



Naviguer dans l'éthique de l'IA : Construire un futur responsable

Résumé

Cet article co-écrit avec ChatGPT explore le sujet critique de l'éthique dans le domaine de l'IA. Reconnaisant les défis de définir des principes éthiques universels, l'article souligne l'importance d'intégrer des cadres largement reconnus pour guider le développement de systèmes d'IA éthiques. Des principes clés tels que l'équité, la transparence, la responsabilité, la confidentialité et les valeurs centrées sur l'humain sont

discutés, car ils constituent des fondements essentiels pour une IA éthique.

L'article met l'accent sur l'importance vitale de la construction de systèmes d'IA éthiques, en soulignant l'influence de l'IA sur les droits de l'homme, la confidentialité et la sécurité, exigeant une délibération prudente et responsable. Il explore le rôle de la confiance, de la conformité juridique et réglementaire, des risques liés à la sécurité et des efforts visant à atténuer les biais et les discriminations dans les systèmes d'IA, autant d'éléments cruciaux pour assurer une mise en œuvre éthique.

Pour évaluer si l'IA, comme ChatGPT, respecte les principes éthiques, l'auteur a engagé une réflexion approfondie avec une IA générationnelle en posant de multiples questions et en recevant des réponses. Elle a observé les efforts des développeurs d'IA pour créer une IA équitable, transparente, inclusive et fiable en se basant sur les meilleures pratiques et les lignes directrices. Néanmoins, l'article encourage fortement les lecteurs d'exercer leur esprit critique et de prendre en compte les biais potentiels lors de l'utilisation des systèmes d'IA.

L'article aborde les risques et les opportunités de l'IA éthique, mettant en évidence des exemples d'IA biaisée et l'importance de donner la priorité à la confidentialité des données. Il présente également des exemples de systèmes d'IA inclusifs, démontrant comment l'IA peut favoriser l'égalité et l'accessibilité.

En conclusion, l'article souligne la nécessité d'une IA éthique pour promouvoir la confiance, l'adoption et un impact positif sur les individus et la société. Il appelle à une conception respectueuse de la confidentialité et à des pratiques d'IA responsables. L'interaction avec l'IA représente une étape prometteuse vers la co-écriture avec l'IA, et Catherine exprime sa gratitude pour l'engagement des lecteurs et les invite à lui envoyer leurs commentaires pour de futurs articles sur les meilleures pratiques et l'audit des systèmes d'IA.

#IAethique #IAresponsable #IAinclusive #AIethics #ResponsibleAI #InclusiveAI

1. Introduction

Parler des principes éthiques est déjà un défi en soi, mais aborder l'éthique dans le domaine de l'IA l'est encore davantage. Tout d'abord, la vision occidentale de l'éthique peut différer de ce qui est considéré comme éthique ailleurs dans le monde. Deuxièmement, il n'existe pas un seul ensemble définitif de principes qui doit être appliqué pour construire un système d'IA éthique. Cependant, il existe plusieurs cadres et ensembles de principes largement reconnus qui fournissent des orientations pour construire des systèmes d'IA éthiques.

Par exemple, l'initiative mondiale de l'IEEE pour les considérations éthiques de l'IA et des systèmes autonomes (1) a élaboré un ensemble de principes de conception alignés sur l'éthique. Ces principes comprennent un large éventail de lignes directrices couvrant différents sujets, notamment la transparence, la responsabilité et l'impact social. De même, le groupe d'experts de haut niveau de l'Union Européenne sur l'IA (2) ou le Partnership on AI (3) ont développé un ensemble d'exigences clés pour des systèmes d'IA fiables, comprenant notamment la supervision humaine, la transparence et la responsabilité. D'autres organisations, telles que le Forum Economique Mondial (WEF) (4) et l'Union Internationale des Télécommunications (UIT - ITU) (5), ont également développé leurs propres plateformes pour promouvoir l'IA éthique de différentes manières.

Dans l'ensemble, il existe de nombreux ensembles de principes, cadres et plateformes variés pouvant être utilisés pour construire un système d'IA éthique, mais les principes spécifiques pertinents dépendront du contexte dans lequel le système d'IA est développé et utilisé. Même si aujourd'hui, le nombre de principes dépend du cadre utilisé, la plupart d'entre eux incluent généralement des principes communs tels que :

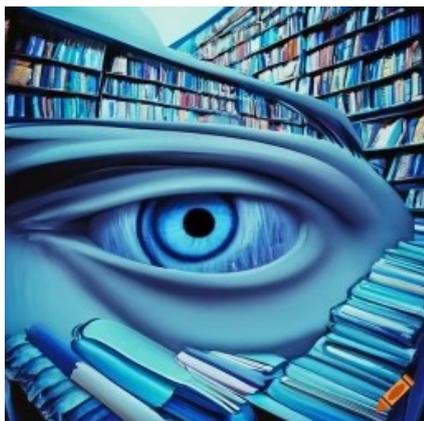
1. Équité : les systèmes d'IA doivent être conçus de manière juste et impartiale, et ne doivent pas discriminer les individus ou les groupes en fonction de facteurs tels que la race, le genre ou l'âge.
2. Responsabilité et transparence : les systèmes d'IA doivent être transparents et explicables, et conçus pour fournir des explications claires et interprétables concernant leurs décisions, ainsi que pour être auditable.
3. Confidentialité : les systèmes d'IA doivent être conçus pour protéger la vie privée des individus et se conformer aux réglementations pertinentes en matière de protection des données.
4. Sécurité : les systèmes d'IA doivent être conçus pour être sûrs et sécurisés, et doivent être soumis à des tests et une validation appropriés pour garantir qu'ils ne représentent pas une menace pour les utilisateurs ou pour la société dans son ensemble.
5. Valeurs centrées sur l'humain : les systèmes d'IA doivent être conçus en tenant compte du respect des valeurs humaines, notamment le respect de la dignité, de la vie privée et de l'autonomie.

Les principes énumérés ci-dessus sont généralement considérés comme les principes de base qui devraient être appliqués dans la conception et le développement de tout modèle d'IA éthique.

Les leaders, les développeurs et les utilisateurs d'IA doivent être conscients que les actions des professionnels de l'informatique transforment notre monde. Pour agir de manière responsable, les professionnels de l'informatique doivent réfléchir aux impacts plus larges de leur travail et œuvrer constamment pour l'intérêt général. Le Code de déontologie et de conduite professionnelle de l'Association for Computing Machinery - ACM (6) vise à inspirer et à guider la conduite éthique de tous les professionnels de l'informatique, y compris les praticiens actuels et futurs, les instructeurs, les étudiants, les influenceurs et tous ceux qui utilisent la technologie informatique de manière significative.

Veillez noter que la perspective occidentale est adoptée dans cet article car, premièrement, je suis européenne, et deuxièmement, ChatGPT a été formé sur des données occidentales provenant d'Internet. Il serait très intéressant de reproduire cette expérience avec une IA générative formée sur un ensemble de données plus diversifié.

Veillez noter que les questions ont été posées à la version 3.5 de ChatGPT en février - mars 2023.



2. Pourquoi est-ce si important ?

La construction de systèmes d'IA éthiques est essentielle car l'IA a le potentiel d'impacter les droits de l'homme, la dignité, la vie privée et la sécurité, que ce soit par des erreurs, des biais ou une utilisation intentionnelle abusive. Prendre en compte les considérations éthiques aide à atténuer ces risques, à prévenir les préjudices et à garantir que l'IA est développée et utilisée de manière à bénéficier à la société dans son ensemble.

- **CONFIANCE** : De nombreux systèmes d'IA reposent sur des données personnelles pour formuler des recommandations et prendre des décisions personnalisées. Si nous ne sommes pas convaincus que nos données sont utilisées de manière responsable, appropriée et conservées en toute confidentialité, nous sommes moins enclins à les partager, ce qui pourrait limiter la précision et l'efficacité des systèmes d'IA. De plus, lorsque nous comprenons comment un système d'IA prend des décisions, notre niveau de confiance augmente et nous pouvons améliorer notre manière d'interagir avec lui.
- **CONFORMITÉ LÉGALE ET RÉGLEMENTAIRE** : Dans de nombreuses juridictions, les lois sur la confidentialité exigent des organisations qu'elles protègent les données personnelles, comme le Règlement Général sur la Protection des Données (RGPD) dans l'Union européenne (7) ou le California Consumer Privacy Act (CCPA) aux États-Unis (8). Ces lois garantissent, dans une certaine mesure, que les données ne sont pas utilisées de manière non autorisée ou inattendue. La construction d'une IA éthique pousse les organisations à se conformer à ces réglementations, à respecter leurs utilisateurs, à éviter les sanctions financières et les dommages à leur réputation.
- **RISQUES LIÉS À LA SÉCURITÉ** : Les données personnelles sont des biens précieux qui peuvent être ciblés par les cybercriminels. Les données personnelles sont souvent sensibles et peuvent être utilisées à des fins malveillantes si elles tombent entre de mauvaises mains. La protection de la confidentialité aide à prévenir les violations de données et autres incidents de sécurité.
- **BIAIS ET DISCRIMINATION** : La construction d'une IA inclusive est cruciale pour une société équitable qui valorise une variété de perspectives et de besoins. Les systèmes d'IA qui utilisent des données personnelles pour prendre des décisions, telles que celles liées à l'emploi ou aux services financiers, peuvent renforcer la discrimination envers certains groupes si le jeu de données est biaisé ou incomplet. À l'opposé, une IA inclusive peut réduire les biais, atténuer les risques de discrimination et améliorer les résultats pour tous en renforçant l'autonomisation, l'accessibilité ou la personnalisation.

Cependant, s'assurer qu'un jeu de données (data set) est inclusive s'avère difficile pour plusieurs raisons :

1. **Biais dans la collecte des données** : les données utilisées pour entraîner les modèles d'IA sont souvent collectées par des êtres humains, et les préjugés et biais humains peuvent influencer quelles données sont collectées et comment elles sont labellisées. Par exemple, un ensemble de données utilisé pour entraîner un algorithme de reconnaissance faciale peut être biaisé envers certaines ethnies ou genres, ou simplement envers les personnes portant des lunettes !
2. **Manque de diversité dans les sources de données** : dans certains cas, les données peuvent ne pas être disponibles pour certains groupes ou peuvent être plus difficiles à collecter. Par exemple, si un ensemble de données est censé représenter une population particulière, mais que les sources de données incluent uniquement des données provenant de certaines régions ou de réseaux sociaux qui ne sont pas représentatifs de toute la population, l'ensemble de données est incomplet.
3. **Complexité de l'identification des biais** : Identifier et éliminer les biais des ensembles de données peut être difficile, surtout lorsque les biais ne sont pas immédiatement évidents ou sont intégrés de

manière complexe dans les données. Cela peut nécessiter beaucoup de temps et de ressources pour mener une analyse approfondie des données afin d'identifier et de corriger les biais.

4. Test insuffisant : Même si l'ensemble de données est diversifié et inclusif, cela peut ne pas être suffisant pour garantir que le système d'IA se comportera de manière équitable et inclusive dans toutes les situations. Les systèmes d'IA sont complexes et peuvent générer des résultats inattendus ou renforcer les biais de manière inattendue, ce qui rend impératif de les tester de manière approfondie dans différentes conditions et situations.
5. Évolution des normes sociales : Les normes sociales et les attentes changent avec le temps, ce qui rend difficile de garantir qu'un ensemble de données restera inclusif et représentatif au fil du temps. Par exemple, un ensemble de données qui est inclusif aujourd'hui peut ne plus l'être demain si les attitudes sociales envers certains groupes changent ou si de nouveaux groupes émergent qui n'étaient pas représentés auparavant dans les données.

3. ChatGPT est-il/elle Ethique ?

Il est toujours très abstrait de réfléchir à ce qu'est l'éthique. C'est pourquoi j'ai décidé de prendre un exemple concret d'un système d'IA, et ChatGPT lui-même s'est imposé comme le candidat évident. Cela m'a amené à poser directement à ChatGPT un ensemble de questions dans le but d'identifier s'il s'agit d'une IA responsable, transparente, inclusive, équitable et fiable. Certaines de mes questions se sont concentrées sur la protection des données, qui est toujours une caractéristique essentielle.

La table ci-dessous souligne les extraits clés des **réponses de ChatGPT** :

Principe	Comment ChatGPT se comporte	Comment cela est réalisé
EQUITE	<p>En tant que modèle de langage IA, je n'ai pas d'émotions ni de désirs.</p> <p>Mon objectif est d'être aussi équitable que possible et de fournir des réponses utiles et informatives à tous les utilisateurs. Cependant, j'ai été entraîné sur un grand ensemble de données textuelles, ce qui signifie que mes réponses ne sont aussi équitables et impartiales que les données sur lesquelles j'ai été formé. Il se peut qu'il y ait des biais dans l'ensemble de données qui pourraient entraîner des réponses biaisées ou injustes. Par exemple, si l'ensemble de données sur lequel j'ai été formé manque de diversité ou d'inclusion, cela pourrait conduire à des réponses biaisées ou injustes.</p>	<p>Mon équipe de développement chez OpenAI s'engage à construire et à maintenir un système d'IA équitable, responsable et éthique, qui soit transparent, impartial et sécurisé.</p> <p>Ils suivent les meilleures pratiques et les lignes directrices pour le développement éthique de l'IA, telles que celles établies par IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems et Partnership on AI.</p>
TRANSPARENCE	<p>En tant que modèle de langage IA, j'ai été conçu pour être aussi transparent que possible. Je suis formé sur un vaste ensemble de données textuelles et j'utilise des algorithmes complexes pour générer des réponses aux entrées des utilisateurs.</p>	<p>ChatGPT a été conçu pour fournir des explications pour certaines de ses réponses, ce qui peut aider les utilisateurs à comprendre comment le modèle prend des décisions. L'équipe d'OpenAI fournit également une documentation qui explique la conception du modèle, y compris les algorithmes utilisés et les sources de données.</p>

INCLUSIVITE	<p>En tant que modèle de langage IA, je suis conçu pour être neutre et inclusif dans mes réponses. Je m'efforce de fournir des réponses précises et informatives à tous les utilisateurs, indépendamment de leur origine ou de leur identité. J'ai été formé sur une vaste quantité de données textuelles provenant de sources diverses, et je suis conçu pour générer des réponses qui reflètent la diversité des perspectives et des expériences contenues dans ces données. Cependant, tout comme tout autre système d'IA, la qualité de mes réponses dépend de la qualité des données sur lesquelles j'ai été formé, et il se peut que je reproduise involontairement des biais ou des inexactitudes présents dans ces données.</p>	<p>Je suis conçu pour privilégier un langage inclusif et non discriminatoire dans mes réponses. J'ai été programmé pour reconnaître et éviter certains types de langage biaisé. Je suis constamment mis à jour et amélioré pour refléter les dernières avancées de la recherche en IA et pour traiter les problèmes ou préoccupations qui surgissent.</p>
FIABILITE	<p>En tant que modèle de langage IA, je suis conçu et formé pour être aussi fiable et précis que possible dans le cadre de mes données d'entraînement et de ma date limite de connaissances en 2021. Comme tout système d'IA, je ne suis pas parfait et je peux commettre des erreurs ou fournir des réponses inexactes dans certaines situations, surtout si les données d'entrée sont contradictoires ou ambiguës. Des erreurs et des réponses inexactes peuvent survenir en raison de divers facteurs, tels que les limites de mes données d'entraînement, les biais dans les données ou les erreurs dans mes algorithmes.</p>	<p>ChatGPT est conçu pour fournir des réponses claires et transparentes, ainsi que des scores de confiance indiquant le niveau de certitude qu'il a dans ses réponses.</p> <p>ChatGPT est soumis à des tests et à des validations continues pour garantir qu'il reste fiable et à jour avec les dernières avancées de la recherche en intelligence artificielle.</p>
SECURITE DES DONNEES	<p>En tant que modèle de langage IA, je suis conçu pour être un système sécurisé et confidentiel qui protège la confidentialité des données de ses utilisateurs.</p>	<p>Lorsque vous interagissez avec ChatGPT, vos conversations sont confidentielles et ne sont pas partagées avec des tiers. De plus, vos informations personnelles ne sont pas stockées ni utilisées à d'autres fins en dehors de vous fournir une réponse à vos questions.</p>



MATIÈRE À RÉFLEXION : *Il est extrêmement intéressant de constater que certaines réponses à mes questions contenaient d'importants points d'attention, mettant souvent en évidence et de différentes manières l'importance de la pensée critique :*

- "Il est important de se rappeler que l'IA est un outil qui peut être utilisé pour soutenir et améliorer la prise de décision humaine, mais qu'elle **ne remplace pas le jugement humain ni la pensée critique**. Il est essentiel que les individus soient conscients des limites et des biais potentiels des systèmes d'IA et les abordent avec un regard critique."
- "Il est important de noter qu'en tant que modèle de langage IA, j'ai des **limitations** et des **biais**. Bien que je sois entraîné sur un grand ensemble de données textuelles, je n'ai peut-être pas accès à toutes les informations ou perspectives sur un sujet donné, et mes réponses ne sont pas toujours parfaites ou

entièrement précises. Lors de la prise de décisions basées sur les informations que je fournis, il est important de **pratiquer une pensée critique et de rechercher plusieurs sources d'information.**"

- "Je suis un modèle d'apprentissage automatique aussi inclusif que les données utilisées pour me former. **Si les données d'entraînement sont biaisées ou manquent de diversité, mes réponses peuvent refléter ce biais ou ce manque de diversité.**"
- "Il est important **d'évaluer en permanence et d'améliorer les données d'entraînement utilisées pour former les modèles d'IA afin de garantir qu'ils sont vraiment inclusifs. Aucun système d'IA n'est parfait** et il y a toujours la possibilité que des biais puissent être introduits de manière involontaire, soit par les données utilisées pour former le modèle, soit par la manière dont les utilisateurs interagissent avec le système. Par conséquent, il est important de surveiller et de tester en permanence les systèmes d'IA pour garantir qu'ils sont aussi inclusifs et impartiaux que possible."
- "Il peut toujours y avoir **des limites à la transparence** de ChatGPT, comme avec tout système d'IA. Par exemple, il peut être difficile pour les utilisateurs de comprendre l'ensemble du processus de prise de décision du modèle, car il implique des calculs complexes et une grande quantité de données. Néanmoins, l'équipe OpenAI travaille à améliorer la transparence et l'explicabilité de ChatGPT et d'autres systèmes d'IA."
- "Il est important de noter que même si ChatGPT est conçu en tenant compte de la protection des données privées, il est toujours important de **faire preuve de prudence lorsque vous partagez des informations sensibles ou des données personnelles avec n'importe quel service en ligne, y compris ChatGPT.** Il est toujours recommandé de lire la politique de confidentialité et les conditions d'utilisation de toute plateforme ou service que vous utilisez, et de **vous assurer que vous êtes à l'aise avec les informations que vous partagez.** "

4. Quels sont les Risques et Opportunités ?

L'IA équitable est un objectif important que de nombreux chercheurs et développeurs s'efforcent d'intégrer dans leurs travaux, car il existe de nombreux exemples d'IA injuste dans diverses applications, dont certains ont reçu une grande exposition dans les médias et réseaux sociaux. Pour cette partie, j'ai demandé à ChatGPT de fournir des exemples de partialité et d'injustice et parmi les nombreuses réponses, voici ma sélection :

1. Des algorithmes de recrutement biaisés : Plusieurs entreprises ont mis au point des outils de recrutement assistés par l'IA, conçus pour trier les CVs et identifier les meilleurs candidats. Toutefois, certains de ces outils se sont révélés biaisés à l'égard des femmes et des minorités. Par exemple, Amazon a mis au point un outil de recrutement qui a ensuite été abandonné parce qu'il s'est avéré discriminant à l'égard des femmes.
2. Les systèmes de reconnaissance faciale qui sont biaisés à l'encontre des personnes de couleur : de nombreuses études ont montré que plusieurs systèmes de reconnaissance faciale ont des taux d'erreur plus élevés lorsqu'il s'agit d'identifier les personnes de couleur, en particulier les femmes. Par exemple, une étude du National Institute of Standards and Technology a révélé que certains systèmes de reconnaissance faciale présentaient des taux d'erreur 10 à 100 fois plus élevés pour les personnes de couleur que pour les personnes de race blanche.
3. Algorithmes de détermination de la peine injustes : Certains systèmes de justice pénale utilisent des algorithmes pour prédire la probabilité de récidive d'un prévenu ou pour déterminer son niveau de risque. Cependant, certains de ces algorithmes se sont révélés biaisés à l'encontre des personnes de couleur. Par exemple, une étude a révélé qu'un algorithme utilisé dans l'État de Floride avait deux fois

plus de chances de signaler à tort des accusés noirs comme présentant un risque élevé que des accusés blancs.

4. Les chatbots sexistes : Certains chatbots conçus pour interagir avec les clients se sont révélés biaisés à l'égard des femmes. Par exemple, une étude portant sur plusieurs chatbots populaires a révélé qu'ils étaient plus susceptibles de répondre aux utilisateurs masculins par des réponses utiles ou informatives, alors qu'ils répondaient aux utilisatrices par des réponses plus condescendantes ou non utiles.

Ce ne sont là que quelques exemples des risques auxquels la société doit faire face lorsqu'elle utilise des systèmes d'IA injustes. Il est donc important d'être conscient de ces problèmes potentiels et de prendre des mesures pour les mitiger lors de la conception et de la mise en œuvre des systèmes d'IA. J'ai ensuite demandé à ChatGPT comment les principes éthiques pouvaient contribuer à la construction d'une meilleure IA.:

- L'IA explicable dans les soins de santé : Dans le domaine des soins de santé, des modèles d'IA sont développés pour aider à diagnostiquer et à traiter les maladies. Cependant, il est important que ces modèles soient transparents et explicables afin que les médecins et les patients puissent comprendre comment les décisions sont prises. Par exemple, le système IBM Watson for Oncology fournit une explication pour chaque recommandation qu'il fait, ce qui permet aux médecins de comprendre le raisonnement qui sous-tend les recommandations du système.
- IA fiable :
 1. Les filtres anti-spam utilisent des algorithmes d'apprentissage automatique pour classer automatiquement les courriels entrants comme légitimes ou comme spam. Ces systèmes sont fiables car ils ont été formés sur de vastes ensembles de données de messages électroniques étiquetés et peuvent identifier avec précision les messages de spam avec un faible taux de faux positifs.
 2. Les assistants vocaux : Les assistants vocaux comme Alexa d'Amazon et Siri d'Apple utilisent des algorithmes de traitement du langage naturel pour interpréter les demandes des utilisateurs et fournir des réponses pertinentes. Ces systèmes sont de plus en plus fiables parce qu'ils comprennent et répondent avec précision à un large éventail d'entrées de l'utilisateur, avec un degré de plus en plus élevé de précision et de cohérence.
 3. Véhicules autonomes : Les véhicules autonomes utilisent des algorithmes d'intelligence artificielle pour analyser les données des capteurs et prendre des décisions sur la manière de faire fonctionner le véhicule de manière sûre et efficace. Ces systèmes sont fiables car ils ont été largement testés et validés dans une série de conditions réelles et sont conçus pour minimiser le risque d'accidents ou d'autres incidents de sécurité.



A ce stade, et alors que je suis moi-même conductrice d'une voiture électrique très élaborée depuis de nombreuses années, j'ai demandé à ChatGPT :

“Es-tu sûr que les véhicules autonomes puissent être classifiés comme une IA fiable ?”

Bien que les véhicules autonomes soient encore en cours de développement et de tests, des efforts considérables sont déployés pour les rendre aussi fiables et sécuritaires que possible grâce à l'utilisation de l'intelligence artificielle.

Les véhicules autonomes dépendent d'une large gamme de capteurs et d'algorithmes d'apprentissage automatique pour naviguer sur les routes, détecter des objets et des obstacles, et prendre des décisions de conduite en temps réel. Afin d'assurer la fiabilité de ces systèmes, les chercheurs et les ingénieurs les soumettent à des tests et à une validation rigoureux, tant dans des environnements simulés que sur les routes du monde réel.



De plus, les organismes de réglementation élaborent des normes de sécurité et des directives pour les véhicules autonomes afin de garantir qu'ils atteignent un seuil minimal de fiabilité et de sécurité. Par exemple, la politique fédérale sur les véhicules automatisés du Département des Transports des États-Unis fournit un cadre pour garantir la sécurité et la fiabilité des véhicules autonomes.

Bien qu'il reste encore du travail à accomplir pour améliorer la fiabilité des véhicules autonomes, ils ont le potentiel de révolutionner les transports et d'améliorer la sécurité sur les routes. Des efforts sont déployés pour s'assurer qu'ils respectent les normes les plus élevées en matière de fiabilité et de sécurité.

De retour sur le grand sujet de la confidentialité des données, je demande 3 fois à ChatGPT de fournir des exemples : *“Donne-moi 2 exemples de systèmes d'intelligence artificielle où des données sensibles ont été divulguées.”*. Voici ceux que j'ai sélectionnés pour vous (le choix était vaste malheureusement) :

1. Violation de données en 2015 sur Ashley Madison : Ashley Madison est un site de rencontres destiné aux personnes mariées ou en couple, et son slogan était "La vie est courte. Ayez une aventure." En 2015, un groupe de pirates informatiques a violé la sécurité du site web et a volé des données sensibles, y compris des noms, des adresses, des numéros de téléphone et des informations de cartes de crédit appartenant à des millions d'utilisateurs. Les pirates informatiques ont menacé de divulguer ces données à moins qu'Ashley Madison ne ferme ses portes. La violation de données était due à une vulnérabilité dans le système de sécurité basé sur l'IA du site web, que les pirates ont pu exploiter pour accéder aux données (9).
2. Problème de confidentialité de données en 2018 sur Strava : Strava est une application de fitness qui permet aux utilisateurs de suivre leurs séances d'entraînement et de partager leurs activités avec d'autres. En 2018, il a été découvert que la fonction de heat map de l'application, qui montrait la localisation et la fréquence de l'activité des utilisateurs, révélait les emplacements de plusieurs bases militaires sensibles dans le monde. Cette information avait été involontairement partagée par des utilisateurs de Strava qui avaient utilisé l'application sur des bases militaires, et le système d'IA de l'application avait utilisé ces données pour générer une heat map (910).
3. Problème de confidentialité de données en 2020 sur Zoom : Pendant la pandémie de COVID-19, de nombreuses personnes se sont tournées vers des plateformes de visioconférence comme Zoom pour communiquer et collaborer à distance. En 2020, la croissance sans précédent de l'utilisation de la visioconférence a entraîné plusieurs menaces pour la sécurité et la vie privée. Il a été découvert que la fonction de fond virtuel basée sur l'IA de Zoom divulguait accidentellement les données des utilisateurs, y compris des images et des vidéos de l'intérieur des maisons des personnes, à d'autres participants aux réunions. Cela était dû à un bug dans la fonctionnalité qui permettait à certains utilisateurs de contourner les paramètres du fond virtuel et de voir l'arrière-plan original de l'utilisateur. Protéger notre vie privée n'est pas toujours entre nos propres mains. La publication d'images de réunions affecte tous les participants, ce qui facilite la collecte de données personnelles, notamment l'âge, le sexe et la relation avec la participation à d'autres réunions (11).

PRIVATE INFORMATION LEAKS FROM VIDEO IMAGES SHARED ONLINE

Accompanying the rapid growth in video conferencing usage, particularly rising during Covid19 pandemic, users started sharing video conferencing collage images on social media. These images contain sensitive private data that can be exploited as follows:

1 Data Collection

Video conferencing collage images are collected from publicly shared posts on social media. Irrelevant images are filtered out using deep learning.



One of the meeting participants uploaded a meeting photo online, and a malicious actor collected the data.

2 Data Extraction

Participant private data such as name, gender, age and geographic location is extracted from the collage images.



The malicious actor extracts metadata from all meeting participants.

3 Networking Connection

Social networks of meeting participants are constructed by cross-referencing repeated users through multiple meetings, utilizing face recognition and background indications.



Using face recognition and background indications, the malicious actor constructs the user's social network.

4 Data Linkage

Based on usernames and faces, the users' social media personal accounts are detected, and additional data is collected.



The malicious actor maps the participants to their social media accounts and then searches for covert connections. For instance, a meeting with employees from other companies can indicate a buyout. Similar private information may be extracted about minor participants in virtual meetings, including their geographic location.

Infographie : Informations privées issues d'images partagées sur système de téléconférence (11)

MATIÈRE À RÉFLEXION : Il est important que les utilisateurs soient conscients des données qu'ils partagent sur telle ou telle plateforme. Il est essentiel de s'assurer que les systèmes d'IA sont conçus en tenant compte de la confidentialité et de la sécurité des données, et qu'ils sont soumis à des tests et à des audits approfondis pour garantir qu'ils ne sont pas vulnérables aux violations de données ou à d'autres problèmes de confidentialité.

Sur une note plus positive, je demande ensuite à ChatGPT de se concentrer sur l'IA inclusive : « **donne-moi trois exemples simples de systèmes d'IA inclusifs** » :

- Technologie de synthèse vocale :** La technologie de synthèse vocale est un exemple de système d'IA inclusif qui peut bénéficier aux personnes ayant des déficiences visuelles, ainsi qu'à celles qui préfèrent les informations auditives. En convertissant le texte écrit en parole, la technologie de synthèse vocale permet aux individus d'accéder à l'information de manière plus accessible et inclusive.
- Technologie de reconnaissance d'images :** La technologie de reconnaissance d'images peut être utilisée de manière inclusive pour aider les personnes ayant des déficiences visuelles à se déplacer dans leur environnement. Par exemple, une application pourrait utiliser la technologie de reconnaissance d'images pour identifier les objets dans l'environnement et fournir des descriptions audio de ces objets à l'utilisateur, l'aidant ainsi à naviguer et à interagir avec le monde qui l'entoure.
- Outils de détection de biais :** Les outils de détection de biais sont des systèmes d'IA conçus pour identifier et signaler les biais potentiels dans les données ou les processus de prise de décision. Par exemple, un outil pourrait être utilisé pour analyser les données de recrutement et identifier des schémas de biais dans le processus de recrutement, tels qu'un nombre disproportionné de candidats rejetés en fonction de leur sexe, de leur origine ou de leur âge. En identifiant ces biais, l'outil peut aider les organisations à prendre des mesures pour les corriger et à créer des pratiques de recrutement plus inclusives.

A la simple mais non évidente question *“What else?” (« et quoi encore ? »)* ChatGPT répondit avec un temps de réponse plus long que précédemment :

Voici quelques exemples supplémentaires de systèmes d'IA inclusifs :

4. **Systèmes de traduction linguistique** : La technologie de traduction linguistique est un exemple de système d'IA inclusif qui contribue à briser les barrières linguistiques et à permettre la communication entre différentes langues et cultures. Cela peut être particulièrement important dans des situations où les individus n'ont pas accès à un interprète ou à un traducteur..
5. **Médecine personnalisée** : La médecine personnalisée est un exemple de système d'IA inclusif qui peut contribuer à résoudre les disparités en matière de santé et garantir que les individus reçoivent des soins de santé adaptés à leurs besoins et à leur situation individuelle. En analysant les informations génétiques et d'autres données d'un patient, l'IA peut aider à identifier des traitements et des thérapies qui sont plus susceptibles d'être efficaces pour cet individu.
6. **Analyse de sentiment** : L'analyse de sentiment est un système d'IA qui peut être utilisé pour analyser les publications sur les médias sociaux, les avis de clients et d'autres formes de communication en ligne afin d'identifier des tendances et des modèles de sentiment. Cela peut aider les organisations à mieux comprendre les besoins et les préférences de leurs clients ou de leurs parties prenantes, et à élaborer des politiques et des pratiques plus inclusives.

Ces quelques exemples illustrent comment l'IA peut être utilisée pour créer des systèmes inclusifs qui bénéficient aux individus de milieux divers et favorisent une meilleure égalité et accessibilité. Ils démontrent comment l'IA peut être utilisée (ou pas) de différentes manières pour lutter contre certaines des disparités sociales et sanitaires présentes dans notre société.

En adoptant une IA inclusive, nous avons l'opportunité de créer un monde où la technologie est conçue et mise en pratique pour servir tout le monde de manière égale et où tous les individus ont accès aux outils et aux opportunités dont ils ont besoin pour s'épanouir.

5. Conclusion

Dans l'ensemble, l'éthique de l'IA est essentielle pour établir la confiance, promouvoir une adoption généralisée, fidéliser les clients, bâtir une réputation positive, et réaliser les avantages potentiels de l'IA de manière durable pour les individus et la société dans son ensemble, sans causer de préjudice ni violer les droits de l'homme. Le respect de la vie privée à chaque étape de la conception et des ensembles de données est un aspect fondamental de l'IA éthique. Les systèmes d'IA qui portent atteinte à la vie privée des individus peuvent éroder la confiance dans l'organisation qui les a créés et nuire aux relations avec les clients, les employés et d'autres parties prenantes.

À mesure que j'ai progressé dans cette nouvelle expérience de co-écriture d'un article avec une machine au lieu d'une équipe de chercheurs humains, la manière dont je formulais mes demandes a évolué et s'est améliorée. Je me suis également retrouvé engagé(e) dans une conversation avec la machine, comme je le ferais avec un collègue (toujours disponible et très poli) via un système de discussion. J'en suis venue à poser la question :

“Est-ce qu'une IA serait moins juste que les humains ? »

Et voici la réponse de ChatGPT :

“ Les systèmes d'IA peuvent présenter des injustices, mais il n'est pas exact de dire qu'ils sont intrinsèquement plus injustes que les êtres humains. Les systèmes d'IA sont programmés et conçus par des êtres humains et ne sont aussi impartiaux et justes que les données et les algorithmes sur lesquels ils sont formés. Si les données et les algorithmes sont biaisés, le système d'IA le sera également.

Les êtres humains, en revanche, ont une plus grande variété de biais potentiels et de préjugés qui peuvent influencer leur prise de décision, y compris des biais conscients et inconscients basés sur des facteurs tels que la race, le genre et d'autres caractéristiques. Ces biais peuvent conduire à la discrimination, à l'injustice et à un traitement inéquitable de certains individus ou groupes.

Cela étant dit, les systèmes d'IA peuvent parfois aggraver ou amplifier les biais humains s'ils sont formés sur des données ou des algorithmes biaisés. De plus, étant donné que les systèmes d'IA sont souvent utilisés pour prendre des décisions à grande échelle, les conséquences de toute injustice ou biais peuvent être beaucoup plus répandues que si un être humain individuel prenait une décision biaisée.

Par conséquent, il est important de veiller à ce que les systèmes d'IA soient conçus et mis en œuvre de manière à favoriser l'équité, la responsabilité et la transparence. Cela implique de prendre conscience des biais potentiels et de prendre des mesures pour les atténuer lors du processus de conception et de mise en œuvre. En faisant cela, nous pouvons contribuer à garantir que les systèmes d'IA sont utilisés pour promouvoir l'équité et la justice, plutôt que de perpétuer les biais et les inégalités existants. ”

Ne s'agit-il pas d'une réponse judicieuse ?

Si vous avez réussi à lire cet article jusqu'à la fin, j'aimerais vous exprimer sincèrement ma gratitude. J'apprécierais grandement si vous pouviez partager vos propres réflexions en me contactant sur catherine@xpertia.ch

Cela me motivera à écrire l'article suivant sur les meilleures pratiques et méthodologies d'audit de systèmes d'IA afin d'évaluer leur niveau d'éthique, ou toute autre sujet IA qui vous intéresserait.

Bibliography

1. IEEE Global Initiative for Ethical Considerations in AI and Autonomous Systems - Ethically Aligned Design: This initiative has developed a set of ethical principles and guidelines for AI and autonomous systems, covering areas like transparency, accountability, and social impact. (Reference: <https://standards.ieee.org/industry-connections/ec/autonomous-systems/>)
2. European Union's High-Level Expert Group on AI: The EU's expert group has developed a set of seven key requirements for trustworthy AI systems, including human oversight, transparency, and accountability. (Reference: <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>)
3. Partnership on AI: This organization brings together industry leaders, researchers, and policymakers to collaborate on AI ethics, fairness, and transparency. (Reference: <https://www.partnershiponai.org/>)
4. WEF - AI for Good: The World Economic Forum's launched a multi-stakeholder initiative that unites industry leaders, governments, academic institutions, and civil society organizations, to champion responsible design and release of AI systems. (Reference: <https://initiatives.weforum.org/ai-governance-alliance/home>)
5. ITU - AI for Good: a global summit organized by the International Telecommunications Union (ITU) in partnership with 40 UN sister agencies and co-convened with the government of Switzerland, with the aim of promoting AI to advance health, climate, gender, inclusive prosperity, sustainable infrastructure, and other global development priorities (Reference: <https://aiforgood.itu.int/about-ai-for-good/>)
6. AI Ethics Guidelines by ACM: The Association for Computing Machinery (ACM) has developed AI ethics guidelines to address ethical concerns in AI research and applications. (Reference: <https://www.acm.org/code-of-ethics>)
7. General Data Protection Regulation (GDPR): This is the EU regulation on data protection and privacy for all individuals within the European Union. It sets guidelines on data privacy and security. (Reference: <https://gdpr.eu/>)
8. California Consumer Privacy Act (CCPA): This is a California state law that gives residents more control over the personal information that businesses collect about them. (Reference: <https://oag.ca.gov/privacy/ccpa>)
9. Data breach on Ashley Madison dating website: news article "Hackers Finally Post Stolen Ashley Madison Data" (Reference accessed on 01.08.2023: <https://www.wired.com/2015/08/happened-hackers-posted-stolen-ashley-madison-data/>)
10. Strava fitness app data privacy issue: incident where the Strava app's heat map feature revealed sensitive military base locations. (Reference accessed on 01.08.2023: <https://www.forbes.com/sites/abrambrown/2022/06/20/strava-fitness-app-israeli-mossad-data-breach-security-hack-segments/?sh=459e82cc68d7>)
11. "Zooming into video conferencing privacy": article by Dima Kagan, Galit Fuhrmann Alpert, and Michael Fire published on **IEEE Transactions on Computational Social Systems**, Institute of Electrical and Electronics Engineers, online ISSN: 2329-924X. (Reference accessed on 01.08.2023: https://www.researchgate.net/publication/366813533_Zooming_Into_Video_Conferencing_Privacy)