Concours national

Promotion de l'Éthique Professionnelle

Co-organisé par le Rotary et la Conférence des Grandes Écoles

**Intelligence artificielle et enjeux éthiques :
Qui est responsable en cas d'erreur ?**

Démarche

L'intégration rapide de l'IA dans des secteurs critiques comme la santé, les transports ou la justice soulève des questions éthiques importantes, en particulier en cas d'erreur. À la différence d'une décision humaine, souvent contextualisée et justifiable, les décisions prises par une IA peuvent être opaques, difficilement explicables et poser un problème d'attribution de responsabilité.

Cette problématique m'a particulièrement interpellé. En tant que futur ingénieur, je serai amené à concevoir, exploiter ou encadrer ces systèmes. Si l'IA offre des perspectives prometteuses, elle soulève aussi des dilemmes inédits : comment garantir que ses décisions restent justes et sûres ? Qui doit être tenu responsable lorsqu'un algorithme cause un dommage ? Les législations actuelles peinent à suivre le rythme des avancées technologiques, rendant ces questions d'autant plus pressantes.

Résumé

On s'intéresse dans cet essai aux limites des réglementations actuelles face à l'essor fulgurant de l'intelligence artificielle. Loin d'être infaillible, l'IA peut engendrer des biais algorithmiques, conduire à des erreurs judiciaires automatisées ou encore influencer des décisions médicales de manière inappropriée. Ces dérives rappellent qu'elle n'est pas neutre, mais qu'elle reflète les choix et les intentions de ceux qui la conçoivent.

Face à cette réalité, plusieurs pistes sont envisagées : renforcer la transparence des algorithmes, établir des audits indépendants, repenser la répartition des responsabilités et adapter le cadre juridique aux spécificités de l'IA.

Références

- [1] Joanna J. Bryson. "The Ethics of Artificial Intelligence : Balancing Risk and Benefit". In : *Science and Engineering Ethics* (2018).
- [2] J. Dastin. "Amazon scraps secret AI recruiting tool that showed bias against women". In : *Reuters* (oct. 2018). url : www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG.
- [3] European Commission. *Proposal for a Regulation on a European Approach for Artificial Intelligence*. Available at : <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence>. 2021.
- [4] Martin Gibert. *Faire la morale aux robots : une introduction à l'éthique des algorithmes*. fr. Montréal : Atelier 10, 2020.
- [5] Frank Pasquale. *New Laws of Robotics : Defending Human Expertise in the Age of AI*. Harvard University Press, 2020.
- [6] Stuart Russell et Peter Norvig. *Artificial Intelligence : A Modern Approach*. Pearson, 2020.

- [7] Sandra Wachter, Brent Mittelstadt et Luciano Floridi. “Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation”. In : *International Data Privacy Law* (2017).

Introduction

Une voiture autonome roule en toute sécurité le long d'une route. Soudain, un imprévu : un piéton surgit sur la chaussée, et sur la seule voie d'évitement, un groupe de cyclistes arrive en sens inverse. L'algorithme doit trancher en une fraction de seconde : sauver le piéton et percuter les cyclistes, ou éviter les cyclistes et faucher le piéton ? Décision dramatique, mais surtout, décision programmée. Alors, qui est responsable ? Le constructeur du véhicule ? L'ingénieur qui a conçu l'algorithme ? Le propriétaire qui s'en remet aveuglément à la technologie ?

Loin d'être un concept nouveau, l'intelligence artificielle trouve ses racines dans les années 1950, lorsque des chercheurs tels qu'Alan Turing et John McCarthy ont posé les bases de machines capables de simuler l'intelligence humaine. Après plusieurs décennies de progrès en dents de scie, les avancées en apprentissage automatique et en puissance de calcul ont permis à l'IA de se déployer massivement dans de nombreux secteurs. Mais plus elle s'intègre à nos vies, plus elle soulève des dilemmes éthiques et juridiques : biais algorithmiques, erreurs judiciaires automatisées, diagnostics médicaux erronés... Des décisions autrefois prises par des humains sont désormais confiées à des systèmes opaques dont les erreurs peuvent avoir des conséquences majeures.

Ainsi, dans un monde où l'IA ne se contente plus d'assister l'humain mais prend des décisions à sa place, qui doit en assumer les erreurs ? L'absence de conscience morale de l'intelligence artificielle nous force à repenser notre conception de la responsabilité et de l'éthique.

Dans un premier temps, nous verrons comment l'IA, bien que prometteuse, demeure une entité opaque aux décisions parfois imprévisibles. Ensuite, nous analyserons les difficultés à attribuer une responsabilité claire entre concepteurs, entreprises et utilisateurs. Enfin, nous explorerons les pistes d'encadrement éthique et juridique permettant de limiter ces risques et d'assurer un développement responsable de l'intelligence artificielle.

1 L'intelligence artificielle : promesses et illusions

L'intelligence artificielle n'est plus une chimère futuriste. Elle est partout, s'infiltrant dans notre quotidien avec une aisance parfois déconcertante. De la reconnaissance faciale sur nos smartphones aux algorithmes qui nous suggèrent la prochaine série à binge-watch, l'IA redéfinit nos manières de produire, de décider et d'interagir. Pourtant, cette révolution technologique oscille entre promesse d'un avenir plus efficace et crainte d'un monde déshumanisé.

1.1 Une révolution technologique entre fascination et prudence

L'essor de l'IA repose sur des avancées spectaculaires en matière de puissance de calcul, d'algorithmes d'apprentissage et d'accès aux données. Son application dans des secteurs tels que la santé, la finance, le transport ou l'industrie engendre des gains d'efficacité remarquables. Qui aurait cru, il y a seulement quelques décennies, que des intelligences artificielles surpasseraient les médecins en radiologie, détectant des anomalies invisibles à l'œil humain ? Ou que des voitures autonomes seraient capables de conduire sur un trajet complet sans quelconque intervention humaine ?

Cependant, cet enthousiasme ne doit pas nous aveugler. Aussi impressionnante soit-elle, l'IA reste une « boîte noire » dont les décisions sont souvent difficilement explicables. Et cette opacité pose un sérieux problème de confiance et surtout de traçabilité de la responsabilité.

Un autre point critique réside dans les biais des données d'apprentissage. L'IA n'est ni neutre ni objective : elle reflète les tendances et préjugés contenus dans les informations qu'on lui fournit. Un exemple frappant est celui de l'algorithme de recrutement d'Amazon, qui, entraîné sur des données majoritairement masculines, a fini par discriminer les candidatures féminines [2]. Ce genre de dérive rappelle que l'IA, loin d'être une entité autonome et impartiale, reste le produit des choix humains qui la façonnent.

1.2 Quand la machine se trompe : les dérives imprévues

L'erreur est humaine... mais elle est aussi algorithmique. Contrairement à l'intuition humaine, les machines appliquent mécaniquement les règles qu'on leur impose, sans la moindre conscience morale. Cette rigidité algorithmique peut entraîner des décisions catastrophiques, et la question de la responsabilité devient alors centrale dans les débats.

Prenons l'exemple du système judiciaire. Aux États-Unis, l'algorithme COMPAS, utilisé pour prédire les risques de récidive, a été pointé du doigt pour ses biais raciaux. Certaines populations se sont retrouvées systématiquement désavantagées, non pas par une intention malveillante, mais parce que l'algorithme avait intégré des données historiquement biaisées. Qui doit être tenu responsable ? Les concepteurs ? Les juges qui s'appuient sur ces recommandations ? Dans le domaine médical, les dérives peuvent être tout aussi préoccupantes. Une IA de diagnostic mal calibrée peut engendrer des erreurs lourdes de conséquences, notamment si les données d'entraînement ne sont pas représentatives de la diversité des patients. Plutôt que de réduire les inégalités en santé, ces modèles risquent de les amplifier, renforçant ainsi un cercle vicieux difficile à briser.

Explicabilité des modèles, validation rigoureuse des résultats, supervision humaine accrue : autant de précautions nécessaires pour éviter que l'IA ne devienne un facteur d'injustice plutôt qu'un outil de progrès. Sur le plan juridique, nos législations actuelles peinent à encadrer ces nouvelles formes d'intelligence qui fonctionnent de manière autonome. Devons-nous envisager une responsabilité partagée entre les développeurs, les entreprises et les utilisateurs ? Ou créer un cadre réglementaire inédit pour les décisions prises par les machines ? La philosophe Adela Cortina met en garde contre une illusion dangereuse : croire que l'IA peut tout résoudre à

notre place. Derrière chaque algorithme, il y a des choix humains, des valeurs implicites et des décisions qui nécessitent une réflexion éthique approfondie. De son côté, Martin Gibert suggère d'enseigner aux machines une forme de morale en s'appuyant sur des dilemmes éthiques comme celui du tramway [4]. Mais jusqu'où peut-on aller ? Peut-on réellement déléguer nos choix moraux à des entités sans conscience ?

2 Une responsabilité en quête de clarté

L'essor fulgurant de l'intelligence artificielle s'accompagne d'un flou juridique qui rend la question de la responsabilité d'autant plus importante. Lorsqu'une erreur survient, qui doit en répondre ? Le développeur qui a conçu l'algorithme ? L'entreprise qui l'exploite ? L'utilisateur qui s'y fie aveuglément ? Ou même le régulateur qui en a validé l'usage ? L'absence de cadre clair alimente la méfiance et ouvre la porte à des abus. Pourtant, un encadrement solide est indispensable : une IA qui prend des décisions autonomes, parfois aux conséquences majeures, ne peut évoluer sans règles précises. Mais face à la vitesse vertigineuse des progrès technologiques, les législations peinent à suivre, toujours un temps de retard sur la réalité du terrain.

2.1 Un écosystème de responsabilités fragmentées

L'intelligence artificielle repose sur un maillage d'acteurs qui, chacun à leur niveau, influencent son développement et son utilisation. Comprendre cette chaîne est un premier pas pour identifier les responsabilités en cas de dysfonctionnement.

Les concepteurs et développeurs sont en première ligne : ce sont eux qui programment les algorithmes, entraînent les modèles et optimisent leur fonctionnement. Malgré leur expertise, anticiper toutes les dérives possibles reste un défi. Un algorithme peut évoluer de manière imprévue, s'adapter à de nouvelles données et, parfois, produire des résultats biaisés ou erronés. Qui doit alors assumer la responsabilité de ces erreurs ? Pour certains, la mise en place de protocoles d'évaluation stricts suffirait à limiter ces risques, mais la réalité montre que les biais émergent souvent bien après la phase de développement initiale. Les entreprises qui intègrent l'IA dans leurs services jouent un rôle tout aussi important. Elles en définissent les usages, mais souvent sans en maîtriser tous les rouages techniques. Un établissement bancaire utilisant un algorithme de notation de crédit a-t-il le devoir d'en comprendre les mécanismes internes ? Dans le secteur médical, un hôpital peut-il être tenu responsable si une IA de diagnostic entraîne un mauvais traitement ? L'absence de normes harmonisées pour ces entreprises complexifie encore davantage la gestion des responsabilités. Certaines firmes adoptent néanmoins des chartes éthiques, mais leur mise en œuvre reste variable selon les juridictions et les intérêts économiques [1].

L'utilisateur final est également concerné. Qu'il s'agisse d'un médecin suivant une recommandation algorithmique ou d'un employé appliquant une décision automatisée, leur rôle dans l'interprétation et l'utilisation des résultats est à prendre en compte. Peut-on exiger d'eux qu'ils remettent systématiquement en question les suggestions de l'IA ? Certains experts suggèrent une formation obligatoire à la lecture et à l'interprétation des résultats produits par les algorithmes, ce qui est déjà le cas dans certaines entreprises [7].

Enfin, une interrogation plus spéculative, mais de plus en plus discutée, porte sur la possibilité de reconnaître une forme de responsabilité propre à l'IA. Certains systèmes, notamment ceux reposant sur l'apprentissage profond, sont capables d'évoluer de manière autonome après leur mise en service. Si une IA commet une erreur ou agit de manière imprévue, est-il pertinent de la considérer comme une entité juridiquement responsable ? Cette hypothèse soulève des questions philosophiques et juridiques fondamentales, car reconnaître une responsabilité propre à l'IA nécessiterait de revoir entièrement les cadres législatifs existants.

L'absence de consensus sur ces questions laisse place à des situations où la responsabilité est souvent diluée entre les différents acteurs, chacun tentant d'éviter d'en assumer les conséquences.

2.2 Une réglementation en retard sur la technologie

L'intelligence artificielle évolue à une cadence qui laisse les législateurs à la traîne. Pensées pour des technologies plus figées, les réglementations actuelles peinent à encadrer des systèmes capables d'apprentissage autonome et d'adaptation en temps réel. Comment légiférer sur une IA dont les décisions ne sont pas entièrement prévisibles ? Comment attribuer la responsabilité lorsque celle-ci repose sur des milliers, voire des millions d'interactions de données ? Dans ce contexte, plusieurs pays tentent d'élaborer des cadres spécifiques, mais leur mise en œuvre reste souvent parcellaire [3].

Aujourd'hui, on tente tant bien que mal d'adapter des cadres juridiques existants à ces nouvelles réalités. Par exemple, la Directive européenne sur la responsabilité des produits défectueux fonctionne bien pour des machines classiques, mais devient bancal face à un algorithme évolutif qui modifie son propre comportement au fil du temps. Faudrait-il créer un nouveau statut juridique pour ces systèmes autonomes ? L'Union européenne tente d'apporter une réponse avec son projet d'Artificial Intelligence Act, qui vise à réguler les IA à haut risque. Mais ces initiatives restent lentes à se concrétiser et risquent d'être dépassées dès leur adoption. Le débat porte notamment sur la classification des IA et la mise en place de certifications obligatoires, un point de friction entre les régulateurs et les entreprises technologiques.

Aux États-Unis, l'approche est plus fragmentée. Certaines juridictions, comme l'État de Californie, ont mis en place des restrictions sur l'usage de la reconnaissance faciale, tandis que d'autres régions n'ont aucune réglementation spécifique. L'absence de cadre fédéral crée des disparités et des incertitudes pour les entreprises et les utilisateurs, freinant l'adoption de règles uniformes. D'autres pays, comme la Chine, adoptent une approche différente, où l'IA est fortement réglementée dans certains domaines stratégiques, mais favorisée dans d'autres.

L'essor des modèles capables d'apprentissage autonome remet en question les cadres de responsabilité traditionnels et nécessite une refonte des lois pour mieux encadrer les usages de ces systèmes. Sans réforme adaptée, les risques d'abus et d'injustices liés aux décisions algorithmiques continueront à proliférer, menaçant la confiance du public envers ces technologies émergentes.

3 Construire un avenir éthique pour l'IA

L'intelligence artificielle s'impose de plus en plus dans nos vies, influençant des décisions aussi cruciales que l'attribution d'un crédit, un diagnostic médical ou même une condamnation en justice. Pourtant, son développement rapide dépasse encore largement les garde-fous législatifs et éthiques censés encadrer son usage. Si l'IA représente un formidable levier d'innovation, elle peut aussi générer des erreurs, des discriminations ou des abus si elle est mal encadrée. La question n'est donc plus de savoir si l'IA doit être régulée, mais comment le faire efficacement, sans freiner le progrès tout en protégeant les citoyens.

Deux leviers majeurs se dessinent : une gouvernance plus responsable des entreprises qui exploitent ces technologies, et un cadre juridique repensé pour répondre aux spécificités de l'IA.

3.1 Une gouvernance responsable pour les entreprises

Les entreprises qui développent et utilisent l'intelligence artificielle ne peuvent plus se contenter d'innover sans se soucier des conséquences. Elles ont une responsabilité directe dans la manière dont ces technologies sont conçues, testées et mises en service. Laisser l'IA se déployer sans contrôle, c'est ouvrir la porte à des décisions biaisées, à des erreurs aux conséquences graves, et à une perte de confiance généralisée.

La transparence des algorithmes est un premier pas incontournable. Trop souvent, les décisions de l'IA sont perçues comme des boîtes noires, incompréhensibles pour les utilisateurs et même pour ceux qui les régulent. Obliger les entreprises à rendre publics les critères de décision de leurs modèles permettrait d'éviter certaines dérives, notamment les discriminations algorithmiques. Ce principe de transparence ne signifie pas forcément dévoiler l'intégralité du code source, mais au moins expliquer les logiques qui sous-tendent les choix de l'IA.

Un autre levier essentiel est la mise en place d'audits et de certifications indépendants. Avant de déployer une IA dans des secteurs critiques – santé, finance, sécurité – des contrôles devraient être systématiques pour s'assurer que ces systèmes respectent des normes éthiques et techniques strictes. Tests anti-biais, vérifications de robustesse, simulations de scénarios extrêmes... Autant de mesures qui permettraient de limiter les risques et d'établir une responsabilité claire en cas de problème [1].

La formation des utilisateurs est une nécessité trop souvent négligée. De nombreuses erreurs liées à l'IA ne viennent pas directement de l'algorithme lui-même, mais d'une mauvaise interprétation de ses résultats par les humains qui s'en servent. Un médecin qui suit aveuglément un diagnostic automatisé, un recruteur qui applique mécaniquement les recommandations d'un algorithme de tri de CV... Ces situations peuvent être évitées avec une meilleure sensibilisation aux limites et aux biais des systèmes d'IA [6].

3.2 Vers un cadre législatif adapté aux spécificités de l'IA

Si les entreprises ont un rôle à jouer, elles ne peuvent être laissées seules à définir les règles. Il semble important d'adapter le cadre juridique pour éviter un système où chacun se décharge de la responsabilité sur un autre acteur.

L'une des propositions les plus radicales, mais aussi les plus controversées, est la création d'une personnalité électronique pour les IA avancées. L'idée étant d'attribuer une forme de responsabilité juridique aux systèmes eux-mêmes, notamment ceux capables d'apprentissage autonome. Si cette mesure peut sembler futuriste, elle pose une vraie question : comment sanctionner un dommage causé par une IA qui a évolué de manière imprévue et dont aucun acteur humain ne peut être directement tenu responsable ?

Une solution plus pragmatique et immédiatement applicable serait d'imposer une assurance obligatoire pour les IA, sur le modèle de l'assurance automobile. Un système d'intelligence artificielle utilisé dans un domaine critique devrait être couvert pour les dommages qu'il pourrait causer, garantissant ainsi une indemnisation rapide des victimes en cas d'erreur. Cela permettrait aussi d'inciter les entreprises à mieux encadrer leurs systèmes, sous peine de primes d'assurance plus élevées si leurs modèles présentent trop de risques [5].

Enfin, la mise en place de comités d'éthique spécialisés pourrait apporter un garde-fou efficace pour superviser les usages sensibles de l'IA. Ces comités, composés d'experts en intelligence artificielle, en droit et en éthique, auraient pour mission de surveiller les applications les plus à risque, d'alerter sur les dérives potentielles et d'accompagner la régulation en temps réel. Contrairement aux régulations traditionnelles, souvent lentes et rigides, ces comités pourraient adapter leurs recommandations au rythme rapide des évolutions technologiques.

Conclusion

Comme la voiture autonome face à un choix impossible, notre société avance à grande vitesse vers un futur où l'IA prend des décisions critiques. Pourtant, nous persistons à croire que nous gardons les mains sur le volant, que nous pouvons maîtriser ces algorithmes qui, en réalité, sont façonnés par des décisions humaines invisibles, éparpillées dans les réseaux de neurones artificiels et des bases de données biaisées. La responsabilité ne peut être une simple équation, répartissant les torts entre concepteurs, entreprises et utilisateurs. L'enjeu est plus profond : sommes-nous prêts à accepter que certaines décisions échappent à toute morale humaine ? Car finalement, ce n'est pas l'IA qui choisit, mais les humains qui lui en donne les règles.