, commerciaux, tout ce qui est possible. Par exemple, cette série de vidéos est aujourd'hui intégrée dans des formations en cours de développement, à la Direction Interministérielle du Numérique qui est en train de fabriquer une formation pour l'ensemble des fonctionnaires en France. J'ai appris qu'elles étaient utilisées par l'Académie du Numérique à l'intérieur du ministère des armées. Et enfin, un autre exemple, c'est ce livre qu'avec Didier Roy et Clémentine Latron nous avons écrit et qui s'adresse dans sa forme aux adolescents, mais sur le fond à tout le monde : les enseignants, les parents, les grands-parents, pour que chacun puisse découvrir un certain nombre de clés afin de comprendre ce nouveau monde.

Marc Deluzet

Merci Pierre-Yves. Si vous avez bien suivi son exposé, il a expliqué pourquoi il faut acheter son livre pour permettre à des jeunes de se former et à leurs parents également. Je vais maintenant donner la parole à Laurent Bibard. Laurent est professeur de gestion et de philosophie à l'ESSEC. Il va revenir sur sa conception de l'intelligence artificielle : l'IA est-elle une intelligence ? Quelle différence y a-t-il entre la machine et l'homme ? La machine va-t-elle remplacer l'humain ? Nous lui avons demandé de traiter la question du pouvoir des intelligences artificielles sur nous-mêmes, en faisant l'hypothèse que plus qu'une prise de pouvoir des machines sur nous-mêmes, nous serions menacés, comme La Boétie l'avait décrit, par un phénomène de servitude volontaire, dans lequel nous nous soumettrions de nous-même à la tyrannie.

Laurent Bibard

Bonsoir à toutes et à tous. Ce que j'ai entendu ce soir me laisse penser que décidément, et c'est la raison pour laquelle j'interviens, l'intelligence artificielle n'est pas une question technologique. Je pense très profondément que c'est une question humaine et que nous ne posons pas les bonnes questions si nous restons impressionnés par l'idée que les technologies nous fabriquent. Nous sommes dans une situation où nous ne réfléchissons pas suffisamment au fait que nous sommes les auteurs des technologies, nous sommes ceux qui les entretenons, nous sommes ceux qui les utilisons et nous sommes ceux qui décidons pour quoi elles sont là. A l'instar de ce que j'ai entendu, par rapport à l'IA, j'ai maintenant le sentiment que nous avons beaucoup cru qu'elle ferait tout pour nous, qu'elle nous remplacerait, qu'elle serait quelque chose qui nous dépasse infiniment et je crois que c'est fondamentalement faux. C'est-à-dire que quand nous sommes dans cette perspective, nous abdiquons notre propre humanité, et que nous devons nous demander à quoi elle sert, ce que nous voulons qu'elle fasse, pourquoi et dans quelle mesure. Et cela rejoint je crois les questions de fond qui ont été présentées préalablement sur les questions d'éducation, pour les jeunes enfants en particulier. Nous avons des technologies, elles sont un excellent outil mais elles ne sont qu'un outil et pas plus que cela. Il est possible de croire qu'elles sont plus que cela, si on se fie à la rapidité de l'information qu'elles font circuler, aux simulations qu'elles rendent possibles, ce qui a été évoqué tout à l'heure, à la possibilité de nous mettre dans de l'illusion. Mais c'est nous qui les utilisons et du coup, la question qui se pose et qui s'impose à mon avis, pour les jeunes générations en particulier, est : comment fait-on pour leur faire comprendre qu'il y a des technologies extrêmement efficaces et qu'elles ne seront là que pour nous.

Il y a un exemple qui est évoqué dans le livre « L'intelligence artificielle n'est pas qu'une question technologique », que j'ai coécrit avec Nicolas Sabouret. Nous relatons une histoire très simple, un témoignage : un utilisateur qui n'est pas spécialiste de l'IA s'interroge avec son épouse pour savoir comment faire pour que leurs enfants ne soient pas obnubilés par l'IA, sans que pour autant ils n'en soient pas exclus. Le couple de parents n'imaginait pas d'enlever l'usage de l'IA pour leurs enfants, ils se demandaient comment faire pour qu'ils ne soient pas exclus mais pas fascinés non plus. Il se trouve que le père a une compétence particulière. Il a mis ses deux enfants devant un ordinateur et devant un Iphone. Il a fait fonctionner les écrans, les enfants ont été fascinés par le fonctionnement des écrans. Puis, il a arrêté le fonctionnement des écrans, il a démonté l'ordinateur et l'iPhone, il a posé toutes les pièces constituant les deux machines, et il leur a dit « voyez ça n'est que cela », puis il a remonté les machines, les a remises en route. Et les enfants n'ont jamais plus été fascinés par les machines.

Cet exemple que je vous évoque, rejoint quelque chose de très important actuellement, qui est la prise de conscience progressive, à laquelle il a été fait allusion précédemment, que les IA consomment une énergie électrique gigantesque et qui exige que nous nous demandions, nous adultes, jusqu'où nous voulons utiliser les IA et pour quoi faire ? Mon propos n'est pas de ne pas utiliser les IA mais de nous demander : pour quelles fonctions sont-elles indispensables ? pour quoi voulons-nous véritablement les utiliser et jusqu'où ? A partir de là, il sera possible de discerner jusqu'où nous tolérons l'utilisation de l'énergie qu'elles consomment, question que nous nous posons un peu seulement maintenant.

Et du coup, je viens à une question tout à fait importante, qui a habité notre passé il y a une dizaine d'années, qui était de fantasmer que l'IA remplacerait l'intelligence humaine, que nous serions débordés par nos technologies. Il est capital de se demander, nous voulons bien être

dépassés peut-être, mais qu'est-ce que cela veut dire pour nos existences et quel sens cela a-t-il ? Je crois qu'on peut dire que cela n'a pas de sens. Nous demander si nos machines, que nous fabriquons, que nous inventons, que nous maintenons, dont nous assurons le fonctionnement, que nous utilisons, vont nous dépasser, n'a pas de sens.

A l'heure actuelle où il y a des personnes qui ont une importance, un rôle politique, pas seulement économique, je pense à quelqu'un comme Elon Musk, qui influence considérablement la planète, ne doit-on pas résister et se demander à quoi cela sert ? Que voulons-nous ? L'enjeu n'est pas de réfléchir, simplement, posément, mais de manière plus décidée à ce que nous voulons faire des technologies. Cela me semble être le débat fondamental. Nous sommes dans une situation économique, politique, morale, peut-être métaphysique, qui nous impose de réfléchir très attentivement par rapport aux enjeux géopolitiques qui concernent l'IA : jusqu'où voulons-nous accepter quoi et de la part de qui ? La question n'est pas du tout secondaire.

Marc Deluzet

En fait, vous nous dites qu'on peut utiliser l'IA pour supprimer des emplois, mais ce n'est pas la machine qui les supprime, plutôt ceux qui mettent en œuvre des technologies d'intelligence artificielle. Nous pouvons prendre dix minutes pour un échange entre vous. Christophe, vous vouliez réagir ?

Christophe Salomon

Je voulais souligner le coût énergétique formidable de l'IA, les personnes en sont conscientes mais elles ne veulent pas regarder les choses en face. Il suffit de constater que Google veut remettre en activité *Three Mile Island*, la centrale nucléaire qui avait été arrêtée, simplement parce qu'il y a des besoins en électricité considérables pour faire fonctionner les IA. Cela est chiffré. Si on regarde la propagation de l'IA, le nombre d'utilisateurs, nous allons vers un problème énergétique majeur. Il faudra choisir entre la voiture électrique et l'IA. Il y a une vraie question de société, à discuter collectivement. La réponse est-elle au niveau national, européen ou au niveau mondial ? C'est est plus compliqué à dire.

Joséphine Corcoral

Je suis entièrement d'accord et alignée. Consommation énergétique, consommation en eau aussi, car tous ces centres de données demandent énormément d'eau pour être refroidis et les centrales de production d'énergie utilisent aussi de l'eau. Les premiers chiffres et ordres de grandeur qui nous parviennent montrent, par exemple, que générer une image sur un logiciel de type *Midjourney* est l'équivalent de la consommation énergétique d'une recharge complète de smartphone et d'une demi-bouteille d'eau. C'est énorme. On en revient à la question du besoin. Nous avons cité de nombreux usages très vertueux de l'IA, il y a donc un véritable enjeu à se questionner sur la pertinence de faire appel à l'IA dans telle ou telle situation. Je pense notamment à des requêtes Google : il est beaucoup plus efficace, pour aboutir à un même résultat, d'utiliser un moteur classique plutôt que ChatGPT qui consomme 10 fois plus d'énergie.

Je voulais aussi réagir à ce qu'a dit Laurent Bibard, sur la vision que l'on projette sur ces outils. Cet après-midi, j'animais une table ronde sur les imaginaires concernant l'IA au salon Génération IA, nous avons la même conversation qui soulignait que l'écosystème tech est très descendant et nous vend un monde, d'un côté très cataclysmique - à coup de robots tueurs,

robots remplaçants, etc... - et d'un autre côté, toujours cette fuite en avant, que la technologie va nous sauver. Ce sont les mêmes qui propagent ces deux narratifs, qui d'un côté vont prêcher un peu la peur et l'effroi et de l'autre côté promettent des solutions, qui sont avec toujours plus de tech. Il y a donc un enjeu à sortir de ces imaginaires et se réapproprier ces outils. Quels imaginaires voulons-nous, quelles technologies avons-nous envie d'utiliser, comment et à quelles fins ?

Pierre-Yves Oudeyer

Que signifie prendre en compte le coût énergétique quand on travaille sur les nouveaux dispositifs ? Je peux donner un exemple dans le domaine de l'éducation. En fait, aujourd'hui, on voit tout un tas de personnes, de startups ou de grandes entreprises qui essaient de développer des tuteurs personnalisés pour leurs apprenants, que ce soit des enfants ou des adultes, avec lesquels on dialoguerait de manière totalement naturelle et qui nous proposeraient des exercices d'apprentissage. D'un point de vue pédagogique, je suis loin d'être sûr que ce soit si pertinent que cela, car la manière dont les systèmes comme ChatGPT sont entraînés fait qu'ils n'ont pas rencontré beaucoup de données d'interactions entre enseignants et apprenants, elles sont peu nombreuses. Même s'il y avait de telles données, même si on était capable de construire quelques démonstrations à petite échelle de ce genre de système qui montreraient que cela fonctionne bien sur le plan pédagogique, si on se met à imaginer développer ce genre d'outils avec tous les élèves, cela ne peut pas fonctionner d'un point de vue énergétique. Il faut penser l'utilisation de ces outils d'une autre manière. Il se trouve que dans le domaine éducatif, il y a d'autres formes d'IA, qui existent depuis longtemps, qui sont déjà déployées dans des logiciels à relativement grande échelle, par exemple pour proposer des séquences d'exercice qui sont optimisés pour chacun des élèves. Et dans ces logiciels, traditionnellement, les exercices eux-mêmes sont écrits à la main par des enseignants. La partie d'IA est celle qui personnalise avec chaque élève. Il y a donc des choses extrêmement positives qui existent. En revanche, fabriquer ce genre de système, cela demande un effort d'écriture des exercices qui est absolument considérable. Pour un prof de math, par exemple, qui va écrire ces exercices, ce qui est intéressant pour lui consiste à dire qu'il y a ces 5 types d'exercices qui sont importants, et donner deux ou trois exemples. Par exemple si on travaille avec des fractions pour l'école primaire, il s'agit de faire 53 variantes avec des fractions, des camemberts, des pizzas, ce qui n'est pas le plus épanouissant. Dans ces systèmes, il est important de le faire à grande échelle. Aujourd'hui, dans mon équipe, nous travaillons sur ce genre de problématique, avec des entreprises du monde de la tech. Nous réfléchissons à utiliser des systèmes d'IA générative, non pas sur le face à face interactif avec les élèves, mais en amont, pour la pré-génération de contenus éducatifs à très grande échelle. Une fois que cela est pré-généré, ce ne coûte pas si cher que cela et il n'y a plus du tout d'IA générative comme précédemment et cela est beaucoup plus frugal. C'est une manière concrète de montrer qu'avoir en tête dès le début le coût énergétique, permet de générer des systèmes de manière complètement différente.

Débat avec la salle

Jeanne-Marie Trougad (Les Apprentis d'Auteuil)

Concernant le Digital Services Act, qui pose comme principe que tout ce qui est illégal hors ligne doit l'être en ligne, est-ce que cela constitue un message fort dans ce débat sur l'IA, ou bien est-ce juste un élément mineur dans un océan de questions philosophiques et éthiques sur l'IA ? Devons-nous nous en réjouir ou nous en inquiéter ?

Alberto Kabuge (Salésien de don Bosco, Union Mondiale des Enseignants Catholiques)

Ma question est plus pratique : aujourd'hui, les jeunes utilisent le GPS et l'IA pour tout. Pour trouver une station de métro à 100m, ils utilisent leur téléphone. En écoutant vos interventions, je me disais : est-ce que ces outils ne nous rendent pas plus faibles ? Celui qui ne sait pas les utiliser devient plus faible. Même pour faire un calcul simple comme 5×5 , ils utilisent une calculatrice. Quels conseils pouvez-vous donner aux parents et éducateurs face à cette situation ? C'est ma préoccupation aujourd'hui, pour dans vingt ou trente ans.

Laurent Bibard

Pour la deuxième question, je donnerais volontiers un avis très simple. Il faut apprendre à se déconnecter régulièrement. Aux jeunes, il faut apprendre à lâcher le téléphone portable, à lâcher toutes les technologies, de temps en temps. Pas toujours, ce n'est pas systématique, ce n'est pas contre les technologies, mais c'est pour les aider à se reconnecter avec une connexion réelle, celle de nous-mêmes avec le monde où nous vivons. Et donc de temps en temps de lâcher sur les connexions.

Si un étudiant ne sait pas où est le métro alors qu'il est à 100 m et qu'il a besoin de ChatGPT ou d'un GPS pour repérer là où il est, c'est catastrophique. Donc il est capital de s'exercer régulièrement à ne pas être connecté aux machines.

Joséphine Corcoral

Si j'ai bien compris votre question, vous faites référence au Digital Services Act (DSA) - le Règlement sur les Services Numériques en bon français - et plus largement à la réglementation européenne sur l'IA. Il faut préciser que ce texte ne couvre pas directement l'IA, il couvre uniquement les services numériques, au sens des plateformes numériques, c'est à dire les réseaux sociaux. Ce règlement définit toute une série d'obligations pour toutes les plateformes, notamment les plus grandes, pour lutter contre ce qu'on appelle le risque systémique : la diffusion de fausses informations et la manipulation de discours publics.

A date, les plateformes comme OpenAI (ChatGPT) ne sont pas concernées par ce règlement, car elles ne dépassent pas encore le seuil d'utilisateurs défini dans le DSA. Donc l'IA n'est pas, peut-être encore, couverte par ce texte. Cependant, d'autres réglementations européennes encadrent l'IA, comme le Règlement sur l'IA - l'AI Act – c'est un règlement « produit » qui vient encadrer les risques de l'IA sous l'angle de ses usages.

La Commission Européenne a établi une matrice de risques avec 4 niveaux, en fonction du degré de risques des usages. Le premier niveau se compose juste d'un déclaratif avec un code de bonne conduite et cela va jusqu'à l'interdiction, par exemple, de la justice prédictive (prédire la probabilité qu'une personne a de commettre un crime est interdit). Le gros du texte concerne les risques élevés, avec les applications ChatGPT et autres IA génératives, pour lesquelles, il y a des obligations très renforcées de transparences.

Est-ce une avancée ? Oui, car l'Europe est le premier continent à mettre en place un cadre législatif clair sur l'IA et le numérique. Mais certains s'inquiètent que cela freine l'innovation.

Personnellement, je préfère qu'on prenne du retard sur l'innovation plutôt que d'avoir des GAFAM européens et des outils toxiques.

Christophe Salomon

Je voudrais dire que le GPS est un outil formidable, tout le monde l'utilise car cela simplifie la vie, c'est un gain de productivité pour tout. Du point de vue éducatif, il faut pouvoir l'éteindre de temps en temps. Il y a des horloges très précises dans le GPS.

Joséphine a présenté les règles mises en place mais elles ne sont pas respectées. Sur le réseau X, il y a dix ans, on comptait 30% de personnes qui étaient climatosceptiques, qui ne croyaient pas à la responsabilité de l'homme sur le réchauffement climatique. Depuis qu'Elon Musk a racheté le réseau, c'est une personne sur deux ! Cela signifie que les personnes sont manipulées. La régulation n'a pas encore agi sur ce type de réseau. J'aurais une question pour Pierre-Yves qui a dit que des films étaient en train d'être produits par l'IA, mais est-ce que ce seront de bons films ?

Pierre-Yves Oudeyer

Juste sur cela rapidement, je pense qu'il y aura des bons films et des très mauvais films. Le fait d'être humain ne garantit pas de réaliser un bon film.

Par rapport à la question précédente, ce qui me semble important est de responsabiliser les apprenants de tous les âges, enfants et adultes, et d'éveiller leur conscience pour leur faire comprendre ce qu'est le sens de l'apprentissage. S'ils utilisent le GPS et ChatGPT, pour faire leurs maths et pour aller plus vite, c'est peut-être qu'ils n'ont pas compris le sens d'apprendre à faire des mathématiques.

Ce n'est d'ailleurs pas spécialement un problème de l'IA. Il y a aujourd'hui beaucoup de manières d'enseigner les langues, les mathématiques et les sciences, dans lesquelles on oublie d'expliquer quel est le sens de ces apprentissages, ce qui fait que les élèves ne s'y intéressent pas et vont prendre le chemin le plus court possible pour arriver à la moyenne qui leur permettra d'avoir l'examen et de se débarrasser de leurs cours juste après. A partir du moment où les enfants ont compris quel est le sens des apprentissages, et davantage encore, quel est le plaisir même qu'ils peuvent y trouver. Il y a du plaisir dans les apprentissages et une éducation est nécessaire pour prendre conscience de ce plaisir et ensuite de s'épanouir. A partir du moment où les enfants auront compris cela, ils sauront eux-mêmes réguler leur utilisation de ces outils. Je reçois régulièrement des étudiants qui sont candidats pour faire un stage dans mon laboratoire, je leur demande systématiquement s'ils utilisent ChatGPT, pour faire quoi et comment. Régulièrement, certains me répondent, je l'utilise pour faire ci, mais je ne l'utilise pas pour faire ça, car c'est trop facile et j'ai envie d'apprendre. Là, je vois que l'étudiant a compris le sens de ses études. C'est sur le sens qu'il faut travailler.

Françoise Meauzé (Fédération des Associations Familiales Catholiques en Europe)

Une question simple, j'ai eu deux activités dans mon existence, une carrière familiale et une carrière associative. La seconde m'a permis de vérifier que les familles, les terrains dans lesquelles nos enfants sont plantés, ont un rôle éducatif absolument impossible à remplacer. Nous sommes maintenant dans une situation où avec l'arrivée de cette modernité, il faut équilibrer les tensions. Parce que cela crée des tensions énormes : savoir que l'on doit toujours être en train de se demander : c'est vrai, ce n'est pas vrai, cela m'aide, cela ne m'aide pas, est-

ce que cela me permet de mieux regarder autour de moi ou est-ce que je vis dans un monde virtuel ? Est-ce je vois mon prochain le plus proche ou celui qui est sur un écran très loin ? Il y a beaucoup de choses qui font que notre humanité risque de s'appauvrir et le seul moyen de ne pas l'appauvrir, c'est de faire comprendre aux enfants dès leur plus jeune âge qui ils sont : ils ne sont pas des produits, ils sont des humains, voulus, je ne dirais pas programmés. On vous dit que si vous avez des enfants, vous en prenez pour dix ans. Ce n'est pas vrai : vous en prenez pour 20 ans aujourd'hui. Quand vous avez la chance d'avoir des petits-enfants qui suivent assez vite, vous n'êtes jamais à la retraite. Les Apprentis d'Auteuil ont compris une chose depuis longtemps, qu'ils pouvaient aider les enfants qui leur étaient confiés, à devenir eux-mêmes, à prendre leurs responsabilités à se mettre au monde, mais que leur tâche avait beaucoup moins de chances de réussir si les familles de ces enfants n'étaient pas parties prenantes.

Laurent Bibard

Je voudrais dire que je souscris totalement à ce que Madame vient de dire, dans le sens suivant : un enfant comme le disent les taoïstes, c'est un point de vue sur le monde qui n'a jamais existé avant et qui n'existera plus jamais après. Cela veut dire que la rencontre avec un enfant est totalement inédite, ineffable, singulière, et que là nous avons tout à apprendre. Du coup, cela libère l'avenir dans un monde où nous croyons beaucoup trop souvent que les technologies épuisent le futur, parce qu'elles seraient censées tout nous donner sous la main, toutes les données. Cela n'est pas vrai. Notre enfance, notre humanité réouvrent tout. Nous, humains, avons voulu en quelque sorte contrôler le monde et cela a du sens qu'il y ait un minimum de contrôle sur le monde. Mais croire qu'il n'y a que le contrôle est tout à fait contre-productif. Parier sur l'enfance et l'avènement des générations futures a profondément du sens.

Joséphine Corcoral

Je voulais réagir sur la question des jeunes usagers et usagères. On sait que les usages numériques et la capacité des familles à accompagner les enfants et les jeunes publics sont très marqués socialement. Je trouve toujours important de le rappeler. Au sein du Conseil National du Numérique, nous avons une sociologue, Dominique Pasquier, qui a travaillé sur l'Internet dans les familles modestes. Ce sont des constats qui se prolongent aujourd'hui avec les nouveaux outils qui se perfectionnent, avec l'IA générative. La capacité à éclairer, faire se questionner, demande d'avoir des compétences mais aussi l'espace mental de se poser ces questions et d'y être soi-même sensibilisé. Cela est très marqué socialement.

Christophe Salomon

Le deuxième conseil est de limiter le temps d'écran pour les enfants. C'est très mauvais pour leur développement. Il y a eu un rapport de l'Académie des Sciences sur ce sujet.

Joséphine Corcoral

Sur le temps d'écran, nous avons de plus en plus de données qui alertent sur les jeunes publics, surtout les très jeunes. Quand on passe le cap de trois à cinq ans, le lien de causalité est plus flou et on sait surtout que l'écran fait écran à d'autres activités qui, elles, sont très bonnes pour le développement de l'enfant, le jeu notamment. Je ne souhaite pas entrer dans un rapport de diabolisation des écrans notamment pour les parents qui sont pétris d'injonctions toute la journée, ce qu'il faut ou ne pas faire avec les enfants et les écrans en rajouterait une couche. Tout dépend de ce que l'on fait en réalité. Il peut y avoir des usages pédagogiques, hyper riches pour les enfants et l'enjeu est de les accompagner. Un élément important dans le

développement cérébral de l'enfant est l'attention conjointe, le fait d'être attentifs à plusieurs sur un même objet entre parents et enfants. Le jeu est un exemple d'attention conjointe. Laisser l'enfant seul sur l'écran est évidemment délétère. Mais accompagner l'enfant et utiliser ensemble l'écran, permet d'entrer dans un processus d'attention conjointe et de tendre vers un usage plus positif. Passer une heure à regarder une vidéo You tube de vulgarisation scientifique n'a pas le même effet que regarder des dessins animés, alimenter pendant une heure une page Wikipedia est encore un autre usage. Ce n'est pas tant une question de temps passé devant l'écran que d'usage.

Stéphane de le Fouchardière (Ancien président du CCIC)

Nous avons beaucoup parlé de l'éducation des jeunes à l'IA, mais qu'en est-il des adultes et des générations plus anciennes ? Dans le métro, je ne vois que de grands enfants, autant d'adultes que de jeunes scotchés à leur téléphone. Ce n'est pas mieux que les petits. Existe-t-il des pratiques éducatives pour les adultes dans les entreprises, ou également pour les retraités ?

Une personne dans la salle

Je voulais réagir à ce qui est dit sur les usages car il y a la question de l'exemplarité. Il n'est pas évident pour un enfant de distinguer l'usage du smartphone pour se repérer, l'usage professionnel, de l'usage récréatif.

Diana Colunga (Fidesco)

Il était question des défis posés par l'IA générative. Avec l'essor des fake news générées par l'IA, notamment les *deepfakes*, où en est l'Europe en termes de protection ? Comment peut-on développer un esprit critique face à cette vague de désinformation, que l'on soit jeune ou moins jeune ? Comment peut-on être acteur dans ce mouvement-là ?

Joséphine Corcoral

Merci pour la question sur l'exemplarité, c'est le point qui m'échappait tout à l'heure. Au Conseil National du Numérique, quand nous étions devant des jeunes publics, nous posions la question : qui estime passer trop de temps sur les écrans ? Tous levaient la main. Et quand on demandait : qui trouve que ses parents passent trop de temps sur les écrans ? Tous levaient la main à nouveau.

Je suis tout à fait d'accord. Aujourd'hui, nous mettons beaucoup d'énergie sur l'éducation des jeunes, mais nous oublions que les adultes aussi ont besoin d'apprendre. C'est pour cela que nous avons lancé des programmes comme Café IA, qui s'adressent à tous les publics : jeunes, parents, professionnels, retraités. Mais il faut aussi des initiatives en entreprise, pour aider les salariés à comprendre et maîtriser ces outils.

Sur la question des contenus générés artificiellement, vous mentionniez aussi les fausses informations et tous les contenus qui altèrent le réel. Où en est-on du droit ? Il y a deux niveaux de réponse : juridique et éducatif. Côté juridique, l'AI Act impose que tout contenu généré par IA soit explicitement marqué comme tel. Mais il y a un problème de faisabilité : c'est techniquement difficile à faire respecter, car les marquages peuvent être retirés. L'autre solution, c'est d'agir sur la diffusion des *fake news*, notamment sur les algorithmes des réseaux sociaux qui les mettent en avant. C'est ce que prévoit le Digital Services Act. »

Pierre-Yves Oudeyer

A côté des régulations juridiques, l'enjeu est celui d'une éducation de masse, non seulement pour les jeunes, mais aussi pour les adultes, les familles, les grands-parents. Nous devons enseigner comment repérer et analyser une *fake news*. L'éducation la plus importante n'est pas une formation technologique pour savoir ce qui se passe sous le capot, mais une éducation aux enjeux sociétaux, culturels, démocratiques. Cette éducation ne s'adresse pas à des personnes très averties, on peut commencer à l'école primaire. Développer des ateliers, discuter ensemble, entre adultes, entre adultes et enfants.

Marc Deluzet

Merci à tous les quatre pour vos interventions et remarques. En conclusion, je soulignerai dans le domaine éducatif qui nous occupait ce soir la question de l'accompagnement : l'accompagnement à l'utilisation des outils d'IA, mais aussi l'accompagnement à l'impact écologique, l'accompagnement sur le sens et la finalité des usages des technologies.

Synthèse et RECOMMANDATIONS

Introduction – Un monde en métamorphose

Depuis la pandémie de COVID-19, le monde vit une profonde transformation. L'intelligence artificielle (IA) en particulier l'IA générative bouleverse les repères éducatifs, sociaux et économiques. Les tables rondes du CCIC & CNFU interrogent l'éthique, le rôle de l'humain et la formation des jeunes.

1. Comment ça marche ? Comprendre l'IA

L'IA regroupe des techniques diverses (symbolique, comportementale, statistique, générative) qui manipulent des données sans conscience ni compréhension. Elle est omniprésente dans notre quotidien (réseaux sociaux, assistants, recommandations) et repose sur des logiques d'optimisation.

Elle peut amplifier des fake news, manipuler nos comportements, ou donner l'illusion d'intelligence. L'éducation à ces mécanismes est donc essentielle.

L'IA et son fonctionnement pose une série de **questions éthiques, des sujets de questionnements** :

- Peut-on parler de vérité si une IA ne comprend pas ce qu'elle dit ?
- Que signifie utiliser une IA sans en révéler l'origine et les sources dans un travail ?
- Quel est l'impact cognitif de ces technologies ?
- L'homme devient-il dépendant d'une machine qu'il alimente volontairement ?
- Comment faire face à l'illusion de pensée (collectives ou individuelle) ?
- Quelle responsabilité pour l'utilisateur ?
- Comment gérer le risque de servitude volontaire ?
- Sensibiliser au phénomène d'amplification des discours de haine via les mécanismes des réseaux sociaux.
- Considérer l'**impact culturel** (chaque société perçoit cerveau et IA différemment), et l'impact environnemental
- Construire un cadre global et **interdisciplinaire**.

- **Éduquer** à la complexité, à la réflexion et à l'évaluation bénéfice/risque.
- Risques liés aux neurotechnologies,
- Nécessité d'encadrer la fusion des données,
- Protection de l'intériorité mentale,
- Définition de nouveaux droits, (numériques)
- Réflexion sur la responsabilité individuelle et collective.
- **Observer le monde tel qu'il est**, pour construire une régulation ancrée dans la réalité.
- Reconnaître la **diversité des cultures juridiques et éthiques** (ex. : bioéthique asiatique non fondée sur les Droits de l'Homme).
- Distinguer **œuvre humaine** et production IA dans les débats sur la propriété intellectuelle.
- Nécessité d'un droit international souple mais effectif,
- Respect de la pluralité culturelle,
- Régulation globale contre les dérives individuelles et industrielles.
- Attention la **dualité de l'IA** : technologie versus application
- Il dénonce une **illusion d'éthique universelle** : toute IA encode nécessairement des valeurs (exemple des IA chinoises excluant Tian'anmen).
- On démontre l'**impossibilité technique** de créer une IA non discriminante au sens absolu (14 définitions mathématiques de l'équité, souvent incompatibles).
- Une **régulation différenciée** : forte pour les applications, souple pour la recherche.
- Des choix politiques clairs sur **les critères de justice et d'équité à implémenter**.
- Une vigilance sur la **dépendance technologique européenne** vis-à-vis des modèles américains et chinois,
- Attention au Biais structurels des IA,
- Opacité des critères d'équité,
- Difficulté de transcrire des valeurs humaines dans des modèles mathématiques, géopolitique des données et des imaginaires

2. Le discernement... et mon âme, dit le robot ?

Cette table ronde explore les enjeux éthiques, philosophiques et spirituels de l'IA. Face à la simulation d'émotion ou d'intuition, il est vital de préserver la liberté humaine, la créativité et la singularité de chacun.

Des voix comme celle de Thierry Magnin appellent à une ‘boussole intérieure’ et rappellent que la machine n’a ni âme, ni inspiration, ni expérience humaine.

Les résultats de l’IA et les mises en garde sur le discernement amènent à une série de **questions éthiques et sujets de questionnements intérieurs et philosophiques** :

- **Perte du silence intérieur et de la réflexion personnelle** dans un monde numérique saturé.
- **Dérive technocratique** : l’algorithme risque de prendre le dessus sur le discernement humain.
- **Rôle central de l’éducation** dans le maintien de l’autonomie morale, du débat et du discernement.
- **Déshumanisation possible** : l’IA ne peut comprendre l’être humain comme une triade « corps-âme-esprit ».
- **Critique du transhumanisme** et de la délégation des décisions aux puissances technologiques.
- **Ambivalence de l’IA** : entre innovations utiles et dangers de perte de contrôle (Prix Nobel 2024).
- **Risques de plagiat, hallucinations et manipulation des images** via l’IA générative.
- **Réductionnisme matérialiste** : la pensée humaine n’est pas un simple traitement de données.
- **Cyborgisation** : fusion homme-machine qui menace l’identité humaine.
- **Confusion machine-humain** : notamment à cause de performances simulant émotions ou conscience alors que la machine n’a aucune conscience
- **Nécessité d’une éducation au prompt et au discernement** : apprendre à dialoguer avec l’IA sans perdre son autonomie.
- **Incarner une pédagogie active** où l’humain garde sa singularité.
- **Différencier intelligence humaine (émotionnelle, incarnée) et intelligence artificielle (statistique, calculatoire).**
- **Remise en cause de la création artistique** par l’IA générative.
- **Danger de l’hubris technicienne** : prétention de l’IA à rivaliser avec l’artiste.
- **Décoloration du goût et affaiblissement de l’esprit critique** : art aseptisé, consensuel.
- **Perte de la transcendance dans la création** : l’IA ne connaît ni l’inspiration, ni l’échec, ni l’incarnation.
- **Préserver la création humaine comme acte unique, imprévisible, traversé par le souffle, la grâce.**

- **Redonner une place centrale à l'émotion, l'émerveillement, le vécu**, que la machine ne peut reproduire.
- **Anthropomorphisme abusif de l'IA** : confusion homme-machine dans l'imaginaire collectif.
- **Désinformation, biais, hallucinations** : nécessité de former aux limites techniques et aux enjeux commerciaux.
- **Alignement problématique des objectifs IA/humains** : exemple de l'araignée IA.

Proposition d'une boussole éducative qui oriente vers :

Nord : Esprit critique pour questionner les résultats.

Sud : Curiosité face à l'innovation sans naïveté.

Est : Responsabilité morale, respect des valeurs.

Ouest : Préservation de l'émerveillement, de l'empathie et de la créativité humaine.

- **La 4ème blessure narcissique** de l'humanité : l'IA remet en cause notre unicité cognitive.
- **Décorrélacion éthique/technique** : l'illusion que l'objet technique est neutre.
- **Danger d'une morale a posteriori** : l'éthique ne doit pas arriver après la conception des outils.
- Revenir à **Hans Jonas** et son **principe de responsabilité** : agir selon une logique de **préservation de la vie humaine authentique**.
- **Éduquer à la sagesse dans un monde technique**, en intégrant l'éthique dès la conception des IA.
- **La machine a-t-elle une conscience ?** (Non, car pas de corps ni d'émotion).
- **La création artistique et la singularité humaine** comme rempart à la robotisation.
- **L'éducation doit-elle mieux intégrer le corps et le dialogue ?**
- **Le rôle de l'UNESCO** : donner des orientations éducatives plus incarnées et universelles.

3. Les moyens pédagogiques

Face à la généralisation de l'IA, l'éducation doit évoluer. Les enseignants doivent être formés, et les élèves accompagnés dans une posture critique et créative.

La 'boussole pédagogique' proposée repose sur 4 axes : esprit critique, curiosité technologique, responsabilité éthique et émerveillement.

Les Mises en garde par rapport à l'éducation des formateurs et des « enseignés ».

Car les manières d'enseigner l'IA doit soulever les **questions éthiques et anthropologiques** :

- Se rappeler **les IA génératives ne sont pas réellement intelligentes** :
 - Elles ne comprennent pas les concepts comme le fait un humain.
 - Elles ne font que **statistiquement associer des mots**.
 - Elles peuvent **inverser des faits** ou **inventer des informations**, car elles ne disposent pas d'une **véritable logique de vérification**. Il faut donc garder un **esprit critique** face aux résultats produits par l'IA.
 - **L'éducation au numérique doit inclure la formation à l'évaluation des réponses des IA.**
 - Un bon scientifique ou un bon économiste sait donner des ordres de grandeur approximatifs- un « utilisateur d'IA » doit faire de même.
 - **La nécessité de diffuser une culture numérique partagée.**
 - Généraliser l'initiative « **Café IA** » : **espaces de discussion autour de l'intelligence artificielle, accessibles à tous et sans barrière technique.**
 - **L'impact environnemental de l'IA, souvent peu traité dans le débat public doit être traité** - L'IA représente aujourd'hui près de **1% de la consommation électrique mondiale**, un chiffre qui va exploser
 - **Les IA génératives sont des technologies de transmission culturelle** : Elles **reproduisent** les connaissances, croyances, mythes et biais des données sur lesquelles elles ont été entraînées. Elles reproduisent **des stéréotypes**.
 - **Montrer les opportunités** :
 - **La science** : les lauréats des Prix Nobel en IA ont montré son **impact en physique, en chimie et en biologie**.
 - **L'éducation** : l'IA peut aider à personnaliser les apprentissages et donner accès à des connaissances aux populations les plus défavorisées.
- ⇒ **L'auto-apprentissage de l'IA peut renforcer de fausses croyances sur son fonctionnement**. Il est donc **essentiel d'accompagner** les jeunes dans leur usage des IA. Nous devons former les élèves à :
1. **Développer leur esprit critique.**
 2. **Améliorer leur métacognition** (*savoir ce qu'on sait et ce qu'on ignore*).
 3. **Se poser la question : pourquoi apprendre ?**
- ⇒ Si un élève comprend **l'intérêt des mathématiques ou de l'histoire**, il ne cherchera pas à tricher avec ChatGPT. Il saura **quand l'IA est une aide et quand elle est un obstacle à son apprentissage**.

- **L'intelligence artificielle n'est pas une question technologique, c'est une question humaine.**
- Nous avons tendance à nous demander **ce que les technologies vont nous imposer**, alors que nous devrions plutôt nous interroger : « **Que voulons-nous faire des technologies ?** »
- **L'IA, Nous devons la démystifier.**
- Nous devons **questionner notre propre servitude volontaire** face aux machines : Pourquoi acceptons-nous si facilement de **nous soumettre à des algorithmes qui orientent nos décisions, nos choix culturels et nos opinions ?** Nous avons le pouvoir de dire **non**.
- Nous devons nous demander : **Quel est le véritable rôle de l'IA dans notre société ? Jusqu'où voulons-nous utiliser ces technologies ? Sommes-nous prêts à en payer le coût énergétique et social ?**

10 recommandations issues du cycle : « Comment éduquer à un usage responsable des Intelligences artificielles ?

Développer l'éducation à l'IA dès le secondaire.

Former les enseignants aux usages et aux enjeux éthiques des IA.

Diffuser une culture du discernement.

Mettre en œuvre les travaux de la chaire UNESCO « Pratiques de la philosophie avec les enfants » qui promeuvent la formation à l'esprit critique et aux valeurs humanistes dès le primaire.

Impliquer les jeunes dans la gouvernance des outils IA.

Favoriser la diversité culturelle et linguistique des données.

Garantir la transparence et la lisibilité des IA.

Protéger la création humaine et l'intuition.

Utiliser l'IA pour renforcer l'inclusion éducative.

Promouvoir des approches interdisciplinaires.

Sommaire

Introduction	2
Première partie : Ethique et IA, de quoi parle-t-on ?	
1-L'IA : de quoi parle-t-on ? et comment ça marche ? Isabelle Chaperon	4
2- L'IA, les neurosciences et l'éthique-Laure Tabouy	9
3-Promouvoir les savoirs et l'éducation, au cœur de la mission de l'UNESCO	12
Chistian Byk	
4-Des technologies au service des <i>soft powers</i>- Thimotée Hunter	15
Débat	21
Deuxième partie : Eduquer à l'éveil de la conscience pour un usage responsable des IA	
Monique Grandjean	27
Thierry Magnin	29
Alexandra Maurice	33
Louis Lourme	36
Débat	39
Troisième Partie : Repères et pistes pour éduquer	
Christophe Salomon	45
Joséphine Corcoral	47
Pierre-Yves Oudeyer	50
Laurent Bibard	54
Débat	56
Synthèse et recommandations	66

