Strasbourg, 24 Juin 2020 CAHAI(2020)06-fin

# COMITE AD HOC SUR L'INTELLIGENCE ARTIFICIELLE (CAHAI)

## Impact de l'intelligence artificielle sur les droits de l'homme, la démocratie et l'État de droit

Rapport par Catelijne Muller, LLM

www.coe.int/cahai

Les opinions exprimées dans cette analyse ne reflètent pas nécessairement la position du CAHAI ou du Conseil de l'Europe.

# Table des matières

**I. Remarques liminaires** ..... 3

**II. L'intelligence artificielle : définition** ..... 4

**III. Impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit** ..... 6

i) IA et respect de l'être humain ..... 6
ii) IA et liberté individuelle ..... 9
iii) IA et égalité, non-discrimination et solidarité ..... 11
iv) IA et droits sociaux et économiques ..... 12
v) IA et démocratie ..... 13
vi) IA et État de droit ..... 15

**IV. Que faire face à l'impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit ?** ..... 16

i) Les droits de l'homme dans le contexte de l'IA ..... 16
ii) Mesures visant la conformité, la réparation et l'obligation de rendre des comptes ..... 16
iii) Protection des structures démocratiques et de l'État de droit ..... 18

**V. Et si les normes actuelles de droits de l'homme, de démocratie et d'État de droit ne nous protègeaient pas assez ?** ..... 19

i) Question zéro ..... 19
ii) Lignes rouges ..... 20
iii) Des droits de l'homme nouveaux ou adaptés ..... 21
iv) Scénarios d'avenir ..... 21

## I. Remarques liminaires

1. L'intelligence artificielle (IA), comme technologie généraliste, a des incidences sur tout le tissu social. En 2017, le Comité économique et social européen, dans ce qui est largement considéré comme le "rapport initial" sur l'impact sociétal plus large de l'IA, identifie, les principaux domaines touchés : la sécurité, l'éthique, le droit et réglementations, la démocratie, la transparence, la vie privée, le travail, l'éducation et l'(in)égalité[1]. Cela signifie que l'IA a un impact sur les droits de l'homme, la démocratie et l'État de droit, à savoir les fondements mêmes des sociétés européennes.

2. En 2019, le Groupe d'experts de haut niveau sur l'IA a présenté ses Lignes directrices en matière d'éthique pour une IA digne de confiance[2]. Ces lignes directrices affirment qu'une IA « digne de confiance » doit être licite, éthique et robuste sur le plan technique et social. S'agissant de l'aspect éthique, elles axent expressément leur approche sur les droits fondamentaux[3]. Bien que certains éléments mentionnés dans ces lignes directrices découlent directement de droits (de l'homme) existants, ils n'ont pas en eux-mêmes de caractère contraignant. Or, on assiste depuis peu à une montée des appels à créer des instruments juridiquement contraignants en matière d'IA, ou à renforcer ceux qui existent. La Commission européenne a annoncé dans son Livre blanc sur l'IA[4] les éléments potentiels d'un cadre législatif, et souligne l'importance d'une IA respectueuse des droits fondamentaux de l'UE et des lois qui en assurent l'application.

3. Le présent document souligne l'impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit. Il identifie les droits de l'homme, tels qu'énoncés dans la Convention européenne des droits de l'homme (« CEDH »), ses Protocoles et la Charte sociale européenne (« CSE »), qui sont actuellement les plus touchés par l'IA ou les plus exposés (chapitre II). Aux chapitres III et IV, notre rapport avance plusieurs stratégies possibles, applicables simultanément si nécessaire. Le chapitre III s'intéresse aux mesures adoptées face à cet impact dans le cadre existant des droits de l'homme, de la démocratie et de l'État de droit. Le chapitre IV examine les stratégies envisageables si ce cadre ne s'avérait pas assez protecteur. Compte tenu de l'évolution rapide des technologies et de la société, le présent document ne saurait prétendre à l'exhaustivité, mais il insiste sur les impacts les plus importants parmi ceux que l'on peut identifier à ce jour.

[1] Avis du CESE sur l'IA et la société (INT/806, 2017).
[2] Lignes directrices en matière d'éthique pour une IA digne de confiance, présentées à la Commission européenne par le Groupe d'experts de haut niveau sur l'IA, 2019.
[3] Les lignes directrices décrivent plus loin sept exigences que devrait respecter une IA digne de confiance (licite, éthique et robuste) : 1) Action humaine et contrôle humain ; 2) Robustesse technique et sécurité ; 3) Respect de la vie privée et gouvernance des données ; 4) Transparence ; 5) Diversité, non-discrimination et équité ; 6) Bien-être sociétal et environnemental ; 7) Responsabilité.
[4] Livre blanc sur l'intelligence artificielle, COM(2020)65 final.

## II. L'intelligence artificielle : définition

4. Une multitude d'applications de l'IA ont déjà cours au sein de nos sociétés : reconnaissance biométrique (y compris faciale), reconnaissance des objets, prédiction des risques et des réussites, (aide à la) prise de décision algorithmique, traduction automatique, systèmes de recommandations, etc. Ces applications ont trouvé leur place dans des secteurs comme les forces de l'ordre, la justice, la gestion des ressources humaines, les services financiers, les transports, la santé, les services publics, etc.

5. L'IA reste une notion très controversée, sans définition universellement reconnue. Néanmoins, on peut grouper ses définitions en deux familles : rationalistes d'une part, centrées sur l'humain d'autre part. L'une des définitions rationalistes les plus connues présente l'IA comme « un agent, créé par des êtres humains, qui prend des décisions et accomplit des actes sur la base de sa perception[5] ». L'une des définitions centrées sur l'humain les plus connues est la réussite au test de Turing, destiné aux systèmes informatiques accomplissant une tâche qui requiert habituellement une intelligence humaine (en particulier l'aptitude à la conversation). Le Groupe d'experts de haut niveau sur l'IA a proposé en 2019 une définition de l'IA[6] :

> *Les systèmes d'intelligence artificielle (IA) sont des systèmes logiciels (et éventuellement matériels) conçus par des êtres humains et qui, ayant reçu un objectif complexe, agissent dans le monde réel ou numérique en percevant leur environnement par l'acquisition de données, en interprétant les données structurées ou non structurées collectées, en appliquant un raisonnement aux connaissances, ou en traitant les informations dérivées de ces données et en décidant de la/des meilleure(s) action(s) à prendre pour atteindre l'objectif donné. Les systèmes d'IA peuvent soit utiliser des règles symboliques, soit apprendre un modèle numérique. Ils peuvent également adapter leur comportement en analysant la manière dont l'environnement est affecté par leurs actions antérieures.*

6. L'IA est souvent décrite comme un ensemble de technologies combinant données, algorithmes et capacité de traitement informatique. Bien que ce soit vrai des systèmes d'IA les plus répandus aujourd'hui, cette description ne concerne qu'une petite partie de l'IA. Le terme d'intelligence artificielle s'étend à de nombreuses applications informatiques ; certaines associent des données et des algorithmes, mais il en existe d'autres non fondés sur les données, comme les systèmes experts ou les systèmes de connaissance, raisonnement et représentation, de planification réactive, d'argumentation, etc.

[5] S. J. Russell, P. Norvig et E. Davis (2010), *Artificial intelligence: A modern approach* (3e éd.), Prentice Hall.
[6] Groupe d'experts de haut niveau de l'UE sur l'IA, *A definition of AI, main capabilities and scientific disciplines*, 2019.

7. La plupart des systèmes d'IA récemment apparus dans nos sociétés sont des exemples d'IA fondée sur les données et ont des incidences particulières sur les droits de l'homme, la démocratie et l'État de droit.

8. Les termes les plus pertinents sont définis ci-après :
   - IA restreinte : systèmes d'IA uniquement capables d'accomplir des tâches restreintes, très spécifiques ;
   - Données massives (archivées) : systèmes d'IA nécessitant un grand volume de données (archivées) pour fonctionner correctement. La qualité, le volume et le contenu des données influencent le fonctionnement des systèmes et entraînent souvent la reproduction et l'amplification des biais, erreurs et lacunes présents dans les données ;
   - Corrélation : beaucoup de systèmes d'IA n'explorent que les relations entre données. Ils n'établissent (ne « voient ») pas de lien de cause à effet entre un cas et une décision, mais effectuent des prédictions sur la base des traits communs avec d'autres cas ;
   - Boîtes noires : beaucoup de systèmes d'IA sont ce qu'on appelle des boîtes noires, c'est-à-dire que les processus (décisionnels) qui s'y déroulent ne peuvent être entièrement expliqués en termes humains ;
   - Bon sens : les systèmes d'IA n'ont pas de « bon sens ». Même capables de reconnaître un chat ou une cellule cancéreuse, ils ne conçoivent pas ce qu'est un chat ou une cellule cancéreuse, mais se contentent d'étiqueter un phénomène spécifique. Ces systèmes sont incapables, par ailleurs, d'utiliser ce qu'ils « savent » des chats ou des cellules cancéreuses pour reconnaître un chien ou une céphalée.

   Toutes ces caractéristiques peuvent rendre l'IA actuelle fragile, instable et imprévisible – mais aussi populaire, et très largement adoptée.

9. Enfin et surtout, les systèmes d'IA sont plus que la somme de leurs composants logiciels. Ils comprennent aussi tout le système sociotechnique qui les entoure. La question de leur gouvernance ne devrait donc pas se focaliser sur la technologie, mais s'étendre aux structures sociales, c'est-à-dire aux organisations, personnes et institutions qui imaginent, développent, déploient, utilisent et contrôlent les systèmes d'IA et aux personnes que ces systèmes touchent – les citoyens dans leurs relations avec l'administration, les consommateurs, les travailleurs, voire la société tout entière.

Définition de l'IA à des fins juridiques

10. Les définitions juridiques sont encore plus difficiles à établir que les définitions purement scientifiques. Elles doivent en effet satisfaire à plusieurs critères[7] (être par exemple

[7] « A Legal Definition of AI », Jonas Schuett, Université Goethe de Francfort, 4 septembre 2019 (« Les définitions juridiques doivent être : 1) inclusives : les buts de la réglementation ne doivent pas concerner un public trop vaste ou trop limité ». (Julia Black, *Rules and Regulators*, *Oxford University Press* , 1997[32] ; Robert Baldwin, Martin Cave et Martin Lodge, *Understanding Regulation: Theory, Strategy, and Practice*, Oxford University Press, 2e édition, 2012) ; « 2) précises : la limite entre les cas qui relèvent de la définition et ceux qui n'en relèvent pas doit être claire ; 3) compréhensibles : ceux auxquels la réglementation s'applique doivent comprendre la définition ; 4) pratiques : les

inclusives, précises, compréhensibles, pratiques et permanentes), dont certains sont obligatoires et d'autres considérés comme des bonnes pratiques[8].

## III. Impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit

11. Il est important d'aborder tout le « cycle de vie de l'IA », c'est-à-dire non seulement sa phase de conception, mais aussi son déploiement et son utilisation. Il faut garder à l'esprit, par ailleurs, que la plupart des applications de l'IA actuellement en usage pourraient avoir d'importants impacts sur les droits de l'homme, la démocratie et l'État de droit ou exacerber et amplifier les impacts existants, affectant simultanément de plus en plus de personnes et des pans de plus en plus vastes de la société.

12. L'IA a des répercussions sur quatre « familles de droits de l'homme » en vertu de la CEDH, de ses Protocoles et de la CSE :
    i) Le respect de l'être humain ;
    ii) La liberté individuelle ;
    iii) L'égalité, la non-discrimination et la solidarité ;
    iv) Les droits sociaux et économiques.

    En outre, l'IA a de forts impacts sur :
    v) La démocratie ;
    vi) L'État de droit.

    Il est important de noter que beaucoup de systèmes d'IA, ou leurs usages, peuvent affecter simultanément la démocratie, l'État de droit et plusieurs droits de l'homme différents. Ils peuvent aussi avoir des effets négatifs sur les droits d'une personne, mais positifs sur ceux d'une autre.

### i) IA et respect de l'être humain

13. Le respect de l'être humain transparaît dans plusieurs droits affirmés par la CEDH, comme le droit à la liberté et à la sécurité (article 5), le droit à un procès équitable (article 6), le principe « pas de peine sans loi » (article 7) et le droit à la vie privée et à l'intégrité physique et mentale (article 8). Nous allons voir comment l'IA peut affecter ces droits.

<u>Liberté et sécurité, procès équitable, pas de peine sans loi (art. 5, 6 et 7 CEDH)</u>

14. Le fait que l'IA puisse perpétuer ou amplifier les biais existants est particulièrement significatif lorsqu'elle est utilisée par les forces de l'ordre et le pouvoir judiciaire. Lorsque la liberté physique ou la sécurité d'une personne sont en jeu, par exemple lorsque l'IA sert à prédire des infractions, à estimer le risque de récidive ou à fixer une peine, le droit à la

professionnels du droit devraient pouvoir déterminer aisément si un cas relève ou non de la définition ; 5) permanentes : une définition juridique ne devrait pas nécessiter de constantes mises à jour ».

[8] L'obligation d'inclusivité découle du principe de proportionnalité, en droit de l'UE (article 5.4 du traité sur l'Union européenne). Toujours en droit de l'UE, les critères de précision et de caractère compréhensible reposent sur le principe de sécurité juridique. La praticité et la permanence sont considérées comme de bonnes pratiques législatives.

liberté et à la sécurité ainsi que le droit à un procès équitable se trouvent exposés. Un système d'IA servant à prédire la récidive ou à fixer une peine peut livrer des résultats biaisés. Or, si ce système est une boîte noire, les professionnels du droit, comme les juges, les avocats et les procureurs, ne peuvent comprendre par quel raisonnement il a abouti à ses résultats, ce qui rend la décision difficile à motiver et donc à contester.

15. L'IA a aussi un impact, bien que moins évident, sur le droit au soupçon raisonnable et sur l'interdiction de l'arrestation arbitraire. Les applications de police prédictive par IA cherchent uniquement des corrélations, sur la base de traits communs avec d'autres cas. Au lieu de soupçonner un crime ou un délit de la part d'un suspect un particulier, les enquêteurs se fondent alors uniquement sur les caractéristiques que le suspect partage avec d'autres (adresse, revenus, nationalité, niveau d'endettement, profession, comportement, comportement de ses amis ou de ses proches, etc.). En outre, les caractéristiques exactes prises en compte par le système d'IA, et leur pondération, demeurent souvent inconnues.

16. Cependant, maniés de façon responsable, certains outils d'IA peuvent s'avérer bénéfiques pour la sécurité. Certaines applications, par exemple, sont capables de « vieillir » les personnes disparues sur les photographies pour augmenter les chances de les retrouver, ou de reconnaître certains objets dans les bagages scannés à l'embarquement dans un aéroport.

Vie privée et familiale ; intégrité physique, psychologique et morale (art. 8 CEDH)

17. Parmi les systèmes d'IA et leurs usages, beaucoup ont de vastes et profondes répercussions sur le droit à la vie privée. Actuellement, les débats dans ce domaine tendent à se centrer sur la protection des données privées et sur la non-discrimination dans le traitement des données personnelles (et non personnelles). Cependant, malgré l'importance de ces aspects, l'impact de l'IA sur notre vie privée dépasse largement le seul champ de nos données personnelles. L'article 8 CEDH protège la vie privée de chaque personne sous un large éventail d'aspects, qu'on peut classer en trois grandes catégories : 1) la vie privée (en général), 2) l'intégrité physique, psychologique ou morale, et 3) l'identité et l'autonomie[9]. Plusieurs usages et applications de l'IA susceptibles d'avoir un impact sur ces catégories n'ont guère attiré l'attention à ce jour.

18. La surveillance (massive) fondée sur l'IA, par reconnaissance faciale par exemple, suppose de saisir, de stocker et de traiter des données personnelles (y compris biométriques, comme nos visages[10]) ; mais elle affecte aussi notre vie privée en général, notre identité et notre autonomie jusqu'à faire de nous des êtres épiés, suivis et identifiés – parfois en permanence. Cela peut avoir un effet « tétanisant », pousser les personnes à se comporter selon une certaine norme et modifier le rapport de pouvoir entre l'État ou

[9] Conseil de l'Europe, *Guide sur l'article 8 CEDH*.
[10] La jurisprudence de la Cour européenne des droits de l'homme (Cour eur. DH) dit clairement que le prélèvement, le stockage et le traitement de telles informations, même pour une courte durée, ont des incidences sur l'article 8 CEDH.

l'entité privée qui utilise la reconnaissance faciale d'une part, et l'individu d'autre part[11]. La doctrine et les précédents juridiques montrent que l'effet tétanisant de la surveillance peut constituer une violation de l'espace privé nécessaire à l'épanouissement de la personne et au débat démocratique[12]. Même si nos visages sont effacés juste après saisie, cette technologie reste une intrusion dans notre intégrité psychologique.

19. La reconnaissance faciale a un impact particulièrement évident sur l'intégrité psychologique et sur le droit « général » à la vie privée ; cependant, un effet similaire pourrait découler du pistage généralisé – en ligne et hors ligne – de tous les aspects de nos vies (via notre comportement en ligne, nos données de localisation et toutes les données issues des objets connectés : montres, applications de santé, haut-parleurs, thermostats, voitures, etc.).

20. D'autres formes de reconnaissance biométrique par IA ont d'encore plus fortes répercussions sur notre intégrité psychologique. Reconnaissance des micro-expressions, de la démarche, du timbre ou du ton de la voix, du rythme cardiaque, de la température, etc. : tous ces procédés sont actuellement appliqués pour évaluer, voire prédire notre comportement, notre état d'esprit et nos émotions.

21. Il faut le noter d'emblée : aucune preuve scientifique solide ne corrobore l'idée qu'il serait possible de « lire » les émotions ou l'état d'esprit d'une personne sur son visage ou dans sa démarche, son rythme cardiaque, le ton de sa voix ou sa température – sans parler de prédire ses comportements futurs. Dans une méta-étude récente, un groupe de scientifiques[13] a conclu que la reconnaissance des émotions par IA pouvait, tout au plus, détecter comment une personne *interprétait* subjectivement telle ou telle caractéristique d'une autre personne. Cela ne coïncide pas nécessairement avec les sentiments réels, et l'IA ne fait qu'étiqueter cette interprétation, par ailleurs fortement dépendante du contexte et de la culture. Les déclarations fantaisistes sur ce thème – l'IA pourrait, par exemple, prédire la réussite à un entretien d'embauche sur la base des micro-expressions ou du ton de la voix – n'ont absolument aucun fondement scientifique.

22. Plus important encore, l'usage généralisé de telles techniques d'IA, par exemple pour les recrutements, dans les forces de l'ordre ou dans l'enseignement, a un impact sur l'intégrité physique, psychologique ou morale des intéressés, et donc sur des aspects constitutifs de leur vie privée.

[11] *Examined Lives: Informational Privacy and the Subject as Object*, Julie E. Cohen, 2000.
[12] Cet « effet tétanisant » se produit lorsque l'exercice légitime d'un droit se trouve inhibé ou découragé. Cet effet a été démontré : les personnes qui se savent surveillées changent de comportement et connaissent une évolution différente. J. Staben (2016), *Der Abschreckungseffekt auf die Grundrechtsausübung: Strukturen eines verfassungsrechtlichen Arguments*, Mohr Siebeck.
[13] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez et S. D. Pollak (2019), « Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements », *Psychological Science in the Public Interest*, 20(1), 1–68.

23. Il faut noter que le RGPD n'encadre que dans une certaine mesure le traitement des données biométriques. Il définit les « données biométriques » comme suit : « données à caractère personnel résultant d'un traitement technique spécifique, relatives aux caractéristiques physiques, physiologiques ou comportementales d'une personne physique, qui permettent ou confirment son identification unique ». La dernière partie de la phrase est cruciale, car une reconnaissance biométrique ne visant pas des fins d'identification (mais par exemple de catégorisation, de profilage ou de reconnaissance des émotions) pourrait ne pas entrer dans la définition du RGPD. De fait, le considérant n° 51 du RGPD affirme que les photographies ne sont considérées comme des données biométriques que « lorsqu'elles sont traitées selon un mode technique spécifique permettant l'identification ou l'authentification unique d'une personne physique ».

24. Beaucoup de technologies de reconnaissance biométrique ne visent pas uniquement à identifier une personne, mais aussi à évaluer son comportement (par exemple en cours) ou à catégoriser les personnes (par exemple pour déterminer leur prime d'assurance sur la base de leur exposition statistique à tel ou tel problème de santé). De tels usages pourraient ne pas relever de la définition des données biométriques (et de leur traitement) en vertu du RGPD.

25. Pour en revenir à la confidentialité des données, les données personnelles et non personnelles ne servent pas seulement à entraîner les systèmes d'IA, mais aussi à profiler et à noter les personnes à des fins aussi variées que la police prédictive, la signature d'un contrat d'assurance, l'octroi de prestations sociales ou la prédiction des performances dans le cadre de procédures d'embauche ou de licenciement. En outre, une énorme quantité de « points de données » sur notre vie quotidienne sont utilisés non seulement pour nous adresser des publicités ciblées, mais aussi pour nous amener vers certaines informations (via des pressions/influences/incitations/coups de pouce) et peser ainsi sur nos choix, affectant par là notre intégrité morale.

ii) IA et liberté individuelle

26. La liberté individuelle est affirmée dans plusieurs articles de la CEDH, comme l'article 10 (liberté d'expression) et l'article 11 (liberté de réunion et d'association). L'IA peut aussi avoir un effet tétanisant sur ces libertés.

Liberté d'expression (art. 10 CEDH)

27. L'article 10 CEDH affirme le droit à la liberté d'expression et d'information, dont la liberté d'avoir des opinions et de recevoir des informations et des idées. L'usage de l'IA pour profiler, étudier, pister et identifier des personnes et pour filtrer, définir, trier ou influencer plus ou moins ouvertement leurs comportements peut avoir non seulement un impact sur le droit à l'intégrité morale, tel que défini plus haut, mais aussi un effet tétanisant sur ces libertés.

28. Appliquée dans l'espace public, la reconnaissance faciale peut porter atteinte à la liberté individuelle d'opinion et d'expression : en ouvrant la possibilité d'identifier tous les membres d'un groupe, elle supprime en effet la protection qu'offre l'« anonymat de la foule ». Cette situation pourrait pousser les intéressés à changer de comportement – à cesser, par exemple, de se rendre à des manifestations pacifiques[14].

29. Il en va de même lorsque toutes nos données sont exploitées pour rendre possibles, via l'IA, des notations, évaluations et mesures de performances (en vue d'octroyer ou de refuser, par exemple, un crédit, un emprunt immobilier, un emploi, une promotion, etc.). Le risque est alors que les intéressés hésitent à exprimer ouvertement certaines opinions, à lire certains ouvrages ou journaux en ligne ou à consulter certains médias en ligne.

30. Concernant le droit de recevoir et de communiquer des informations et des idées, l'IA pose problème lorsqu'elle sert à sélectionner des informations et à personnaliser de plus en plus les contenus en ligne et les actualités adressées à telle ou telle personne. Les moteurs de recherche, systèmes de recommandation de vidéos et agrégateurs d'informations sont souvent opaques non seulement sur les données qu'ils utilisent pour sélectionner ou hiérarchiser les contenus, mais aussi sur les finalités d'une telle sélection[15]. Beaucoup de modèles économiques reposent sur les recettes publicitaires en ligne. Pour que les internautes passent le plus de temps possible sur un site ou une application, il est possible que les systèmes mettent en avant des contenus destinés à cette seule fin (les retenir sur le site), peu importe que ces contenus soient objectifs, avérés, divers ou même présentent un quelconque intérêt.

31. Outre ces considérations commerciales, des motivations politiques ou autres pourraient conduire à optimiser les systèmes d'IA pour qu'ils sélectionnent ou mettent en avant certains contenus en vue d'imposer ou de souligner des opinions en particulier, par exemple dans le cadre d'un processus électoral[16].

32. En outre, l'IA est aujourd'hui de plus en plus capable de produire des contenus médiatiques (sons, images, vidéos) qui imitent l'apparence et/ou la voix de personnes réelles – trucages ultra-réalistes connus sous le nom de « *deep fakes* » –, ce qui permet de tromper le public, dans des buts divers.

33. Tout cela peut engendrer des bulles de filtres et une prolifération d'infox, de désinformation et de propagande qui affectent la capacité à se forger ses propres opinions et à recevoir et communiquer des informations et des idées, et ont donc un impact sur la liberté d'expression.[17]

---

[14] *Privacy Impact Assessment Report for the Utilization of Facial Recognition Technologies to Identify Subjects in the Field*, 30 juin 2011, p. 18.

[15] J. Burrell (2016): « How the machine 'thinks': Understanding opacity in machine learning algorithms », *Big Data & Society*, 3(1), 2053951715622512.

[16] *The Great Hack*, documentaire consacré à Cambridge Analytica disponible sur Netflix.

[17] Rapport du rapporteur spécial des Nations unies sur la promotion et la protection du droit à la liberté d'opinion et d'expression, A/73/348

Liberté de réunion et d'association (art. 11 CEDH)

34. On sait qu'internet et les réseaux sociaux aident à exercer le droit à la liberté d'association et de réunion pacifique. Cependant, le recours à l'IA pourrait aussi compromettre ces libertés, si des personnes ou des groupes de personnes se trouvent automatiquement pistés, identifiés et peut-être même écartés des manifestations ou des mouvements de contestation[18].

35. Comme déjà évoqué, le recours à la reconnaissance faciale, en particulier dans des espaces publics, pourrait décourager la participation à des manifestations et à des rassemblements pacifiques, qui comptent parmi les éléments les plus importants d'une société démocratique. On en a déjà vu des exemples à Hong Kong, où des manifestants se sont mis à porter des masques et à utiliser des lasers pour éviter d'être « pris » par les caméras de reconnaissance faciale.

iii) IA et égalité, non-discrimination et solidarité

Interdiction de la discrimination (art. 14 CEDH, Protocole n° 12)

36. Parmi les droits de l'homme les plus signalés comme menacés par l'IA figurent l'interdiction de la discrimination et le droit à l'égalité de traitement. Comme nous l'avons vu plus haut, dans de nombreux cas, il s'est avéré que l'IA perpétuait ou amplifiait des partis pris inacceptables, voire en créait de nouveaux. Certains de ces partis pris constituent des discriminations.[19] En outre, les systèmes fondés sur les données masquent l'existence de biais et marginalisent les mécanismes de contrôle social qui régissent les comportements humains.

37. Par exemple, Amazon a recruté à l'aide d'une IA favorisant les hommes, car entraînée à partir du profil d'employés « modèles » qui se trouvaient être des hommes. Non content d'exclure les femmes, ce système d'IA avait étudié les caractéristiques de ces employés modèles, comme les formules typiquement utilisées dans leurs CV, et écartait les CV ne présentant pas ces caractéristiques[20].

38. Lorsque les systèmes d'IA ne recherchent, comme c'est largement le cas aujourd'hui, que des corrélations fondées sur des traits communs avec d'autres « cas », ils suscitent très facilement de multiples biais. Le problème est que ces systèmes, même s'ils excellent dans l'identification de schémas, comme les expressions typiquement employées par des salariés méritants, ne comprennent pas le sens de ces expressions, sans parler du sens

[18] *Algorithmes et droits humains – Étude sur les dimensions des droits humains dans les techniques de traitement automatisé des données et éventuelles implications réglementaires*, Conseil de l'Europe, 2018.
[19] Profilage d'affinité et discrimination par association dans la publicité comportementale en ligne, Wachter 2020
[20] J. Dastin (10 octobre 2018), « Amazon scraps secret AI recruiting tool that showed bias against women », Reuters.

du « mérite » ou de la définition d'un « salarié ». Ils sont uniquement capables d'étiqueter un schéma spécifique.

39. Contrairement à une croyance répandue, tous ces biais ne résultent pas de données de faible qualité. La conception de tout artefact résulte en soi d'une série de choix orientés, depuis les entrants jusqu'aux objectifs. Le système a-t-il pour seul but une efficacité optimale, ou tient-il compte des effets sur les travailleurs et sur l'environnement ? Vise-t-il à détecter un maximum de fraudeurs, ou s'attache-t-il à éviter de pointer des innocents ? Tous ces choix sont orientés, d'une manière ou d'une autre, par les partis pris inhérents à la ou aux personne(s) qui les font. En un mot, il serait illusoire de prétendre éliminer tous les partis pris associés à l'IA[21].

iv) IA et droits sociaux et économiques

L'IA intégrée ou liée au lieu de travail

40. Appliquée à des travaux dangereux, pénibles, épuisants, salissants, désagréables, répétitifs ou ennuyeux, l'IA peut présenter un intérêt majeur. Cependant, les systèmes d'IA sont aussi de plus en plus utilisés pour surveiller et pister les travailleurs, répartir les tâches sans intervention humaine et apprécier et prédire le potentiel des travailleurs et leurs performances dans des contextes d'embauche ou de licenciement. De telles applications de l'IA pourraient compromettre le droit à des conditions de travail équitables, à la sécurité et à l'hygiène au travail, ainsi que le droit des travailleurs à s'organiser (CSE, articles 2 et 3). Des salariés constamment surveillés par leur employeur risquent d'hésiter à s'associer au sein d'organisations (article 5). Des systèmes d'IA qui évaluent et prédisent les performances des travailleurs pourraient compromettre le droit à l'égalité de chances et de traitement en matière d'emploi et de profession sans discrimination fondée sur le sexe (article 20 CSE), en particulier si ces systèmes perpétuent des partis pris issus des données ou de leurs concepteurs.

41. Si de plus en plus de tâches et de décisions auparavant opérées par des humains sont transférées à des systèmes d'IA, des compétences nécessaires risquent d'être perdues. Cela pourrait non seulement rendre la main-d'œuvre moins qualifiée, mais aussi créer un risque de défaillance systémique, puisque seuls quelques êtres humains seraient capables de travailler sur les systèmes d'IA et de réagir à une panne.

42. On ignore si l'IA fera perdre ou gagner des emplois et, le cas échéant, dans quelles proportions ; mais pendant la période de bouleversements liée à sa mise en place, l'écart entre une main-d'œuvre de plus en plus vulnérable et les compétences recherchées pourrait entraîner une situation de « chômage technologique[22] ».

[21] Première analyse du Livre blanc sur l'IA, Virginia Dignum, Catelijne Muller et Andreas Theodorou, 2020.
[22] E. Brynjolfsson et A. McAfee (2014), *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*, W. W. Norton & Company.

v) IA et démocratie

43. L'IA peut avoir (et a probablement déjà) un effet négatif sur la démocratie, en particulier sur les aspects suivants : 1) discours social et politique, accès aux informations et influence sur les électeurs, 2) inégalité et ségrégation, et 3) défaillance ou disruption systémique.

Discours social et politique, accès aux informations et influence sur les électeurs

44. Une démocratie en bonne santé suppose des citoyens éclairés, un discours social et politique ouvert et l'absence d'influence souterraine sur les électeurs.

45. À cet égard, la bonne information des citoyens est cruciale. Au sein de nos sociétés de l'information, les citoyens peuvent choisir de ne consulter qu'une petite part des informations disponibles. Les moteurs de recherche, fils d'actualité sur les réseaux sociaux, systèmes de recommandation, et de nombreux sites d'actualités emploient l'IA pour déterminer quel contenu créer et montrer aux utilisateurs (en un mot, les informations sont « personnalisées »). Bien conçu, ce procédé peut aider les citoyens à se repérer dans les flots d'informations disponibles et à améliorer leurs compétences démocratiques, par exemple en leur permettant d'accéder à des ressources en d'autres langues grâce aux outils de traduction[23]. Cependant, si l'IA décide quelles informations afficher, et quels sujets présenter en boucle ou, au contraire, noyer dans le flux des informations en ligne, la représentation des différentes voix et opinions risque de devenir inégale et biaisée.

46. Le suivi et le profilage constants de chaque individu rendent possible la **personnalisation des informations par IA**. Cette infrastructure informationnelle, fondée sur les technologies et mue par des considérations commerciales ou politiques, pourrait amplifier les messages hyperpartisans avec lesquels nous sommes susceptibles d'être d'accord, et fournir un outil d'influence individualisée d'une puissance sans précédent[24]. Elle pourrait ainsi saper les visions communes, le respect mutuel et la cohésion sociale dont la démocratie a besoin pour s'épanouir. Les prédictions personnalisées par IA, si elles gagnent en puissance et en efficacité, pourraient même menacer la capacité de choix et l'autonomie humaines nécessaires pour que les électeurs prennent de véritables décisions[25].

47. Troisièmement, l'IA peut nuire à l'équité des processus électoraux. Des responsables de campagne ou des acteurs étrangers peuvent utiliser (et ont utilisé) des publicités personnalisées pour adresser des messages différents à différents groupes d'électeurs, sans avoir à rendre de comptes sur ce procédé[26]. Toutefois, il faut noter que le doute

[23] R. Schroeder (2018), *Social Theory after the Internet*, UCL Press.
[24] T. Wu (2016), *The attention merchants: The epic scramble to get inside our heads* (1re édition) ; S. Zuboff (2019), *The age of surveillance capitalism: The fight for the future at the new frontier of power*.
[25] M. Taddeo et L. Floridi (2018b), « How AI can be a force for good », *Science*, 361(6404), 751–752.
[26] S. Bradshaw et P. Howard (2019), *Social Media and Democracy in Crisis*, Oxford University Press.

plane encore sur le degré d'influence exact des publicités micro-ciblées[27]. L'IA peut aussi servir à créer et à répandre de la désinformation et des trucages hyperréalistes, sous forme de textes, d'images, de sons ou de vidéos. Pour les citoyens et les journalistes comme pour les institutions publiques, ces trucages s'avèrent difficiles à repérer. Il devient donc de plus en plus facile de tromper et de manipuler le public, ce qui pourrait dégrader le sérieux et la crédibilité des médias et du discours démocratique.

Inégalités et ségrégation

48. On attend très largement de l'IA qu'elle améliore la productivité de l'économie. Toutefois, il est également probable que ces gains de productivité soient inégalement répartis et bénéficient surtout aux plus riches. De même, il est démontré que la sélection des données et les choix de conception, associés au manque de transparence des algorithmes « boîtes noires », perpétuent les partis pris injustes à l'encontre des groupes sociaux défavorisés, comme les femmes ou les minorités ethniques[28]. L'IA pourrait entraîner des inégalités et une ségrégation, et donc menacer le niveau d'**égalité** économique et sociale nécessaire à l'épanouissement de la démocratie.

Risques systémiques

49. Le transfert à l'IA de décisions que seuls des humains étaient auparavant capables de prendre fait peser de nouveaux défis sur la sécurité et la résilience des systèmes sociétaux. En particulier, si de multiples décideurs décentralisés se trouvent remplacés par quelques systèmes centralisés pilotés par IA, les risques systémiques augmentent : une seule panne sur un système peut produire des résultats catastrophiques.

50. Comme le montre l'augmentation des krachs éclair, ces nouveaux risques systémiques sont déjà une réalité sur les marchés financiers, où différents agents d'IA interagissent à une vitesse surhumaine[29]. Le fait que les infrastructures énergétiques essentielles, les réseaux de transports ou les hôpitaux dépendent de plus en plus de décisions prises automatiquement par des agents artificiels crée de nouvelles vulnérabilités, puisqu'une atteinte sur un seul point faible peut avoir des effets dévastateurs. Une fois installés de tels systèmes d'infrastructures essentielles, il devient plus difficile de les remplacer ou de leur trouver des substituts en cas de dysfonctionnement[30].

51. La vision du développement de l'IA comme une course avec la concurrence compromet particulièrement la paix et la sécurité internationales. Dans ces conditions, l'IA promet non

[27] R.M. Bond, C. J. Fariss, J. J. Jones, A. D. I Kramer, C. Marlow, J. E. Settle et J. H. Fowler (2012), « A 61-million-person experiment in social influence and political mobilization », *Nature*, 489(7415), 295–298.
[28] V. Eubanks (2017), *Automating inequality: How high-tech tools profile, police, and punish the poor* (1re édition), St. Martin's Press ; C. O'Neil (2016), *Weapons of math destruction: How big data increases inequality and threatens democracy* (1re édition).
[29] M. P. Wellman et U. Rajan (2017), *Ethical Issues for Autonomous Trading Agents. Minds and Machines*, 27(4), 609–624.
[30] M. Brundage, S. Avin, J. Clark, H. Toner, P. Eckersley, B. Garfinkel, A. Dafoe, P. Scharre, T. Zeitzoff, B. Filar, H. Anderson, H. Roff, G. C. Allen, J. Steinhardt, C. Flynn, S. Ó. Éigeartaigh, S. Beard, H. Belfield, S. Farquhar, D. Amodei (2018), *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*.

seulement d'entraîner des effets secondaires indésirables, mais aussi d'aider des acteurs malveillants, des cybercriminels aux États totalitaires, à resserrer leur emprise sur les populations.

Concentration des pouvoirs numériques

52. Beaucoup d'applications d'IA sont développées et déployées pour une poignée de grands groupes privés, parfois surnommés les géants du Net, les Gafam ou même les *Frightful Five* (« les Cinq qui font peur[31] »). Si quelques acteurs privés, pour lesquels la valeur des actions passe avant le bien commun, concentrent trop de pouvoir politique, l'autorité des États démocratiques peut se trouver menacée.

vi) IA et État de droit

53. Les institutions démocratiques ont une responsabilité particulière à l'égard des individus et de la société, qui se traduit en principes comme la proportionnalité, l'égalité ou la motivation des décisions. D'un côté, l'IA peut rendre ces institutions plus efficaces ; mais de l'autre, elle peut aussi affaiblir leur légitimité procédurale, la confiance dont elles jouissent, et l'autorité du droit.

54. Les tribunaux, les autorités répressives et les administrations pourraient ainsi devenir plus efficaces, mais aussi plus opaques et laisser moins de place aux qualités humaines (choix, autonomie, supervision[32]).

55. De même, alors qu'auparavant, seuls les tribunaux pouvaient dire ce qui constituait ou non un discours de haine illégal, ce sont aujourd'hui avant tout des systèmes d'IA privés qui désignent les messages à retirer des sites de réseaux sociaux[33]. Ces systèmes d'IA empiètent de fait sur l'autorité des juges et du droit, et en général, l'IA peut favoriser le développement de systèmes judiciaires fonctionnant hors des limites et des protections de l'État de droit.

56. Ainsi, les règlements automatiques des litiges en ligne proposés par des entreprises privées reposent sur leurs conditions d'utilisation, et non sur la loi, et n'offrent pas aux consommateurs les mêmes droits et les mêmes garanties procédurales que les juridictions publiques[34].

[31] Google, Apple, Facebook, Amazon et Microsoft. Voir : P. Nemitz (2018), « Constitutional democracy and technology in the age of artificial intelligence », *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180089. A. Webb (2019), *The Big Nine: How the tech titans and their thinking machines could warp humanity*.

[32] J. Danaher (2016), « The Threat of Algocracy: Reality, Resistance and Accommodation », *Philosophy & Technology*, 29(3), 245–268.
J. Weizenbaum (1976), *Computer Power and Human Reason: From Judgment to Calculation*, W. H. Freeman.

[33] J. E. Cohen (2019), *Between truth and power: The legal constructions of informational capitalism*.

[34] J. Susskind (2018), *Future politics: Living together in a world transformed by tech*.

57. Dès 2018, la Commission européenne pour l'efficacité de la justice définissait cinq principes pour l'usage de l'IA dans le système judiciaire, dans un texte intitulé « Charte éthique européenne d'utilisation de l'IA dans les systèmes judiciaires et leur environnement ». Le Groupe d'experts de haut niveau sur l'IA a appelé les organismes publics à respecter, lorsqu'ils conçoivent, achètent ou utilisent l'IA, sept exigences pour une IA digne de confiance. Des principes et exigences similaires devraient être imposés aux services répressifs.

58. Toutefois, l'IA ne fait pas que menacer l'État de droit : elle peut aussi le renforcer[35]. Développée et appliquée de manière responsable, elle peut aider les autorités à détecter les cas de **corruption**[36]. Elle peut aussi servir à détecter les cyberattaques et à s'en protéger[37].

## IV. Que faire face à l'impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit ?

59. Depuis peu, l'impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit attire de plus en plus l'attention, notamment grâce à la parution du Livre blanc sur l'IA de la Commission européenne. Les réponses possibles à cet impact, cependant, restent à ce jour incertaines. Le présent chapitre décrit les stratégies pouvant être adoptées au sein du cadre existant en matière de droits de l'homme, de démocratie et d'État de droit. La liste n'est pas exhaustive ; le but est de favoriser le passage à la phase de débat suivante.

### i) Les droits de l'homme dans le contexte de l'IA

60. Parmi les acteurs (publics et privés) qui conçoivent, déploient et utilisent l'IA, beaucoup ne semblent pas avoir conscience des impacts – potentiels – de l'IA sur les droits de l'homme. Dans un premier temps, il serait très utile de se livrer à un exercice de reformulation, c'est-à-dire de « traduire » dans le contexte de l'IA les droits affirmés dans la CEDH, éventuellement au moyen d'une Convention-cadre.

### ii) Mesures visant la conformité, la réparation et l'obligation de rendre des comptes

61. Pour faire face à l'impact de l'IA sur la démocratie, l'État de droit et les droits de l'homme existants, certains mécanismes actuels visant la conformité, la réparation et l'obligation de rendre des comptes pourraient être étoffés, avec en parallèle la création de nouveaux mécanismes. Cependant, l'utilisation de l'IA est souvent dissimulée ou ignorée, ce qui

[35] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S. D. Langhans, M. Tegmark et F. F. Nerini (2020), « The role of artificial intelligence in achieving the Sustainable Development Goals », *Nature Communications*, 11(1), 1–10.
[36] J. West et M. Bhattacharya (2016), « Intelligent financial fraud detection: A comprehensive review », *Computers & Security*, 57, 47–66. P. Hajek et R. Henriques (2017), « Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods », *Knowledge-Based Systems*, 128, 139–152.
[37] M. Taddeo et L. Floridi (2018a), « Regulate artificial intelligence to avert cyber arms race », *Nature*, 556(7701), 296–298.

rend difficile, voire impossible de savoir s'il existe ou non un impact sur les droits de l'homme, la démocratie et l'État de droit. Toute mesure devrait donc commencer par une obligation de transparence sur l'utilisation de systèmes d'IA susceptibles de se répercuter sur les droits de l'homme, la démocratie et l'État de droit. Cela passerait par un répertoire des systèmes d'IA, qui préciserait la catégorie de risque et le degré de transparence et d'obligation de rendre des comptes pour chaque application.

62. La conformité pourrait donc commencer par une nouvelle culture consistant à intégrer, selon une expression récente, « les droits de l'homme, la démocratie et l'État de droit dès la conception[38] ». Les acteurs qui conçoivent, déploient et utilisent l'IA s'interrogeraient dès le départ sur ses effets possibles sur les droits de l'homme, la démocratie et l'État de droit, et ajusteraient en conséquence la technologie ou ses applications. Cette nouvelle culture pourrait être renforcée par l'obligation (juridique) de réaliser une étude d'impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit.

63. Cette nouvelle culture devrait aussi englober l'obligation de rendre des comptes sur la structure mise en place, mais aussi sur les résultats de cette étude d'impact et sur les décisions prises en matière de conception et de gouvernance à la suite de l'étude d'impact.

64. Dans le contexte de l'impact de l'IA sur les droits de l'homme, la réparation suppose l'accès à la justice et l'existence de recours effectifs. S'agissant de l'accès à la justice, il est peut-être trop tôt pour dire s'il est suffisamment garanti au regard de l'IA et de son impact sur les droits de l'homme. Les tout premiers jugements nationaux concernant l'impact (potentiel) de l'IA sur les droits de l'homme ont été rendus très récemment, et ils concernent un droit en particulier : le respect de la vie privée (article 8 CEDH[39]).

65. Point important, l'accès à la justice est d'autant plus difficile que beaucoup d'applications d'IA ne sont développées et déployées que par une poignée de grands acteurs privés. Ces entreprises dominent à la fois l'élaboration de systèmes d'IA et les écosystèmes au sein desquels ils fonctionnent. Les États sont tenus de protéger les individus et les groupes contre les atteintes aux droits de l'homme commises par des tiers, et la reconnaissance de l'influence des acteurs non étatiques sur les droits de l'homme ne cesse de croître[40]. Aujourd'hui, les géants du numérique sont capables de déterminer, voire de modifier nos structures sociales et même démocratiques ; l'impact de leurs systèmes d'IA, et de l'usage qu'ils en font, sur les droits de l'homme revêt donc une importance considérable. L'IA pourrait offrir une bonne raison de réfléchir à une structure

[38] P. Nemitz (2018), *Constitutional democracy and technology in the age of artificial intelligence*.
[39] Un tribunal néerlandais a jugé qu'une loi ouvrant aux institutions publiques la possibilité d'utiliser l'IA pour prédire les fraudes aux prestations sociales violait le droit à la vie privée, inscrit dans l'article 8 CEDH ; au Royaume-Uni, la Haute Cour de justice, siégeant à Cardiff, a admis que la reconnaissance faciale affectait l'article 8 CEDH, puisqu'elle permet de recueillir des informations « par définition privées », mais a jugé son utilisation légale au regard du principe de proportionnalité. En France, un tribunal a considéré que l'utilisation de la reconnaissance faciale dans des établissements scolaires violait l'article 8 CEDH.
[40] *Business and Human Rights, A Handbook for Legal Practitioners*, Claire Methven O'Brien, Conseil de l'Europe.

qui contraindrait juridiquement les acteurs privés à respecter les droits de l'homme, et à permettre l'accès à la justice s'ils manquent à cette obligation[41]. La question fondamentale est la suivante : faut-il a) reconnaître le pouvoir des entreprises privées utilisant l'IA et s'assurer qu'elles l'exercent de façon responsable, ou b) le contester et réaffirmer le pouvoir de l'État ?

66. Concernant les recours effectifs, Dinah Shelton observe que lorsqu'il s'agit d'IA, ils ne consistent « pas seulement à indemniser la victime ; ils expriment l'opprobre de toute la société sur le fautif » et, par là, « affirment, renforcent et concrétisent les valeurs fondamentales de la société[42] ». Dans son arrêt *Broniowski*, la Cour européenne des droits de l'homme a souligné qu'en vertu du droit international, il était souhaitable que réparation individuelle et réparation générale « aillent de pair[43] ».

67. Pour assurer un recours effectif en cas d'atteinte aux droits de l'homme du fait de l'IA, il faut donc étudier **à la fois** les moyens de réparation individuels et généraux. En outre, les recours devraient être adaptés aux applications de l'IA visées, qui sont multiples : surveillance et identification, profilage, influence, aide à la décision. Pour être effectifs, ils devraient comprendre l'arrêt de la conduite illicite et des garanties pour qu'elle ne se reproduise pas; les États pourraient, par exemple, être tenus d'adopter et d'appliquer une législation contraignante pour éviter de futurs impacts de l'IA sur les droits de l'homme. Il devrait y avoir obligation de réparer les dommages ou les préjudices entraînés par la violation, soit pour un individu, soit pour une communauté. Pour certaines applications de l'IA, assurer simplement un recours effectif pourrait ne pas suffire à traiter l'impact de ces applications sur les droits de l'homme. Des mesures plus larges pourraient s'imposer, comme une interdiction ou des restrictions d'utilisation (voir le chapitre IV).

iii) Protection des structures démocratiques et de l'État de droit

68. Pour éviter les défaillances ou interruptions systémiques dues aux prises de décisions centralisées par IA au sein de structures vitales, et donc prévenir les risques de pannes catastrophiques, il convient de répartir les processus décisionnels au lieu de le centraliser. Ces processus devraient intégrer de solides mécanismes de supervision humaine[44].

69. La supervision d'un système d'IA par des êtres humains garantit que ce système ne sape pas, entre autres effets adverses, l'autonomie humaine. Cette supervision peut être assurée par plusieurs mécanismes de gouvernance, comme ceux des humains « dans la boucle » (*human-in-the-loop*, HITL), « sur la boucle » (*human-on-the-loop*, HOTL) ou

[41] Cela suppose d'aller au-delà des préconisations de la Recommandation CM/Rec(2016)3 du Comité des Ministres du Conseil de l'Europe sur les droits de l'homme et les entreprises (et des Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme).
[42] Dinah Shelton, « The Right to Reparations for Acts of Torture: What Right, What Remedies? », 17(2) *Torture* 96 (2007), 96.
[43] Cour eur. DH, *Broniowski c. Pologne*.
[44] Groupe d'experts de haut niveau sur l'IA, *Lignes directrices en matière d'éthique pour une IA digne de confiance*, 2019.

« aux commandes » (*human-in-command*, HIC). L'approche HITL désigne la possibilité d'une intervention humaine à tous les cycles décisionnels du système, intervention qui n'est dans de nombreux cas ni possible ni souhaitable. L'approche HOTL désigne la possibilité d'une intervention humaine au cours du cycle de conception du système et du suivi de son fonctionnement. L'approche HIC désigne la possibilité de superviser l'activité globale du système d'IA (y compris son impact économique, sociétal, juridique et éthique au sens large) et celle de décider quand et comment utiliser le système dans telle ou telle situation. Cela peut englober la décision de ne pas utiliser de système d'IA dans une situation donnée, de prévoir certains niveaux de latitude humaine lors de l'utilisation du système ou de permettre de passer outre une décision prise par le système.

70. Pour remédier au risque d'inégalité, les pouvoirs publics doivent activement enrayer l'usage des applications d'IA qui aggravent les inégalités.

71. Prévenir les influences sur les élections ou les manipulations du public au travers des informations personnalisées par IA n'est pas une tâche facile. On pourrait envisager une réglementation des campagnes en ligne, soit pour les médias en ligne et sites de réseaux sociaux, soit pour les partis politiques. Cela soulève évidemment des questions de liberté d'expression. La présence d'êtres humains « dans/sur la boucle » et « aux commandes » (voir plus haut) pourrait aider à détecter et à éliminer les influences indésirables sur les électeurs.

72. Pour assurer un usage responsable de l'IA dans les services publics, les passations de marchés publics représentent un outil crucial. La mise à jour des exigences légales associées aux commandes **publiques**, pour y intégrer des critères comme l'équité, l'obligation de rendre des comptes et la transparence sur l'IA, pourrait servir deux fins. Premièrement, elle garantirait que les pouvoirs publics n'utilisent que des systèmes compatibles avec l'État de droit ; deuxièmement, elle inciterait le secteur privé, par intérêt économique, à développer et utiliser des systèmes conformes aux principes de l'État de droit. En outre, l'utilisation de l'IA par les pouvoirs publics devrait être soumise à des mécanismes de supervision, dont des ordonnances judiciaires et la possibilité de saisir des médiateurs.

## V. Et si les normes actuelles de droits de l'homme, de démocratie et d'État de droit ne nous protégeaient pas assez ?

### i) Question zéro

73. Vu le caractère intrusif de certaines applications ou utilisations de l'IA, il pourrait arriver que le cadre actuel en matière de droits de l'homme, de démocratie et d'État de droit ne nous protège pas assez, ou pas à temps. Nous devrons alors peut-être prendre le temps de la réflexion, et chercher la bonne réponse à ce qu'on pourrait appeler la « question zéro » : allons-nous autoriser ce système ou cet usage et si oui, dans quelles conditions ?

On ne peut y répondre sans examiner le système d'IA ou son usage sous tous les angles, avec pour résultat plusieurs « solutions » possibles :

- Le système d'IA ou son usage est placé sous moratoire, interdit (temporairement ou définitivement) ou soumis à restrictions (« lignes rouges ») ;
- Pour parer aux nouveaux impacts négatifs de l'IA, de nouveaux droits de l'homme sont adoptés ;
- Les droits de l'homme existants sont adaptés pour permettre le développement et l'utilisation responsables de l'IA ;
- Un système ou un usage particulier est soumis à un mécanisme de contrôle démocratique spécifique ;
- Les propriétaires privés des puissants systèmes d'IA sont obligés d'aligner leurs structures de développement et de gouvernance de l'IA sur les intérêts de ceux que les systèmes affectent et de l'ensemble de la société, ce qui peut appeler des mesures pour associer les parties concernées (travailleurs, consommateurs, clients, citoyens, responsables politiques, etc.).

74. D'abord et avant tout, l'impact de l'IA doit être étudié à la fois au niveau individuel et au niveau sociétal/collectif, puisqu'il peut toucher aussi bien une personne en particulier que de larges pans de notre société. Deuxièmement, pour décider quelle action est pertinente et proportionnée à la situation, il est important de connaître le contexte, la finalité, la gravité, l'ampleur et la probabilité de l'impact. Pour les applications d'IA qui engendrent des risques inacceptables ou créent la menace d'un préjudice ou d'une défaillance systémique, il convient d'adopter une approche réglementaire fondée sur des principes, dont le principe de précaution. Pour les autres applications d'IA, une approche fondée sur les risques pourrait s'avérer plus adéquate.

ii) Lignes rouges

75. Des lignes rouges pourraient être définies, et appliquées aux systèmes ou usages de l'IA dont les conséquences sont trop lourdes pour qu'ils restent sans contrôle ni réglementation – ou même pour qu'ils soient autorisés. Les applications suivantes pourraient rendre nécessaires une interdiction, un moratoire et/ou de strictes limites ou conditions déterminant des usages exceptionnels et/ou contrôlés :

- Le recours généralisé à la reconnaissance faciale et à d'autres formes de reconnaissance biométrique par des acteurs publics ou privés ;
- La surveillance de masse fondée sur l'IA (via la reconnaissance faciale/biométrique, mais aussi d'autres formes de pistage et/ou d'identification par IA comme la géolocalisation, le comportement en ligne, etc.) ;
- Le pistage, l'évaluation, le profilage, la notation et les incitations, à caractère personnel, physique ou psychologique, via la reconnaissance biométrique et la reconnaissance des comportements ;
- La notation sociale assistée par IA ;
- Les systèmes d'IA dissimulés et les trucages hyperréalistes (*deep fakes*) ;
- Les interfaces IA-êtres humains.

76. L'usage exceptionnel de telles technologies, par exemple à des fins de sécurité nationale, de diagnostic ou de traitement médical, devrait être fondé sur des preuves, nécessaire et proportionné, et n'être autorisé que dans des environnements contrôlés et (si possible) pour une durée limitée.

iii) Des droits de l'homme nouveaux ou adaptés

77. Outre la définition de lignes rouges, on pourrait envisager de créer ou d'adapter les droits de l'homme suivants (liste non exhaustive) :

- Droit des êtres humains à l'autonomie, à la liberté d'action et à l'exercice d'une supervision sur l'IA ;
- Droit à la transparence/au caractère explicable des résultats de l'IA, dont le droit à des explications sur la façon dont l'IA fonctionne, la logique qu'elle suit et les effets de son utilisation sur les intérêts de la personne concernée, même si le système d'IA ne traite pas de données personnelles – s'il traite de telles données, ce droit à l'information existe déjà en vertu du RGPD[45] ;
- Droit distinct à l'intégrité physique, psychologique et morale au regard du profilage et de la reconnaissance des émotions par IA ;
- Droit renforcé à la vie privée, protégeant contre la surveillance de masse assistée par IA ;
- Adaptation du droit à la confidentialité des données pour protéger contre le pistage en ligne des individus, généralisé et à l'échelle de toute une société, utilisant des données personnelles ou non personnelles (ces dernières livrant souvent des indices permettant l'identification).

78. Il ne devrait être permis de déroger à ces droits que dans des circonstances exceptionnelles – à des fins de sécurité par exemple –, dans des conditions strictes et dans le respect du principe de proportionnalité.

iv) Scénarios d'avenir

79. Dans une perspective à plus long terme, des problèmes majeurs susceptibles de se poser à l'avenir font actuellement l'objet de recherches ; les inconnues qui subsistent, et la possibilité d'imprévus catastrophiques, appellent une approche fondée sur les risques. Parmi les problèmes à redouter (mais n'existant pas aujourd'hui), certains rangent l'intelligence artificielle généralisée, la conscience artificielle, les agents moraux artificiels ou la superintelligence, tandis que beaucoup d'autres les jugent irréalistes. Quoi qu'il en soit, il est nécessaire de suivre de près les évolutions, afin de déterminer dans quelle mesure les systèmes qui encadrent actuellement les droits de l'homme, la démocratie et l'État de droit demandent à être adaptés.

[45] P. Nemitz (2018), *Constitutional democracy and technology in the age of artificial intelligence*.

© ALLAI, 2020

ALLAI : Stichting ALLAI Nederland, fondation de droit néerlandais

Les informations figurant dans ce rapport provisoire n'ont qu'un caractère général et indicatif. Ni l'ALLAI ni les membres de son conseil d'administration et de son conseil consultatif, ses employés et agents, les organisations ou personnes qui leur sont associées ou les entités liées (« réseau ALLAI ») n'offrent, à travers ce rapport, des conseils ou services professionnels. Avant de prendre toute décision ou d'entreprendre toute action susceptible d'affecter votre organisation ou votre situation financière, veuillez solliciter un consultant professionnel qualifié. Aucune perte ou dommage de quelque nature que ce soit, directe ou indirecte, subie par une entité ou une personne du fait de son usage du présent rapport, ne peut être imputée à une entité ou personne du réseau ALLAI, dont la responsabilité ne saurait être engagée.