

Researchpaper

## Intelligence artificielle et éthique appliquée : enjeux, défis et perspectives contemporains

Salmane Bourekkadi<sup>1</sup>

*1. Université Ibn Tofail, Kénitra, Maroc*

### PAPERINFO

PaperHistory

Received Dec 2025

Accepted April 2026

### Keywords

Intelligence artificielle,  
Éthique appliquée,  
Gouvernance des  
données,  
Biais algorithmique,  
Innovation responsable

### ABSTRACT

Cet article propose une analyse approfondie des enjeux contemporains liés à l'éthique de l'intelligence artificielle (IA), mettant en lumière les interactions complexes entre innovation technologique et responsabilité sociétale. À travers une approche multidimensionnelle, l'étude examine les fondements théoriques de l'éthique de l'IA, ainsi que les principaux défis liés à la responsabilité, aux biais algorithmiques, à la protection de la vie privée, à la gouvernance des données et à la prise de décision autonome. L'analyse démontre que les systèmes d'IA, tout en offrant des opportunités significatives en termes d'efficacité et de performance, peuvent aussi amplifier les inégalités sociales et introduire de nouveaux risques éthiques. L'article souligne l'importance de mécanismes de gouvernance robustes, fondés sur la transparence, la responsabilité et l'explicabilité, pour garantir l'utilisation éthique et fiable des technologies intelligentes. De plus, l'étude explore les implications économiques et sociales de l'IA, notamment en matière d'emploi et de redistribution de la richesse, ainsi que les défis liés à la diversité culturelle et aux perspectives éthiques mondiales. Enfin, elle propose un cadre conceptuel intégrant les principes d'innovation responsable et de recherche éthique, visant à guider le développement futur de l'intelligence artificielle vers des pratiques durables et inclusives.

## Introduction

L'intelligence artificielle (IA) implique des architectures computationnelles conçues par l'homme, capables d'interagir avec le monde avec un certain degré d'autonomie dans le processus de considération (Pavaloiu & Kose, 2017). Ces systèmes gèrent des phénomènes de signalisation réels et développent des modèles prédictifs pour soutenir la prise de décision autonome. Les applications de l'IA incluent les véhicules autonomes, les véhicules aériens sans pilote (UAV), les assistants au traitement du langage, les systèmes de traitement d'image et les systèmes de recommandation. Les structures politiques régissant les systèmes d'intelligence artificielle influenceront l'organisation future de la vie sur la planète Terre ; les résultats de tels systèmes sont mitigés.

L'éthique appliquée décrit l'examen de questions spécifiques et souvent pressantes du monde réel à travers le prisme de cadres et concepts moraux (philosophiques) établis. Cela soulève la question de savoir si l'intelligence artificielle devrait être un sujet de préoccupation morale. Les systèmes d'IA sont prétendus posséder l'agentivité, la responsabilité et la capacité d'exercer une action morale, car on fait partie d'une corrida, d'un orchestre solo ou d'une équipe de basket-ball en solo, restant ainsi en dehors du discours éthique (C. Müller, 2020). L'idée selon laquelle l'agence éthique est simplement une propriété de l'action elle-même, ou détermine la conséquence morale de l'action elle-même, sans impliquer l'acteur, est défendue par certains agents.

Néanmoins, la conception, la mise en œuvre et le déploiement des IA interagissent avec des structures de prise de décision humaine directement liées à des concepts éthiques spécifiques, des doctrines et des normes prescrites au sein des IA qui sont actuellement tirées des réflexions morales collectives des groupes sociaux et des sociétés. L'intérêt académique et sociétal s'est accru dans ce domaine de manière plus large, apparaissant régulièrement à la fois dans des règlements formels et des lignes directrices informelles (Borenstein & Howard, 2021).

### 2. Fondements de l'IA et de la théorie éthique

L'éthique de l'intelligence artificielle (IA) repose sur un spectre de sujets au sein de l'épistémologie, de l'ontologie et du pluralisme méthodologique. L'épistémologie sous-tend les arguments sur les capacités de représentation des connaissances de l'IA, l'établissement de la vérité et de la justification, ainsi que les considérations associées de biais dans les ensembles de données — si les processus d'apprentissage et de prise de décision sont appropriés en fonction des entrées. L'ontologie, inhérente aux définitions de l'IA, clarifie ce dont l'IA est capable ou non. L'idée générale des technologies actuelles d'IA est l'apprentissage automatique (ML), où un modèle est créé à partir de données d'entraînement et la qualité des données est un facteur crucial pour sa capacité, sa généralisation et ses futurs déploiements. Ainsi, les préoccupations épistémiques mentionnées sont accompagnées de questions de qualité des données et de sources de biais possibles (tant dans la collecte des données que dans l'opération algorithmique). De telles idées font partie de toute question ou considération éthique appliquée inhérente au déploiement des systèmes d'IA, que ce soit sur qui est impacté, comment l'IA peut être régulée ou si l'IA peut renforcer l'agence morale humaine.

Le cadre éthique dans lequel se déroule l'analyse restante nécessite un certain décryptage initial. Une telle base théorique aide à catégoriser les différentes questions examinées, tout en fournissant des éclairages sur les différentes approches de construction et de raisonnement concernant les technologies d'IA. Les principales théories éthiques restent distinctes et distinctes car elles prennent des points de départ différents, et les catégories spécifiques reflètent ces différences innées. Dans tous les cas, cependant, les normes sont hiérarchiques ; ainsi, toute proposition d'IA qui les viole ne peut être justifiée sur d'autres bases. Notamment, la structure d'incitation opérationnelle des humains reste la principale source de tout résultat négatif injustifié.

### 3. Responsabilité dans les systèmes d'IA

Le développement rapide de systèmes d'intelligence artificielle (IA) capables d'accomplir des tâches nécessitant traditionnellement l'intelligence humaine soulève des questions éthiques importantes lorsqu'ils opèrent indépendamment de l'intervention humaine. Une dimension importante de ces questions éthiques est la question de la responsabilité qui émerge avant, pendant et après les phases de conception, production et déploiement des systèmes d'IA. Certaines tâches peuvent impliquer un développeur humain produisant un produit initial suivi d'un système d'IA qui complète le processus. Dans ce cas, un écart peut apparaître quant à la responsabilité qui peut être confiée au système d'IA. Quelles exigences et assurances sont nécessaires pour alerter à la fois l'assignateur de tâche et celui qui exécute la tâche sur le degré d'assumption/attribution de responsabilité ? Qui assume finalement la responsabilité d'une tâche facilement « assumée » par un agent intelligent fonctionnel indépendant ? En l'absence d'un cadre de responsabilité approprié, une partie de responsabilité s'attache abstraitement au

collaborateur initial, étendant la responsabilité tout au long du projet. Quel est le statut du projet en cas d'échec ou de performance sous-optimale ? La question se pose donc quant à la mesure dans laquelle la responsabilité suit le processus automatisé et reste l'agence du développeur initial (Lima & Cha, 2020).

#### 4. Biais, équité et justice sociale en IA

Les systèmes d'IA peuvent perpétuer et amplifier les biais historiques et sociaux, affectant les conditions matérielles et la dignité des individus et des groupes (Leavy et al., 2020). Des biais peuvent apparaître, par exemple, dans la conception, la préparation, le développement, la modélisation et le déploiement d'outils d'IA. Les chercheurs en informatique étudiant les systèmes d'IA sous-estiment et surreprésentent les conditions sociales, expériences et considérations des populations et contextes ; ne reflètent qu'un éventail restreint d'expériences et de valeurs sociales ; et priorisent les défis techniques du développement des systèmes d'IA au détriment des préoccupations socio-éthiques et des bénéfices liés au déploiement des technologies d'IA. Cependant, de nombreux moyens potentiels d'atténuer les biais engendrent également de nouvelles sources de biais ou aggravent de nouvelles formes de discrimination.

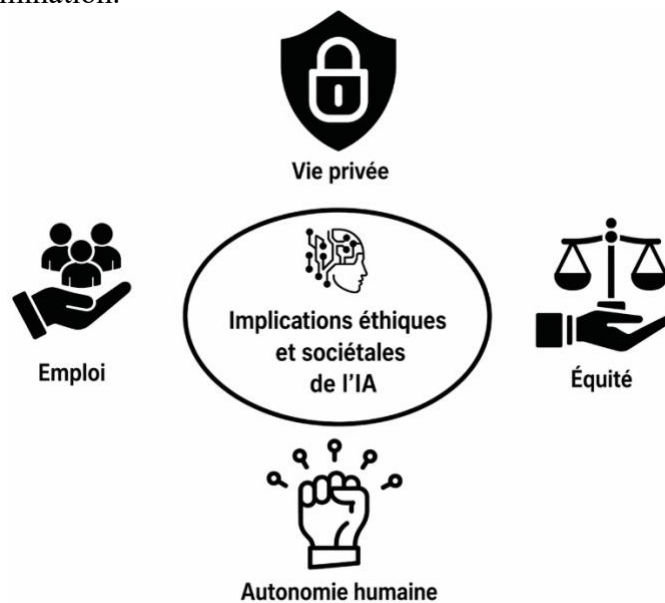


Figure 1. Impacts multidimensionnels de l'intelligence artificielle.

La figure 1 met en lumière la nature profondément ambivalente de l'intelligence artificielle, en tant que moteur simultané de transformation économique et de perturbation sociale, révélant ainsi une tension structurelle entre progrès technologique et impératifs éthiques. D'une part, l'IA favorise des améliorations significatives de la productivité et de l'efficacité organisationnelle grâce à l'automatisation des tâches, à l'optimisation des processus décisionnels et à l'exploitation massive des données, contribuant ainsi à la création de nouvelles opportunités économiques et à l'émergence de secteurs innovants. Cependant, ces avantages s'accompagnent d'effets différenciés qui exacerbent les inégalités, notamment par la polarisation du marché du travail, où les emplois de classe moyenne ont tendance à disparaître au profit de postes hautement qualifiés ou peu qualifiés, et par une concentration accrue de la richesse entre acteurs dotés des capacités technologiques et financières nécessaires pour développer et exploiter ces technologies. De plus, l'IA amplifie les risques de discrimination et d'injustice sociale en reproduisant, voire en renforçant, les biais présents dans les données et les algorithmes, posant de grands défis à l'équité et à la justice. Parallèlement, la dépendance croissante aux données soulève des préoccupations majeures concernant la vie privée, la surveillance et le contrôle des informations personnelles, avec le risque d'une société où les individus sont constamment observés et profilés. En matière de prise de décision, la délégation accrue de pouvoir aux systèmes automatisés remet en question la responsabilité humaine et la transparence des processus, en particulier dans des domaines sensibles comme la santé, la justice et la finance, où l'opacité algorithmique peut entraîner une perte de confiance. Enfin, ces impacts ne sont pas isolés mais interconnectés, générant des risques systémiques pouvant affecter toutes les structures économiques et sociales, rendant essentielle la mise en œuvre de cadres de gouvernance robustes fondés sur la transparence, la responsabilité et la supervision humaine. Ainsi, ce chiffre illustre clairement que l'intelligence artificielle ne peut être comprise uniquement sous le point de vue de la performance technologique, mais doit être intégrée dans une perspective globale d'éthique appliquée visant à concilier innovation, justice sociale et durabilité.

#### 5. Vie privée, surveillance et gouvernance des données

La vie privée, la surveillance et la gouvernance des données en IA concernent la collecte, le partage et la propriété des données associés aux systèmes d'aide à la décision basés sur l'apprentissage automatique. De nombreux systèmes d'IA nécessitent de grandes quantités de données pour entraîner les modèles, ce qui entraîne des compromis de la vie privée. Des biais et des discriminations peuvent résulter de la collecte de données biaisées ou de l'agrégation d'algorithmes et de systèmes d'apprentissage automatique lorsque ces données sont partagées entre de nombreux utilisateurs. Par conséquent, des individus et des gouvernements ont exprimé des objections à la dépendance des modèles d'apprentissage automatique à de grandes quantités de données, car ce volume augmente la probabilité de violations de la vie privée et restreint la liberté d'une personne à contrôler efficacement son comportement. Une automatisation accrue peut également conduire à la création d'une société de surveillance où les individus sont constamment enregistrés et surveillés par des systèmes intelligents. Des inquiétudes ont été soulevées quant à l'échec de la législation existante sur la protection des données à traiter efficacement les défis, car la portée des lois actuelles tend à se concentrer sur la collecte de données en point unique plutôt que sur la reconnaissance de l'effet global de la collecte de données sur la société dans son ensemble.

De nombreux facteurs influencent l'efficacité de la gouvernance des données pour minimiser les effets négatifs de surveillance tout en permettant la mise en place de protections de la vie privée telles que la confidentialité numérique et le contrôle des informations. Les considérations clés incluent la minimisation des données, le consentement, la fatigue du consentement, la confidentialité par conception et la propriété légale des données. La minimisation des données réduit la mesure dans laquelle les données sont collectées et limite le partage de ces données en interne et en externe, tandis que le consentement exige que les individus disposent d'une opportunité éclairée de donner ou de refuser l'autorisation pour cette collecte. La minimisation des données et le consentement dans la législation actuelle peuvent également créer une lassitude du consentement, où les personnes dont le consentement doit être obtenu sont submergées et ont tendance à signer un formulaire de consentement sans trop réfléchir. Les défenseurs de la confidentialité dès la conception de la vie privée considèrent la conception de la vie privée comme très importante.

## **6. Autonomie, prise de décision et supervision humaine**

La notion d'« autonomie » concerne la capacité d'une entité à fonctionner de manière indépendante, sans intervention humaine. L'autonomie implique la liberté de choix ; un agent autonome agit conformément à la volonté de l'agent plutôt qu'à des forces externes ou à des influences contrôlantes. L'autonomie suggère qu'un acteur a atteint un seuil scientifiquement et philosophiquement indéterminé, au-delà duquel l'implication d'une autre partie n'est pas requise. Ces seuils sont généralement formulés en termes de « délégation » ou de « contrôle humain significatif » (Methnani et al., 2021). De nombreux systèmes d'IA impliquent des paradigmes « humain dans la boucle », dans lesquels les décisions prises par le système nécessitent une affirmation humaine ; il existe de nombreux degrés dans lesquels un agent ou un décideur peut trouver une autonomie satisfaisante, allant de la liberté indépendante à l'exigence de consentement formel ou de nombreux comptes de délégation. L'autonomie est un enjeu fondamental de prise de décision en IA, et la supervision humaine représente un aspect critique du contrôle, mais garantir la « supervision humaine » dans les systèmes d'IA opérationnelles s'avère difficile. L'autonomie opérationnelle crée des défis concernant la charge cognitive de l'utilisateur, la confiance dans le système et l'interface par laquelle l'autorité est exercée (Walz & Firth-Butterfield, 2019).

## **7. Sécurité, fiabilité et gestion des risques en IA**

Les systèmes d'IA peuvent avoir des impacts profonds et omniprésents sur le bien-être personnel, la sécurité et la sûreté, soulevant des questions éthiques cruciales sur la gestion des risques liés à leur conception, leur déploiement et leur utilisation. Les défis éthiques posés par l'IA impliquent de nouvelles idées, pratiques et risques, et peuvent souvent être traités plus efficacement non pas par une réglementation formelle, mais avant un déploiement généralisé de la technologie. L'intelligence artificielle (IA) est une technologie numérique importante et en rapide évolution, avec le potentiel d'apporter des bénéfices sans précédent et de modifier profondément chaque aspect de la condition humaine.

Répondre aux attentes des parties prenantes et instaurer la confiance du public dans les systèmes d'IA exige une attention rigoureuse aux questions de sécurité, de fiabilité et de gestion des risques. Les systèmes et applications émergents basés sur l'IA soulèvent des questions importantes. L'automatisation et le fonctionnement autonome impliquent toujours la possibilité d'une défaillance éventuelle ou de conséquences inattendues. Les défis de sécurité liés aux systèmes non IA continuent de s'appliquer à l'IA et créent même de nouveaux dangers qui doivent

être abordés et résolus lors des phases de spécification, de conception et de développement. Une dépendance croissante à des systèmes d'IA de plus en plus puissants, mais souvent ininterprétables, soulève des incertitudes quant au comportement des systèmes dans un éventail croissant de conditions et d'environnements. Des capacités automatisées d'évaluation et de gestion des risques doivent être intégrées à tous les systèmes d'IA.

Les défaillances des systèmes d'IA, qu'elles résultent de problèmes de conception, de mise en œuvre, de déploiement ou opérationnels, peuvent avoir des effets négatifs non seulement sur les parties prenantes largement reconnues, mais aussi sur des parties prenantes inattendues et indirectes, ainsi que sur des systèmes avec lesquels le système d'IA interagit sans aucun contrôle explicite, et pour lesquels la responsabilité peut ne pas être claire. Comprendre et analyser les modes de défaillance potentiels associés aux systèmes d'IA et leurs impacts sociétaux ou des parties prenantes potentielles fournit des éclairages sur des stratégies de conception, de développement et d'atténuation opérationnelle appropriées. Les processus existants de certification de sécurité ou leurs extensions ou adaptations pertinentes fournissent des méthodologies et approches largement applicables. (Guan et al., 2022)

## **8. Implications économiques : travail, inégalités et bien-être**

Les avancées rapides de l'IA devraient avoir des conséquences profondes sur la croissance de la productivité, l'emploi, les inégalités et le bien-être. Alors que certains commentateurs prédisent que l'avènement d'une IA générale capable de surpasser les humains dans toutes les tâches pourrait déclencher une économie d'abondance radicale où aucun n'a besoin de travail, une perspective plus courante envisage que la production entièrement automatisée éliminera en grande partie la demande d'emplois routiniers, déplaçant un grand nombre de travailleurs et augmentant les inégalités. De tels développements nécessiteraient de nouvelles politiques redistributives, peut-être incluant un revenu de base universel et des investissements dans des programmes de reconversion (ABRARDI et al., 2019).

La croissance de la productivité résultant de l'adoption de l'IA pourrait à la fois améliorer le bien-être global et aggraver les inégalités et la polarisation des emplois. De nombreuses professions se caractérisent par une combinaison de tâches dont le degré de routine varie considérablement ; l'automatisation tend à éliminer d'abord les tâches les plus routinières. Ainsi, bien que les travailleurs déplacés puissent être capables d'obtenir de nouveaux emplois dans des professions nécessitant un mélange différent de tâches, de nombreuses tâches routinières restent présentes dans les économies avancées. Dans ces conditions, les travailleurs peuvent trouver un emploi alternatif qui conserve une partie du mélange initial de tâches, mais les emplois de niveau intermédiaire — ceux qui utilisent une combinaison de tâches intermédiaires — devraient subir les plus grandes pertes. Il y a peu de consensus sur l'impact ultérieur sur la demande globale de main-d'œuvre et la répartition des revenus entre les professions.

## **9. Gouvernance, régulation et politique publique**

† La gouvernance, la régulation et la politique publique pour l'IA doivent prendre en compte les différences de compétence dans les normes et l'application (C. Müller, 2020). Le Livre blanc de la Commission européenne adopte un modèle basé sur les risques qui priorise les droits fondamentaux et la sécurité (Mittelstadt, 2019). Les cadres fondés sur des principes cherchent à mettre en œuvre les exigences largement acceptées en matière d'IA. Ces idéaux peuvent renforcer la transparence, la traçabilité et la responsabilité à travers les systèmes d'IA. La gouvernance et la régulation doivent concerner non seulement les capacités et limites technologiques, mais aussi prendre en compte adéquatement le pouvoir et les responsabilités des acteurs corporatifs et institutionnels. Des mécanismes spécifiques pour une conduite conforme, digne de confiance et éthique pourraient favoriser une responsabilité plus large.

## **10. Perspectives mondiales et éthique interculturelle en IA**

Les systèmes d'intelligence artificielle (IA) reflètent les influences sociales, politiques et culturelles de leur création et de leur formation. Ces systèmes reproduisent souvent des biais existants, tels que les biais de genre, qui peuvent marginaliser certains groupes et aggraver les inégalités. Un manque de diversité parmi les concepteurs signifie que la technologie incarne des perspectives étroites. Les principes éthiques intégrés dans l'IA sont donc influencés par des contextes sociaux et culturels spécifiques. De plus, à mesure que l'IA influence de plus en plus les droits humains et les structures sociales, il est crucial de reconnaître que les cadres éthiques et les documents politiques ne sont pas neutres mais reflètent plutôt des visions philosophiques et culturelles particulières. Ces documents dépendent donc du contexte et sont soumis à des interprétations variées selon les sociétés (Roche et al., 2022).

Un écosystème IA sain bénéficiera de multiples perspectives qui remettent en question les récits dominants, élargissent la compréhension des implications éthiques de l'IA et inspirent l'exploration de solutions alternatives. Pourtant, de nombreuses initiatives d'IA reflètent massivement les valeurs occidentales ; certains défenseurs du colonialisme numérique craignent que les grandes entreprises technologiques imposent leurs normes aux pays en développement, sapant ainsi les identités culturelles et adoptant des préférences contraires aux besoins locaux. En plus du design centré sur l'utilisateur, les institutions devraient étudier un design culturellement réactif permettant au développement de s'aligner sur les normes éthiques locales (Mittelstadt, 2019).

### **11. Technologies émergentes : alignement, explicabilité et transparence**

Le développement et l'affinement des techniques d'intelligence artificielle (IA) ces dernières années ont suscité une réflexion sur les implications de leur utilisation (Larsson, 2019). L'attention est de plus en plus portée à la confiance et à la fiabilité — souvent exprimée par les termes alignement, explicabilité et transparence — en particulier dans le contexte des applications de prise de décision. De plus, l'inquiétude grandit quant aux implications sociétales négatives.

Pour établir la confiance dans les systèmes d'IA, il est nécessaire de communiquer la nature et la portée du fonctionnement du système, ainsi que de fournir l'assurance que les résultats escomptés seront atteints (C. Müller, 2020). L'IA explicable offre des degrés variables de compréhension et d'éclairage sur la logique sous-jacente aux conclusions, prédictions et recommandations algorithmiques. Une meilleure compréhension peut aider à percevoir dans quelle mesure les résultats d'apprentissage supervisé ont été alignés comme prévu, à mieux apprécier le degré de ruse implicite dans les deepfakes, et à discerner quels signes constituent des marqueurs de risque dans le diagnostic précoce des maladies à partir des données d'imagerie. En fin de compte, l'objectif est de déterminer si les résultats reflètent ce qui est voulu, plutôt que de rester obscurs et opaques.

L'alignement de l'IA vise à combler le fossé entre les objectifs visés, les modèles utilisés pour les poursuivre et les orientations prises dans le monde réel. Un objectif peut sembler simple, mais les modèles développés pour servir cet objectif divergent fréquemment de la trajectoire prévue ou produisent des résultats contraires aux attentes. Le moyen le plus efficace de renforcer la confiance et de soutenir une gouvernance robuste des systèmes d'IA est une évaluation continue, complète et systématique des objectifs visés ainsi que de la mesure dans laquelle les modèles et résultats correspondants restent alignés.

### **12. Éthique de la recherche et innovation responsable**

L'éthique de la recherche en IA — englobant la conduite responsable, la prise en compte de l'impact sociétal et la diffusion ouverte — a reçu une attention croissante (Samuel et al., 2021). La société est confrontée à des défis complexes aggravés par la pandémie de COVID-19, le changement climatique, des débats polarisés et l'augmentation des inégalités. Des scientifiques issus de disciplines diverses — de la santé à l'économie, des sciences sociales à l'IA — défendent une recherche responsable et proposent diverses approches pour intégrer les considérations sociétales (Kluge Corrêa et al., 2022). Bien que les chercheurs restent engagés dans l'intégrité de la recherche, l'attente de considérer l'impact sociétal et environnemental au-delà de toute diffusion ne cesse de croître.

L'éthique de la recherche concerne généralement la conduite de recherches impliquant des participants humains. L'impact sociétal de la recherche au-delà du milieu académique est de plus en plus reconnu aux niveaux institutionnel et gouvernemental et est intégré dans les critères de financement, d'évaluation et de promotion. La Recherche et l'Innovation Responsables (IRR) favorise l'intégration des considérations sociétales dans le processus de recherche, en soutenant le concept de « science pour et avec la société ». L'émergence du virage numérique a transformé de manière spectaculaire les pratiques de recherche, apportant de nouvelles sources de données, outils méthodologiques, techniques et modèles analytiques tels que l'intelligence artificielle (IA) et la modélisation des données.

Les données numériques remodelent profondément de nombreux paradigmes conceptuels et de recherche. La datafication transforme activités, artefacts et interactions en traces numériques pouvant être capturées, stockées et analysées en temps réel ou proche. Des ensembles de données diversifiés, denses et hétérogènes peuvent offrir des perspectives sans précédent sur la société et la technologie utilisant les données numériques, l'IA et la modélisation des données. La modélisation algorithmique et computationnelle peut permettre à la recherche d'explorer des réponses systémiques, holistiques et émergentes aux défis sociétaux pressants. Cependant, des

implications éthiques émergent concernant la confiance, la responsabilité, la responsabilité, la transparence et l'explicabilité des algorithmes et des analyses. Les systèmes d'IA de plus en plus complexes soulèvent de nouvelles questions éthiques dans les domaines scientifiques et les domaines de recherche.

### **13. Éducation, alphabétisation et engagement public**

L'intelligence artificielle (IA) évolue plus vite que la société ne peut suivre le rythme : « L'apprentissage automatique pourrait représenter une menace encore plus grande pour l'humanité que les armes nucléaires » ((Yu & Yu, 2023)). Les systèmes d'IA approchant une intelligence générale au même niveau que les capacités cognitives humaines devraient émerger dans quelques décennies et devenir la technologie la plus importante pour l'humanité (). Il est urgent de traiter les dilemmes liés aux machines, mais étonnamment peu de travaux substantiels ont été menés sur les politiques éducatives en matière d'éthique de l'IA (). Il est crucial d'accroître la sensibilisation à l'IA responsable et de transmettre les préoccupations éthiques à la société. Les horizons formels se sont réduits dans le discours populaire et professionnel ().

Les gens manquent généralement de connaissances sur la technologie de l'IA, ce qui entrave l'engagement dans des discussions significatives sur l'impact sur les valeurs et les normes. La société n'est pas prête à examiner de manière critique l'IA, la rendant vulnérable à l'exploitation ou aux influences négatives (). L'éducation et la sensibilisation à la littératie IA sont nécessaires pour établir les conditions de débats publics ouverts sur l'impact de l'IA sur les valeurs et normes sociétales. Les gouvernements, les institutions politiques, les citoyens et la société civile doivent participer activement à l'orientation de l'IA ().

### **14. Conclusion**

L'intelligence artificielle (IA) est un sujet intensément débattu sous de nombreux angles tels que la philosophie, les aspects techniques, économiques ou sociétaux, mais le point de vue éthique est l'un des domaines de recherche de premier plan (C. Müller, 2020). L'éthique appliquée est la discipline qui adopte des théories et analyses éthiques, et les étend à divers domaines et domaines de l'activité humaine. Parmi ces applications, de nombreux aspects de l'IA sont de grande importance et de plus en plus étudiés (Pavaloiu & Kose, 2017). Dans la présente analyse, nous laissons de côté les fondements de l'éthique (Hume, Kant, Mill, etc.) pour nous concentrer sur l'état actuel du débat. Aborder la question avec un cadre clair éclaire les diverses facettes de la problématique de l'IA. La pertinence du cadre permet également de situer et d'évaluer les enjeux majeurs de l'IA, ainsi que ses questions éthiques associées, fournissant ainsi un éclairage global et assurant une compréhension globale de l'IA.

L'analyse couvre un ensemble de questions contemporaines aux larges implications pour la gouvernance de l'IA. La discussion éthique sur les systèmes d'IA a essentiellement commencé par deux voies parallèles. La première voie découle d'une appréhension spéculative d'une future émergence de l'Intelligence Artificielle Générale (AGI), dotée de capacités énormes telles que le raisonnement et les croyances personnelles. Dans ce scénario, les considérations éthiques doivent englober l'AGI future elle-même, mais les systèmes d'IA passés et présents introduisent rapidement des possibilités liées à l'éthique qui ne peuvent être négligées. La seconde voie de la discussion commence par un examen approfondi des systèmes d'IA actuels, de leurs applications désormais omniprésentes, et de la série de questions éthiques impactant la société et le monde dans son ensemble. La seconde voie est ouverte à l'exploration de questions multiples et diverses suscitant un intérêt académique sérieux, assurant ainsi une pertinence continue pour la recherche d'une compréhension claire.

Une série de cadres permettant de faire le pont entre l'analyse éthique et le débat actuel sur l'IA est présentée. L'analyse distingue les modèles de base comme se développant sur des principes fondamentaux dans un domaine spécifique (c'est-à-dire l'éthique), et les cadres inter-domaines comparant la situation d'un domaine spécifique à celle d'autres domaines. L'analyse couvre donc des cadres inter-domaines à quatre niveaux comprenant deux catégories : classification des familles éthiques et méthodologie éthique. Un modèle à trois niveaux comprenant uniquement la classification des familles éthiques est également illustré. Un autre cadre correspondant à une question plus fondamentale des hypothèses éthiques de base est également brièvement abordé, car cette question semble assez importante dans le domaine de l'IA. Chaque cadre doit être ajusté sur les diverses caractéristiques liées aux sujets d'IA d'importance primordiale. Une typologie de ces principaux sujets est classée selon la nature des systèmes auxquels correspond les spécifications afin de saisir les fondamentaux de ces caractéristiques ainsi que leur relation avec la famille éthique ou la méthodologie particulière.

### **Références**

- Pavaloiu, A. & Kose, U. (2017). Ethical Artificial Intelligence - An Open Question.
- C. Müller, V. (2020). Ethics of Artificial Intelligence and Robotics.
- Borenstein, J. & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education.
- Lima, G. & Cha, M. (2020). Responsible AI and Its Stakeholders.
- Leavy, S., O'Sullivan, B., & Siapera, E. (2020). Data, Power and Bias in Artificial Intelligence.
- Methnani, L., Aler Tubella, A., Dignum, V., & Theodorou, A. (2021). Let Me Take Over: Variable Autonomy for Meaningful Human Control.
- Walz, A. & Firth-Butterfield, K. (2019). Implementing Ethics into Artificial Intelligence: A Contribution, from a Legal Perspective, to the Development of an AI Governance Regime.
- Guan, H., Dong, L., & Zhao, A. (2022). Ethical Risk Factors and Mechanisms in Artificial Intelligence Decision Making.
- Abrardi, L., Cambini, C., & Rondi, L. (2019). The economics of artificial intelligence: a survey.
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI.
- Roche, C., J. Wall, P., & Lewis, D. (2022). Ethics and diversity in artificial intelligence policies, strategies and initiatives.
- Larsson, S. (2019). The Socio-Legal Relevance of Artificial Intelligence.
- Samuel, G., Chubb, J., & Derrick, G. (2021). Boundaries Between Research Ethics and Ethical Research Use in Artificial Intelligence Health Research.
- Kluge Correa, N., Galvao, C., William Santos, J., et al. (2022). Worldwide AI Ethics: a review of 200 guidelines and recommendations for AI governance.
- Yu, L. & Yu, Z. (2023). Qualitative and quantitative analyses of artificial intelligence ethics in education using VOSviewer and CitNetExplorer.