L'Intelligence Artificielle face au défi de la justice éthique : biais, responsabilité et transparence.

Présentation :

L'intelligence artificielle n'est plus une tendance, un effet de mode ou une nouveauté. Elle est déjà un fait, c'est-à-dire une pratique ancrée dans la vie quotidienne des citoyens, en tant que créateurs ou consommateurs de richesses ou de services. L'angle d'approche retenu ici est celui d'un questionnement des rapports déjà existants mais surtout futurs entre droit et technologie, plus précisément entre l'IA et la justice. "Bon droit a besoin d'aide" pour paraphraser Molière dans La comtesse d'Escarbagnas. En effet, le droit est par principe une science complexe, soumise à l'action humaine et donc à sa faillibilité. L'IA semble de prime abord la bonne idée, voire la solution quant aux échecs actuels auxquels la justice française fait face : illisibilité du droit, engorgement des tribunaux, obésité législative... Pourtant cette solution miracle relève bien des défauts, ayant de possibles conséquences désastreuses sur l'ordre juridique national et même international. Les professionnels du droit, dans un sens très largement entendu, de ceux qui le conçoivent à ceux qui l'appliquent et le défendent, sont en première ligne de cette grande inconnue. L'objectif de la démonstration est avant tout de faire surgir les risques, évidents pour certains mais aussi plus recherchés pour d'autres, mais également de proposer des solutions intéressantes et réalisables dans le but de renforcer ou étayer un système de garantie d'une justice éthique.

Résumé :

L'Intelligence Artificielle (IA) transforme de manière radicale la justice, soulevant ainsi des problèmes éthiques. Cet essai vise à explorer trois volets majeurs de sa prise en compte et de son implication dans notre quotidien : l'apport de l'IA sur les décisions, les risques en matière de biais algorithmiques, et les voies permettant de garantir des usages éthiques et transparents. Dans une première partie, nous avons étudié la façon dont l'IA apporte aux magistrats, l'analyse de la jurisprudence, l'automatisation de certaines procédures administratives et la détection de la fraude. Si l'avancée de l'IA permet d'envisager des gains de temps et d'efficacité, elle provoque de sérieuses interrogations sur la transparence des décisions et l'application de principes fondateurs du droit. Ensuite, nous avons mis en lumière les dérives possibles (biais algorithmiques, opérations sur les systèmes opacifiés). L'IA apparaît comme le vecteur d'un comportement discriminant à l'égard de certains. Le risque d'une justice standardisée où la machine s'introduit à la place de l'individualisation des peines est particulièrement inquiétant et doit être régulé et surveillé. Enfin, nous avons étudié les moyens de mieux encadrer l'IA dans ces milieux. Le renforcement du cadre légal par le biais d'institutions comme la CNIL, des modèles plus transparents, explicables, le maintien de la présence humaine, sont autant de pistes à explorer pour éviter les dérives. Ainsi, il s'agit de garantir que l'IA est un outil qui reste au service de la justice sans remplacer le jugement et l'engagement des humains.

Références bibliographiques

1. **Binns, R. (2018).** "Fairness in Machine Learning: Lessons from Political Philosophy." *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*.
2. **Dwork, C., & Mulligan, D. K. (2013).** "It's Not Privacy, and It's Not Fair." *Stanford Law Review Online*, 66, 35-40.
3. **Rouvroy, A., & Berns, T. (2013).** "Gouvernementalité algorithmique et perspectives d'émancipation." *Réseaux*, 177, 163-196.

Références numériques et rapports officiels

5. **Commission Nationale de l'Informatique et des Libertés (CNIL).** (2021). *Les algorithmes au service de l'action publique : enjeux éthiques et bonnes pratiques*.
<https://www.cnil.fr>
6. **Conseil d'État.** (2022). *Puissance publique et plateformes numériques : accompagner la transition numérique*.
<https://www.conseil-etat.fr>
7. **Parlement Européen.** (2023). *Règlement européen sur l'intelligence artificielle (AI Act)*.
<https://www.europarl.europa.eu>
8. **RGPD (Règlement Général sur la Protection des Données).** (2018).
<https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679>
9. **Article de presse :** "Algorithme de ciblage antifraude dans les CAF : des associations saisissent le Conseil d'État." *Le Monde*, octobre 2024.
<https://www.lemonde.fr>

Face à un monde en perpétuelle évolution géopolitique et scientifique, le développement de l'Intelligence Artificielle (IA) s'établit comme un véritable tournant au sein de notre société et de nos institutions. En effet, dès lors majoritairement utilisée pour un usage récréatif, l'Intelligence Artificielle s'illustre comme un outil majeur au sein de certaines fonctions vitales de notre pays telle la justice ou encore l'économie en étant capable de traiter une large quantité de données, établir un modèle de comportement et en déduire des prédictions.

Se basant sur ces nouveaux enjeux et besoins d'évolutions, la France a dès janvier 2024, avec l'annonce du ministère de la Justice choisie l'utilisation de l'intelligence artificielle (IA) dans milieu de la justice. Ainsi, 4 axes d'utilisations ont été choisis par la Justice : la retranscription des entretiens utiles à l'ensemble des métiers de la justice, l'aide à la recherche des textes juridiques et des jurisprudences, l'interprétariat et traduction permettant d'améliorer la compréhension des justiciables et réduire la barrière de la langue et finalement le résumé des dossiers en synthétisant un volume important de documents juridiques pour un traitement plus rapide.

Ces avancées témoignent du rôle croissant de l'IA dans la prise de décision publique et privée.

Néanmoins, bien que source d'intérêt, cette montée en puissance de l'utilisation de l'Intelligence Artificielle et d'algorithmes complexes soulève des questions éthiques fondamentales. En 2024, une affaire révélée par Le Monde a soulevé l'utilisation par la Caisse nationale des allocations familiales (CNAF) d'algorithmes permettant d'identifier les allocataires qui doivent être contrôlés en priorité en raison d'un plus grand risque de faute au sein de leur dossier. Une note entre 0 et 10 est alors attribué à l'utilisateur. Ainsi, il est légitime de se demander : peut-on confier des décisions de justice pouvant avoir un impact direct sur la vie de personnes à une Intelligence Artificielle ? Quels sont les risques éthiques de discrimination liés aux biais algorithmiques ? Et comment garantir que les algorithmes respectent au mieux les principes fondamentaux du droit et de l'équité ?

Une problématique semble alors émerger de ces constats :

Dans quelle mesure l'IA peut-elle être utilisée de manière éthique et juste dans la prise de décision judiciaire, sans compromettre les principes d'éthique et de transparence ?

Dans un premier temps nous verrons tout d'abord comment l'IA transforme la justice, en offrant des avantages considérables mais aussi des défis majeurs. Puis, dans un second temps, nous analyserons les biais algorithmiques et leurs risques d'iniquité, en mettant en lumière leur impact sur le système judiciaire. Pour conclure, nous explorerons les solutions et régulations mises en place pour garantir un usage éthique et transparent de ces technologies.

Ainsi, dans un premier temps de cette réflexion autour de l'usage de l'Intelligence Artificielle au sein du domaine judiciaire nous allons nous intéresser à la définition et aux enjeux éthiques de l'Intelligence Artificielle appliquée à la justice.

1. Définition et enjeux éthiques de l'IA appliquée à la justice :

Se définissant comme l'ensemble des principes moraux définissant la conduite d'un individu, l'éthique repose sur des valeurs fondamentales telles que la justice, la liberté, l'équité et la dignité. Celle-ci permet alors d'établir les repères nécessaires à la différenciation du bien et du mal, du juste et de l'injuste au regard de notre société et de nos usages et coutumes dans ce que Charles Taylor appelle une « évaluation forte » dans Sources of the self.

Dépassant la simple pensée philosophique, l'éthique se retrouve dans de nombreux domaines scientifiques telle que la médecine et les nouvelles technologies. Ainsi, au sein du milieu de la justice, l'éthique vise à garantir des droits fondamentaux quant à l'utilisation de l'Intelligence Artificielle. Parmi ces droits, on retrouve les notions d'équité, d'impartialité et de transparence des décisions. Ces principes s'appuient alors sur des notions fondamentales de la justice avec l'égalité devant la loi, le droit à un procès entendu équitablement, publiquement et dans un délai raisonnable, par un tribunal indépendant et impartial – Article 6 de la Convention européenne des droits de l'homme – et la protection des données personnelles. Ainsi, le rôle de l'éthique est de garantir, lors de l'intervention de l'Intelligence Artificielle au sein d'une décision judiciaire, l'absence de toute dérive arbitraire.

Maintenant que nous avons vu la définition de l'éthique appliqué à notre domaine de l'Intelligence Artificielle au sein du monde de la justice. Il semble important d'apporter des exemples concrets de cette utilisation de l'IA au sein du monde juridique et de comment ces outils constituent à la fois des opportunités mais également des risques.

Ainsi, l'Intelligence Artificielle de par son fonctionnement et sa réalisation de modèle de comportement peut être une aide précieuse pour l'aide à la décision judiciaire. C'est notamment le cas des logiciels suivants : Predictice outil reposant sur l'IA qui propose de calculer les chances de succès d'une action en justice selon le fondement légal, le montant des indemnités et la durée de la procédure grâce à des algorithmes d'analyse de la jurisprudence – Case Law Analytics qui analyse un dossier pénal à la recherche de vices de forme pour contrôler la conformité d'une procédure. Ces outils permettent alors d'identifier les comportements concernant les décisions rendues par les différentes juridictions et d'adapter la stratégie judiciaire en conséquence. La plateforme Kléros a été également pionnier en la matière notamment dans le domaine de la justice prédictive et de résolution des différends, par l'IA et pour l'IA, en matière de smart contrat spécifiquement.

Les avantages offerts par ces solutions sont alors nombreux et importants pour le milieu judiciaire où le nombre de dossiers est de plus en plus important. Ils permettent alors la réduction du temps de traitement des affaires par l'élimination des vices de forme, la synthèse du dossier et l'optimisation de la stratégie des avocats. Loin d'être un simple gadget, l'utilisation de l'IA concourt directement à la préservation de grands principes comme le droit à un jugement, la clarté et

l'intelligibilité de la loi ou encore la bonne administration de la justice. Cependant, s'accompagnant du revers de la médaille, les risques sont tout aussi nombreux concernant l'utilisation de l'Intelligence Artificielle dans le processus de décision judiciaire. Les magistrats pourraient être influencés par ces prédictions au lieu d'exercer pleinement leur pouvoir souverain de jugement, les grands cabinets d'avocat ayant accès à une base financière plus solide pourrait s'offrir des outils à la pointe pour la gestion de leurs dossiers à contrario des avocats commis d'office favorisant les clients en fonction de leurs revenus, de même en fonction de la nature du dossier et de sa potentielle réussite certains avocats pourraient être tentés de refuser certains clients afin de préserver leurs réputations privant le client du droit fondamental d'être défendu devant un tribunal.

Cette intégration de l'Intelligence Artificielle pose alors de nombreux soucis éthiques majeur pour le fonctionnement de notre justice. Les algorithmes de ces produits sont opaques pour les utilisateurs et les magistrats. Les modèles reposent sur un fonctionnement en « boîte noire » pour l'utilisateur. Or dans le cadre d'une justice équitable, le principe de transparence est essentiel. De plus, ces outils touchent directement au principe d'individualisation de la peine promu par le système français, le juge doit s'adapter au contexte et aux circonstances de l'affaire et ne doit pas uniquement se baser sur un raisonnement statistique établi par une Intelligence Artificielle. Une véritable appréciation humaine est nécessaire pour ne pas déshumaniser le jugement.

L'utilisation de L'IA dans la justice française représente un enjeu important en matière de rapidité et d'accessibilité du droit. Cependant, elle soulève aussi de nombreuses préoccupations éthiques, notamment en ce qui concerne la transparence et le principe d'individualisation du jugement. L'enjeu est donc de trouver un équilibre entre innovation et justice éthique.

2. Les biais algorithmiques et leurs impacts sur l'équité et la justice

Ainsi, nous avons vu dans notre première partie comment l'IA transforme la justice, en offrant des avantages considérables en termes de rapidité et d'analyse de dossiers mais dresse également des défis majeurs en termes d'éthique par un équilibre nécessaire entre humain et automatisation.

Afin de fonctionner correctement l'IA va alors devoir faire face à un défi éthique majeur, la présence possible de biais algorithmiques au sein des modèles de comportement qui risquent de compromettre l'équité des décisions.

Cependant, avant de continuer cette partie il semble nécessaire de définir la notion de biais algorithmique. Un biais algorithme désigne une erreur présente dans les résultats produits par un modèle d'Intelligence Artificielle causés par des erreurs dans la conception du système ou dans le corpus de données utilisées pour entraîner le modèle de notre IA.

L'apparition de ces biais peut être de différentes natures. En effet, nous retrouvons 3 grands types de biais : les biais liés aux données d'entraînement pouvant concerner une IA entraînée sur des dossiers judiciaires possédant des discriminations ou des injustices (Peines variables en fonction du sexe, âge, origine) – les biais de conception liés à la nature humaine des algorithmes (en effet, conçu par des humains, les algorithmes héritent alors des préjugés et biais associés) – les biais d'usage dépendant de la bonne interprétation ou non des résultats obtenus par les professionnels du droit.

Ainsi, nous avons défini les 3 grands types de biais pouvant avoir une influence directe sur l'équité et les décisions de justice. En nous appuyant sur les outils décrit dans la partie 1, nous allons voir

les conséquences possibles d'un biais sur le traitement d'un dossier en matière de décision judiciaire.

Comme décrit précédemment, les outils Predictice ou encore Case Law Analytics permettent à partir d'une analyse de décisions passées d'estimer un futur verdict. Cependant, si cette analyse est biaisée par des données erronées – erreur de jugement, discrimination de sexe, genre, métier, etc. l'IA risque de reproduire l'erreur menant à une prédiction statistique fautive et pouvant imposer un jugement plus lourd à certains profils non pas liés à la nature des faits mais lié à un schéma biaisé si le juge suit la prédiction de l'IA.

C'est notamment le cas du logiciel américain COMPAS - *Correctional Offender Management Profiling for Alternative Sanctions* – majoritairement utilisés par la justice américaine comme aide à la décision. Les résultats obtenus avec les prédictions du programme comparé à des prédictions d'humains sans expérience du domaine judiciaire sont identiques posant alors la question de la justesse de ce logiciel. De plus, un biais concernant les prévenus noirs attribuait des peines plus lourdes qu'à des personnes blanches pour des crimes similaires.

Mais alors, en cas d'erreur et de biais, qui est le responsable ? En effet, si une injustice est commise due à un biais algorithmique qui doit être tenu responsable de cette erreur et sera judiciairement responsable ?

3. Vers une IA éthique et transparente : quelles solutions ?

L'Intelligence Artificielle (IA) joue un rôle croissant dans les décisions judiciaires et économiques, mais son utilisation soulève de nombreuses préoccupations éthiques, notamment en matière de biais, de transparence et de responsabilité. Pour garantir une justice équitable et un usage éthique de l'IA, plusieurs solutions sont envisagées, impliquant à la fois les législateurs, les entreprises, les magistrats et les citoyens.

Dans un premier temps, il est nécessaire de souligner que si l'IA est novatrice, l'État français et donc le spectre judiciaire, n'en sont pas totalement déconnectés. En effet, le droit français a déjà intégré dans son système propre des moyens de contrôler, d'endiguer les dérives que les nouvelles technologies au sens large peuvent causer. Adopté en 2018, le Règlement Général sur la Protection des Données (RGPD) est un texte précurseur qui encadre le traitement des données personnelles en Europe et impose plusieurs règles aux systèmes algorithmiques quant à la transparence et la traçabilité ou encore le droit d'explication (article 22). Au-delà de la simple action nationale, la France s'investit dans les projets européens dont l'AI Act fait partie, afin de réguler au niveau de l'Union les systèmes et les pratiques.

Les réponses issues de ces encadrements existants sont prévisibles mais fondamentaux :

- Interdire certaines pratiques dangereuses (notation sociale, IA prédictive dans le cadre judiciaire).
- Exiger des audits indépendants pour les IA utilisées dans des décisions ayant un impact sur les citoyens.
- Imposer une supervision humaine obligatoire sur les décisions assistées par l'IA.

Dans un deuxième temps, il convient de s'interroger sur la suffisance de ce cadre légal existant, et sur les possibilités d'améliorations.

La question se pose in concreto, car les algorithmes et les systèmes se développent de manière exponentielle, et le droit bien souvent ne peut agir qu'à posteriori.

De nombreux rapports soulignent les besoins en matière d'IA : le Conseil d'État encourage la transparence des algorithmes publics (rapport "Puissance publique et plateformes numériques", 2022), le Conseil national du numérique (CNNum) travaille sur des recommandations pour une IA plus responsable ou encore la CNIL effectue déjà des audits sur certains algorithmes pour s'assurer qu'ils respectent les principes de non-discrimination.

Cependant, les régulations existent en théorie, les intentions louables sont formulées mais bien souvent la mise en œuvre est insuffisante du fait de contournements par les entreprises privées dans une logique de profit.

Une solution envisagée par l'État français et l'Union Européenne réside dans l'élaboration d'un label. Un label de fiabilité et de sûreté, délivré par les institutions étatiques afin de garantir l'absence de vices cachés au sein d'un algorithme. Créer des systèmes nationaux, estampillés république française, assurément transparents envers les utilisateurs et le public, permettrait de favoriser cette transparence et cette fiabilité dont la justice a besoin. Ce label permettrait également d'effectuer des audits indépendants d'algorithmes externes afin de faire contrôler par des experts extérieurs, la sécurité des systèmes publics ou privés. En effet, les informations contenues dans les dossiers juridiques peuvent porter gravement atteinte à la vie privée des personnes ou à leur présomption d'innocence s'ils venaient à être dévoilés, la sécurité des dits systèmes est donc primordiale.

La CNIL et l'ARCEP bénéficient déjà de prérogatives de contrôles mais le manque de moyens empêche une vérification efficace et réellement contraignante. La création d'une agence française associée à des moyens suffisants pourrait permettre la création effective d'un tel label et la mise en application des différentes mesures contraignantes pour faire de l'IA un outil et non une menace pour la justice française.

Outre la répression et le contrôle étatique, les citoyens doivent être informés de leurs droits en matière d'IA et savoir comment contester une décision injuste prise par un algorithme. Les associations et les médias ont aussi un rôle clé à jouer dans la dénonciation des dérives et la mise en lumière des pratiques abusives. Dans le rang des propositions au regard des enjeux de la justice numérique éthique, la création d'un "Défenseur des droits numériques", permettrait de rendre justiciable ces nouveaux différends en recevant et traitant les plaintes liées aux décisions automatisées discriminatoires.

Mais au-delà des administrés qui bénéficient de cette transparence, les professionnels doivent également être formés pour être responsabilisés : l'éthique de l'IA ne dépend pas seulement des régulations, mais aussi de la responsabilité des acteurs qui conçoivent, utilisent et subissent ces technologies. Il est aujourd'hui nécessaire, si l'on veut intégrer l'IA dans la justice, de former les professionnels utilisateurs de tels outils, de démocratiser les guides de bonnes pratiques afin de s'assurer de la bonne utilisation éthique de tels systèmes de sorte à prévenir tout débordement.

Les magistrats et avocats doivent être formés pour comprendre les outils d'IA qu'ils utilisent, afin d'éviter une dépendance aveugle aux algorithmes, les entreprises et les banques doivent être sensibilisées aux risques des biais algorithmiques et encouragées à concevoir des modèles plus équitables. Il serait intéressant d'intégrer des modules d'éthique de l'IA dans la formation des juges, avocats et économistes pour leur permettre d'exercer un regard critique sur ces technologies.

Pour finir, la question de la responsabilité, notamment des concepteurs de systèmes, doit être réinterrogée afin de trancher sur les responsabilités personnelles et collectives en matière de développement d'algorithmes discriminatoires.

Après avoir défini l'IA et les enjeux attachés à son intégration pour une justice éthique, c'est-à-dire faire de l'IA une véritable aide à la justice en préservant ses grands principes ; il a fallu souligner les points noirs d'une utilisation abusive et irresponsable de certains algorithmes en matière de justice.

Pour cela au-delà des 4 axes qui font la colonne vertébrale du développement de l'IA au sein de la justice française (la retranscription des entretiens, l'aide à la recherche des textes juridiques et des jurisprudences, l'interprétariat et traduction et finalement le résumé des dossiers) il semblait opportun de proposer des solutions dont deux majeures méritent une attention particulière : la création d'un label garantissant un algorithme éthique, mais surtout une réelle responsabilisation des acteurs et des administrés dans une logique kantienne dans sa réflexion sur l'éducation estimant que la nature destine l'homme à réaliser lui-même son achèvement, notamment en connaissant ce qui lui est nuisible et en l'enseignant aux autres, qui doit conduire à la pérennisation et au progrès de l'humanité « L'homme ne peut devenir Homme que par l'éducation. Il n'est que ce que l'éducation fait de lui ».

Le rapport Villani (Donner un sens à l'intelligence artificielle, 2018) préconise une IA "éthique par design", intégrant des garde-fous dès sa conception. Passé la fougue du créateur, il faut s'intéresser à la sagesse du régulateur. Si l'IA est source de progrès exceptionnels, elle emporte avec elle de nombreuses menaces. Dans la logique de Villani, n'est-il pas temps de renverser la tendance ? Au lieu de laisser la main aux créateurs des systèmes pour réguler par la suite, n'est-il pas envisageable de prendre les devants, soit en contrôlant la création par des barrières données aux intéressés, soit en étant directement acteur du développement français de l'IA ?