



**TECHNIQUES
DE L'INGÉNIEUR**

Réf. : **AG124 V1**

Date de publication :
10 octobre 2023

Éthique de l'intelligence artificielle : vue d'ensemble

Cet article est issu de : **Innovation | Éco-conception et innovation responsable**

par **Dr Johan ROCHEL**

Mots-clés

technologies | intelligence artificielle | éthique | valeurs

Résumé Cet article présente une vue d'ensemble des questions d'éthique soulevées par l'intelligence artificielle (IA). Ces questions sont structurées sous la forme d'un outil de navigation distinguant trois grandes familles thématiques. La première famille de questions – l'éthique dans la technologie – traite des problématiques liées à la conception et à la réalisation d'outils d'IA. La deuxième famille, « justice sociale », englobe les débats de société quant à l'utilisation à large échelle de ces outils. La troisième famille porte sur les grands récits de société sur les rapports humain-machine. Chaque section propose un résumé des questions que les équipes de développement d'outils d'IA devraient se poser.

Keywords

technologies | artificial intelligence | ethics | values

Abstract This article presents an overview of the ethical issues raised by artificial intelligence (AI). These issues are structured thanks to a navigation tool distinguishing three main thematic families. The first family of issues - ethics in technology - deals with issues related to the design and creation of AI tools. The second family addresses the societal debates about the widespread use of these tools from a social justice perspective. The third family includes questions related to the most important narratives and framings on the relationship between humans and machines. Each section provides a box that summarizes the questions that AI tool development teams should address.

Pour toute question :

Service Relation clientèle
Techniques de l'Ingénieur
Immeuble Pleyad 1
39, boulevard Ornano
93288 Saint-Denis Cedex

Par mail :

infos.clients@teching.com

Par téléphone :

00 33 (0)1 53 35 20 20

Document téléchargé le : **17/11/2023**

Pour le compte : **7200105899 - service public de wallonie // chantal DEVILLEZ // 157.164.187.237**

© Techniques de l'Ingénieur | Tous droits réservés

Éthique de l'intelligence artificielle : vue d'ensemble

par **Dr Johan ROCHEL**

Chercheur
EPFL – Lausanne, Suisse

1. Définitions	AG 124 - 2
1.1 Intelligence artificielle.....	— 2
1.2 Éthique	— 4
2. Un outil de navigation pour les questions éthiques	— 5
3. Éthique dans la technologie	— 5
3.1 Éthique des données	— 5
3.2 Éthique des algorithmes.....	— 6
3.3 Éthique du design	— 7
3.4 Défis de durabilité	— 8
3.5 Défis de contrôle et de responsabilité	— 9
4. Impact sur la société et justice sociale	— 10
4.1 Poser la question de la justice	— 10
4.2 Impact de l'IA sur le marché du travail	— 11
4.3 Une double conversation	— 11
5. Grands récits	— 12
5.1 Les grands récits	— 12
5.2 Les rapports entre l'humain et la machine	— 13
5.3 Les spécificités humaines.....	— 13
5.4 Le sens de la vie humaine	— 13
6. Conclusion.....	— 14
Pour en savoir plus	Doc. AG 124

Depuis fin novembre 2022, le grand public peut pleinement prendre conscience du potentiel des outils d'intelligence artificielle (IA) générative comme ChatGPT. Si les spécialistes du domaine travaillent depuis plusieurs années sur des modèles de langages capables de générer du contenu de manière naturelle (texte, image, son), le lancement public de ChatGPT rentrera certainement dans l'histoire de l'IA comme un moment de démocratisation d'outils encore, selon leurs critiques, trop peu maîtrisés. En parallèle à cette montée en puissance des outils d'IA, les questions éthiques connaissent, elles aussi, une phase de démocratisation.

Ces outils d'IA générative font partie d'un ensemble d'outils, méthodes ou techniques rassemblés dans la grande famille de l'« intelligence artificielle ». Comme définie plus loin, l'éthique pose la question des valeurs et principes dans la conception, la réalisation, l'utilisation et la régulation d'outils numériques. Ces questions de valeurs et principes s'imposent partout et l'éthique de l'IA est en pleine expansion. Les contributions académiques se succèdent à une vitesse impressionnante, le débat public s'empare des questions éthiques et les utilisateurs regardent de manière plus critique les outils proposés. Loin d'être en reste, les entreprises intègrent de plus en plus une dimension proprement éthique dans leurs processus internes. Cette importance grandissante

reflète parfois la conviction profonde des équipes de direction, mais également une ambition de gagner une bonne réputation en se plaçant aux avant-postes d'un mouvement qu'on pressent profond ou, à tout le moins, afin d'éviter des accidents de parcours dommageables pour la relation de confiance avec les clients et les partenaires. Les initiatives des entreprises visent finalement à anticiper les régulations à venir sur les usages de l'IA. À l'exemple de l'Union européenne (UE), la régulation en discussion sur l'IA – visant à établir des règles harmonisées pour les pays de l'UE – va durablement modifier la manière dont les entreprises, les administrations publiques et les particuliers conçoivent et utilisent les outils d'IA.

Le présent article vise à proposer une vue d'ensemble des questions d'éthique de l'IA à destination des spécialistes d'ingénierie. En plus d'une explication générale sur les défis en présence, chaque section se conclut par une série de questions directement pertinentes pour les entreprises/organisations dans leur gestion des outils d'IA.

Le texte est organisé en cinq parties. La première traite des définitions des termes d'intelligence artificielle et d'éthique. La deuxième présente l'approche de l'outil de navigation choisi afin de structurer les questions à se poser. Cet outil fait office de boussole pour trouver son chemin parmi les questions d'éthique ; il propose trois familles de questions éthiques, qui sont développées séparément. La troisième partie traite ainsi de la première famille de questions éthiques, à savoir l'éthique dans la technologie. Les choix éthiques dans la conception et la réalisation des outils d'IA sont au cœur de ses réflexions. La quatrième partie porte sur l'impact des outils d'IA sur les individus, qu'il s'agisse d'une petite équipe de spécialistes ou de la société dans son ensemble pour les outils les plus importants. Ces questions sont abordées avec le concept de justice sociale. La cinquième partie thématise les grands récits de société à travers lesquels nous appréhendons la technologie. Ces récits sont l'occasion de faire un pas en arrière pour interroger la manière dont les outils numériques modifient notre vision de la réalité. Comme on le verra plus loin, ces trois parties forment un tout à considérer dans sa globalité.

1. Définitions

Avant toute chose, il convient de poser les définitions nécessaires à la suite de la lecture, principalement concernant les termes « intelligence artificielle » et « éthique ». Il importe de souligner que ces définitions sont bien souvent au cœur des débats politiques ou scientifiques. Choisir le terme d'« intelligence artificielle » puis sélectionner l'une de ses définitions n'est jamais neutre. Une définition va reposer sur des choix et des évaluations. À sa manière, elle va distinguer ce qui est important de ce qui l'est moins.

1.1 Intelligence artificielle

Afin de définir l'IA, il est possible de démarrer large puis de préciser différentes dimensions.

Pour simplifier, un outil d'IA est défini comme un outil reposant sur un modèle comprenant un ou plusieurs algorithmes qui réalisent un objectif via des opérations statistiques et mathématiques [1] [2]. L'analogie avec la recette de cuisine est souvent utilisée pour expliquer au grand public le fonctionnement d'un algorithme. Comme la recette de cuisine présente les ingrédients et les étapes pour réaliser un plat, un algorithme contient une suite d'opérations statistiques-mathématiques permettant de réaliser un objectif défini. Les opérations les plus courantes portent sur l'identification de certaines données/occurrences statistiques dans un

large ensemble de données, la structuration d'un ensemble selon certains critères ou, comme pour ChatGPT et les « larges modèles de langage » (LLM), l'apprentissage d'une forme de langage dite « naturelle » sur la base d'une analyse d'un immense corpus de textes.

Cette première définition peut ensuite être précisée selon au moins deux axes.

■ Premièrement, il est possible de distinguer entre plusieurs types d'IA. Cette qualification nous emmène brièvement dans l'histoire de l'IA. En tant que discipline, l'IA moderne apparaît dans les années 50. Sa naissance est souvent associée à un séminaire de recherche organisé à Dartmouth en 1956. Le projet de recherche des « pères » de l'IA stipule leur hypothèse selon laquelle « chaque aspect de l'apprentissage ou n'importe quelle caractéristique de l'intelligence peuvent, en principe, être décrits de manière suffisamment précise afin qu'une machine puisse être construite pour la simuler » [3]. Cette ambition d'imiter et de simuler l'intelligence humaine va prendre plusieurs formes.

Pendant longtemps, l'IA fonctionne selon un modèle général symbolique. Dans ce modèle, les différents éléments de la réalité sont représentés au travers de symboles. Les raisonnements humains sont modélisés mathématiquement et appliqués à ces symboles. Cette approche symbolique a notamment permis la création de « systèmes experts », des systèmes qui suivent des règles de décisions prédéterminées. De manière imagée, ces systèmes permettent de prendre des décisions dans un espace déterminé, sur la base de règles et d'options connues.

Aujourd'hui, de vastes ensembles de données et la puissance de calcul nécessaire pour les analyser ont permis l'émergence d'un modèle d'IA dit « connexionniste ». Il repose sur l'analyse, supervisée ou non, d'un très grand nombre de données et sur la capacité d'algorithmes de découvrir et d'exploiter des occurrences statistiques dans ces données.

Ce modèle domine la recherche scientifique et le développement d'outils concrets. L'apprentissage machine (*machine learning*) se décline sur plusieurs méthodes d'apprentissage, comme l'illustre la figure 1.

Dans une **approche supervisée**, on propose par exemple à la machine des millions d'images de chats afin qu'elle identifie ce qui définit un chat (ses quatre pattes, sa queue, ses moustaches, le tout exprimé sous forme de pixels) et comment reconnaître ces caractéristiques. Cette méthode d'apprentissage est dite « supervisée » car les catégories utilisées pour classer les animaux sont connues d'avance, de même que la réponse correcte dans la phase d'entraînement.

Une **méthode non supervisée** est également possible. Dans ce cas, on fournit au programme des millions de données et on le laisse déterminer des catégories d'analyse. La machine cherche alors à repérer des liens au sein de ces données, à les regrouper, les associer et les étiqueter. Les larges modèles de langage comme ChatGPT reposent, en partie, sur cet apprentissage non supervisé. Le programme traite des milliards de mots, analyse leur contexte et en déduit des régularités statistiques qu'il va exploiter pour identifier le mot le plus plausible dans un contexte donné.

Mais cette approche est porteuse de défis importants. Le programme peut faire apparaître des liens statistiques qui n'ont aucun sens entre plusieurs types de données. La nécessité de distinguer entre corrélation et causalité est un problème bien connu en statistiques. Par exemple, mon activité professionnelle et mon niveau de richesse peuvent aisément être reliés par une causalité. En revanche, le taux de mariage et la consommation de beurre par personne n'ont a priori rien à voir ensemble, même s'il est possible de trouver une corrélation statistique entre ces données. Dans les larges modèles de langage, l'approche non supervisée peut conduire à des résultats absurdes ou jugés dangereux.

C'est entre autres pour répondre à ces défis qu'intervient l'humain pour améliorer les résultats de la machine. Dans une perspective d'**apprentissage renforcé**, les interventions humaines visent à corriger certains résultats non satisfaisants. À titre d'exemple, l'entreprise développant ChatGPT, OpenAI, a déclaré fournir des réponses « modèles » sur certaines questions sensibles [4]. On force alors la machine à travailler avec cette réponse

donnée. De même, des interventions humaines peuvent évaluer différentes réponses fournies à une question et donner des indications à la machine sur la réponse la plus souhaitée (par exemple en classant de 1 à 4 une série de réponses et en demandant à la machine de « reproduire » autant que possible les réponses ayant reçu le plus de points). Dans le cas de ChatGPT, il est difficile d'en savoir plus vu le manque de transparence sur les types de réponses données et les instructions données aux équipes de l'étiquetage.

Ces mesures d'entraînement demandent souvent un investissement humain considérable et un travail de « petites mains ». Dans le cas de ChatGPT, le journal *Times* a découvert que ce travail avait été externalisé à une entreprise située au Kenya [5]. Cette chaîne de création de valeur dans les technologies numériques soulève pour elle-même de nombreuses questions, notamment sur les conditions de travail des employés des sous-traitants et leur invisibilisation quasi-totale (au profit d'entreprises présentées comme high-tech dans la Silicon Valley...).

■ Deuxièmement, cette définition très générale de l'IA peut être précisée par le biais des usages visés. Cette définition met en exergue ce que l'IA rend possible.

À titre d'exemple, le projet de Règlement européen sur l'IA définit un système d'IA comme [titre I^{er}, art. 3, al. 1] : « un logiciel qui est développé au moyen d'une ou plusieurs des techniques et approches énumérées à l'annexe I et qui peut, pour un ensemble donné d'objectifs définis par l'homme, générer des résultats tels que des contenus, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels il interagit » [6]. La définition énumère certaines techniques (énoncées dans une annexe, et donc susceptibles d'évoluer au gré des avancées technologiques), mais surtout certains usages. L'approche focalisée sur les usages rendus possibles par un système d'IA – plutôt que sur la technologie utilisée – a l'avantage de mettre en lumière l'impact concret sur les individus. Ainsi, l'ONG Algorithmic Watch ne parle pas de système d'IA, mais plutôt de **processus de décision algorithmique** (ADM, *algorithmic decision-making*) [7].

Le caractère automatique associé à une utilisation à des fins de décision ou d'aide à la décision est, d'après l'ONG, la dimension problématique de l'outil.

Pour notre réflexion éthique à venir, nous pouvons tirer **trois leçons importantes** de ces réflexions sur la définition de l'IA :

- La définition d'un outil d'IA est nécessairement déjà au cœur d'un débat politique sur les défis de l'IA. Mettre en évidence telle ou telle dimension des technologies ou des usages reflète ce que la personne estime être sujet à débat.

- Les outils d'IA sont des artefacts créés par les humains, et qui sont utilisés dans différents contextes. Une discussion uniquement basée sur les caractéristiques techniques de l'outil laisse une immense partie des questions pertinentes dans l'ombre. Il faut considérer ces outils comme des systèmes socio-techniques, c'est-à-dire des outils techniques utilisés dans un certain contexte. C'est la rencontre des dimensions techniques et sociales qui permet de bien identifier les différentes questions. Un même outil peut soulever des questions différentes suivant son contexte d'utilisation, par exemple le type d'institution où il est utilisé, les contraintes qui pèsent sur les utilisateurs ou encore les rapports de pouvoir au sein des équipes [8].

- Cette approche socio-technique nous permet de prendre une saine distance avec la frénésie technologique. La réflexion éthique ne doit pas dépendre de chaque ajustement technique, sauf si les ajustements sont cruciaux pour certaines fonctions. La réflexion vise à rester pertinente au-delà des effets d'annonce sur les « révolutions » techniques en cours. Pour ce faire, garder bien en vue les usages et les effets de ces usages sur les personnes reste une méthode éprouvée.

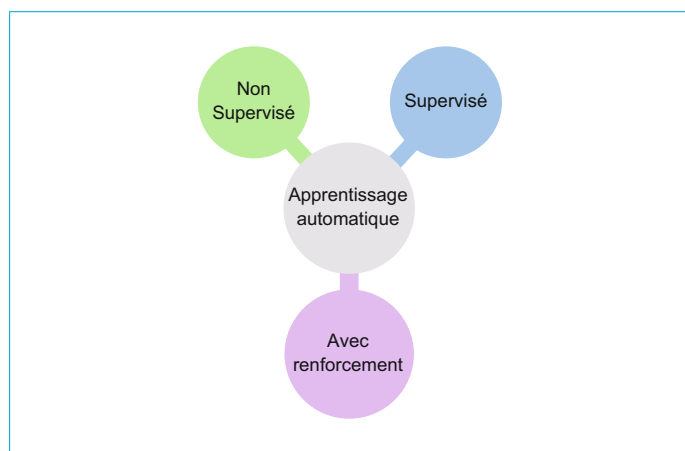


Figure 1 – Vue d'ensemble des domaines de l'apprentissage machine

1.2 Éthique

Le mot « éthique » s'est imposé au quotidien et on le retrouve comme qualificatif sur les produits de consommation, les fruits et légumes, l'artisanat, l'immobilier ou les services technologiques. Cette utilisation inflationnaire du terme doit nous inciter à la prudence. Dire qu'une chose est « éthique » implique de clarifier le standard retenu : éthique par rapport à quelles valeurs ? Il nous faut alors identifier certaines valeurs, principes ou convictions, les définir, puis chercher à établir comment ces standards sont réalisés dans des situations spécifiques.

En tant que discipline des sciences humaines, l'Éthique est une sous-famille de la philosophie. Elle s'intéresse en priorité aux situations de choix et à la justification que ces situations appellent. L'éthique est donc tournée vers l'évaluation et l'action. En matière d'IA, l'éthique va donc interroger l'ensemble de la chaîne de conception, création, réalisation, utilisation et régulation d'outils technologiques. Telle que présentée ici, l'éthique se déploie à trois niveaux :

- Au niveau « micro », elle adopte la perspective des acteurs impliqués dans cette chaîne de création et d'utilisation. Elle pose la question du bien décider et du bien agir à l'aune de leurs valeurs. Au-delà de cette dimension micro, l'éthique interroge sur comment des individus faisant société décident de leurs règles communes.
- Au niveau « meso », elle peut porter sur les organisations intermédiaires, comme l'entreprise et ses choix de valeurs. L'objectif économique de l'entreprise, mais également les manières d'organiser ses activités, relèvent de choix éthiques.
- Au niveau « macro », l'éthique porte sur les choix de société, soit au niveau des États, soit au niveau régional (par exemple l'Union européenne) ou international. Dans ces cas, l'éthique se joint à une analyse de philosophie politique pour investiguer comment les institutions et leurs règles (notamment la régulation des technologies, mais également l'ensemble du cadre juridique) devraient traiter des usages permis par les outils d'IA.

Au vu de cette première définition de l'éthique, il importe de clarifier quatre malentendus courants.

■ D'abord, l'éthique ne se confond pas avec la simple identification des convictions et des intuitions morales que nous avons en tant qu'individus. Cet usage se reflète parfois dans des disputes où chacun est convaincu d'avoir « son éthique » ou « sa morale ». En tant que réflexion structurée, l'éthique tente de clarifier, de structurer et de justifier les convictions individuelles. L'éthique vise ainsi notamment à développer un vocabulaire commun qui nous permet d'échanger sur nos convictions et de les faire évoluer.

■ Ensuite, l'éthique n'est pas l'équivalent d'un code de déontologie professionnelle. Certaines professions, notamment les soignantes et soignants, les avocats et avocats ou encore les ingénieurs et ingénieurs, possèdent des codes de conduite qui énoncent certains principes fondamentaux pour l'exercice de leur profession [9] [10]. Ces codes esquissent ainsi à quoi ressemblent les bonnes pratiques professionnelles du domaine concerné. Ils ont une dimension éthique car ils reposent sur une série de convictions et de valeurs dont les implications dans un domaine d'activité spécifique sont explicitées. Cependant, ils n'épuisent pas les questions éthiques qu'un professionnel devrait se poser.

■ Également, il faut préciser que l'éthique ne travaille pas avec la distinction du « vrai » et du « faux ». Elle n'est pas comparable aux sciences naturelles et à leur ambition de vérité de type 1 + 1 = 2. L'éthique s'intéresse plutôt à la cohérence entre nos valeurs et nos actions. Elle cherche à montrer les pièges de certains raisonnements, à dénoncer des positions philosophiquement intenable ou logiquement contradictoires. Cela explique pourquoi

cet article ne va pas présenter « une » ou « l' » éthique de l'IA, mais des réflexions éthiques sur l'IA. Le lecteur devra ensuite assembler ces différentes réflexions de manière à établir la plus grande cohérence possible avec ses propres convictions et ses décisions.

■ Enfin, l'éthique entretient des liens privilégiés avec le droit. Les concepts utilisés par l'éthique et le droit sont souvent les mêmes (responsabilité, droit, causalité, etc.), mais cette proximité ne doit pas nous tromper. L'analyse juridique fonctionne selon ses propres règles d'interprétation et de procédure. Le droit se construit de manière hiérarchique. Si l'éthique et le droit ont de multiples points communs, il faut donc veiller à ne pas simplement importer une réflexion de l'une dans l'autre. De manière schématique, les réflexions éthiques présentées ici peuvent être utilisées afin de justifier certaines règles juridiques (réflexion éthique comme fondation) ou, à l'inverse, afin de critiquer des règles jugées non satisfaisantes (réflexion éthique comme aiguillon critique).

La régulation sur l'IA de l'Union européenne

En matière de régulation de l'IA, l'Union européenne est un acteur clé du débat global. Le Règlement sur l'IA actuellement en discussion s'intègre dans une vaste stratégie de régulation numérique dont le RGPD (Règlement général sur la protection des données [11]) est l'un des éléments les plus connus. En bref, ce règlement adopte une approche basée sur les risques. Il comprend quatre catégories de risques ; selon la catégorie identifiée, les exigences visant à minimiser les risques évoluent.

Cat. 1 : Outils d'IA interdits

Le règlement mentionne quatre types d'outils interdits : les systèmes d'IA ayant recours à des techniques subliminales pour altérer le comportement des personnes d'une manière qui pourrait causer un préjudice, les systèmes d'IA qui exploitent les vulnérabilités de certains groupes (personnes âgées ou enfants par exemple) afin d'altérer leur comportement, les systèmes d'IA utilisés par les pouvoirs publics qui viseraient à établir un classement de fiabilité des personnes en fonction de leur comportement social, et les systèmes d'identification biométriques à distance utilisés en temps réel (caméras « intelligentes »...).

Cat. 2 : Outils d'IA à haut risque

Cette catégorie inclut notamment les outils d'identification biométriques, les outils utilisés dans les infrastructures critiques, les outils utilisés dans les domaines de la justice, de la police ou encore du contrôle aux frontières. Les outils d'IA utilisés dans les processus RH (notamment recrutement/allocation de tâches) sont également concernés.

Pour tous ces systèmes à haut risque, des conditions strictes de mise sur le marché et de contrôle continu s'appliquent. Les exigences portent principalement sur une analyse de risques, la gouvernance des données, la documentation technique, la transparence, le contrôle humain et des considérations de cybersécurité.

Cat. 3 : Outils d'IA interagissant avec des personnes physiques mais sans risque marqué

Des obligations de transparence envers l'utilisateur s'appliquent.

Cat. 4 : Outils restants

Le règlement encourage ici l'adoption d'un code de conduite destiné à favoriser l'application volontaire d'exigences liées à l'impact environnemental, l'accessibilité pour les personnes en situation de handicap ou la diversité des équipes de développement. Ces codes de conduite que le législateur européen appelle de ses vœux rassemblent de nombreux points à venir de notre discussion sur l'éthique de l'IA.

À retenir

- La définition de l'« intelligence artificielle » peut avoir une composante technique, mais elle n'est pas une question technique. Identifier, nommer et circonscrire une technologie relève d'un débat juridique, éthique, politique et social.
- La régulation des outils d'IA dépend largement de la définition choisie de l'IA. Au-delà du contenu de la législation, c'est l'objet même de cette régulation et les comportements que la technologie rend possibles qui doivent d'abord être établis.
- L'éthique est une réflexion systématique sur les valeurs mobilisées dans un contexte spécifique afin de justifier une décision spécifique. La réflexion éthique porte sur le niveau individuel, sur l'entreprise et sur les règles choisies pour la vie en société.

2. Un outil de navigation pour les questions éthiques

Nous proposons ici un outil de navigation (figure 2) visant avant tout à structurer les questions éthiques autour de l'IA sous forme de grandes thématiques.

La première thématique traite des **questions éthiques dans la technologie**. Elle rassemble les défis qui s'imposent durant le processus de conception, création et réalisation des outils d'IA.

La deuxième thématique rassemble les **questions de justice sociale** liées à l'utilisation des outils d'IA. L'outil est prêt à être vendu, acheté, utilisé librement. Quel sera alors son impact sur la société et comment modifiera-t-il nos comportements ?

La troisième thématique traite des **grands récits que nous mobilisons pour appréhender des outils d'IA**, et plus généralement les technologies numériques. Prendre conscience de ces récits sous-jacents nous aidera à mieux définir et mieux comprendre notre

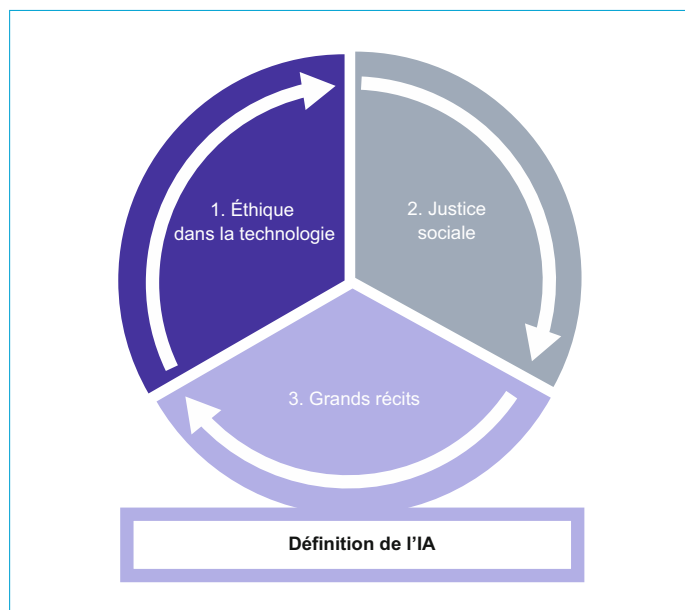


Figure 2 – Outil de navigation des questions d'éthique

rapport aux outils d'IA, notamment la place et la spécificité de l'humain face à la machine.

L'usage de l'outil doit être itératif : les trois thématiques sont complémentaires et doivent être abordées comme un tout. Les réflexions apportées dans l'une des thématiques vont influencer les suivantes. L'usage itératif de l'outil de navigation clarifie que la responsabilité des spécialistes en ingénierie ne s'arrête pas aux questions d'éthique dans la technologie. Même si le présent article porte en priorité sur cette première thématique, il faut souligner que les trois thématiques sont pertinentes.

D'une part, cela amène les spécialistes en ingénierie à sortir de la zone « conception et réalisation » pour aborder de front certains sujets plus politiques et sociétaux. La présente contribution est construite sur la conviction que l'éthique fait pleinement partie du cahier des charges des métiers de l'ingénierie. En ce sens, ce texte peut être lu dans le contexte des changements qui s'opèrent en matière de formation dans les écoles d'ingénieurs [12]. D'autre part, cela signifie également que les non-spécialistes d'ingénierie doivent s'intéresser de plus près à la conception et réalisation des outils technologiques. Ces questions ne peuvent être réservées aux seuls spécialistes techniques. Consommateurs et citoyens doivent demander des comptes aux entreprises et aux organisations qui proposent et utilisent des outils d'IA.

3. Éthique dans la technologie

Cette première thématique contient cinq sous-sections qui forment ensemble une large partie de l'éthique de l'IA : éthique des données, éthique des algorithmes, éthique du design, durabilité, défis de contrôle et de responsabilité.

3.1 Éthique des données

Cette sous-thématique rassemble toutes les questions qui portent sur la **collecte et l'utilisation de données**. Elle fait largement écho au droit de la protection des données, notamment le RGPD en Europe. Ces réflexions spécifiques sur l'éthique des données dans le contexte de l'IA viennent s'ajouter à des considérations plus générales sur la bonne gouvernance des données en entreprise (le cycle de vie des données, c'est-à-dire leur sauvegarde, leur suppression, leur protection) [13].

Dans le débat public, ces questions d'éthique des données sont souvent associées aux problèmes de « biais » des outils d'IA (par exemple envers les femmes ou les minorités). Pour identifier les principaux défis éthiques liés aux données, il est utile de partir de la figure 3 simplifiant le fonctionnement d'un outil d'IA.

Comme expliqué plus haut, de nombreuses données sont utilisées pour entraîner et calibrer les modèles d'IA dite « connexionniste » (par exemple *machine learning*). Dans cette phase d'entraînement, il est de la responsabilité des équipes de conception de poser des questions générales sur les conditions d'obtention et la qualité de leur jeu de données. Parmi les conditions d'obtention, les notions de « consentement » et de « proportionnalité » sont deux des piliers du droit de la protection des données (RGPD, art. 5). D'une part, pour les

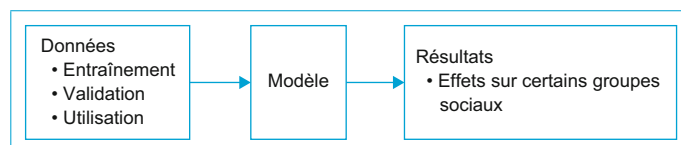


Figure 3 – Schéma de fonctionnement simplifié d'un outil d'IA

données personnelles (qui peuvent rattacher à une personne identifiée ou identifiable), il faut que la personne ait reçu des informations de base (notamment l'utilisation prévue de ses données) et qu'elle ait consenti à cette utilisation. À l'opposé de cette approche, on peut imaginer un outil développé sur la base de données obtenues sans informer les personnes et sans avoir reçu leur consentement [14]. D'autre part, le principe de proportionnalité implique que seules les données nécessaires et utiles à un certain objectif soient utilisées, ni plus, ni moins. Ce principe fonctionne comme un garde-fou face à la tentation d'accumuler le plus possible de données, et d'identifier seulement plus tard comment les utiliser. Cette stratégie, largement appliquée par les géants du numérique dans les années 2000, est clairement prohibée. Ces mêmes principes s'appliquent à l'autre bout de la chaîne si l'outil d'IA crée récolte lui-même des données sur ses utilisateurs.

En plus de ces questions sur les conditions de récolte et d'obtention des données, des scandales dénonçant les biais des outils d'IA ont défrayé la chronique, par exemple des outils favorisant les hommes dans l'obtention d'un crédit ou jugeant les personnes condamnées de couleur noire plus enclines à récidiver [15] [16]. Cette question des biais est souvent intégrée dans un débat plus large portant sur l'équité des outils d'IA (*fairness*) [17]. Cette équité est définie comme une exigence d'éviter que les membres d'un groupe social (par exemple les femmes) ne subissent un désavantage systématique et injustifié. Selon cette approche, les outils d'IA sont équitables s'ils ne sont pas discriminatoires dans leurs effets sur certains groupes sociaux.

Premièrement, cette exigence d'équité a conduit à une réflexion sur la **qualité des jeux de données utilisés à des fins d'entraînement**. Outre les biais individuels que les conceptrices et concepteurs d'IA peuvent posséder (définis comme des préjugés intégrés dans la manière de réfléchir des personnes), la question des biais porte en priorité sur la qualité des jeux de données. Il ne s'agit pas ici de qualité au sens de données manquantes ou non utilisables, mais de qualité dans un sens normatif : le jeu de données satisfait-il des exigences qualitatives ? Ainsi, est-il correct de dénoncer un jeu de données biaisé ? Un ensemble de données documente une partie de la réalité : un certain nombre de catégories a été choisi, puis mesuré. Les catégories choisies (par exemple sexe, âge, lieu d'habitation...) ne reflètent toujours qu'une partie de la réalité. En bref, le reflet du monde qui en ressort est nécessairement biaisé. Un jeu de données est toujours biaisé car il reflète les choix que la personne le créant a réalisés.

La demande ne porte donc pas sur un jeu de données non biaisé (impossible), mais sur l'identification des biais existants et le contrôle de leurs potentiels effets négatifs sur certains groupes sociaux. Ainsi, il importe de se demander si certains groupes sont absents du jeu de données (qui n'apparaît pas dans les données ?) et si certains groupes sont sur- ou sous-représentés par rapport à la réalité. Ce critère de « par rapport à la réalité » est en soi une question importante. Imaginons un outil d'IA utilisé pour analyser des CV et conseiller qui inviter pour un entretien. Si l'outil a été entraîné avec des données reflétant le monde du travail actuel, alors il est plausible que les hommes (ou les parcours typiquement masculins) soient favorisés. Le problème persiste même si la réalité est aussi parfaitement que possible reflétée dans les données. L'idée selon laquelle les données doivent refléter ou représenter la réalité ne devrait donc pas être acceptée comme un pré-supposé neutre. La réalité elle-même est marquée par des injustices et des structures problématiques. Dans l'exemple de l'outil d'analyse de CV, le marché du travail actuel est probablement structurellement injuste. Un outil reflétant la réalité ne ferait que perpétuer ou, pire, aggraver ces injustices. La question principale qui s'impose est donc celle du critère de représentativité choisi pour évaluer les biais. Quelle société idéale (c'est-à-dire par exemple sans injustices structurelles) devrait servir de standard d'évaluation ? Cette question soulève des débats de justice sociale traités dans la deuxième thématique.

L'exigence d'équité va au-delà de l'analyse des jeux de données d'entraînement. Deuxièmement, elle porte sur les **résultats obtenus (output) et leurs effets sur certains groupes (outcome)**.

D'une part, la qualité des résultats d'un outil d'IA peut être différente selon certains critères problématiques. Par exemple, certaines caméras dites « intelligentes » (*smart camera*) repèrent et classifient moins bien les peaux foncées que les peaux claires. La performance de l'outil diffère donc pour les individus selon leur couleur de peau. Garantir l'équité exige des équipes de conception une amélioration des performances pour l'ensemble des groupes sociaux pertinents et une clarification des critères de qualité suffisante avant lancement d'un outil (à partir de quelle performance obtenue l'outil est-il suffisamment bon et pour qui ?).

D'autre part, l'exigence d'équité rappelle qu'un outil d'IA est utilisé dans un contexte spécifique, avec un certain public cible (comme tous les systèmes socio-techniques). Dans certains cas, des « boucles de renforcement » apparaissent (*feedback loop*), conduisant au renforcement de certains effets problématiques. L'exemple classique se réfère à certains outils de prédiction des délits utilisés par la police aux USA, mais également en Europe [18] [19]. Ainsi, sur la base des données d'entraînement (reflétant la réalité et ses injustices, comme expliqué auparavant), l'outil sélectionne les zones où les délits sont les plus probables et suggère aux patrouilles d'y aller le plus souvent possible. En suivant ces recommandations, les patrouilles se déplacent majoritairement dans des quartiers pauvres et, comme elles sont sur place, elles contrôlent les personnes et provoquent certains comportements en retour. Ces outputs (les effets concrets provoqués par l'outil) remontent ensuite dans les prédictions, créant la « boucle de renforcement » et suggérant aux patrouilles d'aller là où, logiquement, ils trouveront quelque chose. Dans leur structure, ces boucles de renforcement ressemblent à des prophéties auto-réalisatrices.

Questions Éthique des données

Les données ont-elles été obtenues en conformité avec les grands principes de protection des données ?

Les personnes concernées ont-elles pu donner leur consentement sur la base d'informations transparentes quant à l'utilisation prévue des données ?

Les données utilisées pour la conception et l'entraînement de l'outil d'IA sont-elles de qualité suffisante ?

Les biais les plus problématiques ont-ils été identifiés et, le cas échéant, corrigés ?

Certains groupes sont-ils défavorisés de manière injustifiée en matière de performance de l'outil ?

Le contexte d'utilisation de l'outil pourrait-il créer des « boucles de renforcement » ?

3.2 Éthique des algorithmes

En plus des questions sur les données utilisées, le **choix des modèles mathématiques utilisés** par les outils d'IA soulève des questions éthiques. Trois questions méritent une attention particulière : l'importance de l'évaluation pour le choix d'outils mathématiques, l'arbitrage entre demandes en tension et le défi de la boîte noire.

La première question est d'ordre général. Au-delà du langage mathématique/statistique utilisé pour formaliser les modèles d'IA utilisés (allant des modèles simples à des modèles d'apprentissage profond), les équipes de développement d'outils d'IA réalisent de nombreux choix normatifs. Elles évaluent l'objectif à atteindre (par exemple prédire un comportement), estiment quel modèle sera le plus efficace pour y arriver et calibrent ensuite ce modèle pour cette tâche particulière. Chacune de ces étapes met en jeu une évaluation des possibles et un choix. Dans une logique de documentation des étapes de conception de l'outil, les équipes devraient être en mesure d'expliquer et de justifier leurs choix [20] [21]. Ainsi, affirmer qu'un modèle a été retenu car il était plus « efficace » va demander d'explicitier, *a minima*, les paramètres retenus pour juger de l'efficacité (pour qui et pour quoi ? par

rapport à quels objectifs ?). De plus, chaque modèle possède son lot d'implications et de limites/faiblesses. Privilégier un modèle plutôt qu'un autre implique de comparer ces limites et faiblesses.

Sur cette base, une deuxième question porte sur les **arbitrages nécessaires dans le choix et l'implémentation d'un modèle**. Choisir le modèle X plutôt que le modèle Y renvoie finalement à la justification proposée pour qualifier X de meilleur/plus adapté à la tâche envisagée. On peut illustrer le défi avec un arbitrage entre complexité et qualité des résultats obtenus. Ainsi, le modèle X est simple à comprendre, mais les résultats qu'il fournit sont de qualité moyenne. Si on lui préfère le modèle Y (très bons résultats, mais difficulté extrême à comprendre comment ces résultats sont produits), alors on aura pris position en faveur de la qualité des résultats au profit de l'explicabilité de ces mêmes résultats.

La troisième question, qui représente le défi principal de cette éthique des algorithmes, porte sur le problème de la « boîte noire ». À partir d'un certain niveau de complexité, il est difficile, voire impossible pour la plupart des individus, de reconstituer comment un outil d'IA produit certains résultats [22]. En revenant à notre schéma simplifié, nous savons ce qui entre dans la machine (données d'utilisation), ce qui en sort (résultats), mais nous sommes dans l'incapacité de reconstruire comment la machine passe de l'un à l'autre. Dans l'exemple d'un outil comme ChatGPT, personne ne peut expliquer pourquoi la machine fournit un certain texte sur la base d'une certaine demande (*prompt*). Les concepteurs sont bien sûr capables d'expliquer le fonctionnement général de l'outil. Mais le bât blesse au niveau des considérations liées à une demande spécifique.

Les enjeux sont de taille. Dans l'exemple de ChatGPT, on veut savoir comment un contenu problématique peut être créé afin de le prévenir. De plus, il y a une exigence procédurale. Si un outil d'IA est utilisé dans le cadre d'une procédure de justice (par exemple un outil de recommandation quant à la dangerosité d'une personne qui souhaiterait obtenir une libération conditionnelle), il est essentiel que la personne concernée puisse comprendre la recommandation de l'outil afin de faire valoir ses droits et, le cas échéant, déposer un recours.

Le défi de l'explicabilité pose en lui-même de nombreuses questions [23] [24]. Ces questions portent sur l'objet qui doit être explicable (le modèle mathématique, l'outil en général, le cas d'application spécifique), le type d'explicabilité (une explicabilité théorique, mais potentiellement très difficile à obtenir en pratique, ou au contraire une explicabilité concrète), le niveau d'explicabilité (pour des spécialistes du domaine ou pour une tierce personne non spécialiste). Les questions portent également sur les distinctions avec, notamment, l'interprétabilité ou l'intelligibilité.

Différentes options concrètes sont en phase de développement ou de test [25]. Ainsi, certains outils offrent une option qui permet de jouer avec certains paramètres pour tester l'effet sur le résultat obtenu (à défaut d'explicabilité complète, une possibilité d'illustrer les changements). D'autres proposent des explications grand public permettant d'identifier les types de données utilisées et le fonctionnement général du modèle. En matière de conseil aux entreprises, un véritable marché de l'*explainable AI* émerge, notamment en prévision des demandes des régulateurs [26]. Les réactions face à un outil comme ChatGPT démontrent la nécessité d'assurer la plus grande transparence et explicabilité possible. Les utilisateurs, mais également les autorités de surveillance et de régulation, doivent être en mesure de comprendre le fonctionnement de l'outil pour en évaluer correctement les risques et les limites. Sans ces informations, il est impossible de juger de la dangerosité ou, à tout le moins, de la pertinence d'un outil dans un contexte particulier. Cette **demande de transparence et d'explicabilité est présentée comme l'un des piliers d'une approche centrée sur la confiance**, telle que développée par exemple par l'Union européenne [27].

Questions Éthique des algorithmes

Les équipes de développement sont-elles capables de justifier leurs choix de modèles (une définition des paramètres de l'efficacité) ?

Peuvent-elles lister les implications et les limites/faiblesses des modèles choisis ?

Le cas échéant, comment les arbitrages entre plusieurs objectifs/valeurs sont-ils réalisés et documentés ?

Quelles mesures peuvent être mises en place pour assurer la meilleure explicabilité possible du modèle ?

3.3 Éthique du design

Dans le cas d'un outil numérique, les questions d'éthique du design portent avant tout sur la conception de l'interface utilisateur, c'est-à-dire ce que voit et peut faire la personne utilisant l'outil d'IA. Deux questions éthiques principales se posent en matière de design : celle de la manipulation et celle de la projection de certaines caractéristiques associées à des groupes sociaux.

Sur le premier point, la réflexion démarre d'une sorte d'idéal : une personne devrait pouvoir utiliser un outil d'IA sans risquer d'être flouée, mal informée, manipulée. Cette personne devrait comprendre et connaître les enjeux et décider en toute connaissance de cause comment utiliser l'outil. En partant de cet idéal, comment évaluer la réalité des expériences utilisateurs ? Premièrement, l'interface peut être caractérisée par des techniques de design visant à amener la personne à réaliser certaines actions de manière plus ou moins consciente. Ainsi, certaines options apparaissent plus colorées et plus grandes, d'autres n'existent pas (seul un choix entre « oui » et « plus tard ») et certains textes sont écrits de manière excessivement compliquée. Ces techniques de design sont rassemblées dans une taxonomie de « *dark patterns* », un ensemble de techniques et pratiques de design utilisées à des fins de manipulation. Le Dark Pattern Detection Project (Allemagne) liste les techniques principales [https://dapde.de/en/dark-patterns-en/types-and-examples-en/]. En matière de données personnelles par exemple, le Laboratoire d'Innovation Numérique de la CNIL propose une classification des *dark patterns* autour de quatre intentions, à savoir pousser l'individu à partager plus que ce qui est strictement nécessaire, influencer le consentement, créer de la friction par rapport aux actions de protection des données, et dérouter l'individu [28].

Il est à noter qu'il est difficile en pratique de tirer une limite claire entre un marketing légitime (interface d'outil d'achat en ligne qui « pousse » aux achats) et une manipulation illégitime [29]. Des indices existent néanmoins : Les objectifs généraux de l'outil/son propriétaire sont-ils ouvertement déclarés ? Une personne attentive peut-elle identifier les tentatives de manipulation ? La liberté d'utilisation de cette personne est-elle véritablement entravée, ou « seulement » rendue plus compliquée ? Du côté des équipes de conception d'outils d'IA, ces questions sont autant d'occasions de réaliser des choix en adéquation avec l'idéal d'autonomie et de liberté en matière d'interface utilisateur. Sur le plan procédural, une logique de documentation et de capacité à justifier pourquoi telle technique de design a été choisie doit s'imposer. Les équipes doivent se comparer à des architectes de choix (CNIL) : elles conçoivent ce que les personnes utilisant l'outil pourront choisir.

Ces réflexions de design responsable sont valables pour l'ensemble des outils numériques. De plus, certains outils d'IA peuvent eux-mêmes être utilisés à des fins de manipulation. Ce n'est plus l'interface de l'outil d'IA qui nous intéresse ici, mais ce que l'outil d'IA rend possible. Premièrement, il n'est pas toujours clair pour la personne utilisant l'outil qu'elle interagit avec un outil d'IA. Certains agents conversationnels (*chatbot*) sont devenus très performants et peuvent facilement tromper une personne sur leur caractère artificiel. Dans ce cas, il semble tout d'abord nécessaire d'avoir une information claire

sur qui participe à une interaction : un humain ou une machine ? Cette information doit être facilement accessible.

Deuxièmement, la puissance des outils d'IA pourrait permettre l'émergence de nouvelles formes de manipulation couplant une multitude de paramètres. Ces nouvelles formes de manipulation s'insèrent dans un débat sur l'économie de l'attention et sur la collaboration entre spécialistes d'IA et spécialistes du cerveau humain (psychologues, neuroscientifiques, spécialistes de marketing) à des fins d'**exploitation systématique des faiblesses cognitives** [30]. L'autre question éthique majeure en matière de design des outils d'IA traite des **représentations que projette l'IA sur les utilisateurs** et le grand public. Le choix de certaines caractéristiques de l'interface (couleurs, illustrations, police de caractères, mais également dénomination de l'outil) projette une certaine image dans le monde. Ainsi la plupart des outils d'IA utilisés à des fins d'assistance sont désignés pour avoir une apparence féminine (nom, voix...). Ces questions apparaîtront encore plus cruciales quand le couplage texte/voix sera plus abouti.

Le recours aux préjugés et clichés offre des parallèles avec le monde de la publicité et du marketing. Comme dans la publicité, les promoteurs d'outils numériques cherchent à correspondre aux habitudes/attentes de leurs clients. Une interface trop éloignée de la réalité vécue par les clients (ou sa représentation) provoque des perturbations [31]. Dans le même temps, cette approche contribue à maintenir et à renforcer des préjugés néfastes pour certains groupes sociaux. Une critique du design numérique par le genre, ou plus largement dans la perspective des groupes minoritaires, a permis de dénoncer certaines pratiques et de faire évoluer les mentalités [32]. Pour certaines voix critiques, l'IA et la robotique pourraient ainsi être un lieu d'expérimentation en vue d'un rapport plus fluide aux genres et moins dépendant de la binarité masculin/féminin qui imprègne la vie sociale. Depuis 2021, Apple a décidé de ne plus configurer par défaut Siri (le programme d'assistance) avec une voix féminine mais de permettre aux utilisateurs de choisir parmi différentes tonalités (pour sa version anglophone) [33].

Questions Éthique du design

L'interface utilisateur contient-elle des éléments susceptibles de ou visant à manipuler la personne utilisant l'outil ?

L'IA est-elle intégrée à un système visant à manipuler les utilisateurs ?

Les choix de design renforcent-ils certains clichés (notamment sous l'angle du genre ou d'autres minorités) ?

3.4 Défis de durabilité

La durabilité et l'éthique entretiennent des rapports qui peuvent être confus. Pour certains, l'éthique est une dimension de la durabilité. Pour certaines analyses de critères ESG (*Environmental, Social, Governance*), l'éthique se retrouve dans le « S » ou dans le « G », selon les approches [34]. À l'inverse, on peut considérer l'éthique comme l'exigence générale, et la durabilité comme l'une de ses dimensions.

Le présent article lui-même pourrait être considéré à la fois comme une contribution au numérique durable (car éthique), et comme une discussion générale couplant éthique et durabilité. Pour notre propos, il n'est pas nécessaire de poursuivre ce débat. Il faut simplement clarifier que cette sous-section intitulée « Durabilité » s'intéresse en priorité à **l'impact environnemental des outils d'IA, notamment sur les ressources énergétiques et naturelles nécessaires à leur fonctionnement**. Cette réflexion s'intègre dans le contexte plus large de l'impact environnemental des technologies numériques (pour une introduction, voir : <https://www.fun-mooc.fr/fr/cours/impacts-environnementaux-du-numerique/>). Cette réflexion gagne en puissance auprès de l'opinion publique et des législateurs nationaux [35].

Les questions éthiques soulevées ici portent principalement sur deux domaines.

Premièrement, la fabrication et la production des infrastructures et des appareils nécessaires aux technologies numériques exigent un certain nombre de ressources naturelles rares. En plus d'être limitées quantitativement, les ressources nécessaires sont inégalement réparties sur la planète, conduisant à des tensions géostratégiques importantes [36]. L'impact environnemental nécessaire à l'obtention de ces ressources, mais également les problèmes sociaux et politiques créés par leur importance stratégique, doivent nécessairement faire partie d'une éthique des outils numériques [37]. Au-delà des ressources les plus rares, la fabrication de l'ensemble du *hardware* nécessaire aux technologies numériques soulève des questions de cycle de vie des matériaux qui sont comparables aux autres domaines d'ingénierie [G 5 550].

Deuxièmement, plus directement en lien avec les outils d'IA, les phases de conception, création et utilisation exigent des ressources énergétiques. Cela concerne notamment l'électricité nécessaire au fonctionnement des serveurs permettant le calcul ou le refroidissement des centres de données [38] [39]. Face à ce constat, un mouvement de conception éco-responsable prend de l'importance [40] [41]. L'objectif fondamental consiste à mesurer et visibiliser l'impact énergétique de chaque étape de conception, d'entraînement et d'utilisation d'un outil d'IA. Au-delà de cette mesure, un effort de minimisation du coût environnemental soulève des questions d'arbitrage avec les fonctionnalités de l'outil. L'objectif de l'éthique de la durabilité consiste alors à visibiliser cet impact et à en faire un paramètre de réflexion et de prise de décision. Dans cette situation, l'impact environnemental vient s'ajouter dans la balance des avantages et inconvénients de l'outil.

Au terme de cette courte réflexion, deux précisions sur les liens entre durabilité et IA s'imposent.

D'une part, ces réflexions d'impact environnemental des outils d'IA n'enlèvent rien au fait que ces mêmes outils représentent des éléments importants dans l'optimisation de notre gestion des ressources rares. À titre d'exemple, les outils d'IA sont utilisés pour optimiser le fonctionnement des réseaux électriques (*smart grid*) [42]. Ces utilisations font elles-mêmes partie d'un large mouvement où l'IA est utilisée à des fins d'amélioration de la vie en société (*AI4GOOD*) [43].

D'autre part, la question de la durabilité pourrait également porter sur des **dimensions d'accès dans le temps**. Le numérique durable ne traite pas que de son impact sur les ressources, mais également de son accessibilité pour le plus grand nombre, maintenant et dans le futur. Cette demande de durabilité porte alors sur la capacité des modèles d'IA (et les données sous-jacentes par exemple) à rester accessibles. Les questions de licence d'utilisation (notamment en *creative commons*), de jeux de données librement utilisables ou encore de reproductibilité des recherches scientifiques en matière d'IA [44] sont au cœur de cette dimension de durabilité.

Questions Éthique de la durabilité

L'entreprise/organisation a-t-elle une vue d'ensemble de la chaîne de création de valeurs pour le matériel (hardware) nécessaire à la création/utilisation des outils numériques ?

Existe-t-il des fournisseurs ayant formulé un engagement environnemental, c'est-à-dire un engagement sur le cycle de vie des appareils ?

En matière de conception des outils d'IA, les principes de l'écodesign ont-ils été pris en compte, notamment l'ambition de choisir les modèles les moins consommateurs d'énergie ?

L'impact environnemental fait-il partie des critères d'évaluation de l'outil ?

3.5 Défis de contrôle et de responsabilité

La dernière sous-section de l'éthique *dans* la technologie porte sur les mécanismes de contrôle et d'*accountability* à mettre en place. La thématique du contrôle soulève la délicate question de l'autonomie. Il nous faut clarifier dans quelle mesure un outil d'IA peut être dit « autonome » et comment, le cas échéant, contrôler cette autonomie.

Brièvement, il faut distinguer entre une autonomie au sens des « ingénieurs » et une autonomie au sens des « philosophes » [45] [46]. L'autonomie des ingénieurs se définit comme la capacité à déterminer les moyens pour réaliser un objectif fixé d'avance. Les normes ISO liées à la robotique nous fournissent un bon exemple de cet usage du concept d'autonomie. L'autonomie y est définie comme « la capacité d'exécuter des tâches prévues à partir de l'état courant et des détections, sans intervention humaine » [47]. Les améliorations en IA vont augmenter cette autonomie car le robot sera toujours meilleur dans le choix et la mise en œuvre des moyens utilisés. En plus de cette autonomie de « moyens » (1^{er} niveau), la dimension philosophique ajoute une capacité de fixer ses objectifs et de les réviser (2^e niveau). La machine n'a pas cette capacité, elle réalise ce que d'autres ont choisi pour elle. L'autonomie des humains est souvent définie en lien avec la conscience de soi. Pour de nombreux auteurs, la conscience que nous avons de notre existence, de nos désirs et de nos valeurs rend possible une autonomie de 2^e niveau [48] [49].

Si l'on accepte cette distinction (tirée aussi à gros traits), un outil d'IA ou une machine utilisant un outil d'IA peut donc être autonome (1^{er} niveau, au sens des ingénieurs), mais non autonome au sens des philosophes (2^e niveau).

Ces définitions posées, on peut revenir à la question du contrôle. Plus la machine est autonome au sens des « moyens », plus la question de son contrôle par ses concepteurs et ses utilisateurs se pose. Cette question se pose à la fois pour les équipes de conception et pour les utilisateurs, rappelant au passage que chaque outil d'IA est un système socio-technique utilisé dans un contexte spécifique. L'équipe de conception peut, dans le design de l'outil, mettre en place des possibilités de contrôle et d'intervention. Elle peut compliquer ou favoriser certains usages par des mesures de design. En complément, l'utilisateur doit mettre en place des mécanismes de contrôle dans le contexte spécifique d'utilisation qui est le sien.

Quelles possibilités d'intervention devraient être prévues dans l'utilisation de l'outil ? Prenons l'exemple d'un outil d'IA utilisé à des fins de décision. Le contrôle le plus strict s'exerce lorsque l'humain fait partie intégrante du processus de décision (*human in the loop*). Autrement dit, c'est l'humain qui prend la décision finale. Mais, même dans ce scénario, il doit gérer des informations fournies par le système pour prendre une décision qui engage sa responsabilité. Le design de l'interaction entre l'humain et cette machine acquiert ici une importance cruciale.

À un deuxième niveau, l'humain peut être placé en situation de contrôler le système sans pour autant intervenir dans le processus décisionnel. L'humain est alors *on the loop*. Il déploie une vue d'ensemble, à la manière du surveillant d'une chaîne de montage. Il peut intervenir en cas de problème manifeste. Néanmoins, le défi porte ici sur l'opportunité réelle d'intervention. Un humain a-t-il le temps de réagir et d'interrompre le processus ? A-t-il les informations nécessaires sur les données et les algorithmes utilisés pour appréhender la décision prise par le système ?

À un troisième niveau, l'humain peut être complètement en dehors du processus décisionnel, sans capacité d'intervenir (*out of the loop*). Dans ce cas, le système interprète seul les données reçues et déclenche une décision sur cette base.

À tous les niveaux, la **question de l'efficacité réelle du contrôle est cruciale**. Des humains ont-ils une véritable possibilité d'exer-

cer un contrôle ou s'agit-il plutôt d'une possibilité largement théorique qui vise avant tout à donner un « visage » humain à un outil ? Ce débat est particulièrement avancé sur les armes autonomes et l'exigence de clarifier qui peut/qui devrait pouvoir intervenir à quel moment (*meaningful human control*) [50].

Plus la machine sera autonome au sens des « moyens » (1^{er} niveau), plus il sera important de définir clairement les responsabilités des uns et des autres. Le terme anglais « *accountability* » traduit cette dimension de la responsabilité : une personne ou une entreprise doit être identifiée pour répondre des actions réalisées par une machine. En effet, comme la machine n'a pas d'autonomie de 2^e niveau, elle ne peut pas être responsable au sens moral ou juridique. Elle n'oriente pas ses actions selon des normes morales et ne peut pas choisir de ne pas faire ce qu'elle est programmée pour faire. En d'autres mots, elle ne fait que réaliser ce qui lui a été demandé. Si la machine n'est pas responsable, cela implique qu'une personne humaine doit l'être. En conséquence, il faut chercher l'humain derrière la machine pour établir les responsabilités. Les différents acteurs qui inventent, conçoivent, produisent et finalement utilisent des outils d'IA peuvent avoir différentes responsabilités. L'allocation des responsabilités le long de cette chaîne de création peut être faite au cas par cas (par exemple en cas d'outils défaillants ayant conduit à un problème). Mais, de manière générale, elle relève de choix politiques que nous appréhendons dans la prochaine partie comme des questions de justice.

Questions Éthique du contrôle/responsabilité

Les équipes de conception ont-elles clarifié comment et à quelles conditions une personne pouvait intervenir dans le fonctionnement de l'outil ?

Quelles formes concrètes peut prendre ce contrôle ?

Ces possibilités de contrôle sont-elles inscrites dans le design de l'outil ou doivent-elles être prévues dans une utilisation concrète, c'est-à-dire chez l'utilisateur ?

Toutes les mesures qui vont permettre l'établissement des responsabilités (notamment la transparence et l'explicabilité) ont-elles été prises ?

Ces 5 sous-sections forment ensemble la thématique des questions d'éthique *dans* la technologie. Il serait erroné de considérer ces questions exclusivement comme des questions techniques qu'on pourrait régler indépendamment du contexte d'utilisation. La responsabilité des équipes de conception s'inscrit toujours dans un cadre particulier, souvent celui d'une entreprise ou d'une institution publique. Ce contexte impose certaines contraintes, qui elles-mêmes découlent de choix politiques généraux (par exemple le système de libre-marché). Les 5 sous-sections ont donc des liens puissants avec les modèles d'affaires des entreprises du numérique [51]. Sur ce point, l'éthique de l'IA entre en dialogue avec l'éthique des affaires [52] [53] [54]. Ces débats sont menés dans le contexte d'une extension de la responsabilité sociale des entreprises (RSE) vers une responsabilité numérique des entreprises (*digital corporate responsibility*), englobant à la fois le développement d'outils dignes de confiance et l'accompagnement structurel des collaborateurs [55] [Le canton de Genève (Suisse) a mandaté la création d'un outil d'auto-évaluation pour les PME ; celui-ci est librement accessible en ligne : <https://digital-responsibility.info>]. Au-delà des modèles d'affaires et de la nécessaire critique éthique, la question des **conditions de compétition** parmi les entreprises numériques se pose également. Le risque de pratiques anti-compétitives – les grandes entreprises profitant d'une position parfois dominante – impacte de plein fouet la capacité de petits acteurs à innover et/ou à modifier certaines pratiques discutables.

À retenir

- L'éthique dans la technologie rassemble toutes les questions éthiques qui se posent au cours du processus de conception et réalisation d'outils technologiques.
- Les sous-thématiques les plus importantes en matière d'IA sont : l'éthique des données, l'éthique des algorithmes, l'éthique du design, les défis de durabilité, les défis de contrôle et de responsabilité.
- Les questions d'éthique dans la technologie s'adressent en priorité aux équipes de conception des outils. Elles ont en effet la capacité et la compétence – et donc la responsabilité – d'intégrer ces questions dans leurs travaux.
- Les consommateurs, les décideurs politiques et le grand public (notamment la société civile) doivent néanmoins s'intéresser de très près à ces questions et ne pas les réserver aux spécialistes techniques.

4. Impact sur la société et justice sociale

La deuxième thématique nous invite à porter notre regard sur l'impact des outils d'IA, ou de l'IA en général, sur la vie en société. Pour résumer cette nouvelle perspective, quittons le laboratoire de recherche & développement pour analyser les comportements et interactions que les outils rendent possibles – en d'autres mots, leur **potentiel de transformation de la société**.

Étant donné leur caractère transversal, les outils d'IA déploient un effet sur l'ensemble des domaines d'activité de la société. L'exemple de ChatGPT illustre bien comment un outil générateur de contenu impacte tous ceux qui créent ou mettent en forme des textes, des images, des films, et ce quel que soit leur domaine d'activité. L'analyse éthique peut rester à un niveau général (par exemple en posant la question de l'adaptation de la législation) ou s'intéresser à des domaines spécifiques (par exemple en croisant l'éthique de la médecine et l'éthique de l'IA).

Cette partie traite trois problématiques principales : une méthode philosophique pour poser la question de la justice, une discussion de l'impact de l'IA sur le marché du travail et quelques remarques générales sur la dimension de long terme des débats sur l'IA. L'objectif général est de présenter aux spécialistes d'ingénierie les méthodes de la philosophie et de l'éthique pour aborder cette question de la justice.

4.1 Poser la question de la justice

L'arrivée de nouveaux outils technologiques dans une société crée des gagnants et des perdants. On assiste à une réallocation au sein d'un groupe des avantages et désavantages, des opportunités et des risques, bref des éléments positifs et négatifs liés à ces outils. Le groupe peut être une société dans son ensemble (un pays), mais également une ville, une entreprise voire, plus large, une entité régionale comme l'Union européenne. De manière très schématique, on peut aborder les différentes implications des outils d'IA comme des questions de distribution.

Pour aborder ces questions, les philosophes ont développé de nombreuses approches afin d'esquisser à quoi ressemblerait une société juste. Parmi toutes celles existantes, il convient d'illustrer en priorité la méthode philosophique [56] [57]. Pour ce faire, on prend l'exemple de l'approche proposée par l'américain John Rawls dans un livre publié en 1971, *Théorie de la justice* [58]. L'auteur y propose une méthode pour réfléchir aux questions de justice spécifiquement adaptée aux situations où nous devons dis-

tribuer certains biens, qu'il s'agisse d'argent, de droits, d'opportunités ou d'autres types de ressources. Rawls nous invite à imaginer une sorte d'assemblée générale de toutes les citoyennes et citoyens, appelée à déterminer les règles de la vie en société. Le mode de travail de cette assemblée est très particulier. En effet, les participants ne disposent d'aucune information sur eux-mêmes. Ils ne savent pas quelles sont leurs caractéristiques personnelles. Ils ignorent s'ils sont riches, beaux, malades, intelligents ou doués pour une activité particulière. De plus, ils ignorent quelle est leur place dans la société, par exemple s'ils ont un statut d'employé ou d'entrepreneur ou s'ils travaillent dans un domaine particulièrement menacé par des évolutions technologiques. Pour décrire ce mode de travail, Rawls parle d'un « voile d'ignorance ». Sa conviction : derrière ce voile, les participants prennent collectivement des décisions plus justes. Ils se montrent plus prudents au moment de faire des choix, car chacun est susceptible d'appartenir à un groupe défavorisé.

Comment répondre au potentiel impact des outils d'IA sur le marché du travail (c'est-à-dire la réponse à un risque inégalement distribué parmi toute la population) ? Si vous êtes multimillionnaire, il vous faudra un sérieux effort d'empathie ou un sens du devoir à toute épreuve pour plaider un renforcement du soutien aux personnes touchées. En revanche, si vous êtes justement dans un emploi à haut risque d'automatisation, une aide publique vous paraîtra de la plus haute importance. Et si vous ne savez pas qui vous êtes, Rawls pense que vous serez précautionneux et que vous rechercherez une forme de sécurité pour le scénario du pire. Votre réflexion commencera par considérer la situation la plus négative possible puis, dans un deuxième temps, vous chercherez à mettre en place des mécanismes pour améliorer cette situation.

Une telle assemblée générale travaillant en conditions de voile d'ignorance n'a bien sûr jamais existé et n'existera pas dans le futur. Cette expérience de pensée doit toutefois nous permettre de réfléchir sur les principes d'une juste répartition sans laisser notre situation personnelle influencer trop fortement notre décision. La méthode proposée par Rawls place au centre des discussions le principe d'égalité et la défense de la liberté de chacun.

Cinquante ans après sa publication, la théorie de la justice proposée par Rawls continue de marquer la philosophie politique. Il n'est pas nécessaire d'adhérer à cette approche pour poursuivre la lecture de ce texte. L'objectif est d'illustrer la contribution de la philosophie comme méthode pour structurer les questions de justice.

Appliquée aux questions liées à l'IA, cette approche peut être utilisée pour débattre des effets au sein d'une entreprise, d'une institution publique ou plus généralement pour l'ensemble de la société [59]. Par exemple dans une entreprise, sans savoir qui je suis dans l'organigramme, serais-je prêt à accepter l'introduction d'un nouvel outil d'allocation des tâches au sein des équipes ? L'approche de Rawls nous oblige alors à clarifier la justification offerte pour ce nouvel outil. Cette justification est-elle partagée par l'ensemble des employés ou certaines catégories de personnes ont-elles des questions particulières à faire valoir ? Le voile d'ignorance oblige à fournir une justification de haute qualité, acceptable par tous.

La question de justice se pose également dans l'accès à la technologie. Les débats autour de la « fracture numérique » illustrent la nécessité de ne pas seulement penser aux usages de la technologie. L'accès est crucial, tant en termes financiers, de compétences que plus largement de ressources pour comprendre et maîtriser certains outils. La justice dans l'accès à la technologie se pose au sein d'un pays, mais également entre les nations. Elle se transforme alors en question de justice internationale. La rhétorique de la « course à l'IA » en matière de contrôle technologique et de développement militaire se décline facilement comme une question de **domination dans l'accès et l'utilisation de certaines technologies** [60].

4.2 Impact de l'IA sur le marché du travail

L'impact sur le marché du travail est traité comme la conséquence sociale la plus importante de l'IA. Sur la face lumineuse, on loue les qualités des outils d'IA comme des forces de libération des tâches répétitives ou ingrates. Ces outils permettent de plus de traiter des masses d'informations de manière simple et rapide, ouvrant d'immenses opportunités économiques et sociales. Sur la face sombre, ces mêmes outils sont dénoncés comme des destructeurs de places de travail, qui « volent » les emplois des humains. Pour mettre de l'ordre dans ces considérations, il nous faut distinguer le constat empirique et les réflexions éthiques autour de la justice sociale.

Le constat empirique tout d'abord, qui mobilise les analyses des sciences économiques, des sociologues ou des sciences politiques : les outils d'IA s'inscrivent et renforcent un **vaste mouvement d'automatisation**. Si cette automatisation était auparavant focalisée sur des tâches répétitives et mécaniques (par exemple une chaîne de montage robotisée), les outils numériques sont aujourd'hui capables de réaliser de plus en plus de tâches intellectuelles. ChatGPT a fait la démonstration que des outils d'IA tout public peuvent formuler du contenu à la manière d'un humain.

Les recherches se succèdent pour tenter d'évaluer le plus précisément possible l'impact de l'automatisation par le biais d'outils numériques sur le monde du travail. Depuis l'étude de Carl B. Frey et Michael A. Osborne en 2013 prédisant que 47 % du total des emplois aux États-Unis allait disparaître, ce type d'analyse s'est imposé dans les universités ainsi qu'auprès des entreprises de conseil en tous genres [61]. D'après la reconstruction historique proposée par Susskind, les modèles développés par les économistes tentent de mieux appréhender l'impact des technologies émergentes sur le marché du travail [62]. L'approche classique considère le marché du travail comme une fonction de production avec une distinction entre différents types d'emplois selon les aptitudes nécessaires. La technologie est alors conçue avant tout comme ayant un impact sur certains emplois avec des aptitudes spécifiques. Cette approche a été enrichie par l'ambition de se focaliser sur les tâches spécifiques réalisées (et non sur les emplois), puis de distinguer le niveau de routine exigé pour chacune des tâches. Ce type de modèle a bien fonctionné avec des machines/technologies dont les effets sur les tâches étaient facilement identifiables. Le défi consiste à intégrer dans ces modèles des technologies dont les capacités exactes ne peuvent être pleinement déterminées *a priori* (par exemple un outil transversal comme une IA générative de texte). De là, deux effets de la technologie sur le marché du travail apparaissent plus clairement : un effet de substitution (pour les tâches, routinières ou non, où la machine est plus productive) et un effet de complément (pour les tâches où la machine va permettre au travail humain de se décupler).

À la lumière d'une réflexion de justice sociale (comme celle de Rawls par exemple), ces effets de substitution, de complément ou encore de création sur le marché du travail peuvent s'analyser comme des défis de politique sociale. **L'idée de transition responsable est cruciale** : les changements technologiques doivent être abordés et anticipés comme des mutations structurelles durables, et non dans une perspective de sauvetage une fois les problèmes en cours.

Deux paramètres apparaissent importants : déterminer quelles mesures d'accompagnement sont pertinentes et légitimes, et aussi qui doit les financer. En matière de mesures, de nombreux niveaux d'action sont envisageables : auprès des personnes directement concernées (mesures de formation continue), dans les domaines d'activité (soutien structurel, fond d'innovation) ou encore dans les structures générales de formation (école, hautes écoles, encouragement structurel à la formation continue). Ces mesures viennent s'intégrer dans une réflexion de politique générale en matière d'employabilité, d'innovation et de compétitivité. En matière de financement, on peut distinguer trois acteurs principaux. Sur le plan éthique, cette question du financement peut s'appréhender

comme une question de justice. La responsabilité est partagée entre les entreprises (formation continue de leur personnel, anticipation des changements technologiques, plan social accompagnant ces changements), les collaborateurs (investissement personnel dans leur formation continue) et les pouvoirs publics (création d'un fonds d'investissement en vue de cette transition, mesures de soutien au titre d'État social).

Au-delà de la macro-prédiction sur l'évolution du marché du travail (souvent résumée au nombre d'emplois menacés), la démocratisation d'outils d'IA soulève des questions portant sur les modifications des tâches qui sont associées à un « métier », comme l'ont bien remarqué les économistes [63]. Ce niveau d'analyse est moins spectaculaire que les grandes annonces sur la fin du travail car il exige de mettre au jour des changements nécessairement contextuels (dans un métier X) et dépendants des outils utilisés. Par exemple, Moradi et Levy identifient certains phénomènes transversaux découlant de l'utilisation d'outils d'IA : la redéfinition des tâches effectuées par un humain ou par une machine (ou par un humain avec une machine) et le calcul des ressources nécessaires à la réalisation (temps, argent, ressources autres), la planification algorithmique des tâches (et la possible dilution de responsabilité pour la hiérarchie), les mécanismes d'incitation personnalisée (par exemple « gamification » des tâches) [64]. Certains de ces usages sont reflétés dans le projet de législation sur l'IA de l'EU et devraient être régulés (annexe 3 et différents exemples de systèmes d'IA à haut risque, par exemple les outils d'allocation automatisée des tâches). Tous ces phénomènes soulèvent des questions qui peuvent avec profit être posées comme des questions de justice dans l'entreprise (par exemple pratiques managériales) et au niveau sociétal (comme objet de régulation). De même, cette méthode philosophique peut être utilisée afin de poser les questions de **redistribution** provoquée par les outils d'IA qui vont au-delà du seul marché du travail.

4.3 Une double conversation

Afin d'aborder ces questions de justice sociale, il est crucial de distinguer entre deux conversations éthiques parallèles. L'une porte sur notre vision d'une société idéale. À quoi ressemblerait une société où les outils d'IA seraient utilisés partout et selon quelles règles cette société devrait-elle être organisée ? Cette interrogation porte sur un futur qu'il nous faut esquisser sous forme de scénarios-objectifs. L'autre discussion vient s'adosser à cette réflexion sur le futur. Une fois les scénarios identifiés quant à l'idéal visé, il nous faut débattre du chemin à parcourir entre l'ici et maintenant et ces futurs possibles. Les défis éthiques ne portent alors plus sur la vision idéale, mais sur la période de transition vers cet idéal. Les questions de politique sociale et d'accompagnement au changement dans les entreprises font partie de cette deuxième conversation.

Pour la conversation à long terme (quelle société voulons-nous ?), deux questions particulièrement importantes en matière de justice sociale sont brièvement mentionnées à titre d'exemples. Premièrement, les réflexions sur le marché du travail mettent en jeu la relation entre emploi, travail, moyens de subsistance et reconnaissance sociale [65]. Dans les sociétés industrielles, ces quatre éléments sont perçus comme un ensemble indissociable : une personne réalise un travail à travers un emploi, reçoit un salaire à cet effet et jouit d'une reconnaissance sociale en lien avec son emploi. Ce quatuor est forcé d'évoluer sous l'effet de la technologie. Le cas échéant, il nous faudrait alors dissocier l'emploi des autres éléments. Le travail ne serait plus exclusivement synonyme d'emploi. L'engagement familial, associatif, bénévole serait revalorisé. Et si le salaire disparaissait ou se réduisait drastiquement comme moyen d'obtenir des ressources, l'idée d'État social devrait profondément se réinventer, par exemple sous la forme d'un revenu de base, inconditionnel ou non. Cette réflexion conduira à **réinventer un statut social** indépendant, ou moins dépendant, des activités professionnelles. L'IA n'est pas la cause profonde de ces débats de justice, mais elle va renforcer et accélérer la nécessité d'y apporter des réponses solides.

Deuxièmement, l'utilisation d'outils d'IA à travers l'ensemble de nos domaines d'activités pose le **défi d'une personnalisation exacerbée**. Comme évoqué plus haut, les outils d'IA ont besoin d'une réalité mise en données. Sur cette base, les outils d'IA vont permettre de créer un bien/service (un film, un emploi, un traitement médical) et de le distribuer de la manière la plus personnalisée possible. Ce mécanisme rendu possible par l'IA touche notre consommation culturelle, notre accès à l'information (par exemple via les plateformes) ou encore le mouvement de fond d'une médecine personnalisée. Pendant quelques années, les outils d'IA étaient « limités » à la transmission personnalisée, dans une logique de *match-making* entre un contenu/un produit et une personne identifiée. Les IA génératives de contenu permettent maintenant de personnaliser le contenu proposé. Si une IA peut facilement créer une image ou un film selon mes désirs, pourquoi produire le même film pour tout le monde ? Nous aurons tous notre film, comme nous aurons un traitement thérapeutique personnalisé.

Outre les opportunités intéressantes de cette personnalisation, les défis sont nombreux. Les domaines qui supposent une solidarité (par exemple une assurance) souffrent d'une personnalisation exacerbée. De même, il existe certains domaines où l'accès à une information partagée peut être positive. Si des groupes importants de personnes évoluent dans des écosystèmes d'informations clos (des « filtres de bulle » portant sur l'ensemble de leurs informations), un danger pour l'idéal de citoyenneté et la démocratie apparaît. Il ne s'agit pas seulement d'une consommation asynchrone ou parcellaire. Beaucoup de personnes regardent le 20 h en ligne, en choisissant les reportages qui les intéressent. Mais le 20 h continue d'exister comme contenu commun, comme le font les livres, les films et les spectacles que nous regardons. Mais l'IA pourrait offrir à chacun « son » contenu culturel et informationnel. Finalement, la sphère du commerce de tous les jours pourrait également être touchée par cette logique. Dans de nombreux domaines, il est déjà admis que tous les consommateurs ne paient pas le même prix pour un objet ou service identique. Le prix évolue selon le comportement passé, la localisation, et des paramètres parfois opaques (réservation d'hôtel ou de billet d'avion en ligne). À voir comment les consommateurs réagiraient si l'ensemble des achats se déroulaient selon cette logique. Le prix fixe du supermarché disparaîtrait complètement pour être fixé selon le profil d'acheteur d'une personne spécifique. Ces évolutions ne sont pas inhérentes à l'IA comme ensemble de technologies, mais elles sont rendues possibles par ces outils.

Ces deux exemples illustrent la dimension de justice sociale et l'importance de penser à la fois la société souhaitée dans le futur (scénario-objectif) et la transition entre l'ici et maintenant et ces scénarios.

À retenir

- Lors de leur déploiement, les outils d'IA provoquent des débats de justice sociale. Ceux-ci traitent des règles à mettre en place au sein des entreprises et/ou de la société (notamment via l'adaptation de la législation).
- Ces questions de justice sociale traitent notamment de l'impact des technologies d'IA sur le marché du travail, principalement les changements en matière d'évolution des missions/tâches réalisées.
- L'éthique et la philosophie politique mettent à disposition un vaste ensemble de théories de la justice mobilisables pour évaluer ces évolutions.
- Lors de l'application de ces théories de justice, deux discussions doivent avoir lieu en parallèle : Quelle société « idéale » souhaitons-nous (projection dans le futur) ? Et Étant donné cet idéal, quels sont les défis de justice pour la transition entre l'ici et maintenant et ce futur souhaité ?

5. Grands récits

La troisième thématique de l'outil de navigation traite des grands récits de société autour de l'IA. Cette partie clarifie la notion de récit, présente trois récits importants autour de l'IA, le rapport entre l'humain et la machine, les spécificités humaines et le sens de la vie humaine.

5.1 Les grands récits

Ces grands récits représentent à la fois le cadre général de nos débats sur l'IA et une histoire collective que nous entretenons [66]. Ils peuvent être définis comme des cadrages (*framings*) et des narrations (*narratives*). Pour percevoir et appréhender le monde qui nous entoure, nous en sélectionnons une partie spécifique en effectuant une opération de cadrage. Ce cadrage relève d'un choix individuel – mon regard sur le réel – mais dépend également de la société dans laquelle j'évolue. La culture au sens large (artefacts culturels, débats de société) offre ainsi un premier cadrage de la réalité.

Parler de « grands récits » ajoute à ce cadrage une dimension dynamique de narration, à nouveau sur le plan individuel (une histoire de vie personnelle) comme au niveau collectif. Ses narrations tentent de donner du sens à la réalité en structurant ses différents éléments et en énonçant certaines valeurs importantes. Les mythes qu'on considère comme fondateurs et les contes et légendes qui transmettent une leçon de vie, mais également les discours politiques, commerciaux et sociétaux qui tentent de donner un cap à notre action, sont autant d'exemples de récits. Sheila Jasanoff, sociologue américaine des sciences, parle d'« imaginaires socio-techniques », qu'elle définit comme des visions de futurs désirables, ancrées collectivement, stabilisées par certaines institutions et portées publiquement. Ces visions racontent des formes de vie en société rendues possibles par certaines technologies, tout en favorisant le développement de futurs possibles [67].

En matière d'IA, trois récits structurent une large partie des débats de société [68]. L'objectif est ici de faire apparaître la dimension éthique de ces récits et de l'imaginaire qu'ils véhiculent. Pour les spécialistes d'ingénierie, ces récits sont à la fois le contexte de leurs activités et le produit de ces activités. À titre individuel, ils ne peuvent échapper aux différents récits de société. Depuis fin novembre 2022, il est par exemple difficile d'éviter la frénésie autour de ChatGPT et des IA génératives. C'est l'occasion de voir s'affronter différents récits, tantôt techno-optimistes (avec des promesses de productivité décuplée et d'heures de travail évitées), tantôt technocritiques (soulignant par exemple les faiblesses flagrantes des modèles) ou encore techno-dystopiques (exigeant de stopper le développement de machines qui finiront par prendre le contrôle de l'humanité). Mais par leurs travaux et la manière dont ils commu-

Questions pour une entreprise/institution

- Gouvernance des outils d'IA :

L'utilisation d'outils d'IA s'inscrit-elle dans une stratégie transparente ?

Cette utilisation est-elle encadrée par des valeurs et principes formulés et accessibles aux équipes et aux clients (comme une charte) ?

Les outils d'IA sont-ils utilisés dans des domaines particulièrement sensibles, tant du point de vue des équipes que des clients/utilisateurs finaux ?

- Accompagnement au changement :

Les équipes sont-elles consultées sur l'utilisation des outils d'IA ?

L'évolution de leurs tâches a-t-elle été ouvertement thématisée (évolution cahier des charges, évolution métiers, promesses de temps de qualité supplémentaire), et des formations adéquates leur ont-elles été proposées ?

- Contribution de l'entreprise aux débats de société :

L'entreprise, ainsi que les autres acteurs de sa branche, contribue-t-elle aux débats de société sur l'impact des technologies d'IA ?

niquent, les ingénieurs entretiennent, enrichissent ou critiquent certains récits. Grâce à leur expertise et leurs activités, ils sont des acteurs écoutés dans ce processus de cristallisation des récits. Cette situation privilégiée leur confère également une responsabilité accrue au moment d'utiliser ou de critiquer ces récits.

5.2 Les rapports entre l'humain et la machine

Le premier récit porte sur la relation générale entre l'humain et la machine. Celle-ci **oscille entre un récit de compétition et un récit de coopération**. Dans la science-fiction, la compétition finit tôt ou tard par une situation où les IA, souvent incarnées au travers de robots, menacent de contrôler et soumettre les humains. Mais au-delà de la science-fiction, la manière dont nous communiquons sur certaines technologies, par exemple dans les médias, entretient ce récit de compétition. Ainsi, ChatGPT « menace » certains emplois. Ce récit de compétition suppose souvent une personnalisation des outils d'IA (ChatGPT n'est pas un outil, il est présenté comme une personne) et une forme humaine (les illustrations d'IA revêtent systématiquement des attributs humains). À l'opposé, nous pouvons considérer les relations machines-humains comme des relations de coopération. Les outils d'IA sont des instruments au service des humains. Ces derniers peuvent les utiliser pour accomplir leurs objectifs. Dans ce récit, les outils d'IA sont de simples instruments, et non des personnes.

Ce grand récit est crucial dans la relation entre une entreprise et ses collaborateurs. L'ambiance générale de travail, mais également les gages de confiance et la cohérence de la communication, impactent la manière dont les outils d'IA sont perçus par les équipes (en tant que concurrents ou en tant qu'outils).

5.3 Les spécificités humaines

Le choix d'un premier grand récit, entre compétition et coopération, conduit ainsi naturellement à un deuxième sur les spécificités de la condition humaine. Que reste-t-il de spécifiquement humain dans un monde mis en données et caractérisé par une présence durable et profonde d'outils d'IA ? En ligne directe avec nos réflexions précédentes, il nous reste une capacité d'autonomie de 2^e niveau, c'est-à-dire la capacité de réviser nos valeurs et nos objectifs. Nous avons l'immense liberté – et la difficile responsabilité qui l'accompagne – d'aménager notre existence. Cette liberté peut ensuite se décliner sous forme de créativité, d'inventivité ou d'imagination. Ce qui distingue l'humain réside ainsi peut-être dans son aptitude à s'éloigner d'une rationalité strictement instrumentale. Une machine fonctionne de la manière dont elle a été conçue et programmée. Pour atteindre un objectif fixé, elle doit mettre en œuvre des décisions dont les conditions de possibilité ont été préétablies. Les humains ont, quant à eux, la liberté de dévier de cette stricte équation entre objectif et moyens mis en œuvre.

À ce titre, l'éthique de l'IA doit également traiter de mouvements de fond qui pourraient menacer la liberté. Les outils d'IA actuelle ont un besoin vital de données pour se repérer et agir dans la réalité. Cette nécessité implique une mise en données du monde. Il doit être possible d'exprimer la réalité sous forme de chiffres afin que la machine puisse ensuite y naviguer. Cette tendance lourde entre dans une tension fondamentale avec l'ambition humaine de liberté. En effet, elle implique une catégorisation, une expression normée de la réalité. En captant un paysage, un outil d'IA ne voit pas d'ensemble, il voit un nombre d'arbres, une inclinaison du terrain, une palette de couleurs ou encore une luminosité spécifique. L'outil a besoin de ces catégories – qui lui sont données ou qu'il identifie seul – pour appréhender la réalité. Prises ensemble, ces catégories vont ainsi peu à peu s'imposer comme étant la nouvelle définition du paysage. Et ce type de démarche se retrouve dans l'ensemble de nos activités sociales, des services de ressources humaines des

entreprises aux parcours scolaires. **C'est le couplage de la mise en données du monde et de l'utilisation à large échelle de processus d'automatisation basés sur ces catégories qui pose un défi immense à la liberté humaine.** Car la catégorisation du monde exerce peu à peu une pression sur ceux dont le comportement s'éloigne des normes et dont la vie ne correspond pas à ces catégories prédéterminées. À cela s'ajoute l'utilisation de certaines catégories problématiques (par exemple l'origine ethnique d'une personne comme catégorie erronée/illégitime) qui viennent encore limiter notre liberté et constituent une difficulté supplémentaire.

En nous confrontant ainsi à la recherche de spécificités humaines, ce récit nous incite à interroger la centralité de l'humain dans notre approche du monde. La question touche à un fondement culturel profond, notamment pour les cultures marquées par les Religions du Livre. En dessous de Dieu, l'humain est la créature choisie pour contrôler la Terre et ses ressources et les utiliser à son service. À l'inverse, si l'homme n'est qu'un animal comme les autres, et qu'en plus des machines parviennent toujours mieux à l'imiter, que reste-t-il ? Face à ces développements, l'humain pourrait apprendre à se détacher de l'image d'une sorte d'Élu de la création. Cette relativisation rejoindrait le débat éthique autour du posthumanisme dont l'ambition affichée consiste à dépasser l'espèce humaine telle que nous la connaissons aujourd'hui [69].

5.4 Le sens de la vie humaine

Cette quête sur les spécificités humaines nous amène vers le choix d'un troisième grand récit portant sur le sens de la vie humaine. Par un effet miroir, l'utilisation accrue d'outils d'IA interroge en retour le sens de la vie humaine, par exemple son **rapport au travail ou à la mort**.

Les outils d'IA sont systématiquement présentés dans le contexte d'une productivité augmentée et d'une capacité de travail décuplée. Mais si nous restons dans un récit de compétition humain/machine (qui présuppose une comparaison légitime entre les deux), la performance humaine ne peut être que déclassée. Tout le monde convient qu'il serait absurde de comparer le nombre de lignes traduites en un jour par un traducteur automatique et un traducteur humain. Si la mesure de performance est quantitative, l'humain n'a aucune chance. La promesse vient peut-être d'une nouvelle division du travail où une mesure quantitative servirait à jauger le travail machines, tandis que des critères qualitatifs reposant sur des capacités que les humains maîtrisent mieux (adaptation, créativité, inventivité) leur seraient réservés.

Au-delà de la reconfiguration du rapport à la performance et au travail au sein de la société, l'IA est souvent associée au vieux rêve d'immortalité. Le corps fatigué mourrait, mais l'esprit continuerait sa vie par le biais d'une transplantation dans un corps mécanisé. Cette promesse de vie éternelle par la technologie se nourrit de la distinction entre le corps et l'esprit qui marque l'histoire de la pensée moderne occidentale. Au XVII^e siècle, le philosophe René Descartes (« *Je pense, donc je suis* ») propose un dualisme qui tente de présenter la matière et l'esprit comme deux substances distinctes [70]. Le corps apparaît alors comme une machine répondant à des lois mécaniques, tandis que l'esprit est le lieu de la liberté.

Ce récit ouvre la voie aux différents projets de téléchargement du cerveau dans une entité artificielle. Mais le défi philosophique de l'identité demeure [71]. Pourquoi supposer que la personne dont le cerveau est transféré vers une IA puis incarnée dans un robot reste la même personne ? N'est-elle pas plutôt une sorte de copie de l'original ? Cette question nous renvoie à plusieurs productions de science-fiction, de *Robocop* à *Ghost in the Shell*, qui mettent en scène les changements qui s'opèrent lorsque l'esprit s'incarne dans une nouvelle enveloppe. Cette question entre d'ailleurs en dialogue avec des traditions religieuses ou spirituelles où l'esprit se réincarne dans un nouveau corps physique.

Questions sur les récits

Sur quels grands récits la communication de l'entreprise/organisation au sujet des machines/IA repose-t-elle (textes, images, couleurs) ?

Les images choisies suggèrent-elles une anthropomorphisation des machines ou un rapport de compétition avec les humains ?

Ce choix est-il conscient et assumé ?

Quels types de promesses (moyen/long terme) sont implicitement ou explicitement formulées à travers la communication choisie ?

Pourrait-on profiter de l'introduction d'outils d'IA pour mener une discussion avec les équipes sur le sens et les spécificités de leurs activités dans l'entreprise/organisation ?

À retenir

- Les débats autour de l'IA reposent sur certains récits de société. Ce sont des histoires collectives que nous, en tant que citoyennes et citoyens, entretenons envers les technologies et leur impact sur l'humain. Ces récits sont au cœur des productions culturelles (films, livres, culture au sens large).
- Ces récits sont eux-mêmes des objets de réflexion éthique : en critiquant ou en promouvant certains récits, nous mettons dans l'ombre, ou au contraire dans la lumière, certaines caractéristiques et implications de l'IA.
- Trois grands récits structurent les débats sur l'éthique de l'IA : le rapport humain-machine (compétition ou instrument), la recherche de spécificités humaines et le sens de la vie dans un contexte de technologies capables de réaliser une bonne partie des activités habituellement réservées aux humains.

6. Conclusion

Au terme de cet article, nous avons une bonne vue d'ensemble des questions d'éthique pertinentes en matière d'IA. Sur la base d'une proposition de structure (outil de navigation), trois thématiques principales sont apparues : les questions d'éthique dans la technologie, les questions de justice sociale et les questions liées aux récits. Pour rappel, l'outil de navigation doit être utilisé de manière itérative et la fin d'un « tour » devrait marquer le début du suivant. Chaque partie contient les questions pertinentes à se poser.

Pour les spécialistes d'ingénierie, la question de la définition et les questions de la première thématique apparaissent comme les plus pertinentes. Leur expertise leur permet de proposer des solutions concrètes. Néanmoins, il n'est pas question de sous-estimer l'importance sur le travail des ingénieurs des récits et des questions de justice sociale. Ces interrogations représentent le cadre général dans lequel nous agissons tous. Consciemment ou non, nous sommes traversés et portés par ces récits et il est essentiel d'en prendre conscience. Face à une communication extrêmement puissante de l'industrie numérique, il est de plus essentiel de garder les pieds sur Terre et de ne pas se laisser emporter par des effets d'annonce.

En conclusion, on ne peut qu'avoir la conviction que l'éthique renforce l'innovation et la rend plus durable. Intégrer les éléments structurés grâce à l'outil de navigation tout au long de la conception d'outils d'IA va minimiser les risques encourus et ouvrir des opportunités pour une entreprise ou une organisation. L'éthique n'apparaît plus en fin de course, comme une instance de contrôle qui confère une sorte de certificat de sortie. Elle fonctionne plutôt comme un auxiliaire critique à tous les niveaux de conception et développement.

Éthique de l'intelligence artificielle : vue d'ensemble

par **Dr Johan ROCHEL**

Chercheur
EPFL – Lausanne, Suisse

Sources bibliographiques

- [1] BARTNECK (C.), LÜTGE (C.), WAGNER (A.) et WELSH (S.). – *An Introduction to Ethics in Robotics and AI*. Springer (2022).
- [2] MÜLLER (V.). – *Ethics of Artificial Intelligence and Robotics*. Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.) (Fall 2020 Edition).
- [3] MCCARTHY (J.), MINSKY (M.), ROCHESTER (N.) et SHANNON (C.). – *A proposal for the Dartmouth summer research project on artificial intelligence*. In *AI Magazine*, 27(4), pp. 12-14 (2006). <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1904>
- [4] OpenAI. – *Introducing ChatGPT* <https://openai.com/blog/chatgpt>
- [5] PERRIGO (B.). – *OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic*, in *TIME* (2023). <https://time.com/6247678/openai-chatgpt-kenya-workers/>
- [6] COMMISSION EUROPÉENNE. – *Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union*. COM/2021/206 final (2021).
- [7] ALGORITHMWATCH. – *ADM Manifeste*. <https://algorithmwatch.ch/fr/vision-mission-valeurs/>
- [8] BAUER (J.M.) et HERDER (P.M.). – *Designing Socio-Technical Systems*. In *Philosophy of Technology and Engineering Sciences*, A. Meijers, pp. 601-630 (2009).
- [9] DIDIER (C.). – *Éthique de l'ingénierie – Un champ émergent pour le développement professionnel*. In *Techniques de l'ingénieur*, pp. 3-4 [AG 102] (2015).
- [10] QUÉBEC. – *Loi sur les ingénieurs. Code de déontologie des ingénieurs*. Éditeur officiel du Québec (2023). <https://www.legisquebec.gouv.qc.ca/fr/pdf/rc/l-9,%20R.%206%20.pdf>
- [11] JOUE. – *Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (Règlement général sur la protection des données, RGPD)* (2016). <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32016R0679>
- [12] SEFI (Société européenne pour la formation des ingénieurs). – *Ethics* [section dédiée à l'éthique]. <https://www.sefi.be/activities/special-interest-groups/ethics-in-engineering-education/>
- [13] SWISS ALLIANCE FOR DATA-INTENSIVE SERVICES. – *Code d'éthique de la valorisation des données* (2020). https://data-innovation.org/wp-content/uploads/2021/01/w-che-Ethical-Codex_A4_Broschure_Grundlagen_FR.pdf
- [14] COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS (CNIL). – *Reconnaissance faciale : sanction de 20 millions d'euros à l'encontre de Clearview AI*. CNIL(2022). <https://www.cnil.fr/fr/reconnaissance-faciale-sanction-de-20-millions-deuros-lencontre-de-clearview-ai>
- [15] KNIGHT (W.). – *The Apple Card didn't "see" gender and that's the problem*. In *WIRED* (2019). <https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/>
- [16] ANGWIN (J.), LARSON (J.), MATTU (S.) et KIRCHNER (L.). – *Machine bias*. In *ProPublica* (2016). <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- [17] FAZELPOUR (S.) et LIPTON (Z.C.). – *Algorithmic fairness from a non-ideal perspective*. In *AAAI/ACM Conference on artificial intelligence, ethics, and society*. ArXiv (2020).
- [18] PRECOBS (Pre-Crime – Observation – System). – *Stadt Zuerich*. https://www.stadt-zuerich.ch/portal/de/index/politik_u_recht/stadtrat/weitere-politikfelder/smartcity/projekte/pre-cobs.html
- [19] ENSIGN (D.), FRIEDLER (S. A.), NEVILLE (S.), SCHEIDEGGER (C.) et VENKATASUBRAMANIAN (S.). – *Runaway feedback loops in predictive policing*. In *Conference on fairness, accountability, and transparency*. ArXiv (2017).
- [20] RICHARDS (J.), PIORKOWSKI (D.), HIND (M.), HOUE (S.) et MOJSILOVIC (A.). – *A Methodology for creating AI FactSheets*. ArXiv (2020).
- [21] MITCHELL (M.) et al.. – *Model cards for model reporting*. In *Conference on fairness, accountability, and transparency*. ArXiv (2019).
- [22] PASQUALE (F.). – *Black Box Society. Les algorithmes secrets qui contrôlent l'économie de l'information*. Fyp Éditions (2015).
- [23] SAMEK (W.), MONTAVON (G.), LAPUSCHKIN (S.), ANDERS (C.J.) et MÜLLER (K.R.). – *Explaining deep neural networks and beyond: A review of methods and application*. In *Proceedings of the IEEE*, 109(3), pp. 247-278 (2021).
- [24] CORTESE (J.F.N.B.), COZMAN (F.G.), LUCCA-SILVEIRA (M.P.) et BECHARA (A.F.). – *Should explainability be a fifth ethical principle in AI ethics?*. In *AI and Ethics*, 3(1), pp. 123-134 (2022).
- [25] MARKUS (A.F.), KORS (J.A.) et RIJNBEEK (P.R.). – *The role of Explainability in Creating Trustworthy Artificial Intelligence for Health Care: A Comprehensive Survey of the Terminology, Design Choices, and Evaluation Strategies*. In *Journal of Biomedical Informatics*, 113(103655) (2021).
- [26] KITE-POWELL (J.). – *Explainable AI is trending and here's why*. In *Forbes* (2022). <https://www.forbes.com/sites/jenniferhicks/2022/07/28/explainable-ai-is-trending-and-heres-why/?sh=14e73b842008>
- [27] COMMISSION EUROPÉENNE. – *Livre Blanc sur l'intelligence artificielle : Une approche européenne axée sur l'excellence et la confiance*. COM(2020) 65, (2020).
- [28] CHATELLIER (R.) et al.. – *La forme des choix : données personnelles, design et frictions désirables*. CNIL (2019). https://www.cnil.fr/sites/cnil/files/atoms/files/cnil_cahiers_ip6.pdf
- [29] THALER (R.H.) et SUNSTEIN (C.R.). – *Nudge*. Vuibert (2022).
- [30] ALOMBERT (A.) et KOKSHAGINA (O.). – *Votre attention, s'il vous plaît !*. Centre national du numérique (2022). <https://cnnumerique.fr/files/uploads/2022/Livres/CNNum-Votre-attention-sil-vous-plait.pdf>
- [31] MACDORMAN (K.F.) (préface, épilogue et traduction). – *La vallée de l'étrange de Mori Masahiro : importance et impact sur l'esthétique et la conception des robots*. e-Phaistos, Revue d'histoire des techniques, 7/2, pp. 1-18 (2019).
- [32] SCHIEBINGER (L.). – *The Robots are Coming! But should they be gendered?*. In *Awis magazine* (2019). <https://taiwan-gist.net/attachments/article/543/Schiebinger%20AWIS%20Robots%202019.pdf>
- [33] FOURNIER-TOMBS (E.). – *Apple's Siri is no longer a woman by default, but is this really a win for feminism?*. In *The Conversation* (2021). <https://theconversation.com/apples-siri-is-no-longer-a-woman-by-default-but-is-this-really-a-win-for-feminism-164030>

- [34] ARCHER (M.). – *The ethics of ESG: sustainable finance and the emergence of the market as an ethical subject*. In *Journal of Global and Historical Anthropology*, 93, pp. 18-31 (2022).
- [35] DIGITAL EU. – *A Green and Digital Transformation of the EU. Ministerial Declaration* (Engagement des pays de l'UE pour une transformation numérique verte) (2021). <https://digital-strategy.ec.europa.eu/fr/news/eu-countries-commit-leading-green-digital-transformation>
- [36] SMAILI (M.). – *Terres rares et métaux critiques, des enjeux de souveraineté sur fond de guerre économique*. In *Revue Défense Nationale*, 855, pp. 67-73 (2022).
- [37] HEFFRON (R.J.). – *The role of justice in developing critical minerals*. In *Elsevier*, 7(3), pp. 855-863 (2020).
- [38] DODGE (J.) et al.. – *Measuring the carbon intensity of AI in cloud instances*. In *Conference on Fairness, accountability, and transparency*, pp. 1877-1894 (2022).
- [39] LUCCIONI (A.S.), VIGUIER (S.) et LIGOZAT (A.L.). – *Estimating the carbon footprint of bloom, a 176b parameter language model*. ArXiv (2022). <https://arxiv.org/pdf/2211.02001.pdf>
- [40] INSTITUT DU NUMÉRIQUE RESPONSABLE. – *GR491, Le guide de référence de conception responsable de services numériques* (2021). <https://gr491.isit-europe.org/>
- [41] FERREBOEUF (H.). – *Déployer la sobriété numérique*. The Shift Project (2020). https://theshiftproject.org/wp-content/uploads/2020/10/Déployer-la-sobriete-numerique_Rapport-complet_ShiftProject.pdf
- [42] DEBUSSCHERE (V.), LEFIEUX (V.), MALLET (P.) et ROCHE (N.). – *Artificial intelligence serving electrical networks*. Think Smartgrids (2022). https://www.thinksmartgrids.fr/wp-content/uploads/2022/10/TSG_Intelligence_Artificielle_EN_vDEF_WEB.pdf
- [43] FLORIDI (L.) et al.. – *AI4People – An ethical framework for a good AI society: opportunities, risks, principles, and recommendations*. *Minds and Machines*, 28, pp. 689-707 (2018).
- [44] KAPOOR (S.) et NARAYANAN (A.). – *Leakage and the Reproducibility Crisis in ML-based Science*. ArXiv (2022).
- [45] HAKLI (R.) et MÄKELÄ (P.). – *Robots, Autonomy, and Responsibility*. In J. Seibt, M. Nørskov and S. Schack Andersen (eds), *What Social Robots Can and Should Do: Proceedings of Robophilosophy*, pp. 145-154 (2016).
- [46] THOMAS (L.). – *L'autonomie de la personne*. In M. Canto-Sperber (ed.), *Dictionnaire d'éthique et de philosophie morale*, PUF, pp. 144-147 (2004).
- [47] ISO. – ISO 8373:2021 – Robotique. (3.2) <https://www.iso.org/obp/ui/fr/#iso:std:iso:8373:ed-3:v1:fr>
- [48] MISSELHORN (C.). – *Grundfragen der Maschinenethik*. 5^e édition, Reclam (2022).
- [49] CHRISTMAN (J.). – *Autonomy in Moral and Political Philosophy*. In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.) (Fall 2020 Edition).
- [50] AMOROSO (D.) et TAMBURRINI (G.). – *Toward a Normative Model of Meaningful Human Control over Weapons Systems*. In *Ethics & International Affairs*, 35, pp. 245-272 (2021).
- [51] ZUBOFF (S.). – *L'âge du capitalisme de surveillance – Le combat pour un avenir humain face aux nouvelles frontières du pouvoir*. Zulma (2020).
- [52] SHAW (W.H.). – *Business Ethics*. 9th ed., Cengage Edition (2016).
- [53] CRANE (A.), MATTEN (D.) et SPENCE (L.J.). – *Corporate Social Responsibility. Readings and Cases in a Global Context*. 2nd ed., Routledge (2014).
- [54] MARZANO (M.). – *L'éthique appliquée*. Collection « Que sais-je », pp. 108-123 (2018).
- [55] LOBSCHAT (L.), MUELLER (B.), EGGERS (F.), BRANDIMARTE (L.), DIEFENBACH (S.), KROSCHE (M.) et WIRTZ (J.). – *Corporate digital responsibility*. In *Journal of Business Research*, 122, pp. 875-888 (2021).
- [56] AUDARD (C.). – *Justice*. In M. Canto-Sperber (ed.), *Dictionnaire d'éthique et de philosophie morale*, PUF, pp. 1001-1010 (2004).
- [57] VANDAMME (P.E.). – *Justice sociale* (A). In Maxime Kristanek (dir.), *L'Encyclopédie philosophique*.
- [58] RAWLS (J.). – *Théorie de la justice*. HUP (1971).
- [59] GABRIEL (I.). – *Toward a theory of justice for artificial Intelligence*. In *Daedalus*, 151(2), pp. 218-231 (2022).
- [60] ARSENAULT (A.C.) et KREPS (S.E.). – *AI and International Politics*. In Justin B. Bullock and others (eds), *The Oxford Handbook of AI Governance*, online edn, Oxford Academic (2022).
- [61] FREY (C.B.) et OSBORNE (M.A.). – *The future of employment: how susceptible are jobs to computerization?*. In *Technological Forecasting and Social Change*, 114, pp. 254-280 (2017).
- [62] SUSSKIND (D.). – *Technological Unemployment*. In Justin B. Bullock and others (eds), *The Oxford Handbook of AI Governance*, online edn, Oxford Academic (2022).
- [63] AUTOR (D.H.). – *The "task approach" to labor markets: an overview*. In *Journal for Labour Market Research*, 46(3), pp. 185-199 (2013).
- [64] MORADI (P.) et LEVY (K.). – *The Future of Work in the Age of AI: Displacement or Risk-Shifting?*. In *The Oxford Handbook of Ethics of AI*, pp. 270-288 (2020).
- [65] KORINEK (A.) et JUELFIS (M.). – *Preparing for the (Non-Existent?) Future of Work*. In *The Oxford Handbook of AI Governance*, online edn, Oxford Academic (2022).
- [66] BONNEUIL (C.) et JOLY (P.B.). – *Sciences, techniques et société*. La Découverte (2013).
- [67] JASANOFF (S.) et KIM (S.H.). – *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*. The University of Chicago Press, p. 4 (2015).
- [68] THE ROYAL SOCIETY. – *Portrayals and perceptions of AI and why they matter*. (2018). <https://royalsociety.org/-/media/policy/projects/ai-narratives/AI-narratives-workshop-findings.pdf>
- [69] HUNYADI (M.). – *Le défi politique du post-humanisme*. In *Études* 2015/3 (mars), pp. 55-64 (2015).
- [70] DESCARTES (R.). – *Discours de la méthode*. Flammarion (2016).
- [71] CHAUVIER (S.). – *La question philosophique de l'identité personnelle*. In *Identité(s)*, pp. 15-27 (2016).

À lire également dans nos bases

DIDIER (C.). – *Éthique de l'ingénierie – Un champ émergent pour le développement professionnel*. [AG 102] (2015).

ROUSSEAU (P.). – *Analyse du cycle de vie (ACV) – Méthodologie, applications et limites*. [G 5 550] (2022).

Annuaire

Dark Patterns Detection Project

<https://dapde.de/en/dark-patterns-en/types-and-examples-en/>

Mooc « Impact Num »

<https://www.fun-mooc.fr/fr/cours/impacts-environnementaux-du-numerique/>

Service d'auto-évaluation en matière de RNE (Responsabilité numérique d'entreprise)

<https://digital-responsibility.info/>

Gagnez du temps et sécurisez vos projets en utilisant une source actualisée et fiable



RÉDIGÉE ET VALIDÉE
PAR DES EXPERTS



MISE À JOUR
PERMANENTE



100 % COMPATIBLE
SUR TOUS SUPPORTS
NUMÉRIQUES



SERVICES INCLUS
DANS CHAQUE OFFRE

- + de 340 000 utilisateurs chaque mois
- + de 10 000 articles de référence et fiches pratiques
- Des Quiz interactifs pour valider la compréhension 

SERVICES ET OUTILS PRATIQUES



Questions aux experts*

Les meilleurs experts techniques et scientifiques vous répondent



Articles Découverte

La possibilité de consulter des articles en dehors de votre offre



Dictionnaire technique multilingue

45 000 termes en français, anglais, espagnol et allemand



Archives

Technologies anciennes et versions antérieures des articles



Info parution

Recevez par email toutes les nouveautés de vos ressources documentaires

*Questions aux experts est un service réservé aux entreprises, non proposé dans les offres écoles, universités ou pour tout autre organisme de formation.

Les offres Techniques de l'Ingénieur



INNOVATION

- Éco-conception et innovation responsable
- Nanosciences et nanotechnologies
- Innovations technologiques
- Management et ingénierie de l'innovation
- Smart city – Ville intelligente



MATÉRIAUX

- Bois et papiers
- Verres et céramiques
- Textiles
- Corrosion – Vieillessement
- Études et propriétés des métaux
- Mise en forme des métaux et fonderie
- Matériaux fonctionnels. Matériaux biosourcés
- Traitements des métaux
- Élaboration et recyclage des métaux
- Plastiques et composites



MÉCANIQUE

- Frottement, usure et lubrification
- Fonctions et composants mécaniques
- Travail des matériaux – Assemblage
- Machines hydrauliques, aérodynamiques et thermiques
- Fabrication additive – Impression 3D



ENVIRONNEMENT – SÉCURITÉ

- Sécurité et gestion des risques
- Environnement
- Génie écologique
- Technologies de l'eau
- Bruit et vibrations
- Métier : Responsable risque chimique
- Métier : Responsable environnement



ÉNERGIES

- Hydrogène
- Ressources énergétiques et stockage
- Froid industriel
- Physique énergétique
- Thermique industrielle
- Génie nucléaire
- Conversion de l'énergie électrique
- Réseaux électriques et applications



GÉNIE INDUSTRIEL

- Industrie du futur
- Management industriel
- Conception et production
- Logistique
- Métier : Responsable qualité
- Emballages
- Maintenance
- Traçabilité
- Métier : Responsable bureau d'étude / conception



ÉLECTRONIQUE – PHOTONIQUE

- Électronique
- Technologies radars et applications
- Optique – Photonique



TECHNOLOGIES DE L'INFORMATION

- Sécurité des systèmes d'information
- Réseaux Télécommunications
- Le traitement du signal et ses applications
- Technologies logicielles – Architectures des systèmes
- Sécurité des systèmes d'information



AUTOMATIQUE – ROBOTIQUE

- Automatique et ingénierie système
- Robotique



INGÉNIERIE DES TRANSPORTS

- Véhicule et mobilité du futur
- Systèmes aéronautiques et spatiaux
- Systèmes ferroviaires
- Transport fluvial et maritime



MESURES – ANALYSES

- Instrumentation et méthodes de mesure
- Mesures et tests électroniques
- Mesures mécaniques et dimensionnelles
- Qualité et sécurité au laboratoire
- Mesures physiques
- Techniques d'analyse
- Contrôle non destructif



PROCÉDÉS CHIMIE – BIO – AGRO

- Formulation
- Bioprocédés et bioproductions
- Chimie verte
- Opérations unitaires. Génie de la réaction chimique
- Agroalimentaire



SCIENCES FONDAMENTALES

- Mathématiques
- Physique Chimie
- Constantes physico-chimiques
- Caractérisation et propriétés de la matière



BIOMÉDICAL – PHARMA

- Technologies biomédicales
- Médicaments et produits pharmaceutiques



CONSTRUCTION ET TRAVAUX PUBLICS

- Droit et organisation générale de la construction
- La construction responsable
- Les superstructures du bâtiment
- Le second œuvre et l'équipement du bâtiment
- Vieillessement, pathologies et réhabilitation du bâtiment
- Travaux publics et infrastructures
- Mécanique des sols et géotechnique
- Préparer la construction
- L'enveloppe du bâtiment
- Le second œuvre et les lots techniques