

Les Normes Éthiques de Conception et de Développement des Applications d'IA à l'ère numérique



■ Prof./ Mohamed Fawzi Wali

Professeur de technologie de l'éducation à la Faculté de Pédagogie
Ancien doyen de la Faculté d'informatique et des systèmes
d'information – Université de Damanhour

■ Résumé:

Le domaine de l'intelligence artificielle (IA) connaît un développement rapide avec la multiplication de ses applications dans divers aspects de la vie, soulevant ainsi des interrogations essentielles et renouvelées quant à la manière de développer et d'utiliser ces applications de manière éthique, responsable et bénéfique pour la société dans son ensemble. L'éthique liée aux applications de l'IA renvoie à un ensemble de principes et de valeurs guidant leur conception et leur usage de manière responsable, visant à en maximiser les bénéfices tout en minimisant les risques potentiels.

La présente recherche vise à identifier les principales normes éthiques encadrant la conception et le développement des applications d'IA à l'ère numérique. Elle aboutit à une liste de normes éthiques composées de quinze critères fondamentaux, parmi lesquels figurent : la responsabilité, la redevabilité, la transparence, l'équité, le respect de la vie privée et la durabilité environnementale.

L'étude recommande de prendre en compte ces normes et leurs indicateurs lors de la conception et du développement des applications d'intelligence artificielle, celles-ci constituant un cadre de référence essentiel pour guider ces processus, en veillant à optimiser les avantages tout en réduisant les risques.

La recherche propose également de mener des études complémentaires sur les répercussions de la prise en compte de ces normes éthiques, aussi bien du point de vue des développeurs que des utilisateurs.



Mots-clés:

normes éthiques, conception et développement des applications d'IA,
Stratégie en IA, Stratégie de lutte contre la corruption.

Introduction

L'intelligence artificielle (IA) est désormais devenue une composante incontournable de notre vie quotidienne. Elle est utilisée à grande échelle et connaît une évolution remarquable. Le Cun, Bengio et Hinton (2015) définissent l'IA comme un ensemble de technologies permettant aux ordinateurs de comprendre le monde qui les entoure et d'interagir avec celui-ci de manière intelligente, en traitant les données, en prenant des décisions appropriées et en développant des modèles prédictifs.

De leur côté, Stone et autres (2016) considèrent l'IA comme un vaste champ de recherche visant à concevoir des systèmes capables de simuler les capacités cognitives humaines, telles que l'apprentissage, le raisonnement, la résolution de problèmes, la prise de décision et la créativité.

Ahmed El-Arabi (2021) souligne quant à lui que l'IA constitue un domaine interdisciplinaire englobant plusieurs branches, telles que l'apprentissage automatique, le traitement du langage naturel, la vision par ordinateur et la robotique. Ce champ vise à concevoir des systèmes intelligents capables d'agir de manière autonome et de prendre des décisions de façon indépendante. Russell et Norvig (2021) précisent, pour leur part, que l'IA est une branche de l'informatique dédiée à la conception et au développement de systèmes intelligents aptes à fonctionner de manière autonome dans des situations complexes et incertaines.

En somme, l'intelligence artificielle s'impose aujourd'hui comme l'une des principales orientations scientifiques contemporaines. Elle constitue un champ de recherche fondamental de l'informatique moderne, s'efforçant de comprendre la nature de l'intelligence humaine, d'en suivre l'attention et les réactions, et de simuler les comportements intelligents afin de formuler des recommandations, résoudre des problèmes ou assister à la prise de décision (Sreenu, 2023).

L'IA possède un potentiel considérable de transformation sociale. Ses applications généralisées à travers divers systèmes intelligents engendrent des bénéfices économiques significatifs, tout en soulevant un ensemble de défis éthiques majeurs. La plupart des chercheurs et développeurs s'accordent sur la nécessité impérieuse de concevoir l'IA de manière fiable, dans le but de servir l'intérêt général. Le schéma ci-dessous illustre les principaux usages des applications de l'IA (Berendt, 2019 ; Cath et autres, 2018 ; Commission européenne, 2020 ; Floridi et autres, 2018 ; Jobin et autres, 2019).

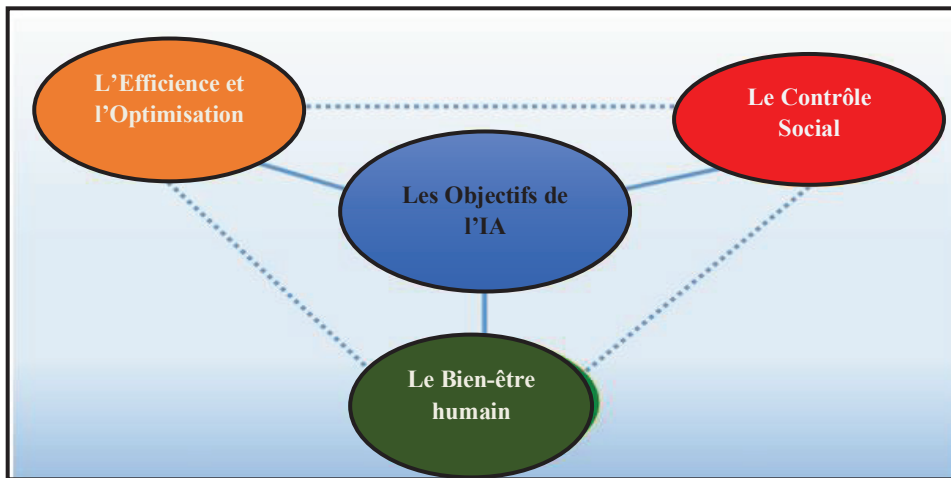


Figure (1) Les Objectifs de l'IA (Berendt, 2019)

La conception et le développement des applications d'IA exigent de surmonter un ensemble de défis d'ordre éthique. Bien que de nombreuses études aient abordé la question du développement des applications d'IA, la littérature demeure lacunaire en ce qui concerne l'identification précise et rigoureuse des normes éthiques encadrant l'usage de ces technologies (Shin & Park, 2019).

L'équilibre entre les différents principes éthiques dans la conception des applications d'IA représente un enjeu fondamental, dans la mesure où il est souvent difficile, voire impossible, de les satisfaire simultanément. Si les applications d'IA visent principalement à alléger la charge des utilisateurs et à faciliter l'exécution de diverses tâches, la mise à disposition de données personnelles suscite de profondes inquiétudes, notamment en matière d'atteinte à la vie privée (Marwick & Lewis, 2017 ; Shin & Park, 2019).

Un groupe d'experts en IA (AI HLEG) mandaté par la Commission européenne a identifié sept exigences fondamentales à satisfaire pour garantir le développement de systèmes d'IA dignes de confiance. Ces exigences sont les suivantes (Shneiderman, 2020) :

- L'autonomie humaine et la supervision
- La robustesse technique et la sécurité
- La protection de la vie privée et la gouvernance des données
- La transparence
- La diversité, la non-discrimination et l'équité
- Le bien-être social et environnemental
- La redevabilité

Par ailleurs, l'Organisation de coopération et de développement économiques (OCDE) recommande de respecter cinq principes éthiques lors de la conception d'applications d'IA (OCDE, 2021) :

- La croissance inclusive, le développement durable et le bien-être
- Le respect des valeurs humaines et la justice
- La transparence et l'explicabilité
- La robustesse, la sûreté et la sécurité
- La redevabilité

Jobin et autres (2019) ont mené une revue de 84 documents issus de différentes régions du monde portant sur l'éthique de l'IA. Les auteurs ont identifié onze principes éthiques universels, dont cinq – la transparence, l'équité, la non-nocivité, la responsabilité et la protection de la vie privée – figurent dans plus de la moitié des études analysées. D'autres principes tels que la bienfaisance, la liberté et l'autonomie sont évoqués dans 41 études respectivement. Enfin, des principes comme la confiance, la durabilité, la dignité ou encore la solidarité apparaissent dans environ un tiers des publications.

Bien que l'ensemble de ces principes éthiques soit, en théorie, souhaitable, leur mise en œuvre effective dans la pratique soulève de nombreuses difficultés. En effet, il est souvent complexe de concilier l'ensemble des impératifs éthiques lors du développement d'un système d'IA (Köbis et autres, 2021).

L'éthique et le droit constituent deux piliers essentiels du cadre normatif qui régit le développement et l'utilisation des applications d'intelligence artificielle. Tandis que l'éthique fournit des lignes directrices assurant la conformité des technologies intelligentes avec les valeurs humaines – telles que la justice, la transparence et la responsabilité –, le droit édicte des règles et des procédures juridiquement contraignantes visant à encadrer l'usage de ces technologies et à prévenir toute dérive, notamment en matière d'atteinte à la vie privée ou de biais discriminatoires. Il convient de souligner que la complémentarité entre exigences éthiques et exigences juridiques est essentielle au développement d'une IA responsable, capable de servir équitablement et en toute sécurité les intérêts de la société (Bryson, 2018).

Il importe également de rappeler que l'éthique de l'IA ne relève pas exclusivement de la responsabilité des développeurs et chercheurs. Elle constitue une responsabilité collective impliquant également les utilisateurs, les entreprises, les gouvernements, les organisations et l'ensemble de la société. C'est par la coopération et l'adhésion partagée aux principes éthiques qu'il sera possible de garantir que l'IA devienne une force positive au service d'un avenir meilleur pour tous.

Dans cette perspective, les gouvernements, les institutions académiques et les organisations technologiques s'efforcent d'élaborer des politiques et des lignes directrices éthiques destinées à encadrer le développement et l'utilisation responsables de l'IA. De nombreuses entités, telles que l'Union européenne ou l'OCDE, ont publié des cadres de référence pour le développement d'une IA éthique (Jobin et autres, 2019). La République arabe d'Égypte, quant à elle, a

adopté une stratégie nationale en matière d'intelligence artificielle, dans une démarche proactive visant à maximiser les bénéfices de ces technologies. Néanmoins, la mise en œuvre concrète de ces principes demeure confrontée à de nombreux défis, ce qui rend nécessaire une mobilisation concertée des chercheurs, des décideurs et des concepteurs afin de garantir la conformité des systèmes d'IA aux valeurs fondamentales de l'humanité (Mittelstadt et autres, 2016).

Problématique de la recherche

Dans le contexte numérique actuel, certaines applications d'IA souffrent d'un manque de prise en compte de la dimension éthique, ce qui peut engendrer de nombreuses conséquences néfastes pour la société. Parmi celles-ci figurent notamment le renforcement des biais et des discriminations à l'égard de certaines catégories sociales, la violation de la vie privée, ainsi que la mauvaise utilisation des données, qui peuvent parfois être collectées sans le consentement préalable des utilisateurs. Cette situation ouvre la voie à des risques tels que le vol de données, le chantage numérique, ou encore l'atteinte à la vie privée.

En outre, les applications d'IA peuvent être exploitées à des fins de désinformation, à travers la production de fausses nouvelles ou de vidéos truquées. Elles peuvent également contribuer à réduire les interactions humaines directes, exacerbant ainsi l'isolement social. Les principaux effets découlant de l'usage actuel des applications d'IA peuvent être synthétisés comme suit :

- **Biais dans les données** : Les préjugés présents dans les ensembles de données peuvent entraîner un apprentissage biaisé des modèles d'IA, produisant ainsi des résultats injustes, notamment des discriminations d'ordre ethnique, social ou économique.
- **Vie privée et sécurité** : Une gestion inadéquate des données sensibles peut provoquer des atteintes à la vie privée ou des fuites d'informations.
- **Qualité et intégrité des données** : Des données de mauvaise qualité, incomplètes ou obsolètes peuvent compromettre la fiabilité des prédictions des modèles et conduire à des décisions erronées.
- **Biais algorithmiques** : Indépendamment des données, les algorithmes eux-mêmes peuvent intégrer des biais lorsqu'ils ne sont pas conçus avec rigueur, engendrant des résultats discriminatoires.
- **Automatisation biaisée** : Une confiance excessive accordée aux recommandations de l'IA, supposées plus précises qu'elles ne le sont en réalité, peut induire des erreurs de jugement.
- **Défaillances systémiques** : Les systèmes d'IA peuvent échouer ou se comporter de manière imprévisible face à des environnements dynamiques ou à des scénarios inédits.
- **Problèmes de passage à l'échelle** : À mesure que les systèmes d'IA se généralisent, leur maintenance et leur performance dans des contextes variés deviennent plus complexes.

- **Attaques adverses** : Les systèmes d'IA, notamment dans les domaines sensibles comme la cybersécurité, sont vulnérables à des attaques consistant à manipuler les entrées pour générer des sorties incorrectes.
- **Inégalités et discrimination** : Les applications d'IA peuvent aggraver les inégalités existantes ou générer de nouvelles formes de discrimination si elles ne sont pas régulées de manière éthique.
- **Transparence et redevabilité** : L'opacité des processus décisionnels dans certains systèmes d'IA peut rendre difficile l'identification des responsabilités en cas d'erreur ou de préjudice.
- **Contrefaçon et désinformation** : Les contenus générés par l'IA, tels que les deepfakes, peuvent être utilisés pour diffuser des informations trompeuses, portant atteinte à la qualité de l'information publique.
- **Propriété intellectuelle** : L'utilisation abusive de données ou de contenus protégés par des droits d'auteur soulève des problématiques juridiques relatives à la propriété intellectuelle.

Face à ces enjeux, il apparaît nécessaire de définir un ensemble de normes éthiques devant encadrer la conception et le développement des applications d'IA. Cette exigence s'impose en raison de la nature évolutive de cette technologie et de sa complexité croissante, appelant à l'élaboration de fondements méthodologiques garantissant la conformité des systèmes d'IA aux principes éthiques fondamentaux tels que la transparence, l'équité et la responsabilité.

C'est dans cette perspective que s'inscrit la présente recherche, qui vise à identifier et formuler une série de normes éthiques susceptibles de guider le développement de systèmes d'intelligence artificielle respectueux des valeurs humaines et aptes à minimiser les risques potentiels associés à leur utilisation.

Objectifs de la recherche

La présente étude vise à :

1. Identifier les principales normes éthiques applicables à la conception et au développement des applications d'intelligence artificielle dans divers domaines. Cela inclut des questions telles que la vie privée, l'équité, les biais, la responsabilité juridique et les impacts sociaux de cette technologie.
2. Étudier les méthodes et stratégies permettant d'aborder ces enjeux de manière responsable, visant à concilier l'évolution technologique rapide avec la protection des valeurs humaines fondamentales⁽¹⁾.
3. Proposer des solutions et recommandations destinées à orienter le développement et l'utilisation des applications d'IA dans une optique de

(1) Les valeurs humaines fondamentales désignent un ensemble de principes éthiques universels qui orientent le comportement des individus et des sociétés. Elles englobent des notions telles que la dignité humaine, la justice, l'égalité, la liberté, la solidarité et la paix. Ces valeurs constituent les fondements du vivre-ensemble pacifique et de la coopération entre les personnes. Elles contribuent à l'édification de sociétés prospères, fondées sur le respect mutuel et la valorisation de la diversité (Schwartz, 1992).

développement durable et de réduction des risques de corruption systémique.

4. Analyser la situation actuelle relative à l'éthique de la conception et du développement des applications d'IA, en tenant compte de la nature du problème, de ses causes, de ses effets, des politiques antérieures, des acteurs impliqués (positivement ou négativement), ainsi que des indicateurs favorables ou préoccupants observés dans le contexte actuel.

Hypothèses de la recherche

La recherche actuelle cherche à examiner les hypothèses suivantes :

1. Il existe une corrélation entre la prise en compte des normes éthiques dans la conception et le développement des applications d'IA, et la réduction des problèmes et défis liés à une utilisation irresponsable de ces technologies.
2. Il existe une corrélation entre l'existence de normes éthiques claires encadrant la conception, le développement et l'utilisation des systèmes d'IA, et la garantie d'un usage responsable et bénéfique de cette technologie pour la société.

Questions de recherche

Cette étude s'attache à répondre aux questions suivantes :

1. Quels sont les principaux aspects de la situation actuelle en matière d'éthique de la conception et du développement des applications d'IA ?
2. Quels sont les concepts éthiques fondamentaux à prendre en compte lors de la conception, du développement et de l'utilisation des systèmes d'IA ?
3. Quels sont les principaux défis éthiques résultant de l'utilisation de l'IA dans divers secteurs ?
4. Quelles sont les normes éthiques essentielles qui devraient encadrer la conception et le développement des applications d'IA?

Méthodologie de la recherche

Cette recherche adopte une approche descriptive et analytique, fondée sur l'examen des études antérieures et des rapports scientifiques consacrés à l'éthique de la conception et du développement des applications d'intelligence artificielle. Elle s'appuie également sur l'analyse de cas concrets d'utilisation de l'IA dans divers contextes.

Définitions des concepts clés

- **Intelligence artificielle (IA)** : Capacité d'une machine à imiter les facultés cognitives humaines, telles que l'apprentissage, la résolution de problèmes, la prise de décision, l'adaptation à de nouvelles situations, ainsi que l'exécution d'autres tâches requérant normalement une intelligence humaine (Naqvi, 2020).
- **Éthique de l'intelligence artificielle** : Ensemble de principes et de valeurs visant à encadrer le développement et l'usage de la technologie de l'IA de manière responsable et éthique. Ces principes incluent notamment la

transparence, la redevabilité, l'équité, le respect de la vie privée et la sécurité, avec pour objectif de garantir que l'IA soit utilisée au bénéfice de l'humanité tout en minimisant les risques (Taddeo & Floridi, 2018).

- **Normes éthiques de conception et de développement des applications d'IA** : Ensemble de principes et de valeurs devant être pris en compte lors de la conception et du développement de ces applications, afin d'en assurer un usage responsable et éthique. Ces normes englobent des enjeux tels que la transparence, la redevabilité, la justice, la protection de la vie privée et la sécurité, et visent à orienter les processus de développement pour garantir que l'IA serve les individus sans leur porter préjudice (Jobin et autres, 2015).

Principaux axes de la recherche

Afin de souligner l'importance d'un usage éthique des applications d'intelligence artificielle (IA) dans les différents aspects de la vie, plusieurs axes de réflexion ont été retenus. Ceux-ci permettent d'analyser les principaux défis associés à leur déploiement, notamment ceux d'ordre éthique. Une attention particulière est également portée à l'expérience de l'État égyptien dans l'élaboration de sa stratégie nationale pour l'IA, en soulignant les modalités de valorisation de cette stratégie dans la mise en œuvre de la stratégie nationale de lutte contre la corruption. L'objectif est d'aboutir à la formulation d'un ensemble de normes éthiques susceptibles d'encadrer la conception et l'utilisation des applications d'IA dans différents secteurs. Les axes sont détaillés comme suit :

I. L'importance de l'usage éthique des applications d'IA

Le domaine de l'éthique de l'intelligence artificielle connaît un développement remarquable, marqué par une orientation vers l'établissement de cadres normatifs plus globaux et flexibles. Parmi les tendances dominantes figurent la promotion d'une IA responsable, la mise en place de mécanismes de contrôle et d'audit, la coopération internationale et l'implication des citoyens (Hagendorff, 2020 ; UNESCO, 2021).

L'expansion de l'usage des applications d'IA aura inévitablement une influence majeure sur les compétences requises sur le marché du travail futur. Les individus devront dès lors développer de nouvelles aptitudes adaptées à l'ère numérique, telles que la pensée critique, la résolution de problèmes, la collaboration, la communication, la créativité, l'innovation, l'adaptabilité et l'apprentissage continu.

L'IA peut contribuer à améliorer les conditions de vie en augmentant la productivité, en créant de nouvelles opportunités professionnelles et en réduisant les inégalités socio-économiques grâce à un accès plus équitable à l'éducation et à la formation (Harari, 2016 ; O'Neil, 2016).

En outre, l'IA peut favoriser une efficacité accrue, entraînant une baisse des coûts et générant ainsi des bénéfices économiques tangibles pour la société, avec un

impact positif sur la qualité de vie des individus. La Commission européenne prévoit que l'intelligence artificielle pourrait se diffuser dans de nombreux secteurs et fonctions, stimulant ainsi la productivité et soutenant une croissance économique durable (Craglia, 2018).

L'éthique de l'intelligence artificielle renvoie à un ensemble de principes et de valeurs guidant le développement et l'utilisation responsables de cette technologie. Ces principes incluent notamment la justice, la transparence, la responsabilité et la sécurité (Bryson & Floridi, 2019 ; Floridi, 2020 ; Hagendorff, 2020 ; UNESCO, 2021 ; Taddeo & Floridi).

Quant aux lois encadrant l'IA, elles constituent un ensemble de règles et de réglementations élaborées par les gouvernements et les organismes de réglementation pour encadrer son usage. Ces lois visent à trouver un équilibre entre la stimulation de l'innovation et la protection de la société contre les risques potentiels que comporte cette technologie (Goodman & Flaxman, 2017).

La distinction fondamentale entre éthique et droit réside dans leur caractère contraignant : l'éthique repose sur des principes directeurs non obligatoires, tandis que le droit s'impose par la force normative. L'éthique se fonde sur des valeurs morales, variables selon les individus et les cultures, et oriente vers ce qui « devrait » être fait dans une optique de bien commun et de justice. Par contre, le droit délimite ce qui « doit » être fait, dans le cadre d'un ordre juridique défini, et s'accompagne de sanctions en cas de non-respect.

Selon Flanagan (2024), cette distinction peut se décliner de la manière suivante :

- Les décisions éthiques s'appuient sur des convictions personnelles et des normes culturelles, leur conférant un caractère subjectif et évolutif, tandis que les lois s'appliquent uniformément, indépendamment des différences idéologiques ou culturelles.
- Les violations éthiques peuvent entraîner des répercussions sociales ou professionnelles (telles que la perte de réputation), sans nécessairement impliquer de conséquences juridiques, contrairement à la transgression des lois, qui expose à des sanctions claires (amendes, incarcération).
- L'éthique est plus réactive face aux évolutions sociales et technologiques, alors que le droit, souvent rigide, nécessite des procédures complexes pour être modifié, d'où un décalage possible avec la réalité.
- L'éthique intervient comme un cadre de référence dans les domaines où la législation est absente ou peu claire, notamment dans les secteurs émergents tels que l'intelligence artificielle ou le génie génétique. Elle permet de combler les vides juridiques et d'orienter les comportements dans des contextes non encore régis par des règles formelles, jouant ainsi un rôle complémentaire au droit dans l'élaboration des normes sociales.

En conséquence, l'éthique et le droit forment ensemble la pierre angulaire de tout cadre régulateur de l'intelligence artificielle. L'éthique définit les principes à

respecter (respect de la vie privée, équité, sécurité des usagers...), tandis que le droit établit les règles contraignantes et les responsabilités légales applicables aux usages de cette technologie.

Les principales notions constitutives d'une éthique de l'IA sont les suivantes (Bostrom, 2014 ; Zuiderveen et autres, 2018 ; O'Neil, 2016 ; Taddeo & Floridi, 2018 ; Zuboff, 2019 ; Al-Garadi et autres, 2020) :

- **Responsabilité et redevabilité** : la responsabilité renvoie à l'attribution des conséquences des décisions prises par les systèmes d'IA à un acteur déterminé — qu'il s'agisse du développeur, de l'utilisateur ou, dans certains cas, du système lui-même, tandis que la redevabilité implique la mise en place de mécanismes permettant de tenir les parties concernées responsables des actions et comportements de ces systèmes.
- **Transparence et explicabilité** : les processus décisionnels de l'IA doivent être compréhensibles, vérifiables et interprétables.
- **Justice et équité** : les systèmes d'IA doivent garantir une égalité de traitement, sans discrimination ni partialité.
- **Vie privée et protection des données** : il convient de respecter la confidentialité des données personnelles et d'en garantir la sécurité.
- **Sécurité et fiabilité** : les systèmes doivent garantir la protection des individus et de la société contre tout dommage potentiel pouvant résulter de l'usage des systèmes d'intelligence artificielle. Cela inclut également la nécessité de sécuriser ces systèmes face aux cyberattaques et aux tentatives de manipulation.

Un schéma récapitulatif illustrera ces principes éthiques de gestion des applications d'IA.

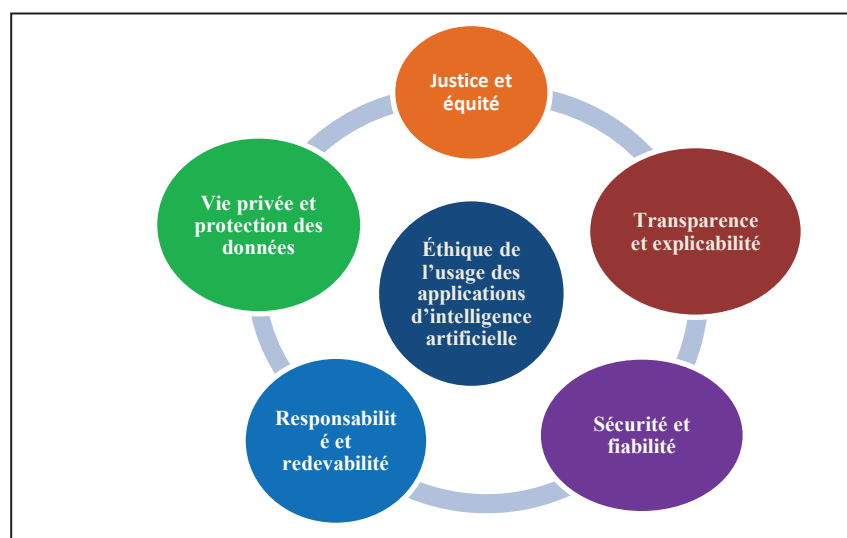


Figure (2) Éthique de l'usage des applications d'intelligence artificielle (Taddeo & Floridi, 2018)

De nombreuses voies permettent de renforcer une utilisation éthique des applications d'intelligence artificielle. Parmi celles-ci figurent : l'élaboration de normes éthiques claires, l'intégration de l'éthique dans la conception et le développement des systèmes, la formation des professionnels du secteur, la sensibilisation des utilisateurs, ainsi que l'encouragement de la recherche scientifique. Prendre en compte la dimension éthique dans la conception et le développement des systèmes d'IA constitue une exigence fondamentale afin de garantir un usage responsable de cette technologie, au service de l'intérêt collectif.

La figure suivante synthétise les principales modalités susceptibles de promouvoir une approche éthique dans le recours aux applications d'IA :

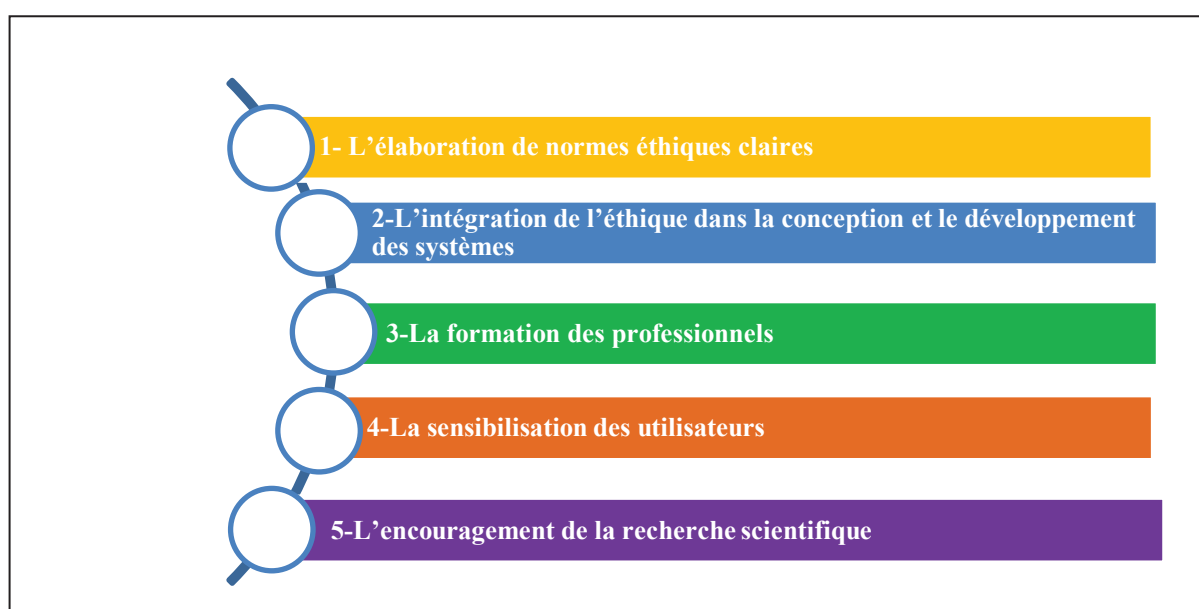


Figure (3) Moyens de renforcement d'un usage éthique des applications d'IA
(Al-Garadi et al., 2020)

Ainsi, l'éthique de l'usage des applications d'IA constitue un élément crucial pour garantir le développement et la mise en œuvre de ces technologies d'une manière qui favorise la justice, assure la sécurité et profite à l'ensemble de la société. Un comportement éthique à l'égard de l'intelligence artificielle peut être renforcé par la diffusion des normes éthiques, tout en tenant les utilisateurs pleinement responsables des informations qu'ils extraient des outils d'IA et de la nécessité de vérifier leur fiabilité (McKendrick, 2022).

II- Les principaux défis éthiques liés à l'utilisation des applications d'IA

L'intelligence artificielle peut jouer un rôle ambivalent dans la propagation des fausses nouvelles et de la désinformation. D'une part, elle peut être exploitée pour produire et diffuser des fausses informations de manière rapide et efficace, rendant leur détection et leur vérification plus difficiles. D'autre part, elle peut également servir au développement d'outils et de technologies permettant

d'identifier et de combattre les fausses nouvelles (Marwick & Lewis, 2017 ; Wardle & Derakhshan, 2017).

Parmi les défis éthiques les plus importants auxquels les médias sont confrontés à l'ère de l'intelligence artificielle figurent les techniques de falsification, qui permettent de créer des vidéos ou des images truquées d'un réalisme troublant. Ces contenus peuvent être utilisés à des fins de désinformation ou pour nuire à la réputation d'individus ou d'institutions. Les techniques de «deepfake» suscitent de nombreuses préoccupations éthiques, notamment en ce qui concerne l'atteinte à la vie privée, la manipulation médiatique et la diffamation.

L'usage des applications d'IA soulève de nombreux défis éthiques, parmi lesquels figurent (Mittelstadt et autres, 2016 ; Obermeyer et autres, 2019 ; Rahwan et autres, 2019 ; Brundage et autres, 2020) :

- **Les biais** : les systèmes d'intelligence artificielle peuvent intégrer des biais présents dans les données utilisées pour leur apprentissage, ce qui engendre des résultats discriminatoires ou injustes.
- **La détermination et l'attribution de la responsabilité** : se pose la question de savoir qui est responsable des dommages causés par les systèmes d'intelligence artificielle.
- **La transparence** : certains systèmes prennent des décisions à travers des processus si complexes qu'ils deviennent difficiles à comprendre ou à expliquer.
- **La violation de la vie privée** : la collecte et l'analyse massive de données personnelles soulèvent des inquiétudes majeures en matière de respect de la vie privée.

Le schéma suivant résume les principaux défis éthiques liés à l'usage des applications d'IA :

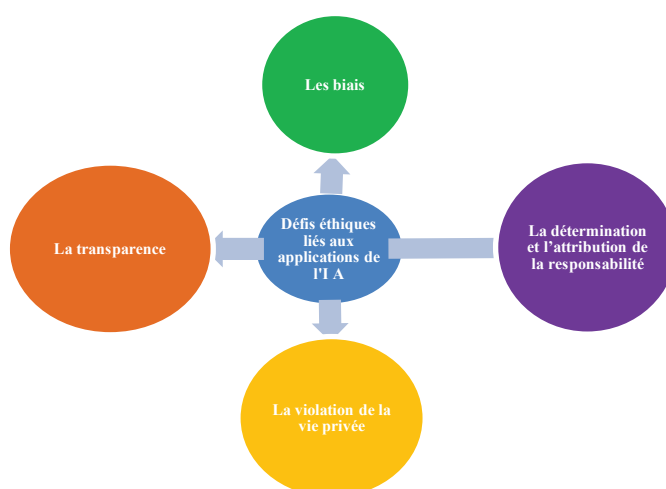


Figure (4) Les défis éthiques liés à l'usage des applications d'IA (Brundage et autres, 2020)

La question de la responsabilité en cas de dommages causés par l'utilisation des applications d'intelligence artificielle demeure complexe. Il est en effet difficile, dans de nombreux cas, d'identifier de manière certaine l'entité responsable, dans la mesure où l'IA ne constitue ni une personne physique ni une personne morale, mais un système technique fondé sur des algorithmes et des données (Bostrom, 2014).

Plusieurs parties peuvent potentiellement être tenues pour responsables des erreurs de l'intelligence artificielle, notamment les développeurs, les fabricants, les exploitants et les propriétaires. Face à la complexité de la détermination des responsabilités, il est nécessaire d'élaborer des cadres juridiques clairs et efficaces encadrant l'usage de cette technologie. Ces cadres devraient définir explicitement les responsabilités, inclure des mécanismes de réparation des dommages causés par les erreurs de l'IA, et aborder d'autres enjeux cruciaux comme la protection des données, la transparence et la redevabilité.

Certains États ont déjà commencé à élaborer des modèles de responsabilité éthique et juridique dans le domaine de l'intelligence artificielle, parmi lesquels l'Union européenne, les États-Unis et Singapour.

Par ailleurs, les mécanismes de prise de décision des systèmes d'intelligence artificielle nécessitent un haut degré de transparence. Les utilisateurs comme les concepteurs doivent être en mesure de comprendre le fonctionnement de ces systèmes et la manière dont ils parviennent aux résultats fournis. La transparence est essentielle pour instaurer la confiance, identifier les erreurs et les biais, et améliorer les performances (Doshi-Velez & Kim, 2017).

Les biais dans la conception des applications d'IA peuvent avoir des conséquences graves, tant pour les individus que pour les sociétés. Lorsqu'un système intelligent est programmé à partir de données déséquilibrées ou biaisées, ces biais peuvent se refléter dans les décisions prises par l'organisation, entraînant des discriminations à l'encontre de certaines catégories, telles que les minorités ethniques ou sociales. Cette problématique peut accentuer les inégalités et les discriminations dans des domaines sensibles comme l'emploi, la santé ou la justice pénale, où des décisions injustes pourraient être prises sur la base de modèles de données subjectifs. Il est donc impératif de faire face à ce problème en développant des applications plus équitables et en améliorant la qualité des données utilisées, afin de garantir un usage responsable et juste de l'intelligence artificielle.

L'équité dans l'accès aux applications d'IA et leur utilisation représente également un enjeu fondamental. Tous les individus, quels que soient leurs antécédents ou leurs capacités, doivent pouvoir bénéficier de cette technologie. L'équité implique ici de garantir un accès universel et un usage équitable de ces outils (Goodman & Flaxman, 2017).

Un autre défi majeur auquel les sociétés sont confrontées à l'ère numérique est la recherche d'un équilibre entre sécurité et libertés individuelles. D'un côté, il est nécessaire de protéger les personnes et les communautés contre les risques sécuritaires, mais d'un autre côté, il convient de respecter les droits et libertés des individus. Pour atteindre cet équilibre, il faut adopter des lois et règlements clairs, notamment en matière de transparence, de redevabilité, de contrôle et de développement.

Les questions éthiques liées à la collecte et à l'utilisation des données dans les systèmes d'intelligence artificielle doivent également être soigneusement prises en considération. Les données doivent être collectées de manière transparente et responsable, et utilisées exclusivement aux fins pour lesquelles elles ont été initialement recueillies. L'absence de respect des principes éthiques dans la collecte et l'usage des données peut engendrer de nombreux risques, notamment en matière de violation de la vie privée. Par exemple, la surveillance constante des individus sans leur consentement peut restreindre leur liberté personnelle et modifier leur comportement sous l'effet d'un sentiment de surveillance permanent. D'autres risques incluent le vol d'identité ou de données personnelles, pouvant donner lieu à une utilisation illégale des informations dans des activités de fraude financière (Zuboff, 2019).

L'impact de l'IA sur le travail suscite un large débat. Certains estiment qu'elle entraînera une suppression massive d'emplois, tandis que d'autres y voient une opportunité de création de nouveaux postes. Il est probable que l'intelligence artificielle numérise certaines tâches, en particulier les tâches routinières et répétitives. Parallèlement, elle devrait également générer de nouvelles opportunités d'emploi, notamment dans les domaines du développement, de la conception, de la maintenance des systèmes d'IA, ainsi que dans l'analyse de données et d'autres secteurs fondés sur les technologies avancées (Agrawal et autres, 2018 ; Brynjolfsson & McAfee, 2014 ; Ford, 2015).

III- La Stratégie nationale égyptienne pour l'intelligence artificielle

La question de la prise de décision éthique à travers l'usage des applications d'intelligence artificielle suscite un débat intense dans les milieux scientifiques et intellectuels. Certains estiment que l'IA, de par sa nature technologique, reste limitée au cadre logique et aux données sur lesquelles repose l'analyse (Data analysis), et qu'elle manque de la capacité à saisir la complexité et la variabilité des valeurs éthiques. D'autres considèrent, au contraire, qu'avec le développement de ses capacités de calcul et d'analyse, l'IA pourrait, dans certaines situations, prendre des décisions éthiques plus pertinentes que celles des humains, notamment lorsqu'il s'agit de traiter d'énormes volumes de données et d'agir rapidement.

La programmation des valeurs humaines représente ainsi l'un des défis majeurs auxquels sont confrontées les applications d'IA. En effet, les valeurs éthiques sont

souvent implicites, évolutives, et influencées par une pluralité de facteurs culturels, sociaux et personnels. Leur traduction en règles numériques programmables s'avère particulièrement complexe, au risque de simplifier à l'excès ou de déformer les valeurs humaines authentiques.

L'intérêt de questionner la capacité de l'intelligence artificielle à prendre des décisions éthiques se fait particulièrement sentir dans des domaines sensibles tels que la médecine, la justice et la sécurité, où les décisions prises par ces technologies peuvent avoir un impact profond sur les individus et sur la société. Dans le secteur médical, par exemple, l'IA peut être amenée à arbitrer l'allocation de ressources médicales limitées ou à recommander des traitements dans des cas complexes. Dans le domaine judiciaire, elle peut contribuer à l'analyse des preuves et à la formulation de recommandations à l'intention des juges. Dans le champ sécuritaire, l'IA peut servir à la surveillance des frontières et à l'identification des menaces potentielles.

C'est dans ce contexte que l'Égypte a lancé sa stratégie nationale pour l'intelligence artificielle, avec pour objectif de tirer parti de cette technologie avancée au service du développement durable dans divers secteurs. Cette stratégie vise à renforcer les capacités nationales dans le domaine de l'IA, à encourager la recherche et l'innovation, et à construire l'infrastructure nécessaire à la mise en œuvre efficace des applications d'IA. Elle entend également promouvoir la coopération internationale et l'échange d'expertise avec les pays les plus avancés en la matière (Ministère des Communications et des Technologies de l'information, 2021).

La Stratégie nationale repose sur plusieurs axes prioritaires, parmi lesquels : le développement des compétences et expertises en intelligence artificielle, la stimulation de l'investissement dans les entreprises émergentes spécialisées dans ce domaine, ainsi que le déploiement de l'infrastructure numérique indispensable au soutien des projets liés à l'IA. Elle met en outre l'accent sur la nécessité de renforcer la confiance dans les usages de l'IA, à travers la garantie de la transparence, la redevabilité et le respect des normes éthiques tout au long des processus de développement et de déploiement technologique (Agence centrale pour la mobilisation publique et les statistiques, 2022).

Parmi les objectifs stratégiques de cette stratégie figurent également : l'amélioration de la qualité des services publics, la stimulation de l'innovation dans les secteurs économiques, la modernisation de l'enseignement et du système de santé, ainsi que le renforcement de la cybersécurité. À travers cette stratégie, l'Égypte ambitionne de devenir un pôle régional de référence en matière d'intelligence artificielle et de contribuer activement au développement de cette technologie à l'échelle mondiale (Conseil national pour l'intelligence artificielle, 2023).

IV- Mécanismes de mobilisation des applications de l'IA dans la lutte contre la corruption

L'Égypte a lancé sa Stratégie nationale de lutte contre la corruption dans le but de promouvoir l'intégrité et la transparence au sein des différentes institutions de l'État, et de combattre toutes les formes de corruption financière et administrative. Cette stratégie met l'accent sur le développement des cadres législatifs et institutionnels de lutte contre la corruption, sur le renforcement du rôle des organes de contrôle, ainsi que sur la consolidation de la coopération internationale dans ce domaine. Elle vise également à sensibiliser la société à l'importance de la lutte contre la corruption et à mobiliser la société civile en tant qu'acteur clé dans ce processus (Autorité de Contrôle Administratif, 2023).

La Stratégie nationale égyptienne de lutte contre la corruption repose sur plusieurs axes fondamentaux, parmi lesquels : le renforcement de la transparence et le développement des mécanismes d'audit et de contrôle financier, la dynamisation du rôle des organes de contrôle dans la détection des crimes de corruption et la poursuite de leurs auteurs, la réforme de la fonction publique, notamment en améliorant les procédures de nomination et de promotion, la mise en œuvre de la redevabilité en cas de violations ou de dépassements.

Parmi les objectifs stratégiques de cette stratégie figurent : la réduction de la prévalence de la corruption dans tous les secteurs, le renforcement de la confiance dans les institutions de l'État, l'amélioration du climat des affaires et l'attractivité des investissements étrangers, ainsi que la réalisation du développement durable.

À travers cette stratégie, l'Égypte aspire à s'imposer comme un pays pionnier dans la lutte contre la corruption, tout en contribuant aux efforts internationaux déployés dans ce domaine.

Par ailleurs, l'IA joue un rôle de plus en plus déterminant dans la lutte contre la corruption en Égypte. Ses outils permettent d'analyser d'immenses volumes de données afin de détecter des schémas suspects ou de prévoir des irrégularités potentielles. À titre d'exemple, des techniques d'apprentissage automatique peuvent être mobilisées pour examiner les données financières et administratives et détecter les pratiques frauduleuses ou les détournements de fonds publics. De même, l'IA peut être utilisée pour l'analyse des contrats publics et le suivi des projets afin de s'assurer de la transparence des processus et de prévenir toute forme de manipulation ou de corruption (Ministère des Communications et des Technologies de l'information, 2023).

Les outils d'intelligence artificielle exploitables dans la lutte contre la corruption sont nombreux et diversifiés. Certains se basent sur l'analyse des mégadonnées (*big data*) pour détecter les anomalies ; d'autres utilisent des technologies de traitement automatique du langage naturel (TALN) pour analyser des textes et repérer des messages suspects. Par ailleurs, des outils de vision par ordinateur

peuvent servir à l'analyse d'images ou de vidéos en vue d'identifier des falsifications ou des manipulations visuelles.

De plus, l'IA permet le développement de systèmes d'alerte précoce capables de détecter les risques de corruption à un stade anticipé, facilitant ainsi la mise en place de mesures préventives. À cet égard, l'Autorité de Contrôle Administratif a pu, grâce à l'analyse par IA de ses bases de données, démanteler un réseau criminel spécialisé dans la falsification de documents officiels. Ces documents permettaient frauduleusement à certains citoyens de bénéficier des avantages des programmes sociaux réservés aux personnes en situation de handicap, notamment l'obtention de pensions exceptionnelles ou l'importation de véhicules sans paiement de droits de douane (Autorité de Contrôle Administratif, 2023).

Les outils d'IA contribuent à la lutte contre la corruption en Égypte de plusieurs manières :

- en améliorant l'efficacité des organes de contrôle dans la détection des irrégularités,
- en réduisant les risques de corruption par la prévision proactive des comportements suspects,
- et en renforçant la redevabilité et la lutte contre l'impunité, notamment par la traçabilité des fonds publics et la récupération des avoirs détournés.

L'Égypte s'efforce de consolider l'usage de l'intelligence artificielle dans la lutte contre la corruption en développant les compétences locales dans ce domaine et en favorisant la coopération entre les secteurs public et privé dans le développement et le déploiement de ces outils (L'Agence centrale d'Audit, 2021).

V- Le cadre théorique de la recherche

L'un des cadres théoriques pouvant fonder cette recherche est celui de la théorie de la gouvernance éthique de l'IA (Ethical AI Governance Theory). Cette théorie postule que le développement et le déploiement des systèmes d'intelligence artificielle doivent s'inscrire dans un cadre global de gouvernance intégrant des principes éthiques tels que la justice, la transparence et la redevabilité. Selon cette approche, les institutions et les autorités de régulation ont la responsabilité d'élaborer des politiques rigoureuses afin d'assurer la conformité des systèmes intelligents aux valeurs humaines et sociétales (Floridi et autres, 2018).

La théorie souligne également la nécessité de développer des systèmes d'intelligence artificielle explicables, tout en évitant les biais susceptibles d'influencer les résultats qu'ils produisent (Jobin, Ienca & Vayena, 2019).

Elle soutient en outre le principe de « l'éthique par la conception » (*Ethics by Design*), qui consiste à intégrer la réflexion éthique à toutes les étapes du cycle de vie des systèmes d'intelligence artificielle – depuis la collecte des données jusqu'à leur mise en œuvre concrète (Mittelstadt, 2019).

Les recherches démontrent que la mise en œuvre de cette théorie contribue à renforcer la confiance du public dans les technologies d'IA et à réduire les risques

sociaux liés à leur utilisation, notamment en ce qui concerne les atteintes à la vie privée ou les décisions injustes à l'encontre de certains groupes sociaux (Morley et autres, 2020).

Par ailleurs, cette théorie est étroitement liée aux lignes directrices élaborées par des organismes internationaux tels que l'UNESCO ou la Commission européenne, visant à définir des normes juridiques et éthiques garantissant un usage respectueux des droits humains et des valeurs collectives. À ce titre, la théorie de la gouvernance éthique de l'intelligence artificielle constitue un cadre conceptuel pertinent pour comprendre les normes éthiques relatives à la conception et au développement des applications de l'IA à l'ère numérique (Commission européenne, 2020).

VI- Analyse de la situation actuelle en matière d'éthique des systèmes d'IA

La transparence constitue l'un des piliers fondamentaux des normes éthiques encadrant le développement et les usages de l'intelligence artificielle. Néanmoins, sa mise en œuvre en Égypte se heurte à des obstacles notables, en particulier le manque de divulgation des algorithmes employés dans les différents secteurs. Cette absence de transparence peut engendrer un déficit de redevabilité et compromettre la qualité des processus décisionnels. Selon les analyses de Zuboff (2019) et de Brey & Dainow (2023), la consolidation de la transparence requiert l'adoption de politiques explicites contraignant les développeurs d'applications à révéler les mécanismes de fonctionnement de leurs systèmes intelligents ainsi que les effets qu'ils peuvent produire sur les utilisateurs — ce qui demeure, à ce jour, en phase de structuration dans l'environnement numérique égyptien.

En ce qui concerne le principe de justice, des inquiétudes croissantes émergent à propos des biais des systèmes d'IA, qui peuvent influencer les décisions dans des domaines sensibles tels que l'emploi, l'éducation ou les services publics. Les travaux d'Anderson (2007) et de Bostrom (2014) indiquent que ces biais trouvent souvent leur origine dans des ensembles de données qui ne représentent pas équitablement toutes les catégories sociales, ce qui engendre des résultats déséquilibrés. En Égypte, les efforts pour garantir l'équité des données utilisées dans l'apprentissage des modèles intelligents demeurent limités, ce qui appelle à des interventions juridiques et techniques pour encadrer les normes de développement et assurer une utilisation équitable des applications.

S'agissant de la vie privée et de la cybersécurité, la dépendance croissante aux applications d'intelligence artificielle en Égypte suscite des préoccupations quant à la protection des données personnelles et à leur utilisation responsable. Selon Cath (2018), ainsi que Brey et Dainow (2023), renforcer la sécurité des données exige l'adoption de lois claires et de procédures efficaces pour protéger l'information contre les violations ou les usages abusifs. Bien que l'Égypte ait promulgué des lois relatives à la protection des données personnelles, leur mise en

œuvre effective reste confrontée à des difficultés, notamment en raison du développement rapide des technologies fondées sur l'analyse des mégadonnées.

VII- Élaboration d'une liste de normes éthiques pour la conception et le développement des applications d'IA

Les normes éthiques encadrant la conception, le développement et la production des applications d'IA revêtent aujourd'hui une importance capitale, dans la mesure où cette technologie s'est intégrée de manière profonde et durable dans notre quotidien. Ces applications doivent être conçues de manière à garantir le respect des droits humains et des libertés fondamentales, en évitant toute forme de discrimination ou de partialité à l'encontre d'un groupe social donné. Par ailleurs, elles doivent faire preuve de transparence et être soumises à des mécanismes de redevabilité, de sorte que les utilisateurs soient informés du fonctionnement des systèmes et des modalités de prise de décision qui y sont associées (Anderson & Anderson, 2007).

Les normes éthiques applicables à la conception des applications d'IA couvrent un large éventail d'enjeux, notamment la vie privée, la sécurité, la responsabilité et la transparence. Les systèmes doivent être conçus de manière à protéger les données personnelles et la vie privée des utilisateurs. De plus, une répartition claire des responsabilités doit être établie afin de déterminer qui est imputable en cas de dommages résultant de l'utilisation de ces applications (Bostrom, 2014).

Il est également impératif que ces applications soient élaborées sans introduire de discrimination ni de biais à l'encontre de quelque groupe que ce soit. Elles doivent garantir équité et justice, en évitant toute forme d'injustice ou de traitement inégal. À cet effet, les systèmes doivent être rigoureusement testés afin de s'assurer qu'ils ne contiennent ni préjugés ni erreurs susceptibles de générer une discrimination (O'Neil, 2016).

La responsabilité sociale constitue en outre l'une des pierres angulaires des normes éthiques de conception des applications d'IA. Les concepteurs doivent prendre en compte les répercussions sociales potentielles de leurs systèmes et être prêts à assumer la responsabilité des éventuels préjudices découlant de leur utilisation (Russell & Norvig, 2021).

Voici les étapes proposées pour intégrer les principes éthiques dans le processus de conception et de production des applications d'IA, selon Brey et Dainow (2023) :

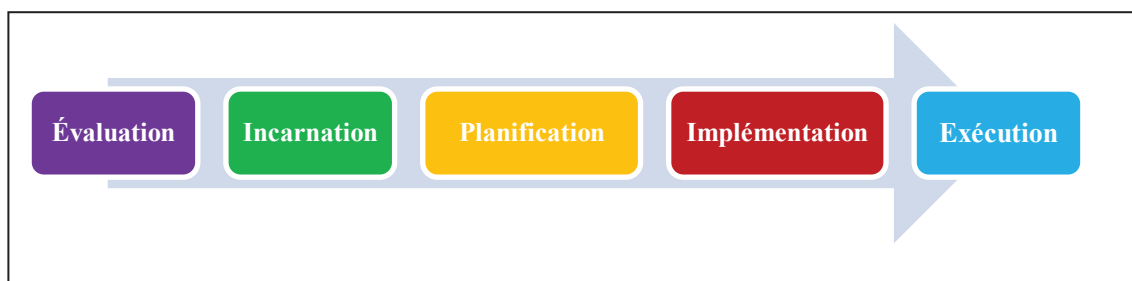


Figure (5) Étapes d'intégration des normes éthiques dans la conception et la production des applications d'IA (selon Brey & Dainow, 2023)

- **Évaluation** : la première étape consiste à évaluer les objectifs du système vis-à-vis des valeurs éthiques fondamentales, afin d'éviter toute atteinte à celles-ci.
- **Incarnation** : les valeurs éthiques sont traduites en indicateurs concrets, qui doivent être présents dans le système d'intelligence artificielle concerné.
- **Planification** : cette phase vise à convertir les exigences générales en matière de conception éthique en actions spécifiques à mettre en œuvre durant la conception. Cela inclut des fonctions du système, des structures de données ou encore des dispositifs organisationnels.
- **Implémentation** : il s'agit de déterminer les modalités pratiques d'intégration de chaque exigence éthique. Cette phase permet de créer un cadre générique de développement facilitant l'intégration des normes éthiques aux différentes étapes de la conception et de l'exécution. Dans certains cas, cette phase doit être répétée.
- **Exécution** : cette dernière étape constitue l'application effective du principe de conception éthique. Bien qu'elle paraisse simple dans sa formulation, elle représente la partie la plus conséquente du travail de développement.

Certaines exigences peuvent nécessiter l'ajout de nouvelles fonctionnalités, d'autres agissent comme des contraintes limitant les fonctions autorisées. De nombreuses exigences éthiques requièrent aussi des procédures organisationnelles supplémentaires. Par exemple, afin d'éviter les biais algorithmiques, une évaluation formelle des données doit être effectuée en amont de leur utilisation.

Étapes d'élaboration de la liste des normes éthiques pour la conception et le développement des applications d'IA

L'élaboration de la liste des normes éthiques encadrant la conception et le développement des applications d'IA a suivi les étapes suivantes :

a. Définition de l'objectif de la liste :

Il s'agit de construire une liste rigoureuse de normes éthiques destinée aux programmeurs et développeurs d'applications d'intelligence artificielle, dans une optique de bénéfice partagé avec les utilisateurs finaux.

b. Dérivation des normes et de leurs indicateurs :

Une première version de la liste a été élaborée à partir de la littérature existante, notamment les travaux de Jobin et autres (2019) ainsi que Brey & Dainow (2023).

c. Vérification de la validité de la liste :

La version préliminaire de la liste (voir Annexe 1) a été soumise, sous forme de questionnaire (voir Annexe 2), à un panel d'experts en informatique et en technologies numériques. Ces évaluateurs ont été invités à se prononcer sur :

- la pertinence des domaines principaux de la liste des normes;
- la capacité des normes à répondre aux objectifs visés ;
- la qualité linguistique et terminologique des formulations ;
- les suggestions d'ajouts, de suppressions ou de reformulations.

d. Vérification de la fiabilité de la liste des normes:

La liste a été révisée à la lumière des remarques des experts. Le degré de consensus a été mesuré à l'aide de la formule de Cooper :

Taux d'accord = nombre d'accords / (nombre d'accords + nombre de désaccords) × 100

Les normes pour lesquelles le taux d'accord atteignait au moins 85 % ont été retenues. Celles en deçà ont été écartées. Plusieurs indicateurs ont été reformulés ou supprimés selon les observations formulées (voir Annexe 3).

e. Version finale de la liste :

Après intégration de l'ensemble des modifications proposées, la version définitive de la liste comprend **15 normes éthiques** et **73 indicateurs** (voir Annexe 4).

Résultats de la recherche

La recherche visait à tester un ensemble d'hypothèses relatives aux normes éthiques encadrant le développement des applications d'IA. Les résultats sont les suivants :

Résultats relatifs à la première hypothèse de la recherche :

Cette hypothèse postule l'existence de normes éthiques fondamentales devant être prises en compte lors du développement d'applications d'IA. L'analyse des études antérieures a permis de confirmer cette hypothèse, en identifiant 15 normes éthiques principales, parmi lesquelles figurent la responsabilité, la redevabilité, la transparence, la justice, la protection de la vie privée et la durabilité environnementale.

Les résultats indiquent que le respect de ces normes contribue significativement à l'amélioration de la fiabilité et de la performance des applications. En effet, la conformité aux principes de responsabilité, redevabilité et transparence renforce la confiance des utilisateurs, tandis que l'attention portée à la justice et à la vie privée permet de réduire les risques juridiques et éthiques.

Résultats de la vérification de la seconde hypothèse de la recherche :

La seconde hypothèse de la recherche souligne la nécessité de conduire davantage d'études afin de mieux comprendre les effets des normes éthiques sur les développeurs et les utilisateurs. L'analyse théorique des résultats des recherches et études pertinentes a confirmé la validité de cette hypothèse, tout en insistant sur l'impératif d'approfondir l'étude des effets que peut produire la prise en compte des normes éthiques sur l'expérience des utilisateurs, ainsi que des modalités de réponse des développeurs aux exigences de conformité à ces normes.

Réponses aux questions de recherche :

- En réponse à la première question portant sur l'identification des principales normes éthiques devant être respectées lors du développement des applications d'IA, la recherche a permis d'établir une liste de quinze (15) normes, parmi lesquelles figurent notamment : la responsabilité, la redevabilité, la transparence, l'équité, la protection de la vie privée, et la durabilité environnementale.
- Concernant la manière d'assurer le respect de ces normes par les applications d'IA, l'étude analytique des travaux antérieurs a montré qu'il est indispensable d'élaborer des politiques claires encadrant le processus de développement, de mettre en œuvre des tests périodiques permettant de vérifier la conformité des applications aux normes éthiques, et d'impliquer les utilisateurs ainsi que les experts dans l'évaluation de ces applications.
- Quant à l'identification des conséquences du respect ou du non-respect de ces normes, les résultats ont révélé que la conformité aux normes permet d'améliorer la fiabilité des applications et de renforcer la confiance des utilisateurs. En revanche, leur négligence peut engendrer des problèmes d'ordre juridique, éthique, voire commercial, tels que la perte de la confiance des clients ou l'apparition de biais dans les décisions prises par les systèmes d'IA.

Les résultats de cette recherche reflètent l'orientation mondiale vers un développement responsable de l'intelligence artificielle. Les gouvernements, les institutions académiques et les entreprises technologiques s'emploient en effet à élaborer des normes éthiques pour encadrer ce domaine. L'accent mis sur la transparence et la justice s'explique par les inquiétudes croissantes liées aux biais et aux impacts inévitables de certaines applications, tandis que l'intégration de la durabilité environnementale traduit une prise de conscience accrue des effets de la technologie sur l'environnement. La recherche souligne que le respect de ces normes ne saurait se limiter à des considérations théoriques : il doit être traduit en politiques claires et en procédures techniques tout au long des phases de conception et de développement. Cela nécessite une collaboration étroite entre développeurs, utilisateurs et décideurs afin de garantir un équilibre entre innovation et responsabilité.

Principales recommandations et conclusions :

À la lumière de l'analyse des études antérieures et de l'identification des normes éthiques applicables à la conception et à l'utilisation des applications d'intelligence artificielle, les recommandations suivantes peuvent être formulées :

- Les organisations internationales et les gouvernements doivent œuvrer à l'élaboration de normes éthiques claires et globales encadrant la conception, le développement et l'utilisation des systèmes d'intelligence artificielle, de manière à garantir l'équité et l'intégrité de ces technologies. Il est essentiel que ces systèmes soient transparents et explicables, afin de permettre une compréhension des mécanismes à l'œuvre et de favoriser une prise de décision éclairée quant à leur usage. Des mécanismes clairs de redevabilité doivent également être mis en place afin de répondre à tout dommage ou conséquence négative pouvant en découler.
- Des mesures strictes doivent être adoptées pour protéger les données personnelles et prévenir toute forme d'exploitation ou de mésusage de celles-ci dans le cadre des applications d'intelligence artificielle. Cela passe par l'adoption de politiques de confidentialité rigoureuses et de mécanismes de contrôle efficaces. Pour promouvoir un usage éthique et responsable de ces technologies, il est indispensable de renforcer la coopération internationale entre gouvernements, organisations internationales et société civile, en vue de formuler des politiques unifiées orientant l'IA vers l'intérêt général.
- Il convient d'impliquer le grand public dans les discussions relatives à l'éthique de l'IA, notamment à travers l'organisation de débats ouverts et d'activités de sensibilisation. Cette démarche vise à renforcer la conscience collective autour des enjeux liés à ces technologies et à en garantir un usage responsable et durable, bénéfique pour l'ensemble de la société.
- Enfin, il est nécessaire de mener davantage de recherches afin d'explorer les normes éthiques du point de vue des développeurs et des utilisateurs, ce qui permettrait de mieux comprendre les défis qui leur sont associés. Il importe également de proposer des solutions pratiques pour mettre en œuvre de manière effective chaque principe éthique dans les systèmes d'intelligence artificielle, dans le but de réduire l'écart entre les principes théoriques et leur application concrète, et de favoriser le développement de systèmes intelligents responsables.

Bibliographie

I. Références arabes :

- **Ahmed Al-Arabi** (2021), L'intelligence artificielle : la prochaine révolution technologique, Muasast Al-Fekr Al-Araby.

- **Agence centrale pour la mobilisation publique et les statistiques** (2022), Stratégie nationale de l'intelligence artificielle en Égypte : vision prospective, Le Caire : Agence centrale pour la mobilisation publique et les statistiques.
- **Agence centrale d'audit** (2021), Rapport sur le rôle de l'intelligence artificielle dans la lutte contre la corruption, Le Caire : Agence centrale d'audit.
- **Conseil national pour l'intelligence artificielle** (2023), Rapport sur les progrès réalisés dans la mise en œuvre de la stratégie nationale pour l'intelligence artificielle, Le Caire : Conseil national pour l'intelligence artificielle.
- **Autorité de Contrôle Administratif** (2023), Stratégie nationale de lutte contre la corruption, Le Caire : Autorité de Contrôle Administratif.
- **Autorité de Contrôle Administratif** (2023), L'Autorité de contrôle administratif démantèle un réseau criminel spécialisé dans la falsification de documents officiels. <https://aca.gov.eg/News/5217.aspx>
- **Ministère des Communications et des Technologies de l'information** (2021), Stratégie nationale de l'intelligence artificielle en Égypte, Le Caire : Ministère des Communications et des Technologies de l'information.
- **Ministère des Communications et des Technologies de l'information** (2023), Stratégie nationale de l'intelligence artificielle, Le Caire : Ministère des Communications et des Technologies de l'information.

II- Références étrangères :

- Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction Machines: The Simple Economics of Artificial Intelligence. Harvard Business Review Press .
- Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I & Guizani, M. (2020). A survey of machine and deep learning methods for internet of things (IoT) security. IEEE Communications Surveys & Tutorials, 22(3), 1646-1685.
- Anderson, M. (2007). Machine ethics: Creating an ethical intelligent agent. AI & Society, 22(4), 477-493.
- Anderson, M., & Anderson, S. (2007). Machine ethics: New directions. Oxford University Press.
- Berendt, B (2019) AI for the common good?! Pitfalls, challenges, and ethics pen-testing. Paladyn, Journal of Behavioral Robotics 10(1): 44–65. DOI: 10.1515/pjbr-2019-0004.
- Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.
- Brey, P., & Dainow, B. (2023). Ethics by design for artificial intelligence. AI and Ethics, 1-13.
- Brundage, L., Avin, S., Clark, J., Toner, M., Eckersley, P., Garfinkel, S., & Anderson, R. (2020). Toward trustworthy AI development: Aligning theoretical and practical perspectives. arXiv preprint arXiv:2004.07218.
- Brynjolfsson, E & McAfee, A. (2014). The Second Machine Age: Work Progress and Prosperity in a Time of Brilliant Technologies. W. W. Norton & Company .

- Bryson, J. (2018). The artificial intelligence of the ethics of artificial intelligence: An introductory overview for policymakers. *Cambridge Handbook of Artificial Intelligence*, 316-334.
- Bryson, J. J., Taddeo, M & Floridi, L. (2019). Ethics of artificial intelligence. *The Oxford handbook of digital ethics*, 109-139.
- Cath, C. (2018). Governing artificial intelligence: Ethical, legal, and societal issues. *Philosophy & Technology*, 31(4), 611-627.
- Cath, C; Wachter, S; Mittelstadt, B. (2018) Artificial intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics* 24(2): 505–528. DOI: 10.1007/s11948-017-9901-7.
- Craglia M, Annoni, A & Benczur. P. (2018). Artificial intelligence : a European perspective. Publications Office of the European Union, Luxembourg
- Doshi-Velez, F & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
- European Commission. (2020). White paper on artificial intelligence: A European approach to excellence and trust. Publications Office of the European Union.
- Flanagan, W. (2024). Ethics vs Laws: 5 Key Differences Experts Want You to Know in 2024. Available at: <https://www.infonetica.net/articles/ethics-vs-laws>
- Floridi, L. (2019). The ethics of artificial intelligence: Principles, challenges, and opportunities. *Minds and Machines*, 29(4), 569-589.
- Floridi, L. (2020). The ethics of artificial intelligence. Oxford University Press.
- Floridi, L; Cowls, J; Beltrametti, M. (2018). AI4People - an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and machines* 28(4): 689–707.
- Ford, M. (2015). *Rise of the Robots: Technology and the Threat of a Jobless Future*. Basic Books.
- Goodman, B & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, 38(3), 50-57.
- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30, 99-120.
- Harari, Y. (2016). *Homo Deus: A Brief History of Tomorrow*. Harper.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.
- Köbis, N; Starke, C & Rahwan, I. (2021). Artificial intelligence as an anti-corruption tool (AI-ACT). Potentials and pitfalls for top-down and bottom-up approaches.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Machanavajjhala, a; Korolova, A & Sarma, A. (2011) Personalized social recommendations - accurate or private? *Proceedings of the VLDB Endowment (PVLDB)* 4(7).

- Marwick, A. E & Lewis, R. (2017). Media Manipulation and Disinformation Online. Data & Society Research Institute.
- McKendrick, J. (2022). 7 Steps to More Ethical Artificial Intelligence. Available at: <https://www.forbes.com/sites/joemckendrick/2022/06/10/7-steps-to-more-ethical-artificial-intelligence/>
- Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501-507.
- Mittelstadt, B. D., Allo, P., Taddeo, M & Wachter, S. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679.
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods, and research to translate principles into practices. *Science and Engineering Ethics*, 26(4), 2141-2168.
- Naqvi, A. (2020). Artificial intelligence for audit, forensic accounting, and valuation: A strategic perspective. John Wiley & Sons. <https://doi.org/10.1002/9781119601906>
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366 (6464), 447-453 .
- OECD (2021). Recommendation of the council on artificial intelligence.
- O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishing Group.
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C & Rahwan, T. (2019). Machine behavior. *Nature*, 568(7753), 477-486.
- Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 25, 1 pp. 1-65). Academic Press.
- Shin, D. & Park, Y. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior* 98: 277–284.
- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction* 36(6): 495–504.
- Sreenu, D. (2023). How can artificial intelligence accelerate scientific research? <https://scientificnirvana.com/how-can-artificial-intelligence-accelerate-scientific-research>
- Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., & Walsh, T. (2016). Artificial intelligence and life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA.
- Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, 361(6404), 751-752

- UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence.
- Wardle, C & Derakhshan, C .(2017). Information Disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe report, DGI (2017) 9 .
- Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Public Affairs.
- Zuiderveen, F. J., Möller, J., S., Fathaigh, R., Irion, K., Dobber, T, & Hildebrandt, M. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. Utrecht Law Review, 14(2), 69-201791.

Annexe (1) – Version préliminaire de la liste des normes éthiques pour la conception et le développement des applications d'IA

Domaine éthique	Indicateurs	Références
Transparence	– Sensibilisation des utilisateurs au fonctionnement de l'application et à la manière dont les décisions sont prises, de sorte qu'ils soient capables de comprendre les raisons sous-jacentes à ces décisions. – Divulcation claire à toutes les parties prenantes de l'objectif du système, de ses capacités, de ses limites, de ses avantages, de ses risques et des décisions qu'il prend. – Conception des systèmes d'intelligence artificielle de manière à permettre aux individus d'examiner, de contester ou de demander la modification des activités de l'IA. Cela inclut la mise en place de processus organisationnels permettant aux opérateurs de recevoir et d'évaluer les demandes des parties externes. – Mise en œuvre de mesures permettant de tracer et de surveiller le système d'intelligence artificielle afin de garantir la transparence et la responsabilité. – Prise de décisions sur des questions éthiques, telles que la manière d'éliminer les biais d'un ensemble de données. – Documentation des processus et des	(O'Neil, 2016; Floridi, 2019; Brey & Dainow, 2023)

	outils utilisés dans les décisions éthiques, de manière à comprendre comment les normes éthiques ont été respectées, ce qui permet d'effectuer des audits, de contester les décisions du système ou de corriger tout problème éthique survenant après son déploiement.	
Clarté (Explicabilité)	– Possibilité d'expliquer les raisons des différentes décisions prises par le développeur et le système intelligent. – Capacité du système à justifier ses décisions. – Mise à disposition de mécanismes permettant l'interprétabilité et l'explication des décisions et des données utilisées.	(Zuboff, 2019; Brey & Dainow 2023)
Redevabilité	– Mise en place de mécanismes de redevabilité en cas de dommages causés par l'application, avec possibilité de compensation. – Supervision humaine des cycles de fonctionnement et de la prise de décision du système. – Évaluation des risques avant déploiement, et existence de procédures d'atténuation dès la mise en service. – Auditabilité des systèmes d'IA par des entités indépendantes, incluant les décisions et les outils de développement.	Floridi, 2019 ; Shneiderman, 2020 ; Russell & Norvig, 2021
Précision	– Entraînement rigoureux du système pour garantir une précision très élevée. – Réduction des erreurs au minimum, voire leur élimination.	Zuiderveen et al., 2018 ; Naqvi, 2020
Justice	– Conception des applications sans discrimination ni biais envers aucun groupe.	Anderson, 2007 ; O'Neil, 2016 ; Anderson &

	– Évitement des biais algorithmiques, y compris dans les données, les modèles et les algorithmes. – Accessibilité universelle des systèmes d'IA et égalité des fonctionnalités pour tous les utilisateurs, indépendamment de leurs capacités, croyances, préférences ou centres d'intérêt. – Conception visant à éviter les effets sociaux négatifs, en particulier sur les groupes vulnérables.	Berendt, 2019
Vie privée	– Conception des applications garantissant la protection de la vie privée des utilisateurs. – Sécurisation des données personnelles contre les accès non autorisés.	Anderson, 2007 ; Bostrom, 2014 ; Anderson
Autonomie des systèmes	– Formation appropriée des développeurs du système intelligent. – Capacité du système à fonctionner sans supervision humaine.	Cath, 2018 ; Brey & Dainow, 2023
Supervision humaine / agentivité humaine	– Supervision humaine obligatoire pour les décisions sensibles ou critiques. – Conception des systèmes pour qu'ils assistent les humains sans les remplacer. – Ne pas priver les individus de leur pouvoir de décision personnel. – Ne pas réduire les libertés fondamentales (expression, mouvement, information...). – Éviter les conceptions favorisant l'addiction aux services fournis. – Laisser la décision finale sur les questions personnelles, éthiques ou	Obermeyer et al., 2019 ; Brundage et al., 2020 ; Brey & Dainow, 2023

	politiques à l'humain, le système ne pouvant que recommander.	
Sécurité	– Conception des applications garantissant la sécurité des utilisateurs et évitant tout usage nuisible ou contraire à l'éthique. – Respect des droits des personnes concernées lors du traitement des données personnelles. – Mise à disposition de mécanismes assurant la légitimité, l'équité et la transparence du traitement des données personnelles. – Mise en œuvre de mesures pour protéger les droits des titulaires des données par le biais de dispositifs techniques (ex. : anonymisation) et organisationnels (ex. : contrôle d'accès) . – Droit des individus à retirer leur consentement concernant l'utilisation de leurs données. – Collecte, stockage et traitement des données de façon vérifiable par des humains.	Bostrom, 2014 ; Al-Garadi et al., 2020 ; Russell & Norvig, 2021
Fiabilité et cybersécurité	– Protection des systèmes contre les cyberattaques et les manipulations. – Garantie de la continuité de service et de l'absence d'interruptions dues aux menaces. – Maintien de la sécurité des services pour protéger les utilisateurs et les données.	Craglia et al., 2018 ; Brey & Dainow, 20
Responsabilité sociale	– Sensibilisation des développeurs et des producteurs aux impacts sociaux de leurs applications.	Anderson & Anderson, 2007 ; Russell &

	– Prise en compte des effets potentiels des applications sur la société. – Engagement constant à concevoir des applications au service de l'intérêt général.	Norvig, 20
Valeurs humaines	– Respect des droits humains par les systèmes d'intelligence artificielle. – Prise en compte de la diversité et des différences entre individus et communautés. – Garantie de l'autonomie des individus dans leurs décisions sans restrictions injustifiées.	Köbis et al., 2021 ; Brey & Dainow, 20
Bien-être humain, social et environnemental	– Contribution positive des systèmes d'IA aux individus, à la société et à l'environnement. – Assurance que l'utilisation de l'IA renforce les droits humains. – Contribution de l'IA au bien-être général. – Prévention d'usages de l'IA qui saperaient les droits humains et le bien-être.	Brynjolfsson & McAfee, 2014 ; Brey & Dainow, 2023
Durabilité et responsabilité environnementale	– Garantie que les systèmes d'IA sont sûrs à utiliser, sans risque de nuire ou de causer une dégradation importante de la santé physique ou mentale des parties prenantes (utilisateurs, clients, titulaires des données, autres). – Prise en compte des principes de durabilité environnementale dans le développement de l'IA. – Prévention des impacts négatifs de l'IA sur la qualité de la communication, les	Jobin et al., 2019 ; Floridi, 2019 ; Brey & Dainow, 2023 ; Flanagan, 2024

	interactions sociales, le flux d'information, les relations sociales et les processus démocratiques.	
--	--	--

Annexe (2) : Liste des membres du jury

N°	Nom	Université	Faculté	Fonction
1	Inji Abdel-Maboud Bayoumi	Université de Sadate City	Informatique et Intelligence Artificielle	Maîtresse de conférences en informatique
2	Ahmed Abdelhamid Saleh	Université de Damanhour	Informatique et Systèmes d'Information	Maître de conférences en informatique
3	Racha Mohamed Montasser	Université de Damanhour	Informatique et Systèmes d'Information	Maîtresse de conférences en systèmes d'information
4	Amr Fathi El-Sayed	Université de Damanhour	Informatique et Systèmes d'Information	Maître de conférences en informatique
5	Mai Helmi Chahine	Université de Damanhour	Informatique et Systèmes d'Information	Maîtresse de conférences en technologies de l'information
6	Ahmed Hamed Atia	Université de Damanhour	Informatique et Systèmes d'Information	Maître de conférences en informatique
7	Mohamed Abdel-Latif Sayed El-Ahl	Université de Damanhour	Informatique et Systèmes d'Information	Maître de conférences en informatique



Annexe (3) : Liste des modifications proposées par les membres du jury

Premier évaluateur:

- Ajout d'un nouvel indicateur sous le critère de **transparence**.
- Ajout d'un nouvel indicateur sous le critère de **l'autonomie des systèmes**.
- Ajout d'un nouvel indicateur sous le critère de **la supervision humaine**.
- Ajout d'un nouvel indicateur sous le critère de **bien-être humain**.
- Ajout d'un nouveau critère (**adaptabilité**) accompagné de deux indicateurs.

Deuxième évaluateur:

- Reformulation des indicateurs **1, 2, 3 et 5** sous le critère de la **transparence**.
- Ajout d'un nouvel indicateur sous le critère de **redevabilité**.
- Ajout d'un complément à l'indicateur 1 dans le critère de **précision**.
- Ajout d'un nouvel indicateur sous le critère de **précision**.
- Reformulation du cinquième indicateur dans le domaine de la **supervision humaine**.
- Ajout d'un indicateur dans le critère de **responsabilité sociale**.

Troisième évaluateur:

- Pour le critère de précision, il est écrit : « Réduire les erreurs au minimum, de manière quasi inexistante. » Il est suggéré de définir un taux d'erreur acceptable.
- Un possible conflit a été relevé entre deux formulations : celle du critère d'autonomie (« Capacité du système à fonctionner sans supervision humaine») et celle de la supervision humaine (« Garantir la supervision humaine pour les décisions critiques ou professionnelles »). Il est proposé de reformuler le critère d'autonomie comme suit : « Capacité du système à fonctionner sans supervision humaine, sauf pour les décisions critiques ou professionnelles. »
- Pour le critère de l'autonomie humaine, il est écrit : « Concevoir ou utiliser des systèmes d'intelligence artificielle pour créer une dépendance au système ou à ses services. » Il est proposé de reformuler ainsi : « Ne pas concevoir ou utiliser des systèmes d'intelligence artificielle de manière à créer une dépendance au système ou aux services qu'il propose. »

Quatrième évaluateur:

- Reformulation de l'indicateur 1 sous le critère de **transparence**.
- Reformulation de l'indicateur 3 sous le critère de **transparence**.
- Ajout d'un nouvel indicateur sous le critère de **clarté**.
- Ajout d'un nouvel indicateur sous le critère de **précision**.
- Ajout d'un nouvel indicateur sous le critère de **protection de la vie privée**.
- Ajout d'un nouvel indicateur sous le critère de **l'autonomie**.
- Ajout d'un nouvel indicateur sous le critère de **sécurité**.
- Ajout d'un nouvel indicateur sous le critère de **fiabilité**.
- Ajout d'un nouvel indicateur sous le critère de **responsabilité sociale**.
- Ajout d'un nouvel indicateur sous le critère de **valeurs humaines**.
- Ajout d'un nouvel indicateur sous le critère de **bien-être humain**.
- Ajout de deux indicateurs sous le critère de **durabilité et responsabilité environnementale**.

Cinquième évaluateur:

- Ajout d'un nouvel indicateur au critère de **protection de la vie privée**.
- Reformulation d'un des indicateurs du critère de **l'autonomie**.

Sixième évaluateur:

- Reformulation des indicateurs du critère de **redevabilité**.
- Reformulation des indicateurs du critère de **justice**.

Septième évaluateur:

- Reformulation des indicateurs du critère de **sécurité**.
- Reformulation des indicateurs du critère de **durabilité et responsabilité environnementale**.
-

Annexe (4) : Liste finale des normes éthiques pour la conception et le développement des applications d'IA

Domaine éthique	Indicateurs	Références
Transparence	– Sensibiliser les utilisateurs au fonctionnement de l'application intelligente et à son processus décisionnel, afin qu'ils puissent comprendre les services intelligents ainsi que les raisons sous-jacentes aux décisions prises. – Divulguer de manière claire à toutes les parties prenantes l'objectif du système intelligent, ses capacités, ses limites, ses avantages, ses risques, ainsi que les décisions qu'il prend. – Permettre l'audit du système, la possibilité de questionner ses décisions ou de les contester, y compris par des procédures organisationnelles permettant aux utilisateurs d'évaluer les services fournis par le système intelligent. – Mettre en place des mesures permettant de suivre et de contrôler le système d'IA afin de garantir la transparence et la redevabilité. – Prendre des décisions concernant les questions éthiques, comme la manière d'éliminer les biais dans certaines catégories de données. – Documenter les processus et les outils de développement liés aux décisions éthiques, de manière à rendre compréhensible le respect des normes	(O'Neil, 2016; Floridi, 2019; Brey & Dainow, 2023)

	éthiques, ce qui permet l'audit, la contestation des décisions du système, ou la correction de problèmes éthiques après le déploiement. – Informer clairement les utilisateurs sur les types de données collectées et les finalités de cette collecte.	
Clarté (Explicabilité)	– Être en mesure d'expliquer les raisons des décisions prises, que ce soit par le développeur ou par le système intelligent. – Garantir la compréhensibilité des décisions prises par le système d'IA et la capacité à expliquer pourquoi une décision particulière a été prise. – Mettre à disposition un mécanisme permettant d'expliquer les décisions, les données utilisées et d'assurer l'interprétabilité des décisions prises. – Développer des systèmes d'IA explicable (XAI – Explainable AI) afin de garantir que les décisions soient compréhensibles pour tous.	(Zuboff, 2019; Brey & Dainow 2023)
Redevabilité	– Mettre en place des mécanismes de redevabilité en cas de dommages causés par l'utilisation de l'application, avec la possibilité d'obtenir une indemnisation en cas de préjudice. – Garantir une supervision humaine des cycles de fonctionnement et des processus décisionnels des systèmes d'intelligence artificielle. – Intégrer, dans le processus de déploiement du système d'IA, une évaluation des risques, accompagnée de mesures d'atténuation prêtes à être mises en œuvre dès la mise en service du système. – Assurer la possibilité d'audit des systèmes d'intelligence artificielle par des organismes indépendants ; l'audit ne doit pas se limiter aux décisions prises par le système, mais inclure également les processus et les outils utilisés lors du développement. – Suivi continu des systèmes intelligents afin	(Floridi, 2019; Shneiderman, 2020; Russell & Norvig, 2021)

	de détecter d'éventuels risques imprévus pouvant survenir à l'avenir.	
Précision	– Former intensivement le système intelligent afin qu'il fonctionne avec une très grande précision, ce qui nécessite un entraînement sur des mégadonnées multimodales. – Utiliser des techniques d'apprentissage automatique auditables pour garantir l'absence d'erreurs non détectées. – Réduire les erreurs au minimum, avec un taux d'erreur n'excédant pas 1 %. – Employer des indicateurs de performance pour évaluer la précision des modèles utilisés dans les applications d'IA.	(Zuiderveen, et al., 2018; Naqvi, 2020)
Justice	– Concevoir les applications de manière à garantir l'absence de discrimination ou de partialité envers toute catégorie de personnes. – Éviter les biais algorithmiques dans les systèmes d'intelligence artificielle, notamment ceux présents dans les données d'entrée, les modèles et la conception des algorithmes. – Rendre les systèmes d'intelligence artificielle accessibles au niveau mondial, dans la mesure du possible, en assurant les mêmes fonctionnalités et avantages à tous les utilisateurs, indépendamment de leurs capacités, croyances, préférences ou centres d'intérêt. – Concevoir les systèmes d'IA, autant que possible, de manière à éviter les effets sociaux négatifs sur les groupes sociaux, en particulier les groupes protégés	(Anderson, 2007; O'Neil, 2016; Anderson & Berendt, 2019)
Vie privée	– Concevoir les applications de manière à protéger la vie privée des utilisateurs. – Sécuriser les applications pour garantir la protection des données personnelles contre tout accès non autorisé. – Utiliser des techniques telles que le chiffrement et l'apprentissage fédéré (Federated Learning) pour protéger les	(Anderson, 2007; Bostrom, 2014 & Anderson)

	données. – Le droit à l'oubli (Right to be Forgotten) est un principe juridique qui permet aux individus de demander la suppression de leurs données personnelles des bases de données des entreprises ou des fournisseurs de services numériques, lorsqu'il n'existe aucune raison légale de les conserver. Ce principe, issu du RGPD européen, renforce la protection de la vie privée en ligne.	
Autonomie des systèmes	– Former les développeurs du système intelligent de manière appropriée. – Assurer la capacité du système à fonctionner sans supervision humaine, sauf dans les décisions critiques ou professionnelles. – Garantir aux utilisateurs un contrôle total sur leurs données personnelles. – Obliger les développeurs à fournir des outils de supervision permettant une intervention humaine en cas d'urgence	(Cath, 2018; Brey & Dainow, 2023)
Supervision humaine/ agentivité humaine	– Garantir la présence d'une supervision humaine pour toutes les décisions sensibles ou importantes. – Concevoir des systèmes d'IA pour assister les humains dans leur prise de décision, plutôt que de les remplacer totalement. – Établir une liste des décisions sensibles qui doivent rester sous supervision humaine complète. – Ne pas concevoir ni utiliser les systèmes d'IA de façon à empêcher les individus de prendre par eux-mêmes les décisions qu'ils sont censés pouvoir prendre. – Ne pas concevoir ou utiliser l'IA d'une manière qui réduit les libertés fondamentales humaines, y compris la liberté de mouvement, de réunion, d'expression, et d'accès à l'information. – Ne pas concevoir ou utiliser les systèmes d'IA pour induire une dépendance au système ou aux services qu'il fournit. – Les systèmes d'IA ne doivent pas prendre	(Obermeyer et al., 2019; Brundage et al., 2020 & Brey & Dainow, 2023)

	les décisions finales sur des questions importantes de nature personnelle, éthique ou politique, mais peuvent formuler des recommandations. – Réaliser des évaluations régulières pour garantir le respect continu des normes éthiques.	
Sécurité	– Concevoir les applications de manière à garantir la sécurité des utilisateurs et de la société, et à éviter tout usage nuisible ou contraire à l'éthique. – Respecter les droits des personnes concernées lors du traitement des données personnelles. – Mettre en place des mécanismes expliquant comment la légalité, la justice et la transparence sont assurées lors du traitement des données personnelles par l'IA. – Mettre en œuvre des mesures pour protéger les droits des personnes concernées, à travers des solutions techniques (comme l'anonymisation) et des dispositifs organisationnels (tels que les systèmes de contrôle d'accès). – de retirer leur consentement à l'utilisation de leurs données personnelles. – Collecter, stocker et traiter les données de manière à permettre leur audit humain. – Mettre en place des mécanismes de réponse rapide en cas de failles de sécurité détectées.	(Bostrom, 2014; Al-Garadi et al., 2020; Russell & Norvig, 2021)
Fiabilité et cybersécurité	– Protéger les systèmes contre les attaques informatiques et les manipulations. – Assurer la continuité du service et éviter toute interruption due aux menaces. – Maintenir la sécurité du service pour protéger les utilisateurs et les données. – Obliger les entreprises à disposer de plans d'urgence en cas de défaillance du service.	(Craglia et al., 2018; Brey & Dainow, 2023)
Responsabilité	– Sensibiliser les développeurs et producteurs aux impacts sociaux de leurs applications. – Prendre en compte les effets potentiels des	(Anderson, &

sociale	applications sur la société. – S'efforcer constamment de concevoir des applications qui servent le bien commun. – Informer sur les avantages et les risques liés à l'IA. – Mettre en place des mécanismes pour recueillir les retours des utilisateurs et améliorer les systèmes en conséquence.	Anderson 2007; Russell & Norvig, 2021)
Valeurs humaines	– Veiller à ce que les systèmes d'intelligence artificielle respectent les droits humains. – Prendre en considération la diversité et les différences entre individus et communautés. – Garantir l'autonomie des individus dans leurs prises de décisions, sans contraintes injustifiées. – Former les développeurs à concevoir des systèmes respectant les valeurs humaines.	(Köbis et al., 2021; Brey & Dainow, 2023)
Bien-être humain, social et environnemental	– Veiller à ce que les systèmes d'intelligence artificielle bénéficient aux individus, à la société et à l'environnement. – S'assurer que l'usage de l'IA renforce les droits humains. – Contribuer au bien-être général grâce à l'IA. – Empêcher toute utilisation de l'IA qui porterait atteinte aux droits humains et au bien-être. – Concevoir des systèmes accessibles aux personnes en situation de handicap ou aux groupes marginalisés. – Imposer aux entreprises de réaliser des évaluations d'impact environnemental pour leurs systèmes.	(Brynjolfsson & McAfee, 2014; Brey & Dainow, 2023)

Durabilité et responsabilité environnementale	– Garantir que les systèmes d'intelligence artificielle soient sûrs à utiliser et qu'ils n'aient pas tendance à causer des dommages ou une dégradation significative de la santé ou du bien-être physique ou mental de toutes les parties concernées, y compris les utilisateurs, les clients, les propriétaires de données et les autres parties affectées. – Veiller à ce que le développement de l'intelligence artificielle respecte les principes de durabilité environnementale. – Veiller à ce que les systèmes d'IA n'affectent pas négativement la qualité de la communication, les interactions sociales, le flux d'informations, les relations sociales ou les processus démocratiques. – Élaborer des normes garantissant que les systèmes d'IA soient respectueux de l'environnement. – Encourager l'utilisation d'énergies renouvelables dans le fonctionnement des centres de données.	(Jobin et al., 2019; Floridi, 2019; Brey & Dainow, 2023; Flanagan, 2024)
Adaptabilité	– Capacité d'adaptation aux changements. – Garantir la capacité du système à s'adapter à différents contextes sans perdre en efficacité ou en précision.	Salehie, M., & Tahvildari, L. (2009). (TAAS)