

L'IA À L'ÉPREUVE DE L'ÉTHIQUE

Youssef Ben Othmen & Rym Rafrafi



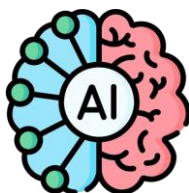
L'IA bouleverse nos sociétés, modifie nos interactions, nos modes de vie, et pose de nouveaux défis majeurs : **comment préserver la vie privée, éviter les biais technologiques, et redéfinir la responsabilité dans un univers où les machines prennent des décisions autonomes ?**

Ces interrogations soulèvent des **débats philosophiques et éthiques** sur la nature de l'intelligence, de la conscience et de la moralité.

Cet exposé, articulé autour des « pros and cons » de la question, **YBO et RR jouerons le jeu et rapporterons des idées qui sont parfois les nôtres ou celles des autres, donc qui ne reflètent pas nécessairement nos positions personnelles.**

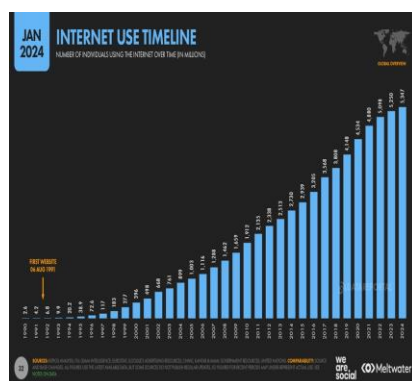
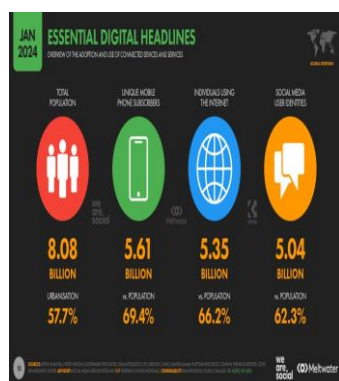
Notre but est de **mettre en lumière les enjeux éthiques et philosophiques** posés par l'IA et d'explorer les risques comme les bénéfices pour notre avenir collectif.

Liens d'intérêt, déclarations et remerciements



Pas de LI

Remerciements Dr BANI



Expansion du numérique et IA dans plusieurs domaines dont la santé



Enthoven le glorieux veut nous rassurer, L'IA n'aurait pas la capacité de « problématiser » ni de « philosopher » ce qui signerait la suprématie de l'esprit humain !?

Depuis beaucoup de critiques ont expliqué la défaite de ChatGPT et les algorithmes ont bien « appris » et évolué

L'exposé de Pr Hantouss démontre la multitude des horizons positifs de l'IA en médecine et dans les technologies et sciences en général, à quel point que contrairement à ce qu'a suggéré Enthoven, l'avenir de l'homme n'est plus envisagé qu'avec l'IA, (peut-on arrêter le progrès et a-t-on le droit de le faire ?)

Graal, Feu de Prométhée, ou Boîte de Pandore?



Sera-t-elle son Graal promesse de maîtrise du monde, de capacités décuplées, sans maladie, repoussant la mort ?

L'IA constitue un tournant dans l'histoire de l'humanité tout comme le feu, la pierre et l'écriture. Elle est le feu de Prométhée qu'il a dérobé aux Dieux de l'Olympe, mais Prométhée a payé le prix cher.

N'est-ce pas de l'arrogance humaine ? Car on parle d'un ensemble de technologies capables de simuler certaines fonctions cognitives humaines avec pour **ambition sous-jacente**, formulée dès les **années 1950** par des pionniers comme Alan Turing, de concevoir des **machines capables de "penser" comme voir mieux que l'homme**.

L'homme Soulèvera-t-il la Colère des dieux de l'Olympe, attisera-t-il le courroux de Némésis déesse de la vengeance et de la justice distributive (à chacun son dû) ?

Christianisme : rédemption / Islam « law te3ala9at himatoulmar i bil 3archi lé néléh? »

L'homme n'a-t-il pas créé un Frankenstein qui causera sa perte et ouvert la boîte de Pandore avec tous les risques dont on va vu certains avec Pr Hantouss et dont d'autres sont si bien démontrés par les œuvres dites de science-fiction (MATRIX)?



Dans *L'Homme augmenté – Futurs de nos cerveaux*, Raphaël Gaillard explore l'idée que l'intelligence humaine et

l'IA ne devraient pas être en confrontation mais en hybridation. Il anticipe une ère où les technologies d'interface cerveau-machine permettront une fusion entre capacités humaines et intelligence artificielle. Selon lui, cette fusion pourrait non seulement prolonger nos facultés mentales, mais aussi ouvrir des possibilités inédites dans les domaines de la mobilité, de la communication des pensées, et même du traitement des émotions.

Gaillard compare cette nouvelle forme d'hybridation avec celle qu'a représentée l'apparition de l'écriture et de la lecture, ayant transformé l'humanité et marqué la transition de la Préhistoire à l'Histoire. Il perçoit ainsi l'IA comme une extension des capacités humaines qui, plutôt que de nous menacer, pourrait nous préparer à relever les défis modernes avec des outils cognitifs améliorés. En somme, l'ouvrage appelle à une approche proactive, pour intégrer et utiliser l'IA de manière judicieuse afin d'en tirer des bénéfices tout en se préparant aux implications psychologiques et éthiques de cette coévolution avec les machines

Dans *L'homme diminué par l'IA*, Marius Bertolucci examine les effets délétères de l'intelligence artificielle sur l'humain, en particulier en matière d'autonomie, de jugement, et de créativité. Il soutient que la dépendance croissante à l'IA pour des tâches de décision et d'analyse tend à réduire la capacité humaine à exercer un contrôle critique et une pensée autonome. Plutôt que de seulement amplifier les capacités humaines, Bertolucci argumente que l'IA risque de standardiser les comportements, de limiter la spontanéité et d'instaurer une forme d'aliénation subtile au quotidien. Pour Bertolucci, l'impact de l'IA sur l'homme est un affaiblissement de son humanité même. Ce phénomène va au-delà de la simple question technologique et pose des enjeux éthiques profonds, liés à la préservation des valeurs d'autonomie et de singularité humaine dans un monde où l'efficacité est primordiale. Cette critique contribue à un débat sur la gouvernance éthique de l'IA, où la technologie ne devrait pas prendre le dessus au point de remplacer la capacité de jugement humain mais, idéalement, la renforcer sans la supplanter.

Toute la question est de savoir si L'IA augmente ou diminue l'Homme

Aux sources philosophiques de l'IA: une vision mécaniste de l'homme

- Qu'est-ce que l'homme? Esprit et corps!
- La théorie de l'animal-machine (Descartes)
- La théorie de l'homme-machine (De La Mettrie)
- De la théorie de l'homme-machine à l'intelligence artificielle: une mutation technoscientifique fondamentale
- Quelle issue à cette mutation ?

L'intelligence artificielle (IA) puise ses racines dans une longue tradition philosophique et scientifique. Dès l'Antiquité, les penseurs s'interrogent sur la nature de l'intelligence et du raisonnement. Des philosophes comme **Aristote** et **Descartes** ont exploré la possibilité de modéliser la pensée humaine.

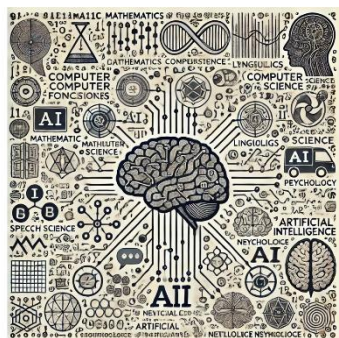
Pour Descartes, seule l'âme humaine était capable de pensée consciente. Descartes, avec son **dualisme**, a distingué entre le corps matériel et l'esprit immatériel. Toutefois, il a posé indirectement la question de savoir si la pensée humaine pouvait être reproduite par des mécanismes. En revanche, des philosophes comme **Hobbes**, avec son **mécanicisme**, ont soutenu que tout comportement humain pouvait être compris en termes de mouvements matériels. En parallèle, des penseurs comme **Diderot** et **La Mettrie** ont défendu une vision **matérialiste**, affirmant que l'esprit était une extension de la matière.

Cette dichotomie entre le dualisme et le matérialisme influence encore les débats contemporains sur la nature de l'intelligence et de la conscience, posant la question de savoir si l'IA, en tant que simulation de l'intelligence humaine, pourrait être considérée comme un prolongement de notre esprit ou simplement comme un outil dénué de conscience.

Origines philosophiques et
interdisciplinaires de l'IA (La
faute à Descartes... et Skinner)

- Philosophie de l'esprit
 - Matérialisme, dualisme, mécanisme
- Logique, mathématiques et informatique
- Psychologie
- Sciences cognitives
- Neurosciences
- Linguistique

Les avancées dans ces disciplines ont permis de définir des modèles qui aident à simuler certains aspects de l'intelligence humaine, comme la résolution de problèmes ou l'apprentissage.



Au XXe siècle, ces questions se sont concrétisées avec l'émergence des sciences cognitives, alimentées par des disciplines comme la **psychologie**, la **linguistique**, les **neurosciences** et les **mathématiques**.

L'IA ne s'appuie **pas uniquement sur les mathématiques ou l'informatique**. Elle émerge de l'intersection de plusieurs domaines, notamment la **philosophie de l'esprit**, qui explore les processus mentaux, et les **sciences cognitives**, qui cherchent à modéliser la perception et la prise de décision. Elle partage également des connexions avec des disciplines comme la **linguistique**, les **neurosciences** et l'**ingénierie**.

Cette interdisciplinarité reflète la complexité de la tâche à accomplir : recréer, au moins partiellement, le fonctionnement de l'esprit humain. **Les avancées dans ces disciplines ont permis de définir des modèles qui, bien qu'imparfaits, aident à simuler certains aspects de l'intelligence humaine, comme la résolution de problèmes ou l'apprentissage.**

Le post humain: défis numériques aux SHS

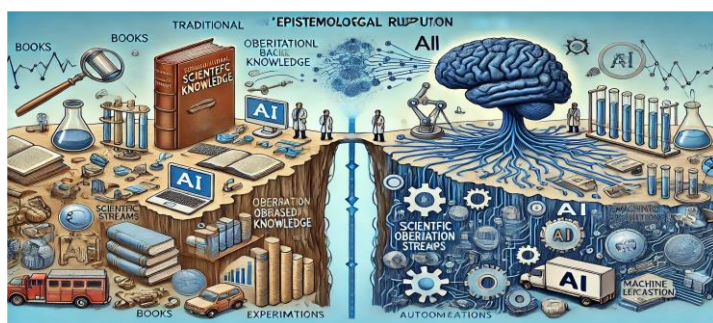
- « *Homo numericus* »!
- Descartes et Al-Khwarizmi: animal-machine/homme-machine, intelligence humaine/intelligence artificielle!
- IA et impératif économique : Cohen
- L'« homme » du post-humanisme est-il encore humain?
- Rôles des SHS et responsabilité pour l'avenir de l'humain!



Peut-on considérer l'IA comme une science théorique, fondée sur des principes mathématiques et logiques, ou comme une discipline expérimentale et empirique ? Si l'on opte pour cette dernière approche, l'IA se situe alors du côté des **sciences de la nature**, qui étudient des phénomènes naturels observables et quantifiables (la terre, le corps...), ou des **sciences de la culture**, qui s'intéressent aux comportements et aux productions humaines ?

L'étude des relations entre l'IA et les sciences humaines et sociales illustre non seulement l'intérêt historique que l'IA suscite pour ces disciplines, mais également son utilisation croissante et les transformations qu'elle engendre, souvent désignées comme le « **tournant calculatoire** ». Ainsi, l'apport de l'IA ne se limite pas à des applications pratiques mais au **rapprochement entre IA et sciences humaines**

Rupture épistémologique



L'IA introduit dans le domaine de la génération des connaissances tant celles de la nature que de la culture introduit ce que Gaston Bachelard qualifie de **rupture épistémologique**, signifiant une dissociation entre l'évidence première de l'observation et les faits scientifiques issus de l'expérimentation.

Dans le cas de l'IA, cette rupture réside dans la capacité des machines à traiter et à interpréter des données complexes de manière autonome, remettant ainsi en question notre rapport traditionnel à la connaissance, qui s'appuie sur l'expérimentation et la causalité.

Cela nous amène à introduire la notion d'**explicabilité**.

Cette explicabilité fait de plus en plus défaut avec le deep learning et les réseaux de neurones, rendant cette IA de plus en plus autonome et forte, faisant écho avec le mythe du Terminator ou de la Matrice

IA faible → IA forte ou Intelligence Augmentée?

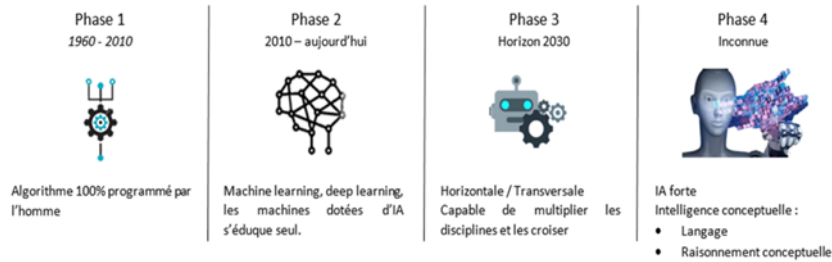


Figure 1 : Evolution de l'IA depuis les années 1960

Auteur : Raminoson-goni Sitaka, 02/01/2020

L'IA forte se réfère à l'idée que les machines pourraient un jour atteindre une forme d'intelligence égale, voire supérieure, à celle des êtres humains, incluant des capacités de conscience et d'auto-perception. Ce concept repose sur l'hypothèse que le cerveau humain, en tant que système de traitement d'information, peut être reproduit en totalité par des algorithmes.

Par ailleurs, l'IA a également nourri l'imaginaire d'un avenir « transhumain ». Le transhumanisme envisage l'IA comme un instrument clé pour transcender les limitations biologiques humaines, en fusionnant les technologies avancées avec le corps humain. Cette perspective est source de controverses, car elle pose des questions profondes sur l'identité humaine. (sera développé par Mme Derouiche)

Le concept d'intelligence augmentée est important dans les discussions contemporaines autour des usages éthiques de l'IA, en particulier pour éviter des peurs liées à l'IA forte, et recentrer les débats sur des solutions pratiques et humaines.

Au-delà de l'irrationalité des peurs suscitées par l'imaginaire collectif vis à vis du transhumanisme et de son assimilation à l'IA, la réticence de ces nouvelles voies de connaissance et de gouvernance pose la question éthique de « non-assistance à humanité en danger ». Est-il éthique de ne pas utiliser les programmes d'aide à la décision médicale en matière d'oncologie et d'imagerie ? Peut-on faire l'économie des modèles de surveillance climatologique et sismique ? les exemples sont nombreux.... Plutôt que de viser à créer une intelligence artificielle "forte", de nombreux chercheurs et penseurs se tournent vers le concept d'intelligence augmentée. Cette approche vise à améliorer ou assister les compétences humaines dans des domaines spécifiques, sans chercher à les remplacer ou dénaturer. Contrairement à l'IA forte, l'intelligence augmentée est déjà utilisée dans des applications concrètes, notamment dans le domaine médical, financier et industriel. Elle permet aux experts humains de prendre des décisions éclairées grâce à des analyses massives de données.

Cette forme d'IA ne se préoccupe pas de moralité ou des conséquences de ses actions, qui sont des prérogatives humaines.

Problèmes éthiques+++

- **Équité et biais des algorithmes**
 - conclusions fausses et discriminatoires.
 - Importance de la transparence
- **Responsabilité et IA autonome**
 - Qui est responsable des erreurs de l'IA ? (santé, justice, transports).
 - l'explicabilité comprendre les décisions d'IA complexes.
 - Agentivité : IA sans réelle conscience ni intention humaine.
- **Vie privée et surveillance**
 - Menaces potentielles à la vie privée.
 - Explicabilité l'utilisation des données, politiques.
- **Impact sur l'emploi et transformation des compétences**
 - Risque de perte d'emplois, besoin de clarifier les choix d'automatisation.
 - Transition vers de nouvelles compétences
- **Enjeux philosophiques : identité et déshumanisation**
 - Risque de déshumanisation : IA opaque
 - Questions d'identité : les limites de l'IA par rapport à l'humain.
 - Ethique transhumaniste : augmentations technologiques.



Comme l'a souligné Paul Dumouchel, l'IA est avant tout un outil qui permet d'optimiser le traitement de l'information et d'exécuter certaines tâches plus efficacement. Cette forme d'IA ne se préoccupe pas de moralité ou des conséquences de ses actions.

Elle est un outil, ni plus ni moins, destiné à assister les capacités humaines sans prétendre les surpasser.

Bien que l'intelligence augmentée semble être une voie plus pragmatique et prometteuse, elle n'est pas sans défis, notamment sur les plans éthique et social, l'intelligence augmentée soulève avec encore plus d'acuité des questions éthiques importantes.

- Équité et biais des algorithmes
- Données biaisées mènent à des conclusions discriminatoires.
- Importance de la transparence pour éviter discrimination et inégalités.
- Régulation et techniques anti-biais pour améliorer la fiabilité des IA.
- Responsabilité et IA autonome
- Problème central : Qui est responsable des erreurs de l'IA ? Notamment critique en santé, justice, transports.
- Nécessité de l'explicabilité pour tracer et comprendre les décisions d'IA complexes.
- Agentivité : Capacité d'action autonome des systèmes IA, sans réelle conscience ni intention humaine.
- Vie privée et surveillance
- IA et collecte de données : Menaces potentielles à la vie privée.
- Explicabilité essentielle pour comprendre l'utilisation des données.
- Régulations (ex : RGPD) pour informer et protéger les utilisateurs.
- Impact sur l'emploi et transformation des compétences
- Automatisation : Risque de perte d'emplois, besoin de clarifier les choix d'automatisation.
- Transition vers de nouvelles compétences : Importance de la compréhension de l'IA et de l'explicabilité.
- Solutions pour un avenir inclusif : Formation axée sur l'IA et les compétences techniques.
- Enjeux philosophiques : identité et déshumanisation
- Risque de déshumanisation : IA opaque pouvant altérer l'autonomie humaine.
- Questions d'identité : Explicabilité permettant de distinguer les limites de l'IA par rapport à l'humain.
- Ethique transhumaniste : Impact sur les valeurs humaines, intégrité et identité face aux augmentations technologiques.

Les principes éthiques de l'IA



Programmer une IA pour obtenir des décisions éthiques est une tâche complexe qui exige une compréhension profonde de la nature de l'éthique et de ses nuances. L'éthique oscille entre des principes universels et des valeurs subjectives, influencées par le contexte culturel, politique, et social.

On peut quand même dresser **les principes éthiques pour une IA digne de confiance**

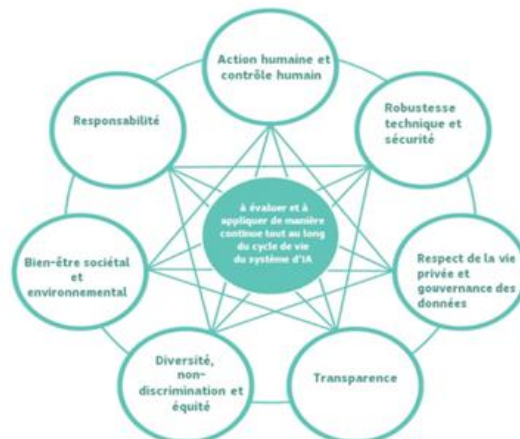


Figure 2: interrelation des sept exigences: elles revêtent toutes une importance égale, elles se soutiennent mutuellement et devraient être appliquées et évaluées tout au long du cycle de vie d'un système d'IA.

Vers une gouvernance éthique

Gouvernance IA doit allier principes universels et adaptation contextuelle.

Transparence et explicabilité essentielles.

Continue depuis le besoin le pourquoi le comment jusqu'aux conséquences et suivi

Portée par tous les intervenants ! mais qui ? Ni les concepteurs et firmes Ni les « Comités d'éthique »

Le but final de « l'éthique planétaire » ou « éthique de l'humanité » est de « surmonter l'impuissance de l'humanité à se constituer en humanité ». (Edgar Morin)



Parallèlement à la croissance du lavage de l'éthique, sa condamnation a conduit à une tendance au « dénigrement de l'éthique ». Cela consiste en une banalisation de l'éthique et de la philosophie morale, désormais comprises comme des outils distincts ou des structures sociales préformées telles que des comités d'éthique, des programmes d'auto-gouvernance ou des groupes de parties prenantes. Sans relation avec les politiques ni les lois.

Concepts clés : Ethics Washing & Ethics Bashing

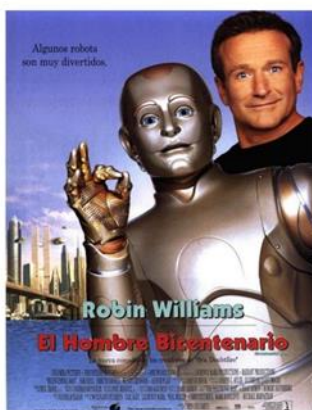
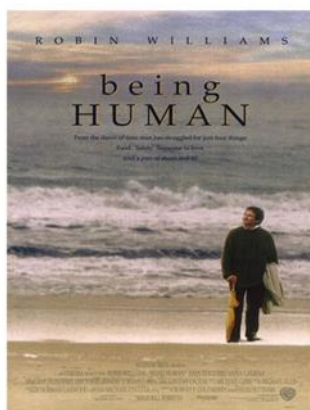
Ethics Washing : éthique en vitrine, sans vraie régulation ni éthique.

Ethics Bashing : critique de l'éthique, peut être excessive pour détourner l'attention des véritables problèmes de justice sociale et de régulation.

Besoin de vraies mesures : audits réguliers et indépendants, transparence

Dépasser l'éthique comme simple outil d'image ou de défense contre la régulation. Au contraire, **l'éthique doit être intégrée dans chaque étape de la gouvernance technologique, et considérée comme un élément clé des politiques de développement, plutôt que comme une simple vitrine.**

L'aiguillage éthique, au cœur du process IA et avec tous les intervenants permet de « surmonter l'impuissance de l'humanité à se constituer en humanité ». (Edgar Morin)



Pour plus d'informations sur les balises qui font notre humanité je vous recommande ces deux films de Robbin Williams ; **Humain est synonyme de faillible et mortel**

Références

- Jean-Gabriel Ganascia. Intelligence artificielle et épistémologie. Allers-retours indispensables. JECIS, Jean-Sébastien Vayre; Gérald Gaglio; Manuel Boutet; Lise Arena, Jun 2022, Nice, France. Ffhal03760357f
- Justin-Éric BOILEAU, Ilona BOIS-DRIVET, Hannes WESTERMANN, Jie ZHU; Rapport sur l'épistémologie de l'intelligence artificielle (IA), document de travail n° 32, Laboratoire de cyberjustice
- https://www.google.com/search?sca_esv=9177e7b6d39e290e&q=epist%C3%A9mologie+de+l%27intelligence+artificielle&tbm=vid&source=inms&fbs=AFQNm0CbCVgAZ5mWEJDg6aoPVCBgTiosgQSuz8MmAdio07UCIdZtLazlRgowYID0nDsqEIN7XYIyS3u8TzHmWPP2pnW7G_IJrzTNDMziXf_SeamZvsmQEpVWAo5N9Kl0kgaUhXfR1pvCaBMGis3J2_YASoH3vq235tGT22Lup522AIHeF4&sa=X&ved=2ahUKEwi4tsXhivKlAxUpR_EDHUqkAJgQQ0pQJegQIEBAB&biw=1366&bih=633&dpr=1#fpstate=ive&vid=cid:50f46ec2,vid:6f5PHistvCs,st:0
- Elettra Bietti: From Ethics Washing to Ethics Bashing: A Moral Philosophy View on Tech Ethics. JOURNAL OF SOCIAL COMPUTING ISSN 2688-5255 05/06 pp266-283 Volume 2, Number 3, September 2021
- Alban Leveau-Vallier. Intelligence artificielle et intuition. Philosophie. Université Paris 8 - VincennesSaint-Denis, 2023. Français. ffNNT : ff.fttel-04015572f