UN CADRE JURIDIQUE POUR LES SYSTÈMES D'INTELLIGENCE ARTIFICIELLE

Étude de faisabilité d'un cadre juridique pour le développement, la conception et l'application de l'intelligence artificielle, basé sur les normes du Conseil de l'Europe sur les droits de l'homme, la démocratie et l'État de droit

DGI (2021)04

Adopté par le CAHAI
le 17 décembre 2020

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

UN CADRE JURIDIQUE POUR LES SYSTÈMES D'INTELLIGENCE ARTIFICIELLE

Étude de faisabilité d'un cadre juridique pour le développement, la conception et l'application de l'intelligence artificielle, basé sur les normes du Conseil de l'Europe sur les droits de l'homme, la démocratie et l'État de droit

Étude du Conseil de l'Europe
DGI (2021)04

Adopté par le CAHAI
lors de sa 3e réunion plénière
le 17 décembre 2020

Conseil de l'Europe

Édition anglaise :
A legal framework for AI systems

Couverture et mise en page : Service de la production des documents et des publications (SPDP), Conseil de l'Europe
Photos : Shutterstock

Cette publication n'a pas fait l'objet d'une relecture typographique et grammaticale de l'Unité éditoriale du SPDP.

Imprimé dans les ateliers du Conseil de l'Europe

Table des matières

1. INTRODUCTION GÉNÉRALE

1. Le Conseil de l'Europe est la principale organisation de défense des droits de l'homme du continent et le gardien des droits de 830 millions d'Européens. Il n'a eu de cesse d'accompagner les transformations de notre société depuis 1949, en veillant à ce que le développement (y compris les évolutions technologiques) soit toujours placé sous le signe des droits de l'homme, de la démocratie et de l'État de droit. Certains de ses instruments juridiques sont devenus des normes européennes ou mondiales reconnues, réconciliant innovation et régulation au profit de l'être humain[1].

2. Plus spécifiquement, dans le domaine du numérique, les avancées de ces dernières décennies ont radicalement transformé la société en fournissant de nouveaux outils pour la communication, la consommation de l'information, l'administration publique, l'éducation, et de nombreuses autres facettes de notre vie quotidienne. Grâce à la détection de schémas et tendances dans des quantités massives de données, en s'appuyant sur les méthodes statistiques, les systèmes algorithmiques offrent maintenant la possibilité de reconnaître des images ou des sons, de rationaliser les services ou les produits et de réaliser des gains d'efficience considérables dans l'exécution de tâches complexes. Ces produits et services, que l'on qualifie communément d'intelligence artificielle (IA[2]), peuvent favoriser la prospérité humaine et le bien-être individuel et social en étant porteurs de progrès et d'innovation. Les États membres conviennent que la croissance économique est un objectif majeur des politiques publiques et considèrent l'innovation comme l'une de ses composantes clés. Dans le même temps, on assiste à une montée des préoccupations concernant les effets négatifs de différents types d'applications de l'intelligence artificielle sur l'Homme et la société. La discrimination, l'avènement d'une société de surveillance, l'affaiblissement de l'action humaine (*human agency*), la distorsion de l'information, l'ingérence électorale, l'exclusion numérique, l'économie de l'attention, figurent parmi les préoccupations concrètes qui sont exprimées.

3. Il est dès lors essentiel que les normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'État de droit soient effectivement ancrées dans les cadres législatifs appropriés par les États membres. Si les instruments généraux internationaux et régionaux existants en matière de droits de l'homme, y compris la Convention européenne des droits de l'homme (CEDH), restent applicables dans tous les domaines de la vie, y compris en ligne et hors ligne et quelle que soit la technologie, il convient d'élaborer une réponse juridique du Conseil de l'Europe, visant à combler les lacunes juridiques[3] des textes en vigueur et adaptée aux enjeux spécifiques soulevés par les systèmes d'IA, sur la base de vastes consultations multipartites. Cela a déjà été fait par le passé pour des procédés industriels innovants tels que les produits pharmaceutiques, la biomédecine ou l'industrie automobile. En outre, une telle réponse juridique pourrait également favoriser et influencer les technologies d'IA conformément aux normes susmentionnées.

4. Par conséquent, le 11 septembre 2019, le Comité des Ministres a établi un comité ad hoc sur l'intelligence artificielle (CAHAI) chargé d'examiner, sur la base de larges consultations multipartites, la faisabilité et les éléments potentiels d'un cadre juridique pour le développement, la conception et l'application de l'intelligence artificielle, fondé sur les normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'État de droit. La présente étude de faisabilité tient compte des normes du Conseil de l'Europe relatives à la conception, au développement et à l'application des technologies numériques dans les domaines des

---

1. Voir à cet égard la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (« Convention 108 », STE n° 108) et son protocole (« Convention 108+ », STCE n° 223) ; la Convention pour la protection des Droits de l'Homme et de la dignité de l'être humain à l'égard des applications de la biologie et de la médecine, STE n° 164 (« Convention d'Oviedo ») ; la Convention sur la cybercriminalité, STE n°185 (« Convention de Budapest ») ; la Convention relative à l'élaboration d'une Pharmacopée européenne, STE n° 50.
2. La section 2 clarifie davantage l'utilisation de ce terme aux fins de l'étude de faisabilité. Afin d'éviter toute forme d'anthropomorphisme et d'inclure toutes les technologies englobées sous le terme générique d'intelligence artificielle, les termes « systèmes d'IA », « applications d'IA », « solutions d'IA » seront généralement préférés dans la présente étude de faisabilité pour désigner des systèmes algorithmiques fondés, indifféremment, sur l'apprentissage automatique (*machine learning*), l'apprentissage profond (*deep learning*), les systèmes à base de règles comme les systèmes experts ou toute autre forme de programmation informatique et de traitement des données. La notion de « système algorithmique » s'entend telle que définie dans l'annexe à la Recommandation CM/Rec(2020)1 du Comité des Ministres, à savoir « des applications qui, au moyen souvent de techniques d'optimisation mathématique, effectuent une ou plusieurs tâches comme la collecte, le regroupement, le nettoyage, le tri, la classification et la déduction de données, ainsi que la sélection, la hiérarchisation, la formulation de recommandations et la prise de décision. En s'appuyant sur un ou plusieurs algorithmes pour remplir leurs missions dans les environnements où ils sont mis en œuvre, les systèmes algorithmiques automatisent les activités de manière à permettre la création de services adaptables à l'échelle et en temps réel. »
3. Comme précisé au paragraphe 5.4 de la présente étude de faisabilité.

droits de l'homme, de la démocratie et de l'État de droit, ainsi que des instruments juridiques internationaux – universels et régionaux – existants qui sont pertinents. Elle tient également compte des travaux menés par d'autres organes du Conseil de l'Europe ainsi que des travaux en cours au sein d'autres organisations régionales et internationales (Nations Unies – Unesco, HCDH, UIT, OMPI, OMS – Union européenne, OCDE, OSCE, G7/G20, Banque mondiale, Forum économique mondial). Enfin, elle intègre une perspective de genre et entend contribuer au renforcement de la cohésion sociale dans nos sociétés, de même qu'à la promotion et à la protection des droits des personnes vulnérables, comme les personnes handicapées ou les jeunes.

2. CHAMP D'APPLICATION D'UN CADRE JURIDIQUE DU CONSEIL DE L'EUROPE SUR L'IA

5. À ce jour, il n'existe pas une définition unique de l'intelligence artificielle au sein de la communauté scientifique. Ce terme, qui est entré dans notre quotidien, recouvre un vaste ensemble de sciences, théories et techniques mises en œuvre en vue de réaliser des machines capables de reproduire les capacités cognitives de l'être humain. Le terme peut donc englober toute automatisation résultant de cette technologie, ainsi que des technologies précises telles que l'apprentissage automatique ou l'apprentissage profond basé sur des réseaux de neurones.

6. De même, les diverses organisations internationales qui ont travaillé sur l'intelligence artificielle ne sont pas parvenues à s'entendre sur une définition. Le groupe d'experts indépendants de haut niveau constitué par la Commission européenne a par conséquent publié un document complet sur la question[4]. L'observatoire « AI Watch » de la Commission européenne a aussi conduit une étude approfondie en vue de proposer une définition opérationnelle et une taxonomie de l'IA[5]. La Recommandation sur l'intelligence artificielle adoptée par le Conseil de l'OCDE comporte un préambule définissant ce qu'il faut entendre par « système d'IA », « cycle de vie d'un système d'IA », « connaissances en matière d'IA », « acteurs de l'IA » et « parties prenantes »[6]. L'Unesco a produit une étude préliminaire dans laquelle il est fait référence aux « machines basées sur l'IA » et à « l'informatique cognitive »[7], ainsi qu'un projet de recommandation sur l'éthique de l'intelligence artificielle qui envisage les systèmes d'IA comme « des systèmes technologiques capables de traiter l'information par un processus s'apparentant à un comportement intelligent, et comportant généralement des fonctions de raisonnement, d'apprentissage, de perception, d'anticipation, de planification ou de contrôle »[8].

7. S'agissant des instruments non contraignants publiés jusqu'à présent sur ce sujet par le Conseil de l'Europe, on n'y trouve de la même façon aucune définition uniforme de l'intelligence artificielle. La Recommandation du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme[9] définit la notion de « systèmes algorithmiques », qui peut couvrir tout ou partie des applications de l'IA. La Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques[10] ne donne pas de définition et utilise indifféremment des concepts tels que « technologies », « systèmes fondés sur les données », « outils d'apprentissage automatique », en fonction des points à considérer. La Commissaire aux droits de l'homme[11], le comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (T-PD)[12] et la Commission européenne pour l'efficacité de la justice (CEPEJ)[13] utilisent une définition générique à peu près semblable, qui désigne un ensemble de sciences, théories et techniques.

---

4. GEHN IA, *A Definition of AI : Main Capabilities and Disciplines* [Définition de l'IA : principales capacités et disciplines scientifiques], avril 2019.
5. AI Watch, Service commun de recherche, *Defining artificial intelligence : towards an operational definition and taxonomy of artificial intelligence*, février 2020.
6. OCDE, Recommandation du Conseil sur l'intelligence artificielle, juin 2019.
7. Unesco, Étude préliminaire sur les aspects techniques et juridiques liés à l'opportunité d'un instrument normatif sur l'éthique de l'intelligence artificielle, mars 2019.
8. Unesco, Avant-projet de recommandation sur l'éthique de l'intelligence artificielle, septembre 2020.
9. Conseil de l'Europe, Comité des Ministres, Recommandation CM/Rec(2020)1 du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme, avril 2020.
10. Comité des Ministres, Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques, février 2019.
11. Commissaire aux droits de l'homme, Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme, mai 2019.
12. Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (T-PD), Lignes directrices sur l'intelligence artificielle et la protection des données, janvier 2019.
13. CEPEJ, Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, décembre 2018.

8. En résumé, on peut conclure que le terme «IA» est utilisé comme un mot «foure-tout» pour diverses applications informatiques basées sur différentes techniques, qui présentent des capacités communément et actuellement associées à l'intelligence humaine. Ces techniques peuvent consister en des modèles formels (ou systèmes symboliques) ainsi qu'en des modèles fondés sur des données (systèmes basés sur l'apprentissage) qui reposent généralement sur des approches statistiques, y compris par exemple l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage de renforcement. Les systèmes d'IA agissent dans le monde réel ou numérique en percevant leur environnement par l'acquisition de données, en analysant certaines données structurées ou non structurées, en raisonnant sur les connaissances récoltées ou en traitant les informations dérivées des données, et sur cette base, décident des meilleures actions à prendre pour atteindre un certain objectif. Ils peuvent également être conçus pour adapter leur comportement dans le temps en fonction de nouvelles données et améliorer leurs performances en vue d'atteindre un certain objectif.

9. Les membres du CAHAI et les personnes associées à ses travaux en qualité d'observateur ou de participant ont également différentes approches concernant la (nécessaire) définition de l'intelligence artificielle, du fait de leurs diverses traditions et cultures juridiques. Un consensus s'est toutefois dégagé sur la nécessité d'aborder les systèmes d'IA d'une manière technologiquement neutre (c'est-à-dire quelle que soit la technologie sous-jacente utilisée), de façon à prendre en considération toutes les technologies permettant une prise de décision automatisée qui peuvent être englobées sous ce terme générique, y compris le contexte sociotechnique dans lequel elles s'inscrivent. Un équilibre devrait en outre être trouvé entre, d'une part, une définition trop précise d'un point de vue technique, qui pourrait dès lors devenir obsolète sur le court terme, et, de l'autre, une définition qui serait trop vague, laissant ainsi une grande marge d'interprétation, pouvant entraîner une application non uniforme du cadre juridique.[14]

10. Partant, un futur cadre juridique relatif à l'intelligence artificielle devrait adopter une définition simplifiée et technologiquement neutre de son objet, afin de couvrir les pratiques ou les cas d'application où le développement et l'utilisation de systèmes d'IA, ou plus généralement de systèmes décisionnels automatisés, peuvent avoir un impact sur les droits de l'homme, la démocratie et l'État de droit, et en tenant compte de toutes les implications socio-techniques des systèmes[15].

3. OPPORTUNITÉS ET RISQUES DÉCOULANT DE LA CONCEPTION, DU DÉVELOPPEMENT ET DE L'APPLICATION DE L'INTELLIGENCE ARTIFICIELLE POUR LES DROITS DE L'HOMME, L'ÉTAT DE DROIT ET LA DÉMOCRATIE.

3.1. Introduction

11. Comme souligné dans plusieurs documents du Conseil de l'Europe, dont des rapports récemment adoptés par l'Assemblée parlementaire (APCE)[16], les systèmes d'IA sont en train de transformer profondément nos vies et ont un impact sans précédent sur le tissu social et le fonctionnement des institutions. Leur utilisation peut permettre de générer des bénéfices substantiels dans de nombreux domaines tels que la santé, les transports, l'éducation ou l'administration publique, et ouvre des possibilités prometteuses pour

14. Quelques membres du CAHAI ont souligné l'importance de délimiter plus clairement la définition de l'IA qui devrait être utilisée aux fins du cadre juridique. Cette délimitation - qui aidera également à définir la portée du cadre juridique - devrait être soigneusement étudiée par le groupe Cadres juridiques du CAHAI et pourrait également faire partie des consultations des parties prenantes envisagées.

15. Si l'on examine d'autres instruments juridiques du Conseil de l'Europe traitant de domaines scientifiques, l'on peut noter que la Convention sur les droits de l'homme et la biomédecine (Convention d'Oviedo, STE n°164) ne définit pas non plus son objet. La Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel («Convention 108» et «108+»), pour sa part, définit la notion de «traitement de données», qui découle de l'utilisation des technologies de l'information, sans mentionner explicitement les objets techniques, tels que les algorithmes. La Convention définit néanmoins son objet, les «données à caractère personnel», permettant ainsi de déterminer si une opération de traitement relève ou non de son champ. Toute nouvelle réglementation du Conseil de l'Europe ne doit pas contrevenir aux instruments existants du Conseil de l'Europe tels que la Convention sur les droits de l'homme et la biomédecine, la Convention 108 et 108+.

16. Voir par exemple les rapports ci-après de l'Assemblée parlementaire du Conseil de l'Europe: «La nécessité d'une gouvernance démocratique de l'intelligence artificielle», «Le rôle de l'intelligence artificielle dans les systèmes de police et de justice pénale», «Prévenir les discriminations résultant de l'utilisation de l'intelligence artificielle», «Les interfaces cerveau-machine: nouveaux droits ou nouveaux dangers pour les libertés fondamentales?», «Intelligence artificielle et santé», «Intelligence artificielle et marchés du travail», «Aspects juridiques concernant les "véhicules autonomes"». Voir également la recommandation de la Commissaire aux droits de l'homme, «Décoder l'intelligence artificielle: 10 mesures pour protéger les droits de l'homme», et la Recommandation CM/Rec(2020)1 du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme.

l'humanité tout entière. En même temps, le développement de systèmes d'IA n'est pas sans présenter des risques **réels**, en particulier lorsque leur utilisation porte atteinte aux droits de l'homme, à la démocratie et à l'**État de droit**, **à savoir les fondements mêmes de nos sociétés européennes**.

12. Les systèmes d'IA devraient être vus comme des « **systèmes sociotechniques** ». Leur impact – quelle que soit la technologie sur laquelle ils se basent – dépend non seulement de la conception des systèmes, mais aussi de la façon dont ils sont développés et utilisés dans un cadre plus large. À cet égard, il y a lieu de prendre en considération les données utilisées, la finalité des systèmes, leur fonctionnalité, leur degré de précision et l'ampleur de leur **déploiement**, outre le contexte organisationnel, sociétal et juridique plus large dans lequel ils sont mis en œuvre[17]. Leurs conséquences, positives ou négatives, sont aussi fonction des valeurs et du comportement des personnes chargées de leur développement et de leur déploiement, d'où l'importance de garantir l'existence d'une responsabilité humaine. Les systèmes d'IA présentent cependant des caractéristiques décisives qui les distinguent d'autres technologies s'agissant de leurs incidences, positives ou négatives, sur les droits de l'homme, la démocratie et l'**État de droit**[18].

13. Premièrement, **l'échelle, la connectivité et la portée des systèmes d'IA** peuvent amplifier certains risques qui sont aussi inhérents à d'autres technologies ou au comportement humain. Les systèmes d'IA sont capables d'analyser des volumes sans précédent de données précises (y compris des données à caractère personnel extrêmement sensibles) beaucoup plus rapidement que les personnes. Cela peut conduire à les utiliser d'une manière qui perpétue ou amplifie des biais injustes[19], lesquels sont aussi fondée sur de nouveaux motifs de discrimination en cas de discrimination dite « par procuration »[20]. L'importance accrue de la discrimination par procuration dans le contexte de l'apprentissage automatique peut soulever des questions d'interprétation sur la distinction entre discrimination directe et indirecte ou, en fait, sur l'adéquation de cette distinction telle qu'elle est traditionnellement comprise. De surcroît, les systèmes d'IA sont sujets à des marges d'erreurs statistiques. Même si la marge d'erreur d'un système appliqué à des millions de personnes est proche de zéro, des milliers de personnes peuvent encore être affectées négativement en raison de l'ampleur du déploiement et de l'interconnectivité des systèmes. D'autre part, l'échelle et la portée des systèmes d'IA impliquent également qu'ils peuvent être utilisés pour atténuer certains risques et biais qui sont inhérents à d'autres technologies ou au comportement humain, et pour surveiller et réduire les taux d'erreur humaine.

14. Deuxièmement, **la complexité ou l'opacité** de nombreux systèmes d'IA (notamment dans le cas de l'apprentissage automatique) peut rendre impossible, y compris pour leurs développeurs, la compréhension de leur fonctionnement ou du cheminement ayant abouti au résultat. Cette opacité, combinée au fait que plusieurs acteurs interviennent à différents stades du cycle de vie du système, complique encore l'identification du ou des responsables d'éventuelles conséquences négatives des applications d'IA, réduisant ainsi la responsabilité humaine et l'obligation de rendre des comptes (accountability).

15. Troisièmement, certains systèmes d'IA peuvent se **recalibrer** eux-mêmes grâce à la rétroaction (*feedback*) et à l'apprentissage par renforcement (*reinforcement learning*).Toutefois, si un système d'IA est reformé sur des données résultant de ses propres décisions qui contiennent des biais injustes, des erreurs, des inexactitudes ou d'autres déficiences, cela pourrait créer une boucle de rétroaction vicieuse qui peut conduire à un fonctionnement discriminatoire, erroné ou malveillant du système et qui peut être difficile à détecter.

3.2. Les opportunités de l'intelligence artificielle

16. Les systèmes d'IA peuvent avoir un impact très positif sur la société. En tant que levier de développement socio-économique mondial, ils peuvent contribuer à atténuer certains des problèmes auxquels le monde est confronté et atteindre les objectifs de développement durable des Nations unies[21]. Les systèmes d'IA peuvent optimiser les procédés agricoles, révolutionner les transports et les modes de vie urbains, aider à atténuer les effets du changement climatique ou prédire les catastrophes naturelles et faciliter un meilleur accès à l'information et à la connaissance.

17. Voir aussi la Recommandation Rec/CM(2020)1 du Comité des Ministres.
18. Ces trois facteurs sont interdépendants et se renforcent mutuellement. Étant donné l'évolution rapide de la technologie et les inconnues quant à ses usages futurs, cette liste n'est ni exhaustive ni définitive, mais devra être actualisée sans cesse.
19. Par « biais injuste », on entend une violation du droit à l'égalité et à la non-discrimination dans un contexte d'application spécifique de la technologie de l'IA.
20. Voir par exemple l'étude du Conseil de l'Europe par F. Zuiderveen Borgesius, *Discrimination, artificial intelligence, and algorithmic decision-making*, 2018 ; *Affinity Profiling and Discrimination by Association in Online Behavioural Advertising*, Wachter 2020.
21. Voir "*The role of artificial intelligence in achieving the Sustainable Development Goals*", Nature, https://www.nature.com/articles/s41467-019-14108-y.

17. En effet, les systèmes d'IA peuvent fournir des capacités intelligentes dans de nombreux domaines qui sont utiles aux individus et à la société dans son ensemble, et compte tenu de leur efficacité et de leurs effets à grande échelle, être utilisés pour aider à surmonter certains des obstacles posés par la disponibilité limitée des capacités cognitives et décisionnelles humaines. Ils peuvent améliorer considérablement l'efficacité des pratiques industrielles existantes, contribuer au développement de nouvelles applications industrielles et renforcer leur sécurité. Les systèmes d'IA peuvent également conduire à la création de nouveaux services, produits, marchés et industries, qui peuvent accroître considérablement le bien-être des citoyens et de la société dans son ensemble et être utilisés pour soutenir des applications et des services socialement bénéfiques. Les solutions d'IA peuvent également améliorer la cybersécurité en permettant de détecter les comportements malveillants et d'automatiser la première réponse aux cyber-attaques (de faible niveau).

18. L'un des attributs les plus significatifs des systèmes d'IA est leur incidence potentielle sur la santé des individus et sur les systèmes de santé. Cela englobe l'amélioration des diagnostics et des traitements, l'amélioration de la santé fœtale, les diagnostics prédictifs, ainsi que ainsi que la prédiction et la surveillance avancées des épidémies et des maladies chroniques. Certaines opportunités générées par les systèmes d'IA peuvent aussi être observées dans le contexte de la pandémie de Covid-19. Des systèmes d'IA sont déployés pour étudier le virus, accélérer la recherche médicale, mettre au point un vaccin, détecter et diagnostiquer les infections, prédire l'évolution du virus, et assurer un échange rapide des informations. De plus, dans d'autres domaines

19. De plus, dans d'autres domaines, les systèmes d'IA peuvent transformer la portée et la manière dont la recherche est menée, et peuvent être utilisés pour faire progresser et accélérer les découvertes scientifiques qui profitent à la société dans son ensemble. Au-delà de la recherche, les systèmes d'IA peuvent également être utilisés pour améliorer les possibilités d'éducation en permettant des approches d'apprentissage personnalisées et en augmentant la disponibilité de l'éducation à une plus grande échelle.

20. Enfin, plus généralement, les systèmes d'IA peuvent promouvoir et renforcer les droits de l'homme et contribuer à faire en sorte qu'ils soient respectés et effectivement appliqués, par exemple en détectant des décisions (humaines ou automatisées) biaisées, en surveillant les modes de représentation de différents groupes (comme les femmes dans les médias) ou en analysant les structures discriminatoires au sein des organisations. Ils peuvent aussi renforcer l'*État de droit* et la démocratie en améliorant l'efficience des procédures administratives, en aidant les pouvoirs publics à mieux répondre aux besoins et en libérant du temps pour aborder les questions complexes et importantes. Les systèmes d'IA peuvent aussi aider les autorités et d'autres acteurs publics à mieux identifier les besoins et préoccupations des citoyens, contribuant ainsi *à l'élaboration de politiques plus* efficace.

3.3. Impact sur les droits de l'homme, la démocratie et l'État de droit

21. Malgré ces bénéfices, le recours croissant aux systèmes d'IA dans tous les domaines de la vie publique et privée soulève aussi des enjeux majeurs au regard des droits de l'homme, de la démocratie et de l'*État de droit*[22]. Nous allons les examiner tour à tour ci-après. Les droits de l'homme étant une composante essentielle de l'*État de droit* et de la démocratie, l'analyse des défis que posent les systèmes d'IA pour l'*État de droit* et la démocratie doit nécessairement être lue conjointement avec l'analyse de l'impact de l'intelligence artificielle sur les droits de l'homme.

3.3.1. Impact sur les droits de l'homme

22. Le développement et l'utilisation de systèmes d'IA ont une incidence sur un large éventail de droits de l'homme.[23] Les principaux aspects sont exposés brièvement ci-dessous, en mettant plus particulièrement l'accent sur les droits énoncés par la Convention européenne des droits de l'homme (CEDH) et ses protocoles, ainsi que par la Charte sociale européenne (CSE).

Liberté et sûreté ; Procès équitable ; Pas de peine sans loi (art. 5, 6 et 7 CEDH)

23. Le risque de voir les systèmes d'IA faciliter ou amplifier des biais injustes, comme indiqué plus haut, peut constituer une menace pour le droit à la liberté et à la sûreté, combiné au droit à un procès équitable

22. Voir le rapport établi par Catelijne Muller (CAHAI(2020)06-fin), *Impact de l'intelligence artificielle sur les droits de l'homme, la démocratie et l'État de droit*.

23. Voir par exemple l'étude de la FRA (Agence des droits fondamentaux de l'UE) sur la technologie de reconnaissance faciale et les droits fondamentaux (2019) : https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf. Voir également le nouveau rapport de la FRA, «Getting the Future Right - Artificial Intelligence and Fundamental Rights in the EU», Luxembourg : Office des publications de l'Union européenne, 14 décembre 2020,

(art. 5, 6 et 7 CEDH), lorsque ces systèmes sont utilisés dans des situations susceptibles de mettre en jeu la liberté physique des individus ou la sûreté de la personne (domaines de la justice et du maintien de l'ordre par exemple). *À titre d'illustration, l*es systèmes d'IA servant à prédire la récidive se fondent sur les caractéristiques que le suspect partage avec d'autres (adresse, revenus, nationalité, niveau d'endettement, profession). Une telle approche suscite des préoccupations quant au respect du principe d'individualisation de la peine et d'autres aspects essentiels du droit à un procès équitable[24]. En outre, du fait de l'opacité du système d'IA, il peut être impossible de comprendre le raisonnement mis en œuvre pour arriver à un résultat donné. Dès lors, il devient difficile, voire impossible, d'assurer le plein respect du principe de l'égalité des armes, de contester la décision ou de demander réparation. Cependant, maniées de façon responsable, certaines applications d'IA peuvent aussi rendre le travail de la justice et des forces de l'ordre plus efficace et, par conséquent, avoir une incidence positive sur les droits susmentionnés. Cela nécessite des efforts supplémentaires pour renforcer les compétences des acteurs judiciaires dans leur connaissance et leur compréhension des systèmes d'IA et de leur application.

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

24. La protection offerte par l'article 8 de la Convention englobe un large éventail d'aspects de la vie privée, qui peuvent être divisés en trois grandes catégories : i) la vie privée d'une personne (en général), ii) son intégrité physique, psychologique ou morale et iii) son identité et son autonomie[25]. Plusieurs applications de l'intelligence artificielle peuvent affecter tous ces aspects. C'est notamment le cas lorsque l'on traite des données à caractère personnel (pour identifier ou surveiller des individus par exemple), mais de telles situations sont également possibles sans traitement de données personnelles. En ce qui concerne les applications de l'IA à caractère invasif, il s'agit en particulier des systèmes de reconnaissance faciale ou d'autres paramètres biométriques tels que les micro-expressions du visage, la démarche, (le timbre de) la voix, le rythme cardiaque ou la température[26]. Au-delà de leur utilisation aux fins d'identification, ces données peuvent aussi servir à analyser et prédire le comportement d'une personne, ou à l'influencer. Il est ainsi procédé à des profilages ou à des catégorisations des individus pour des finalités diverses et dans différents contextes, de la police prédictive aux assurances[27]. De nombreux éléments indiquent par ailleurs que l'utilisation de la technologie de reconnaissance biométrique peut entraîner une discrimination fondée notamment sur la couleur de la peau et/ou du sexe, lorsque le biais dans l'algorithme ou l'ensemble de données de base n'est pas suffisamment pris en compte.[28]

25. De plus, le recours généralisé aux techniques de traçage porte une atteinte considérable à la vie privée (en général), *à l'identité et à l'autonomie en créant une* situation où les individus sont constamment surveillés, suivis, identifiés et influencés, ce qui affecte aussi leur intégrité psychologique et morale. Cela peut pousser les personnes à se comporter selon une certaine norme, modifiant ainsi les rapports de pouvoir entre l'État ou l'entité privée qui utilise les technologies de surveillance et les individus (ou un groupe d'individus)[29]. Le pistage indiscriminé – 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 intelligents, thermostats, voitures, etc.) est de la même façon susceptible de porter atteinte au droit à la vie privée et notamment à l'intégrité psychologique. Le droit à la vie privée implique

24. L'utilisation problématique des systèmes d'IA (tels que le système COMPAS utilisé aux États-Unis) a été démontrée par plusieurs études, notamment l'étude de Dartmouth *The accuracy, fairness, and limits of predicting recidivism* par Julia Dressel et Hany Farid, Science Advances 17 jan 2018, Vol. 4, no. 1, DOI : 10.1126/sciadv.aao5580. En même temps, en utilisant des approches plus responsables, certaines études ont indiqué que les systèmes d'IA peuvent également aider à améliorer les prévisions. Voir, par exemple, *The limits of human predictions of recidivism*, Zhiyuan Lin, Jongbin Jung, Sharad Goel et Jennifer Skeem, Science Advances,14 Feb 2020, Vol. 6, no. 7, DOI : 10.1126/sciadv.aaz0652.
25. Voir Conseil de l'Europe, *Guide sur l'article 8 de la Convention européenne des droits de l'homme*.
26. Il ressort clairement de la jurisprudence de la Cour européenne des droits de l'homme que la collecte, la conservation et le traitement de telles données, même pour une courte durée, relèvent du champ de l'article 8 de la Convention.
27. Il est à noter qu'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 via d'autres données biométriques. Voir à ce propos l'étude de Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M. et Pollak, S. D. (2019).
28. Voir par exemple l'étude du MIT par Joy Buolamwini (2018) : https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212 ; l'étude de l'Institut National Américain des standards et de la technologie (*US National Institute of Standards and Technology*) sur la reconnaissance des visages (2019) : https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf ; Study by FRA (EU Agency for Fundamental Rights) on Facial recognition technology and fundamental rights (2019), pages 26-27 : https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf.
29. Voir Muller, C., CAHAI(2020)06-fin, par. 18 ; Cohen, Julie E., *Examined Lives : Informational Privacy and the Subject as Object*, 2000.

pourtant un droit à une sphère privée exempt de surveillance par l'IA, nécessaire au développement personnel et à la démocratie [30].

Liberté d'expression ; Liberté de réunion et d'association (art. 10 et 11 CEDH)

26. L'utilisation de systèmes d'IA – en ligne et hors ligne – peut porter atteinte à la liberté d'expression, de même qu'à la liberté de réunion et d'association.[31] Les applications d'IA peuvent être utilisées pour intervenir dans l'espace media avec une grande efficacité, et modifier substantiellement les interactions humaines. L'immense potentiel de l'internet et des plates-formes de réseaux sociaux a été mis à profit par les citoyens pour s'organiser et exercer leurs droits de réunion pacifique et d'association. Dans le même temps, la surveillance par intelligence artificielle peut compromettre ces droits dans la mesure où elle s'appuie sur des systèmes automatisés de pistage et d'identification des individus (ou groupes d'individus) concernés ; cela peut même conduire à les exclure d'une manifestation ou d'un rassemblement contestataire[32]. Le pistage personnalisé des individus – dans le monde virtuel et dans la vie réelle – est par ailleurs susceptible d'entraver l'exercice de ces droits en diminuant la protection conférée par « l'anonymat de groupe ». Cela peut dissuader la participation aux manifestations pacifiques et, plus généralement, conduire à ne pas exprimer ouvertement ses opinions et à éviter de regarder certains médias ou de lire certains ouvrages ou journaux.

27. Les systèmes d'IA peuvent en outre porter atteinte au droit de recevoir ou de communiquer des informations ou des idées lorsque l'IA est utilisée dans les médias (sociaux) en ligne et les sites d'actualités pour sélectionner des informations ou présenter les contenus en fonction des centres d'intérêt ou des habitudes de l'internaute. Cela peut également renforcer des normes sociales dépassées, notamment les stéréotypes fondés sur le sexe, et alimenter la polarisation et l'extrémisme en créant des « chambres d'écho » et des « bulles filtrantes ».[33] Les moteurs de recherche, systèmes de recommandation et agrégateurs d'actualités manquent souvent de transparence tant en ce qui concerne les données qu'ils utilisent pour sélectionner ou hiérarchiser le contenu, mais aussi en ce qui concerne la finalité de la sélection ou de la hiérarchisation spécifique[34] qu'ils peuvent utiliser pour la promotion des intérêts financiers et politiques. Les systèmes d'IA peuvent être utilisés pour sélectionner et hiérarchiser les contenus dans le seul but de faire rester les internautes le plus longtemps possible sur un site, peu importe que ces contenus soient ou non objectifs, fondés sur des données factuelles, variés ou pertinents. En outre, l'information est de plus en plus « faussée » par la production de contenus médiatiques de synthèse imitant l'apparence ou la voix de personnes réelles en utilisant ce qu'on appel du « deep fakes ». Cette technologie permet déjà de manipuler ou de générer des contenus visuels et audios ayant un potentiel de tromperie sans précédent, à tel point qu'il devient difficile de distinguer le vrai du faux. Cela peut considérablement entraver la capacité de chacun à former et exprimer librement ses opinions et à recevoir ou communiquer des informations ou des idées, et pourrait provoquer une érosion de la société de l'information[35]. Par ailleurs, les plateformes en ligne se tournent de plus en plus vers les systèmes d'IA pour identifier, signaler, déclasser et supprimer les contenus qui enfreignent leurs conditions de service. Les inexactitudes des systèmes d'IA peuvent avoir pour conséquence que des contenus légitimes - protégés par le droit à la liberté d'expression - soient signalés ou supprimés par erreur. Cela est particulièrement complexe pour les contenus qui nécessitent une compréhension des nuances et du contexte, en rapport avec des domaines tels que le discours de haine et la désinformation. Enfin, comme les plateformes en ligne ont revendiqué des audiences et des revenus publicitaires, certains médias d'information traditionnels ont eu du mal à survivre. Les menaces qui pèsent sur la viabilité des médias d'information, liées à la consommation de nouvelles et d'informations par le biais des plateformes en ligne, présentent un risque pour un écosystème médiatique libre, indépendant et pluraliste.

---

30. Entre ici en jeu la notion d'« effet dissuasif » (*chilling effect*), c'est-à-dire le fait de créer des inhibitions ou de décourager l'exercice légitime d'un droit. Des études ont montré que les personnes se sachant surveillées commencent à modifier leur comportement et évoluent différemment. Staben, J., *Der Abschreckungseffekt auf die Grundrechtsausübung : Strukturen eines verfassungsrechtlichen Arguments*, Mohr Siebeck, 2016.
31. Pour l 'impact de l'IA sur la liberté d'expression, voir UNESCO, 2019, *Steering AI and Advanced ICTs for Knowledge Societies : A Rights, Openness, Access, and Multi-stakeholder Perspective* sur https://unesdoc.unesco.org/ark :/48223/pf0000372132.
32. Conseil de l'Europe, *Algorithmes et droits humains, Étude sur les dimensions des droits humains dans les techniques de traitement automatisé des données et éventuelles implications réglementaires*, Éditions du Conseil de l'Europe, 2018.
33. Voir par exemple l'étude réalisée par les chercheurs de Carnegie Mellon : https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study
34. Burrell, J., "*How the machine 'thinks' : Understanding opacity in machine learning algorithms*", *Big Data & Society*, 3(1), DOI : 2053951715622512, 2016.
35. Nations Unies, Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, A/73/348. Les effets de ce phénomène sur la démocratie seront examinés plus loin.

Égalité et non-discrimination (art. 14 CEDH, Protocole 12)

28. L'impact des systèmes d'IA sur l'interdiction de la discrimination et le droit à l'égalité de traitement est l'un des aspects les plus largement abordés. Comme indiqué plus haut, l'intelligence artificielle peut servir à détecter et atténuer les biais humains. En même temps, elle peut aussi contribuer à perpétuer ou amplifier les biais et les stéréotypes[36], le sexisme, le racisme, la discrimination fondée sur l'âge, les discriminations fondées sur divers motifs et d'autres discriminations injustes (y compris les discriminations fondées sur des critères indirects ou intersectionnels[37]), ce qui crée un nouveau défi à la non-discrimination et à l'égalité de traitement.

29. Il existe de multiples facteurs de risques de discrimination, par exemple si l'échantillon d'apprentissage est biaisé (lorsque l'ensemble de données n'est pas représentatif ou comporte des inexactitudes, par exemple) ou si des biais sont introduits dans la conception même de l'algorithme ou dans sa fonction d'optimisation, en raison d'une conception biaisée de l'algorithme ou de sa fonction d'optimisation (par exemple, en raison des stéréotypes ou préjugés conscients ou inconscients des développeurs), en raison de l'exposition à un environnement biaisé une fois qu'il est utilisé, ou en raison d'une utilisation biaisée du système d'IA. À titre d'illustration, compte tenu des discriminations subies par les femmes dans le passé, en droit et en fait, les bases de données historiques pourront manquer de données suffisamment équilibrées sous l'angle des sexes. Si une telle base de données est ultérieurement utilisée par des systèmes d'IA, cela peut donner lieu à des décisions biaisées et perpétuer ainsi une discrimination injuste. Il en va de même pour d'autres groupes traditionnellement vulnérables, exclus ou marginalisés. En outre, la sous-représentation des groupes susmentionnés dans le secteur de l'intelligence artificielle pourrait encore contribuer à amplifier ce risque[38]. Des mesures visant à favoriser la diversité des effectifs du secteur de l'IA en termes de sexe, d'appartenance ethnique et d'origine sociale pourraient contribuer à atténuer certains de ces risques.

30. En outre, lorsque la transparence des processus décisionnels des systèmes d'IA n'est pas assurée, et en l'absence de toute obligation de signalement ou d'exigences en matière d'audit, l'existence de tels biais peut aisément passer inaperçue ou même être occultée[39], marginalisant ainsi les mécanismes de contrôle social qui régissent généralement les comportements humains.[40]

Droits économiques et sociaux (art. 2, 3, 5, 11, 12, 13 et 20 CSE)

31. Les systèmes d'IA peuvent offrir des avantages considérables lorsqu'il s'agit d'effectuer des travaux dangereux, pénibles, épuisants, désagréables, répétitifs ou ennuyeux. Cependant, l'adoption généralisée de l'IA dans tous les domaines de la vie quotidienne crée aussi de nouvelles menaces pour les droits économiques et sociaux. Les systèmes d'IA sont 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. Dans certaines situations, cela peut également avoir des conséquences préjudiciables pour le droit des travailleurs à un salaire décent, car leur salaire peut être déterminé par des algorithmes de manière irrégulière, incohérente et insuffisante.[41] En outre, ils peuvent aussi servir à détecter et contrer des tentatives de syndicalisation des travailleurs. Ces applications peuvent compromettre le droit à des conditions de travail équitables, à la sécurité et à l'hygiène dans le travail et à la dignité dans le travail, de même que le droit syndical. Des biais injustes dans les systèmes d'analyse prédictive

---

36. Cela peut notamment inclure les stéréotypes de genre, qui sont «des idées préconçues selon lesquelles les hommes et les femmes se voient attribuer arbitrairement des caractéristiques et des rôles déterminés et limités par leur sexe». Voir à ce sujet : https://rm.coe.int/prems-093718-fra-gender-equality-strategy-2023-web-a5-corrige/16808e0809
37. A cet égard, voir aussi la note de bas de page 154.
38. Selon une étude de l'AI Now Institute (*Discriminating systems, Gender, Race, and Power in AI*), en 2019, les femmes représentaient 15 % seulement des personnels de recherche sur l'IA chez Facebook et 10 % chez Google. Il n'y a pas de données publiques concernant les travailleurs trans ou d'autres minorités sexuelles. S'agissant des Noirs, le tableau est encore pire. Ainsi, 2,5 % seulement des effectifs de Google sont noirs, tandis que Facebook et Microsoft en sont chacun à 4 %. Sachant que ce déséquilibre est un sujet de préoccupation depuis des décennies et vu les investissements consentis pour tenter d'y remédier, la situation actuelle dans ce domaine est alarmante.
39. C'est ce que montre, par exemple, le faible nombre de plaintes adressées aux autorités responsables, y compris les organismes nationaux pour l'égalité de traitement (*NEB*), mais aussi les affaires judiciaires.
40. Dans le domaine de la législation anti-discrimination, des règles spécifiques sur le partage de la charge de la preuve sont généralement utilisées dans le but de compenser ce manque de transparence.
41. Ce problème peut être lié à une question plus générale liée au modèle de l'économie « des petits boulots » (*gig economy*) ou des plate-formes de crowdsourcing (des services de livraison aux services d'enrichissement des données), qui est généralement basé sur le modèle du « travailleur en tant qu'entrepreneur indépendant » (au lieu d'un employé généralement plus protégé), selon lequel les travailleurs n'ont souvent pas accès aux allocations de chômage, aux congés de maladie, aux vacances et aux prestations sociales générales.

appliquée au recrutement, utilisés pour sélectionner les candidats ou prédire la performance, peuvent aussi compromettre l'égalité (hommes-femmes notamment) en matière d'emploi et de travail.

32. En outre, les systèmes d'IA peuvent, par exemple, être utilisés dans le cadre de décisions de sécurité sociale, auquel cas le droit garanti par l'article 12 de la Charte sociale européenne - qui dispose que tous les travailleurs et leurs personnes à charge ont droit à la sécurité sociale - peut être affecté. En effet, l'administration de la sécurité sociale s'appuie de plus en plus sur les systèmes d'IA, et les décisions prises dans ce contexte peuvent avoir un impact significatif sur la vie des individus. Des problèmes similaires se posent lorsque les systèmes d'assurance sociale sont déployés dans le cadre des administrations chargées de l'éducation ou de l'attribution des logements.

33. En outre, lorsque les systèmes d'IA sont utilisés pour automatiser les décisions concernant la fourniture de soins de santé et d'assistance médicale, cette utilisation peut également avoir un impact sur les droits consacrés par les articles 11 et 13 de la Charte, qui disposent respectivement que toute personne a le droit de bénéficier de mesures lui permettant de jouir du meilleur état de santé possible et que toute personne ne disposant pas de ressources suffisantes a droit à une assistance sociale et médicale. Les systèmes d'IA peuvent, par exemple, être utilisés pour déterminer l'accès des patients aux services de soins de santé en analysant les données personnelles des patients, telles que leurs dossiers de soins de santé, les données relatives à leur mode de vie et d'autres informations. Il est important que cela se fasse dans le respect non seulement du droit à la vie privée et à la protection des données personnelles, mais aussi de tous les droits sociaux énoncés dans la charte susmentionnée, dont l'impact a jusqu'à présent reçu moins d'attention que celui sur les droits civils et politiques.

3.3.2. Impact sur la démocratie

34. Le développement et l'utilisation de systèmes d'IA peuvent aussi avoir une incidence sur le fonctionnement des institutions et processus démocratiques, de même que sur le comportement social et politique des citoyens et de la société civile dans son ensemble[42]. Les systèmes d'IA peuvent être utilisés pour améliorer la qualité de la gouvernance (obligation de rendre des comptes, réactivité et efficacité des institutions publiques), aider à lutter contre la corruption et promouvoir le pluralisme. Ils peuvent permettre d'élargir l'espace démocratique en décentralisant les systèmes d'information et les plates-formes de communication. En outre, ils peuvent améliorer la manière dont les citoyens et la société civile dans son ensemble s'informent sur les processus politiques et les aider à y participer à distance, en facilitant l'expression politique et en fournissant des canaux de communication avec les acteurs politiques. En même temps, les systèmes d'IA sont susceptibles de donner lieu à des utilisations qui, délibérément ou non, font obstacle à la démocratie. [43]

35. Une démocratie qui fonctionne repose sur un discours social et politique ouvert, ainsi que sur l'absence d'influence indue des électeurs ou de leur manipulation. Comme indiqué ci-dessus, les technologies d'IA peuvent être utilisées pour interférer dans l'espace médiatique (social) en ligne à des fins de gains financiers ou politiques privés plutôt que dans l'intérêt public.[44] Si la propagande et la manipulation ne sont pas nouvelles, les outils basés sur l'IA ont amplifié leur échelle et leur portée, et ont facilité une itération rapide pour renforcer leur capacité à influencer les gens. Ils permettent de mener des campagnes de désinformation à grande échelle mais ciblées, par le biais de comportements inauthentiques coordonnés, par exemple par le biais de contenus profondément falsifiés, de faux comptes, du microciblage illegal des électeurs et de la polarisation du débat public. En outre, elles peuvent menacer de compromettre la capacité d'action humaine et l'autonomie nécessaires à la prise de décisions significatives par les électeurs, qui sont au cœur de la création d'institutions légitimes.[45] En conséquence, certaines utilisations de l'IA peuvent miner à la confiance dans les institutions démocratiques et entraver le processus électoral.

36. Plus généralement, ces risques peuvent être accrus par la concentration du pouvoir entre les mains de quelques plates-formes privées avec une réglementation limitée jusqu'à présent, alors qu'elles sont devenues

---

42. Pour de plus amples informations sur l'impact des systèmes d'IA sur la démocratie, voir le rapport de l'Assemblée parlementaire du Conseil de l'Europe sur « La nécessité d'une gouvernance démocratique de l'intelligence artificielle » (Doc. 15150).
43. Maja Brkan, '*Artificial Intelligence and Democracy' : Delphi* – Revue Interdisciplinaire des technologies émergentes 2 *(Interdisciplinary Review of Emerging Technologies 2)*, no. 2 (2019) : 66–71 https://doi.org/10.21552/delphi/2019/2/4.
44. Si les technologies de filtrage automatisé des contenus peuvent contribuer à limiter l'affichage de contenus illégaux ou autrement problématiques, dans certaines situations, leur utilisation peut également restreindre les discussions sur l'égalité des sexes ou les préoccupations liées aux discours haineux, comme l'a par exemple constaté l'étude du Parlement européen sur la modération des contenus en ligne (2020), https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652718/IPOL_STU(2020)652718_EN.pdf p 59.
45. Voir également le rapport sur le traitement des données à caractère personnel par et pour les campagnes politiques : L'application de la Convention 108 modernisée du Conseil de l'Europe» par Colin J. Bennett, https://rm.coe.int/t-pd-2020-02rev-political-campaigns-en-clean-cjb-/1680a01fc3.

de facto partie intégrante de la sphère publique. De surcroît, des partenariats public-privé sur l'usage de l'IA dans des domaines sensibles, comme le maintien de l'ordre ou le contrôle des frontières, peuvent brouiller les frontières entre les intérêts et responsabilités des États démocratiques, d'une part, et ceux des sociétés privées, de l'autre. Cela soulève notamment des questions quant à l'obligation des institutions publiques de rendre des comptes pour les décisions prises par le biais de solutions d'IA fournies par des acteurs privés.[46]

37. Enfin, l'impact de l'IA sur les droits de l'homme peut faire obstacle à la démocratie d'une manière plus générale. L'utilisation par les gouvernements de systèmes d'IA pour contrôler les citoyens, par exemple en opérant un filtrage automatique des informations (censure) ou en mettant en place une surveillance (de masse) par intelligence artificielle, sape les fondements de la démocratie, le libre arbitre des citoyens et érode les libertés politiques – notamment d'expression, d'association et de réunion.

38. Jusqu'à présent, les institutions publiques qui ont eu recours aux systèmes d'IA l'ont fait principalement pour soutenir des décisions administratives standardisées. La perspective de voir les institutions publiques s'en remettre à l'intelligence artificielle pour décider des politiques serait très problématique si elle remplaçait un dialogue entre la majorité et la minorité ou si elle n'était pas soumise à un débat démocratique. En outre, le recours croissant à l'IA pourrait considérablement affecter la nature des pouvoirs de l'État (législatif, exécutif et judiciaire) et les relations qu'ils entretiennent entre eux.

3.3.3. Impact sur l'État de droit

39. En plus d'avoir un impact sur les droits de l'homme et la démocratie, les systèmes d'IA peuvent également avoir une incidence sur l'État de droit.[47] L'État de droit prescrit que tous les pouvoirs publics agissent dans les limites fixées par la loi, conformément aux principes de la démocratie et des droits de l'homme, et sous le contrôle de tribunaux indépendants et impartiaux. Lorsqu'ils sont utilisés de manière responsable, les systèmes d'IA peuvent être utilisés pour accroître l'efficacité de la gouvernance, y compris des institutions juridiques telles que les tribunaux[48], ainsi que des services répressifs et des administrations publiques. [49] En outre, les systèmes d'IA peuvent aider les agences à identifier la corruption au sein des entités publiques[50], ainsi qu'à détecter et à se défendre contre les cyberattaques.[51]

40. L'État de droit exige le respect de principes tels que la légalité, la transparence, la responsabilité, la sécurité juridique, la non-discrimination, l'égalité et une protection judiciaire efficace - qui peuvent être menacés lorsque certaines décisions sont déléguées à des systèmes d'IA. En outre, les systèmes d'IA peuvent également avoir une incidence négative sur le processus d'élaboration des lois et l'application du droit par les juges.[52] Des inquiétudes ont également été exprimées quant aux effets négatifs possibles de certaines applications d'IA utilisées dans les systèmes judiciaires ou dans des domaines connexes.[53] Une telle utilisation pourrait remettre en cause le droit à un procès équitable inscrit à l'article 6 de la CEDH[54], dont des éléments tels que le droit à un pouvoir judiciaire indépendant et impartial, le droit à un avocat ou le principe de

46. Concernant les obligations des entreprises du secteur privé en matière de respect des droits de l'homme, voir « Les principes directeurs des Nations unies sur les entreprises et les droits de l'homme qui soulignent les obligations des entreprises du secteur privé en matière de respect des droits de l'homme » : https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf.
47. Voir par exemple Mireille Hildebrandt, "*Algorithmic Regulation and the Rule of Law", Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences* 376, no. 2128, 2018, 20170355. https://doi.org/10.1098/rsta.2017.0355.
48. Les systèmes d'IA peuvent soutenir le travail des professionnels du droit, par exemple en les aidant à accomplir des tâches complexes telles que l'analyse et la structuration des informations sur les affaires et les documents juridiques, la transcription des procès-verbaux des procédures judiciaires, la promotion de la classification automatisée des documents, ce qui permet d'éliminer beaucoup de temps de traitement pour les tribunaux, les registres d'état civil et les bureaux territoriaux, ou la fourniture d'informations juridiques par l'intermédiaire de chatbots.
49. Danaher, J., « The Threat of Algocracy : Reality, Resistance and Accommodation », *Philosophy & Technology*, vol. 29 (3), 2016, p. 245-268.
50. West, J. et Bhattacharya, M., « Intelligent financial fraud detection : A comprehensive review », *Computers & Security*, vol. 57, 2016, p. 47-66 ; Hajek, P. et Henriques, R., « Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods », *Knowledge-Based Systems*, vol. 128, 2017, p. 139-152.
51. Taddeo, M. et Floridi, L. (2018a), « Regulate artificial intelligence to avert cyber arms race », *Nature*, vol. 556 (7701), 2018, p. 296-298.
52. En favorisant l'émergence de tendances quantitatives d'analyse des décisions judiciaires, le processus traditionnel d'application du droit par le juge pourrait être mis en péril. Voir la Charte éthique européenne de la CEPEJ sur l'utilisation de l'IA dans les systèmes judiciaires et leur environnement, §35. Voir par exemple G. Buchholtz, "*Artificial Intelligence and Legal Tech : Challenges to the Rule of Law*" dans T. Wischmeyer, T. Rademacher (eds.), Regulating Artificial Intelligence, Springer (2020).
53. Voir la Charte éthique européenne (de la CEPEJ) sur l'utilisation de l'IA dans les systèmes judiciaires et leur environnement, qui fait spécifiquement référence aux risques découlant des systèmes d'anticipation des décisions judiciaires en matière civile, administrative et commerciale, des systèmes d'évaluation des risques en matière pénale, et de l'utilisation de systèmes d'IA sans garanties appropriées dans le cadre d'un règlement extrajudiciaire des litiges. Parmi ces risques, la CEPEJ note les risques d'« effet performatif » et de délégation de responsabilité, ainsi que le manque de transparence des décisions judiciaires.
54. Ainsi que le droit à un recours effectif consacré par l'article 13 de la CEDH.

l'égalité des armes dans les procédures judiciaires sont des éléments clés qui sont également essentiels à la mise en œuvre effective de l'État de droit.

41. En outre, les entreprises sont confrontées à une pression accrue, y compris par le biais de la réglementation, pour prendre des décisions sur la légalité du contenu qui est présenté sur leur plate-forme. Depuis que les plateformes de médias sociaux sont devenues la nouvelle « place publique », leurs propres conditions de service fixent essentiellement les règles de la manière dont la liberté d'expression se manifeste en ligne, mais avec moins de garanties que dans les contextes publics plus traditionnels. Il est toutefois essentiel que les États puissent continuer à assumer leur responsabilité en matière de protection de l'État de droit et qu'ils le fassent.

3.4. Une approche contextuelle et fondée sur le risque pour régir l'IA

42. Il ressort des explications qui précèdent que certaines applications des systèmes d'IA présentent une série de risques pour les droits de l'homme, la démocratie et l'État de droit. Ces risques dépendent toutefois du contexte d'application, de la technologie et des acteurs concernés. Afin de contrer tout étouffement de l'innovation en matière d'IA socialement bénéfique et de garantir que les avantages de cette technologie puissent être pleinement exploités tout en s'attaquant de manière adéquate à ses risques, le CAHAI recommande qu'un futur cadre juridique du Conseil de l'Europe sur l'IA poursuive une approche basée sur les risques ciblant le contexte d'application spécifique.[55] Cela signifie non seulement que les risques posés par les systèmes d'IA doivent être évalués et revus de manière systématique et régulière, mais aussi que toute mesure d'atténuation, qui est élaborée plus en détail au chapitre 7, doit être spécifiquement adaptée à ces risques. Outre l'approche fondée sur les risques, il convient d'envisager, le cas échéant, une approche de précaution[56], y compris des interdictions potentielles.[57] Cela peut, par exemple, être le cas lorsqu'un certain système d'IA dans un contexte spécifique présente un niveau de risque significatif associé à un niveau élevé d'incertitude quant à la réversibilité du dommage. Une telle approche peut contribuer à garantir que les risques spécifiques posés par le développement et l'utilisation de l'IA soient traités, tout en garantissant que les bénéfices générés par cette technologie innovante puissent être récoltés et ainsi améliorer le bien-être des individus et de la société.

43. Les applications d'IA qui promeuvent, renforcent et augmentent la protection des droits de l'homme, de la démocratie et de l'État de droit, devrait être encouragées. Toutefois, lorsqu'il ressort d'une évaluation des risques spécifique au contexte qu'une application d'IA peut présenter des risques « importants » ou inconnus pour les droits de l'homme, la démocratie ou l'État de droit, et qu'aucune mesure d'atténuation appropriée n'existe dans les cadres juridiques existants pour atténuer ces risques de manière adéquate, les États doivent envisager l'introduction de mesures réglementaires supplémentaires ou d'autres restrictions pour l'utilisation exceptionnelle et contrôlée de l'application et, le cas échéant, une interdiction ou un moratoire (lignes rouges).[58] La mise en place d'un accord international sur les utilisations problématiques de l'IA et les lignes rouges peut être essentielle pour anticiper les objections concernant les désavantages concurrentiels et pour créer des conditions de concurrence claires et équitables pour les développeurs et les diffuseurs d'IA.[59] Les systèmes de reconnaissance biométrique à distance - ou d'autres applications de suivi basées sur l'IA - qui risquent de conduire à une surveillance de masse ou à une notation sociale, ou la manipulation secrète d'individus basée sur l'IA, sont des exemples d'applications susceptibles de tomber sous le coup des lignes rouges, chacune ayant un impact significatif sur l'autonomie des individus ainsi que sur les principes et libertés démocratiques fondamentaux. L'utilisation exceptionnelle de ces technologies doit être spécifiquement

55. Cela serait également conforme à l'approche adoptée par l'Union européenne dans son Livre blanc sur l'IA de février 2020, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_fr.pdf

56. L'approche de précaution est généralement utilisée dans le contexte d'innovations ayant un potentiel (important) de dommages lorsque des connaissances scientifiques étendues sur la question font (encore) défaut. Pour plus d'informations, voir par exemple une communication de la Commission européenne sur le principe de précaution, Bruxelles, 2.2.2000, COM(2000) 1 final, https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:52000DC0001&from=EN

57. À cet égard, on peut se référer au chapitre 7.2, qui indique comment les États membres peuvent mettre en œuvre une approche de la gouvernance de l'IA fondée sur les risques. Le CAHAI peut examiner cette question plus en détail dans le cadre de ses travaux futurs.

58. L'une des intentions de la construction d'un accord international sur les lignes rouges est de prévenir les désavantages concurrentiels. Les lignes rouges sous forme de moratoires pourraient dans certains cas être surmontées lorsque des dispositions peuvent être prises pour garantir des méthodes appropriées pour développer une IA digne de confiance (légale, éthique et solide), par exemple lorsque l'évaluation préalable, le contrôle continu, les procédures de certification ou les processus de développement standardisés peuvent assurer des garanties appropriées pour sauvegarder les droits de l'homme, la démocratie et l'État de droit.

59. Voir par exemple Pekka Ala-Pietilä et Nathalie Smuha, «*A Framework for Global Cooperation on Artificial Intelligence and its* Governance (September 2020)», https://ssrn.com/abstract=3696519

prévue par la loi, nécessaire dans une société démocratique et proportionnée à l'objectif légitime, et autorisée uniquement dans des environnements contrôlés et (le cas échéant) pour des périodes limitées. D'autre part, lorsqu'une certaine application d'un système d'IA ne présente aucun risque pour les droits de l'homme, la démocratie ou l'État de droit, elle doit être exemptée de toute mesure réglementaire supplémentaire.[60] Lors de l'évaluation du risque posé par un système d'IA, il convient de se demander si l'utilisation d'un système d'IA peut entraîner un risque plus élevé que la non-utilisation de l'IA.

44. Une évaluation contextuelle et périodique des risques résultant du développement et de l'utilisation de l'IA est nécessaire, compte tenu de la nature spécifique du contexte des avantages et des risques liés à l'application de l'IA. En tant que technologie transversale, la même technologie d'IA peut être utilisée à différentes fins et dans différents contextes, et les conséquences positives ou négatives de la technologie en dépendront fortement.

4. LES TRAVAUX DU CONSEIL DE L'EUROPE DANS LE DOMAINE DE L'INTELLIGENCE ARTIFICIELLE À CE JOUR

45. L'impact significatif des technologies de l'information sur les droits de l'homme, la démocratie et l'État de droit a conduit le Conseil de l'Europe à élaborer des mécanismes pertinents, contraignants et non contraignants, qui se complètent et se renforcent mutuellement. Nous allons les examiner ci-après, de même que la jurisprudence de la Cour européenne des droits de l'homme sur les nouvelles technologies.

4.1. Travaux dans le domaine de la protection des données à caractère personnel

46. La « Convention 108 »[61], modernisée en 2018 par un protocole d'amendement[62] (« Convention 108+ »), fixe des normes internationales qui garantissent aux individus le droit au respect de la vie privée et à la protection des données à caractère personnel, indépendamment des évolutions technologiques. Elle exige en particulier que le traitement de données sensibles[63] ne soit autorisé qu'à la condition que des garanties appropriées, venant compléter celles de la Convention, soient prévues par la loi. Elle confère aux personnes concernées le droit d'avoir connaissance du traitement de leurs données à caractère personnel et de ses finalités, ainsi qu'un droit de rectification lorsque des données sont traitées contrairement aux dispositions de la convention. Le protocole d'amendement a ajouté de nouveaux principes tels que la transparence (art 8), la proportionnalité (art 5), l'obligation de rendre des comptes (accountability) (art 10), les analyses d'impact (art 10) et le respect de la vie privé dès la conception (art 10). En ce qui concerne les droits des personnes, ont notamment été introduits le droit de toute personne de ne pas être soumise à une décision l'affectant de manière significative, qui serait prise uniquement sur le fondement d'un traitement automatisé de données, sans que son point de vue soit pris en compte[64], et le droit d'obtenir connaissance du raisonnement qui sous-tend le traitement de données, lorsque les résultats de ce traitement lui sont appliqués (art. 9). Ces nouveaux droits sont particulièrement importants dans le cadre du profilage[65].

47. Même s'il n'est pas spécifique aux applications de l'intelligence artificielle, le cadre juridique construit autour de la Convention 108 reste pleinement applicable à cette technologie, dès lors que les données traitées relèvent du champ de ce texte. Des lignes directrices et un rapport publiés en 2019 ont précisé les principes directeurs à appliquer tant par le législateur et les décideurs que par les développeurs, les fabricants et

60. En effet, il est également possible que le développement et l'utilisation d'un système d'IA particulier dans un certain contexte n'ait pas nécessairement d'impact - positif ou négatif - sur les droits de l'homme, la démocratie ou l'État de droit.
61. Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, (« Convention 108 », STE n° 108).
62. Protocole d'amendement à la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, STCE n° 223.
63. Pour une liste complète des données sensibles, se reporter à l'article 6 de la Convention 108+.
64. La Convention précise qu'une personne ne peut exercer ce droit si la décision automatisée est autorisée par une loi à laquelle le responsable du traitement est soumis et qui prévoit également des mesures appropriées pour sauvegarder les droits et libertés et les intérêts légitimes de la personne concernée.
65. Voir à cet égard la Recommandation CM(Rec(2010)13 sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage et son exposé des motifs. Le comité consultatif de la Convention 108 poursuit son travail de mise à jour de cette importante recommandation.

les prestataires de services[66]. Un instrument juridique sur les applications de l'IA devra par conséquent tenir pleinement compte de ces acquis et les compléter (c'est-à-dire en se concentrant sur les lacunes de protection qui subsistent), par exemple en incluant dans son champ les opérations de traitement ne portant pas uniquement sur des données à caractère personnel et en élargissant son champ d'application à la prévention des atteintes aux autres droits de l'homme et en incluant les atteintes à la société (et non pas seulement pour l'individu).

4.2. Travaux dans le domaine de la cybercriminalité

48. Divers systèmes d'IA sont susceptibles de créer des risques majeurs dans le domaine de la cybercriminalité et sont déjà très utilisés pour commettre ce type d'infractions, qui vont des attaques informatiques par « déni de service distribué » (*Distributed Denial of Service* ou DDoS), orchestrées à grande échelle, au fait de scanner les vulnérabilités des systèmes pour en exploiter les failles, en passant par l'ingénierie sociale, l'usurpation d'identité ou la cybercriminalité autonome, commise par des machines. La Convention sur la cybercriminalité (Convention de Budapest) constitue un instrument important en ce qui concerne l'incrimination des infractions commises contre des systèmes informatiques et au moyen de systèmes informatiques, la mise en œuvre des pouvoirs et procédures prévus aux fins d'enquêtes dans des affaires de cybercriminalité, appliqués à la collecte des preuves électroniques de toute infraction pénale, en veillant à ce que ce pouvoir soit encadré par toutes les sauvegardes de l'État de droit, et l'instauration d'une coopération internationale effective[67]. Un nouveau protocole à la convention, relatif à une coopération renforcée sur la cybercriminalité et les preuves électroniques, est en préparation et pourrait être disponible en 2021[68]. La Convention de Budapest et ses dispositions sont pleinement applicables aux actes commis ou facilités par des systèmes d'IA. En outre, les traités généraux du Conseil de l'Europe dans les domaines de la lutte contre la criminalité et le terrorisme[69] peuvent également être applicables aux infractions commises à l'aide de technologies d'IA.

4.3. Travaux dans le domaine des systèmes algorithmiques

49. Le Comité des Ministres a adopté une Déclaration sur les capacités de manipulation des processus algorithmiques[70] en février 2019 et une Recommandation sur les impacts des systèmes algorithmiques sur les droits de l'homme[71] en avril 2020. Des études et rapports sur les dimensions des droits humains dans les techniques de traitement automatisé des données[72] et sur le thème « Responsabilité et IA »[73] ont par ailleurs été élaborés par des comités d'experts et autres organes spécialisés. Enfin, l'élaboration d'un instrument normatif (ou d'orientations adressées par le Comité des Ministres aux États membres sous une autre forme) sur les impacts des technologies numériques sur la liberté d'expression[74] est en cours.

---

66. Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (T-PD), Lignes directrices sur l'intelligence artificielle et la protection des données, janvier 2019, et T-PD, Rapport sur l'intelligence artificielle. Intelligence artificielle et protection des données : enjeux et solutions possibles, janvier 2019. Lignes directrices sur la protection des personnes à l'égard du traitement des données à caractère personnel à l'ère des mégadonnées, Janvier 2017
67. Convention sur la cybercriminalité, STE n°185. Voir aussi une note d'orientation adoptée en décembre 2014.
68. Voir également le Protocole additionnel à la Convention sur la cybercriminalité, relatif à l'incrimination d'actes de nature raciste et xénophobe commis par le biais de systèmes informatiques, STE n° 189.
69. Comme, par exemple, la Convention du Conseil de l'Europe sur la prévention du terrorisme, la Convention européenne pour la répression du terrorisme et la Convention européenne d'entraide judiciaire en matière pénale, avec leurs protocoles pertinents.
70. Comité des Ministres, Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques, Decl(13/02/2019)1, 13 février 2019. La déclaration attire l'attention des États membres « sur la nécessité de bien évaluer le besoin d'adopter des mesures réglementaires ou autres plus strictes afin de garantir une surveillance appropriée et démocratiquement légitime du dessin, du développement, du déploiement et de l'utilisation des outils algorithmiques, en vue de mettre en œuvre une protection efficace contre les pratiques déloyales et les abus de position monopolistique ».
71. Comité des Ministres, Recommandation CM/Rec(2020)1 du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme, 8 avril 2020. La recommandation, pour sa part, invite les États membres à « revoir leurs cadres législatifs et leurs politiques, ainsi que leurs propres pratiques en matière d'acquisition, de conception, de développement et de déploiement en cours de systèmes algorithmiques pour s'assurer qu'ils sont en accord avec les lignes directrices énoncées dans l'annexe à la présente recommandation ».
72. Voir l'étude produite par le Comité d'experts du Conseil de l'Europe sur les intermédiaires internet (MSI-NET) sous l'autorité du Comité directeur sur les médias et la société de l'information (CDMSI) – Étude sur les dimensions des droits humains dans les techniques de traitement automatisé des données et éventuelles implications réglementaires, DGI(2017)12.
73. MSI-AUT, Responsabilité et IA – Étude sur les incidences des technologies numériques avancées (dont l'intelligence artificielle) sur la notion de responsabilité, sous l'angle des droits humains, DGI(2019)05, septembre 2019.
74. Voir les travaux en cours du Comité d'experts sur la liberté d'expression et les technologies numériques (MSI-DIG).

4.4. Travaux dans le domaine de la justice

50. La Commission européenne pour l'efficacité de la justice (CEPEJ) a adopté en décembre 2018 la Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires[75], qui énonce cinq principes fondamentaux (respect des droits fondamentaux, non-discrimination, qualité et sécurité, transparence, neutralité et loyauté, maîtrise par l'utilisateur) pour l'utilisation des systèmes d'IA dans ce domaine. La CEPEJ étudie actuellement l'opportunité et la faisabilité d'un cadre de certification ou labellisation pour les produits d'intelligence artificielle utilisés dans les systèmes judiciaires. Le Comité européen de coopération juridique (CDCJ) travaille à l'élaboration de lignes directrices visant à assurer la compatibilité de ces mécanismes avec les articles 6 et 13 de la Convention européenne des droits de l'homme. Le Comité européen pour les problèmes criminels (CDPC) a entamé une réflexion sur le thème « IA et droit pénal » et pourrait proposer la création d'un nouvel instrument juridique spécialisé[76].

4.5. Travaux dans le domaine de la bonne gouvernance et des élections

51. Le Comité européen sur la démocratie et la gouvernance (CDDG) prépare une étude sur les répercussions de la transformation numérique – notamment de l'IA – sur la démocratie et la gouvernance. L'étude examine l'impact de l'IA sur les élections, la participation des citoyens et le contrôle démocratique. Dans le chapitre consacré à la gouvernance, elle dresse un état des lieux de l'utilisation de l'IA dans l'administration publique en Europe et analyse l'usage qui en est fait au regard des 12 principes de bonne gouvernance démocratique.

52. La Commission de Venise a aussi publié un rapport sur les technologies numériques et les élections[77] et élabore actuellement des principes pour une utilisation des technologies numériques respectueuse des droits de l'homme dans les processus électoraux.

4.6. Travaux dans le domaine de l'égalité des genres et de la non-discrimination

53. La Recommandation CM/Rec(2019)1 du Comité des Ministres sur la prévention et la lutte contre le sexisme recommande aux États membres d'intégrer une perspective d'égalité des genres dans toutes les politiques, programmes et recherches en rapport avec l'intelligence artificielle, afin d'éviter les risques potentiels de perpétuation du sexisme et des stéréotypes de genre, et d'examiner comment l'intelligence artificielle pourrait contribuer à éliminer les écarts entre les sexes et le sexisme.

54. Des travaux sont en cours dans le domaine de l'égalité et de la non-discrimination[78] pour donner suite à l'étude exhaustive commandée par l'ECRI intitulée « Discrimination, intelligence artificielle et décisions algorithmiques »[79].

55. La Commission européenne contre le racisme et l'intolérance (ECRI) suit les affaires de discrimination liées à l'intelligence artificielle et à la prise de décision algorithmique (ADM), qui relèvent de son mandat, et, le cas échéant, formule des recommandations pertinentes pour combler les lacunes législatives ou autres afin de prévenir la discrimination directe ou indirecte liée à l'intelligence artificielle et à la prise de décision algorithmique.

4.7. Travaux dans les domaines de l'éducation, de la jeunesse et de la culture

56. La Recommandation du Comité des Ministres visant à développer et à promouvoir l'éducation à la citoyenneté numérique[80] invite les États membres à adopter des mesures politiques et réglementaires en matière d'éducation à la citoyenneté numérique, à évaluer leur impact à intervalles réguliers, et à assurer ou faciliter la mise en place d'une formation initiale et continue appropriée sur l'éducation à la citoyenneté numérique à l'intention des enseignants et des autres professionnels de l'éducation, pour ne citer que

75. CEPEJ, Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, CEPEJ(2018)14, décembre 2018.
76. Voir CDPC(2020)3Rev, Étude de faisabilité quant à un futur instrument du Conseil de l'Europe sur l'intelligence artificielle et le droit pénal.
77. Commission de Venise et DG1, Rapport conjoint de la Commission de Venise et de la Direction de la société de l'information et de la lutte contre la criminalité, Direction générale Droits de l'homme et État de droit (DGI) sur les technologies numériques et les élections, CDL-AD(2019)016, 24 juin 2019.
78. Voir également la conférence du 25e anniversaire de l'ECRI et sa feuille de route sur l'égalité effective (septembre 2019).
79. Voir l'étude commandé par l'ECRI sur : « Discrimination, intelligence artificielle et prise de décision algorithmique » (2018), rédigé par l'expert indépendant Frederik Zuiderveen Borgesius. Voir à cet égard également le §81 de cette étude.
80. Recommandation CM/Rec(2019)10 du Comité des Ministres aux États membres visant à développer et à promouvoir l'éducation à la citoyenneté numérique.

quelques-unes des mesures recommandées. Sur cette base, le Comité des Ministres a chargé le Comité directeur pour les politiques et pratiques éducatives (CDPPE) d'étudier les incidences de l'intelligence artificielle et d'autres nouvelles technologies sur l'éducation en général et sur leur utilisation dans ce secteur en particulier. Une recommandation du Comité des Ministres traite spécifiquement des droits de l'enfant dans l'environnement numérique[81]. En outre, le Comité de la Convention 108 a également adopté des lignes directrices sur la protection des données personnelles des enfants dans un cadre éducatif[82].

57. Différentes activités ont été organisées depuis octobre 2018 concernant l'IA et l'art, la créativité et le patrimoine culturel, qui ont démontré l'impact croissant des systèmes d'IA sur ces trois domaines et ont souligné la nécessité d'une implication directe des professionnels de la création et de la culture dans les développements des systèmes d'IA et les politiques qui s'y rapportent. De plus, Eurimages a publié une étude sur l'impact des technologies prédictives et de l'IA sur le secteur audiovisuel, y compris les éventuelles mesures spécifiques à mettre en place pour garantir la liberté d'expression et la diversité culturelle[83].

4.8. Les travaux de l'Assemblée parlementaire du Conseil de l'Europe

58. L'Assemblée parlementaire du Conseil de l'Europe (APCE) a adopté, le 28 avril 2017, une recommandation intitulée « La convergence technologique, l'intelligence artificielle et les droits de l'homme »[84]. Le 22 octobre 2020, l'APCE a adopté sept rapports portant sur la nécessité d'une gouvernance démocratique de l'intelligence artificielle, le rôle de l'intelligence artificielle dans les systèmes de police et de justice pénale, la discrimination causée par l'IA, les menaces pour les libertés fondamentales, les défis médicaux, juridiques et éthiques dans le domaine de la santé, les conséquences sur les marchés du travail, et les aspects juridiques concernant les « véhicules autonomes ». Ils s'accompagnaient de recommandations adressées au Comité des Ministres et de résolutions.

59. Dans le cadre de cette étude de faisabilité, le rapport de l'APCE sur la nécessité d'une gouvernance démocratique de l'intelligence artificielle est d'une grande importance. Ce rapport propose notamment que le Comité des Ministres soutienne la rédaction d'un instrument juridiquement contraignant régissant les applications de l'IA, éventuellement sous la forme d'une convention du Conseil de l'Europe[85].

4.9. Les travaux du Congrès des pouvoirs locaux et régionaux du Conseil de l'Europe

60. Le Congrès des pouvoirs locaux et régionaux du Conseil de l'Europe a, ces dernières années, travaillé de diverses manières sur des questions liées à l'intelligence artificielle. Récemment, les membres de la commission de la gouvernance ont eu un échange de vues sur un rapport en cours de préparation : « Les villes intelligentes : les défis pour la démocratie », qui sera publié au cours du second semestre 2021.

4.10. Les travaux de la Commissaire aux droits de l'homme

61. La Commissaire aux droits de l'homme a publié, le 14 mai 2019, une recommandation, « Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme »[86]. Elle propose aux autorités nationales une série de recommandations pratiques pour 10 grands domaines d'action, déclinés comme suit : Évaluation de l'impact sur les droits de l'homme ; Consultations publiques ; Mise en œuvre des normes des droits de l'homme dans le secteur privé ; Information et transparence ; Contrôle indépendant ; Non-discrimination et égalité ; Protection des données et respect de la vie privée ; Liberté d'expression, liberté de réunion et d'association, et droit au travail ; Possibilités de recours ; Promotion de la connaissance et de la compréhension de l'intelligence artificielle.

---

81. Recommandation CM/Rec(2018)7 du Comité des Ministres sur les lignes directrices relatives au respect, à la protection et à la réalisation des droits de l'enfant dans l'environnement numérique.
82. Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, Convention 108, Lignes directrices : Protection des données personnelles des enfants dans un cadre éducatif, novembre 2020.
83. Eurimages, Étude sur l'impact des technologies prédictives et de l'IA sur le secteur audiovisuel, y compris les éventuelles mesures spécifiques à mettre en place pour garantir la liberté d'expression et la diversité culturelle, décembre 2019 [en anglais].
84. Recommandation 2102 (2017).
85. APCE, Commission des questions politiques et de la démocratie, rapport sur la nécessité d'une gouvernance démocratique de l'intelligence artificielle, Doc. 15150, 24 septembre 2020.
86. Commissaire aux droits de l'homme, Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme, mai 2019.

4.11. Les travaux du Conseil de l'Europe dans le domaine de la jeunesse

62. La Stratégie du Conseil de l'Europe pour la jeunesse 2030 fait référence à l'IA dans le cadre de la priorité stratégique « accès des jeunes aux droits » avec un accent particulier sur « l'amélioration des réponses institutionnelles aux questions émergentes affectant les jeunes, les droits des personnes et leur passage à l'âge adulte, tels que, [...] l'intelligence artificielle, l'espace numérique [...] ». Sur cette base et sur la base de la recommandation CM/Rec(2016)7 sur l'accès des jeunes aux droits, le département de la jeunesse continuera à promouvoir et à soutenir une approche coordonnée pour améliorer l'accès des jeunes aux droits dans tous les domaines politiques pertinents, y compris la gouvernance de l'IA, et continuera à promouvoir l'alphabétisation en matière d'IA et à doter les jeunes des aptitudes, compétences et connaissances nécessaires pour participer à la gouvernance de l'IA et bénéficier des technologies en développement.

4.12. La jurisprudence de la Cour européenne des droits de l'homme relative aux technologies de l'information

63. Il n'existe pour l'heure aucune jurisprudence de la Cour européenne des droits de l'homme (la Cour) concernant les systèmes d'IA. Dès lors, le CAHAI ne pouvait pas s'appuyer sur des décisions de la Cour directement relatives à la technologie de l'intelligence artificielle. Pour le moment, il n'y a pas non plus d'affaires pendante devant la Cour.

64. La jurisprudence existante en lien avec ce sujet concerne les algorithmes en général et des violations des articles 8 (vie privée) ou 10 (liberté d'expression) de la Convention et, de manière plus résiduelle, de l'article 14 (non-discrimination) dans des affaires portant, par exemple, sur la surveillance de masse[87], la responsabilité éditoriale des plates-formes[88] ou les interférences électorales[89].

65. Dans l'affaire *Sigurður Einarsson et autres c. Islande*[90], le parquet avait eu recours à des techniques statistiques de traitement des données pour examiner une masse considérable de documents et établir des preuves dans une affaire économique et financière. La question soulevée en l'espèce concernait la possibilité, pour la défense, d'avoir accès aux données d'où avaient été extraites les éléments de preuve à charge.

66. D'autres décisions de la Cour se sont intéressées aux conséquences de mécanismes algorithmiques utilisés pour prévenir la commission d'infractions. La Cour avait jugé en 2006 dans l'arrêt *Weber et Saravia c. Allemagne*[91] que les abus éventuels des pouvoirs de surveillance de l'État étaient encadrés par des garanties adéquates et effectives et qu'en tout état de cause, l'Allemagne disposait en la matière d'une marge d'appréciation relativement large.

67. S'agissant de la surveillance de masse de la population au moyen d'algorithmes, qui pourrait potentiellement inclure des outils d'IA, il y a lieu de suivre deux affaires qui ont été renvoyées devant la Grande Chambre (date de la dernière audience : 10 juillet 2019) : *Centrum för rättvisa c. Suède*[92] et *Big Brother Watch et autres c. Royaume-Uni*[93].

5. CARTOGRAPHIE DES INSTRUMENTS APPLICABLES À L'INTELLIGENCE ARTIFICIELLE

5.1. Instruments juridiques internationaux applicables à l'intelligence artificielle

68. Les instruments généraux, internationaux et régionaux, de protection des droits de l'homme, notamment la CEDH, la déclaration universelle des droits de l'Homme et la Charte des droits fondamentaux de l'Union européenne, s'appliquent à tous les domaines de la vie, y compris en ligne et hors ligne et quelle que soit la technologie utilisée et donc également aux systèmes d'IA. Se pose cependant la question de savoir si ces instruments, appliqués séparément ou conjointement, suffisent à répondre aux défis que posent les

87. *Big Brother Watch et autres c. Royaume-Uni*, 13 septembre 2018 (arrêt de chambre) – affaire renvoyée devant la Grande Chambre en février 2019.
88. *Delfi AS c. Estonie*, 16 juin 2015 (Grande Chambre).
89. *Magyar Kétfarkú Kutya Párt c. Hongrie*, 23 janvier 2018 – affaire renvoyée devant la Grande Chambre en mai 2018.
90. *Sigurður Einarsson et autres c. Islande*, 4 juin 2019 (deuxième section).
91. *Weber et Saravia c. Allemagne*, 29 juin 2006, requête n° 54934/00.
92. *Centrum för rättvisa c. Suède*, 19 juin 2018, requête n° 35252/08, renvoyée devant la Grande Chambre en février 2019 – audience tenue le 10 juillet 2019.
93. *Big Brother Watch et autres c. Royaume-Uni,* 13 septembre 2018, requêtes n^{os} 58170/13, 62322/14 et 24960/15 – affaire renvoyée devant la Grande Chambre en février 2019 – audience tenue le 10 juillet 2019.

systèmes d'IA et à garantir, tout au long du cycle de vie de ces derniers, le respect des normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'État de droit. Il n'existe à l'heure actuelle aucun instrument juridique qui s'applique de manière spécifique et globale aux enjeux que soulèvent les systèmes d'IA – ou plus généralement la prise de décision automatisée – pour la démocratie, les droits de l'homme et l'État de droit. Un certain nombre d'instruments juridiques internationaux traitent en revanche pour partie de certains aspects indirectement liés aux systèmes d'IA.

69. À cet égard, le CAHAI a pris note, lors de sa 2e réunion plénière, de l'analyse des instruments internationaux juridiquement contraignants réalisée par un consultant indépendant[94]. Cette analyse se base sur un examen des instruments juridiques contraignants et non contraignants dans quatre domaines essentiels (protection des données, santé, démocratie et justice), et a été complétée par un aperçu des instruments du Conseil de l'Europe dans d'autres domaines. Elle a fait ressortir que plusieurs instruments juridiques internationaux protégeaient déjà les droits de l'homme en général[95], les droits de groupes spécifiques qui peuvent également présenter des vulnérabilités dans le contexte de l'IA[96] et certains droits fondamentaux auxquels l'IA est susceptible de porter atteinte. Ceux-ci englobent, par exemple, le droit à la non-discrimination[97] et le droit au respect de la vie privée et à la protection des données à caractère personnel[98], notamment dans le cadre de leur traitement automatisé.

70. Le Protocole modifiant la Convention originale pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (Convention 108+), déjà mentionné ci-dessus, revêt une importance particulière. Ce protocole a non seulement modernisé l'instrument historique de 1980, mais il a également permis une cohérence totale avec le règlement général de l'UE sur la protection des données[99]. Il a introduit, par exemple, des exigences de transparence et de responsabilité, ainsi que des droits de protection pour les personnes concernées qui sont soumises à des processus décisionnels automatisés. Ce protocole n'est pas encore entré en vigueur.[100]

71. Par ailleurs, outre les instruments d'application horizontale, il existe un certain nombre d'instruments juridiques internationaux qui portent sur des secteurs ou des domaines spécifiques qui peuvent être indirectement liés à l'IA ou aux processus de prise de décision automatisés. Ces instruments couvrent des domaines aussi divers que la cybercriminalité[101], la (bio)médecine[102] et l'aviation[103]. Enfin, certains instruments juridiques concernent les droits procéduraux – comme la transparence[104] et l'accès à la justice[105] – qui peuvent s'avérer utiles pour contrôler et assurer la protection des droits matériels ou pour traiter d'aspects liés à la responsabilité applicable à certains préjudices[106].

72. Reconnaissant la pertinence de ces différents instruments juridiques pourraient présenter dans le contexte de la réglementation de l'IA, le CAHAI partage néanmoins les conclusions de l'analyse selon lesquelles ces instruments n'offrent pas toujours les garanties adéquates pour répondre aux défis posés par les systèmes d'IA. Ce point fera l'objet d'une analyse plus approfondie dans la sous-section 5.4 ci-dessous.

---

94. Voir CAHAI (2020)08-fin, Analyse des instruments internationaux juridiquement contraignants, rapport élaboré par Alessandro Mantelero, Université de Turin.
95. Citons la Convention européenne des droits de l'homme (STE n°5) et ses Protocoles ; la Charte sociale européenne (STE n° 163) ; la Déclaration universelle des droits de l'Homme ; et la Charte des droits fondamentaux de l'Union européenne.
96. Voir par exemple la Convention relative aux droits de l'enfant et la Convention relative aux droits des personnes handicapées. Voir également la Charte européenne des langues régionales ou minoritaires (STE n° 148) qui peut indirectement favoriser la prise en compte des langues minoritaires lors de la conception d'applications d'IA.
97. Voir par exemple la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes et la Convention sur la cybercriminalité et son Protocole additionnel.
98. Voir par exemple la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (STE n°108), le Règlement général de l'Union européenne sur la protection des données (2016/679) et la Directive « Police » de l'UE (2016/680).
99. Le règlement général sur la protection des données (UE) 2016/679 (RGPD).
100. Le protocole n'entrera en vigueur que lorsqu'il aura été ratifié, accepté ou approuvé par toutes les parties au traité STE 108, ou le 11 octobre 2023 si le protocole compte 38 parties à cette date.
101. Voir par exemple la Convention sur la cybercriminalité (STE n° 185). Concernant l'UE, voir par exemple le Règlement sur la cybersécurité (2019/881) et la Directive 2016/1148 sur la sécurité des réseaux et des systèmes d'information (Directive NIS).
102. Voir par exemple la Convention sur les droits de l'homme et la biomédecine (STE n° 164). Voir également le Règlement de l'UE relatif aux dispositifs médicaux (2017/745) et le Règlement relatif aux dispositifs médicaux de diagnostic in vitro (2017/746).
103. Voir par exemple la Convention de Chicago relative à l'aviation civile internationale.
104. Voir la Convention du Conseil de l'Europe sur l'accès aux documents publics (STE n° 205).
105. Voir par exemple la Convention européenne sur l'exercice des droits des enfants (STE n° 160) et la Convention européenne d'entraide judiciaire en matière pénale (STE n°30).
106. Voir par exemple la Convention européenne sur la responsabilité du fait des produits en cas de lésions corporelles ou de décès (STE n° 91) et la Directive de l'Union européenne sur la responsabilité du fait des produits défectueux et la Directive « Machines ».

73. Le besoin croissant d'un cadre de gouvernance plus complet et plus efficace pour répondre aux nouveaux défis et opportunités soulevés par l'IA a été reconnu par un certain nombre d'acteurs intergouvernementaux au niveau international. À ce jour, la plupart des initiatives en la matière se limitent à des recommandations non contraignantes[107]. Il convient de mentionner que la Commission européenne a annoncé l'élaboration d'une proposition législative visant à relever les défis en matière de droits fondamentaux liés à la garantie d'une IA digne de confiance, dont la publication est prévue au premier trimestre 2021[108].

5.2. Lignes directrices éthiques applicables à l'intelligence artificielle

74. Ces dernières années, plusieurs entreprises du secteur privé, établissements universitaires et organisations du secteur public ont publié des principes, des lignes directrices ainsi que d'autres instruments de droit souple pour une utilisation éthique de l'IA[109]. À cet égard, le CAHAI a pris note, à sa 2e réunion plénière, du travail de cartographie réalisé par deux consultants indépendants[110] qui ont examiné 116 documents, principalement élaborés en Europe, en Amérique du Nord et en Asie, portant sur une « IA éthique ». Cette étude a montré que les lignes directrices actuelles sur l'éthique en matière d'IA s'accordaient sur certains principes généraux, mais – dans la mesure où ils donnent des conseils pratiques – elles ont tendance à être en profond désaccord sur les détails de ce qui devrait être fait en pratique. S'agissant notamment de la transparence de l'IA, qui est le principe le plus fréquemment traité dans ces lignes directrices, il était difficile d'établir si la transparence devait passer par la publication du code source, en rendant les données de formation algorithmiques accessibles ou vérifiables (tout en tenant compte des lois applicables en matière de protection des données) ou par d'autres moyens. La résolution du défi que représente l'application de ces principes dans la pratique et l'examen des interdépendances et des compromis potentiels avec d'autres propriétés souhaitables ont donc été considérés comme une question importante à traiter par les décideurs politiques.

75. Selon la cartographie réalisée, comparativement aux pays du reste du monde, les documents de droit souple établis dans les États membres du Conseil de l'Europe semblent mettre davantage l'accent sur les principes éthiques de solidarité, de confiance et de fiabilité et faire plus rarement référence aux principes de bienfaisance et de dignité. Les principes de respect de la vie privée, la justice et la loyauté sont les principes pour lesquels les différences sont les moins marquées entre les pays membres et observateurs du Conseil de l'Europe et le reste du monde et donc ceux qui affichent le plus haut degré de constance transgéographique et transculturelle.

76. S'agissant des répercussions sur l'élaboration des politiques, l'étude a montré que les lignes directrices en matière d'éthique s'avèrent être de précieux outils permettant d'exercer une influence sur les décisions publiques relatives à l'IA et d'orienter le développement des systèmes d'IA en faveur du bien-être social. Elle a toutefois souligné que les initiatives de droit souple ne sauraient remplacer les instruments de gouvernance contraignants. Dans certains cas, en raison du fait que les intérêts de ceux qui développent et commercialisent la technologie et de ceux qui pourraient en subir les conséquences négatives ne sont pas toujours pleinement alignés, il est particulièrement à craindre que les acteurs du secteur privé de l'IA n'adoptent des mesures d'autoréglementation pour contourner ou écarter les mesures de gouvernance contraignantes édictées par des autorités (inter)gouvernementales. Les instruments juridiques non contraignants et les initiatives d'autorégulation peuvent toutefois jouer un rôle important en complétant la gouvernance contraignante, en particulier lorsque les intérêts des différents acteurs sont mieux alignés et lorsqu'il n'existe aucun risque important d'effets négatifs sur les droits de l'homme, la démocratie et l'État de droit.[111]

---

107. Ainsi, l'OCDE a adopté une recommandation du Conseil sur l'intelligence artificielle contenant un certain nombre de principes éthiques (voir https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449), lesquels ont inspiré les principes sur l'IA axés sur le facteur humain adoptés par le G20 dans une déclaration ministérielle rendue publique en juin 2019 (voir https://www.mofa.go.jp/files/000486596.pdf). En outre, l'UNESCO travaille actuellement à l'élaboration d'une recommandation (non contraignante) sur une IA éthique (voir UNESCO, avant-projet de recommandation sur l'éthique de l'intelligence artificielle, septembre 2020, https://unesdoc.unesco.org/ark:/48223/pf0000373434.) Si le projet de recommandation de l'UNESCO dans sa version actuelle évoque l'impact de l'IA sur les droits de l'homme et l'État de droit, il ne s'attarde pas sur les défis qu'elle pose pour la démocratie.
108. La Commission européenne attire tout particulièrement l'attention sur les risques pour les droits fondamentaux, pour la sécurité et le bon fonctionnement du régime de responsabilité. Voir le Livre blanc sur l'intelligence artificielle, publié en février 2020, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_fr.pdf.
109. Parmi les initiatives récentes, citons « Les lignes directrices en matière d'éthique pour une IA digne de confiance » publiées en avril 2019 par le Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle, mis en place par la Commission européenne, et la « Liste d'évaluation pour une IA digne de confiance » (ALTAI), leur outil opérationnel publié en juillet 2020.
110. Voir CAHAI (2020)07-fin, Lignes directrices sur l'éthique en matière d'IA : situation en Europe et dans le monde, rapport élaboré par Marcello Ienca et Effy Vayena.
111. Une gouvernance contraignante efficace nécessite toutefois un instrument signé et ratifié par un nombre suffisant d'États afin de garantir des conditions de concurrence équitables au niveau transfrontalier, compte tenu notamment de la nature transfrontalière des produits et services d'IA.

77. Le CAHAI partage les conclusions générales de l'étude de cartographie et considère que l'on pourrait envisager d'inclure les principes communs qu'elle a permis de dégager des différentes lignes directrices éthiques dans les réflexions du CAHAI sur l'élaboration d'un cadre juridique sur l'IA. Le respect des droits de l'homme, qui ont été mentionnés dans seulement un peu plus de la moitié des documents de droit souple examinés, devraient être au cœur de tout futur instrument juridique sur l'IA fondé sur les normes du Conseil de l'Europe. En outre, l'étude de cartographie pourrait être utilisée comme base pratique pour mettre en œuvre des cadres éthiques dans les États membres de manière harmonisée.

5.3. Aperçu des instruments, politiques et stratégies nationaux relatifs à l'intelligence artificielle

78. L'analyse de la consultation électronique réalisée auprès des membres, observateurs et participants du CAHAI sur ce point[112] a fait ressortir que quatre États membres ont adopté des cadres juridiques spécifiques sur des systèmes d'IA précis concernant l'essai et l'utilisation de voitures et d'entreprises autonomes. Deux États membres sont en train d'élaborer des cadres juridiques sur l'utilisation de systèmes d'IA dans les domaines du recrutement et de la prise de décision automatisée par les autorités publiques.

79. L'adoption de chartes éthiques et de documents de droit souple au niveau national semble plus courante ; ces instruments portent sur des questions telles que la robotique, la reconnaissance faciale, l'utilisation d'une « IA éthique » dans les services publics et les processus électoraux et l'utilisation des données personnelles et non personnelles. Un État membre a lancé un programme de certification volontaire de l'IA et deux autres ont officiellement adopté des cadres éthiques non contraignants, internationaux ou européens, sur l'IA. Au total, 12 États membres et quatre États observateurs ont adopté un ou plusieurs des instruments mentionnés plus haut. Différents types d'institutions telles que des conseils nationaux, des comités, des institutions publiques spécialisées dans l'IA et des entités gouvernementales ont été responsables de leur développement.

80. Trente États membres et quatre États observateurs ont mis en place des stratégies et des politiques relatives aux systèmes d'IA. Élaborées sur la base de plans d'action pluriannuels, assorties dans certains cas de programmes de financement ambitieux, elles ont pour objectif d'accroître la confiance dans cette technologie et de promouvoir son utilisation, de renforcer les compétences nécessaires à sa conception et son développement, de soutenir la recherche et de stimuler le développement des entreprises. Les États ont très souvent associé des experts des secteurs public et privé ainsi que des universitaires à l'élaboration de ces projets. Les systèmes d'IA font dans la plupart des cas l'objet de stratégies ciblées ; mais il arrive aussi qu'ils relèvent de politiques sectorielles plus générales relatives à l'économie et aux technologies numériques. Le développement et l'utilisation de systèmes d'IA sont également pris en compte dans les stratégies sectorielles concernant l'agriculture, la justice électronique, les services publics, la santé, l'environnement, l'éducation, la sécurité et la défense, la mobilité ou les données.

81. Pour finir, la consultation a fait ressortir que sept stratégies nationales avaient souligné la nécessité de promouvoir le développement de l'IA dans le respect des règles d'éthique et des normes internationales en matière de droits de l'homme.

5.4. Avantages, inconvénients et limites des instruments internationaux et nationaux existants et des lignes directrices éthiques sur l'intelligence artificielle

82. Ce tour d'horizon des instruments, politiques et stratégies a montré qu'un certain nombre de dispositions d'application plus large s'appliquent déjà au développement et à l'utilisation de systèmes d'IA. Faute d'instruments internationaux juridiquement contraignants axés sur l'IA permettant d'apporter une réponse juridique spécifique, un travail considérable d'interprétation des cadres juridiques existants à la lumière des problématiques liées à l'IA et de formulation de règles non contraignantes pour contextualiser les principes énoncés dans les instruments existants a été fourni[113]. Cependant, les instruments juridiques existants ayant été adoptés avant l'avènement de l'IA, leur efficacité à répondre de manière adéquate et ciblée aux défis posés par les systèmes d'IA est souvent plus limitée puisqu'ils ne sont pas adaptés à leurs spécificités. Par exemple une étude intitulée « Discrimination, intelligence artificielle et décisions algorithmiques » commandée par l'ECRI a souligné que, bien que les instruments juridiques internationaux et nationaux existants dans le domaine de la non-discrimination s'appliquent à l'utilisation des systèmes d'IA et peuvent dans certains

112. Voir le document CAHAI (2020) 09 rev 2, sur la consultation électronique des membres, observateurs et participants du CAHAI, dans lequel figurent les réponses communiquées au 30 septembre 2020.

113. Voir par exemple le document T-PD(2019)01 Lignes directrices sur l'intelligence artificielle et la protection des données ; CEPEJ. 2019. Charte éthique européenne d'utilisation de l'intelligence artificielle (IA) dans les systèmes judiciaires et leur environnement.

cas déjà fournir un certain niveau de protection, ils présentent encore certaines limites[114]. L'analyse réalisée par une experte indépendante pour le compte du CAHAI sur l'impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit dressait le même constat concernant les autres droits[115], et l'étude du Conseil de l'Europe sur l'IA et la responsabilité a spécifiquement souligné les limites des dispositions existantes en matière de droits de l'homme pour assurer une protection complète.[116]

83. En outre, bien qu'ils se superposent et se renforcent mutuellement, le nombre et la diversité des instruments rendent difficile l'interprétation et l'application cohérente et globale au contexte de l'IA, ce qui entraîne des niveaux inégaux de protection. Si certains instruments de droit souple (par exemple les lignes directrices éthiques) contiennent des principes spécifiquement adaptés au développement et à l'utilisation de systèmes d'IA, ceux-ci restent cependant non contraignants et peuvent être limités en termes d'efficacité, en ce qui concerne le respect des droits de l'homme, de la démocratie et de l'État de droit, étant donné que leur mise en œuvre dépend entièrement de la bonne volonté des personnes concernées. En outre, les lignes directrices éthiques n'ont par ailleurs pas la même dimension universelle que les normes fondées sur les droits de l'homme et se caractérisent par une diversité d'approches théoriques[117], ce qui limite leur utilité. Le CAHAI observe donc que, si la réglementation de l'IA ne souffre d'aucun vide juridique, il existe néanmoins un certain nombre de lacunes juridiques de fond et de procédure[118].

84. Premièrement, les droits et obligations formulés dans les instruments juridiques existants ont tendance à être formulés de manière large ou générale, ce qui n'est pas problématique en soi, mais peut dans certains cas soulever des difficultés d'interprétation dans le contexte de l'IA. En outre, ils n'abordent pas explicitement certaines questions spécifiques à l'IA, ce qui entrave leur leur application effective aux défis que soulèvent les systèmes d'IA tout au long de leur cycle de vie.[119] Il a été indiqué qu'une traduction ou une concrétisation des droits de l'homme existants dans le contexte des systèmes de IA[120] par le biais de

---

114. L'étude examine les instruments existants dans le contexte de la non-discrimination et indique que nombre d'entre eux peuvent déjà offrir une certaine protection contre la discrimination fondée sur l'IA. Toutefois, elle souligne également leurs limites, car l'IA ouvre également la voie à de nouveaux types de différenciation injuste qui échappent aux lois actuelles, ce qui suggère la nécessité d'une réglementation (sectorielle) supplémentaire pour protéger l'équité et les droits de l'homme dans le contexte de l'IA. Ces limites concernent, par exemple, le fait que ces instruments ne s'appliquent pas si un système d'IA invente de nouvelles classes qui ne correspondent pas aux caractéristiques protégées par ces instruments (c'est-à-dire le sexe ou l'ethnicité), pour différencier les personnes. Une telle différenciation peut néanmoins être injuste (par exemple, une discrimination des prix fondée sur l'IA pourrait conduire à ce que certains groupes de la société paient toujours plus). Dans une autre étude écrite par le même expert, il est suggéré que les législateurs « envisagent d'introduire une clause générale de discrimination » qui peut servir de filet de sécurité pour ces écarts et que les personnes peuvent « invoquer directement devant les tribunaux nationaux, en ce qui concerne la discrimination exercée par les autorités publiques et les organismes privés » (plutôt que d'ajouter de nouveaux motifs ou de nouvelles exemptions aux lois antidiscriminatoires fermées existantes. Voir J. Gerards et F. Zuiderveen Borgesius, *"Protected grounds and the system of non-discrimination law in the context of algorithmic decision-making and artificial intelligence"*, novembre 2020, SSRN : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3723873. D'autres études soulignent également les limites de la législation actuelle en matière de non-discrimination, et en particulier le fait que, même si ces lois s'appliquent, la complexité et l'opacité du processus décisionnel des systèmes d'intelligence artificielle rendent pratiquement impossible pour les personnes injustement discriminées d'en avoir connaissance et de le contester devant les tribunaux. Il a donc été suggéré que la législation sur la non-discrimination devrait être renforcée par des mesures pouvant « inclure l'introduction et l'élargissement des mécanismes d'action collective dans les procès antidiscriminatoires, des « audits » d'Amnesty International ou même des compétences supplémentaires des agences de lutte contre la discrimination et/ou de protection des données » afin de combler ce manque de connaissances et de garantir l'application effective du droit à la non-discrimination. Voir A. Tischbirek «*Intelligence artificielle et discrimination» : Discriminating Against Discriminatory Systems*» dans T. Wischmeyer, T. Rademacher (eds.), Regulating Artificial Intelligence, Springer, 2020.
115. Voir CAHAI(2020)06-fin, rapport élaboré par Catelijne Muller.
116. Étude du Conseil de l'Europe DGI(2019)05, Rapporteur Karen Yeung, Responsabilité et IA, septembre 2019. La nouvelle étude de la FRA décrit également l'impact de l'IA sur plusieurs droits de l'homme et propose certaines mesures pour remédier aux limitations existantes («*Getting the Future Right - Artificial Intelligence and Fundamental Rights in the EU*», 14 décembre 2020, https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights).
117. Comme l'a relevé le rapport de l'experte indépendante (CAHAI (2020)08-fin) cité plus haut.
118. C'est également la conclusion à laquelle parvient la Commission européenne dans son étude d'impact initiale sur un éventuel règlement européen sur l'IA, en précisant que « Bien que les dommages potentiels ci-dessus ne soient pas en soi nouveaux ou nécessairement liés à l'IA uniquement, l'analyse préliminaire de la Commission dans le Livre blanc indique qu'un certain nombre de risques spécifiques et importants sont en jeu en ce qui concerne l'IA et ne sont pas couverts de manière adéquate par la législation existante », https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Requirements-for-Artificial-Intelligence. Il convient également de noter à cet égard le texte adopté par la Commission permanente, agissant au nom de l'Assemblée, le 22 octobre 2020 (voir Doc. 15150, rapport de la commission des affaires politiques et de la démocratie).
119. Cela a, par exemple, été souligné par le rapport susmentionné de la FRA sur l'IA et les droits de l'homme (note de bas de page 114), qui, à cet égard, a particulièrement noté que ceux qui développent et utilisent l'IA sont souvent incertains quant à l'applicabilité des lois existantes en ce qui concerne l'IA.
120. Comme le fait le Règlement général sur la protection des données de l'UE en ce qui concerne la protection des données personnelles.

dispositions plus spécifiques, pourrait contribuer à remédier à ce problème[121]. Il faudrait pour cela reformuler les droits de l'homme généraux en droits plus spécifiques et concrets dont pourraient se réclamer ceux qui sont exposés à des systèmes d'IA et à leurs décisions. Ainsi, *le droit à un procès équitable* pourrait être formulé de manière plus concrète comme le droit de contester les éléments de preuve fournis par un système d'IA et d'y avoir accès[122]. Un autre moyen d'y parvenir serait de définir des obligations spécifiques à respecter ou des exigences auxquelles devraient se soumettre ceux qui développent ou déploient des systèmes d'IA. Par exemple, le droit à la non-discrimination pourrait donner lieu à une obligation de diligence raisonnable pour analyser et atténuer, tout au long du cycle de vie des systèmes d'IA, le risque de partialité injuste. Faute d'une telle concrétisation des droits existants dans le contexte des applications d'IA, et des obligations claires pour les développeurs et les déployeurs de systèmes d'IA de garantir le respect de ces droits, les individus pourraient ne pas bénéficier de la protection complète et effective de ceux-ci[123]. Le CAHAI considère que les normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'État de droit pourraient fournir une base appropriée pour l'élaboration de dispositions plus spécifiques pour assurer une protection efficace contre les risques posés par l'application pratique de certains systèmes d'IA.

85. Il convient ensuite de relever qu'un certain nombre de principes essentiels nécessaires au respect des droits de l'homme, de la démocratie et de l'État de droit dans le contexte de l'IA ne sont actuellement pas protégés par la loi. Cela concerne par exemple la nécessité d'exercer un *contrôle et un suivi humains*[124] suffisant des applications de l'IA, de veiller à leur *robustesse technique* et de garantir leur *transparence*[125] et leur *explicabilité*[126], notamment lorsqu'elles ont des répercussions juridiques ou d'autres effets importants sur les individus. Plusieurs études ont mis en lumière l'absence de dispositions légales visant à protéger ces principes dans les instruments existants[127]. Il est important de noter que le respect de ces principes est aussi une condition préalable au respect des droits matériels, au vu de l'opacité des systèmes d'IA et des choix réalisés par l'humain dans leur conception et leur utilisation[128]. En l'absence de transparence ou d'explication d'une

121. Voir documents CAHAI(2020)06-fin et CAHAI (2020)08-fin, cités plus haut. Voir également Karen Yeung, Andrew Howes et Ganna Pogrebna (Université de Birmingham), 'AI Governance by Human Rights–Centered Design, Deliberation, and Oversight : An End to Ethics Washing', in The Oxford Handbook on Ethics of AI (eds. M. D. Dubber, F. Pasquale, and S. Das), 2020, DOI : 10.1093/oxfordhb/9780190067397.013.5 ; Nathalie A. Smuha (KU Leuven), 'Beyond a Human Rights-Based Approach to AI Governance : Promise, Pitfalls, Plea', in Philosophy and Technology, 2020, https://doi.org/10.1007/s13347-020-00403-w.
122. Une telle concrétisation pourrait non seulement clarifier pour les sujets juridiques les droits dont ils disposent dans le cadre des systèmes d'IA, mais garantirait également la prévisibilité des dimensions substantielles couvertes et l'égalité devant la loi lorsque le droit à un procès équitable est appliqué par les juges. Sans une telle concrétisation, le risque existe que tous les juges n'interprètent pas le droit au sens large comme impliquant ce droit plus concret, ce qui conduirait à une protection inégale des individus.
123. Par ailleurs, les personnes chargées d'appliquer et d'interpréter les droits existants (comme les juges) pourraient ne pas disposer d'orientations suffisantes sur la manière de procéder dans le contexte de l'IA ce qui, selon la juridiction, pourrait conduire à des normes de protection inégales/inadaptées.
124. Cela pourrait également inclure des dispositions visant à minimiser les risques qui pourraient résulter d'une altération ou d'une interférence non qualifiée avec les systèmes d'IA.
125. S'agissant de la transparence des processus automatisés de décision, il convient de noter que la protection limitée qu'offre la Convention 108+ en la matière ne s'applique qu'au traitement des données personnelles ; or les systèmes d'IA peuvent avoir des effets négatifs sur les individus et les sociétés même lorsqu'ils traitent des données non personnelles.
126. La Convention 108+ et le règlement général sur la protection des données de l'UE ne contiennent pas de droit explicite à une explication pour la personne concernée, et il est fortement contesté que ce droit puisse y être implicitement lu et dans quelle mesure. Voir par exemple Sandra Wachter et autres, "*Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*", 7 International Data Privacy Law 2, 76-99 (2017), doi :10.1093/idpl/ipx005. Voir aussi Andrew D. Selbst et Julia Powles, *Meaningful information and the right to explanation*, International Data Privacy Law 7(4), novembre 2017, p233-242, https://doi.org/10.1093/idpl/ipx022.
127. Voir documents CAHAI(2020)06-fin et CAHAI (2020)08-fin, cités plus haut. Voir aussi Automating Society Report 2020, AlgorithmWatch, https://automatingsociety.algorithmwatch.org/report2020/european-union/ ; Voir également le Livre blanc sur l'IA de la Commission européenne, 19 février 2020, COM(2020) 65 final, page 11 : « *Le principal résultat de ce retour d'information est que les régimes législatifs ou réglementaires existants tiennent déjà compte d'un certain nombre d'exigences, mais que, dans de nombreux secteurs économiques, celles qui concernent la transparence, la traçabilité et le contrôle humain ne sont pas spécifiquement couvertes par la législation en vigueur* ».
128. Ce n'est que lorsque la traçabilité de l'IA est assurée, par exemple grâce à la documentation ou à la journalisation des informations concernées, qu'un système peut faire l'objet d'un audit indépendant et qu'il est possible de vérifier dans quelle mesure il risque, par exemple, de porter atteinte au droit à la non-discrimination. En outre, si le processus de décision du système d'IA n'est pas expliqué, les individus ne pourront pas contester une décision et demander réparation. À cet égard, le Livre blanc de la Commission européenne sur l'IA fait plus généralement remarquer, page 14 , que « *Les particularités qui caractérisent de nombreuses technologies de l'IA, notamment l'opacité (« effet de boîte noire »), la complexité, l'imprévisibilité et le comportement partiellement autonome, peuvent rendre difficile la vérification de la conformité aux règles du droit de l'UE en vigueur destinées à protéger les droits fondamentaux et peuvent entraver le contrôle de l'application de celles-ci* ». Cela s'applique également aux dispositions relatives aux droits de l'homme qui figurent dans les autres instruments juridiques existants, car ils ne sont actuellement pas adaptés aux problèmes spécifiques posés par l'IA. Cependant, il existe plusieurs exemples de « modèles supplémentaires » et d'autres méthodes pour comprendre comment une décision a été prise. Les modèles complémentaires sont de plus en plus courants, tout comme l'utilisation de systèmes d'IA plus interprétables (voir https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-ai/).

décision importante prise par un système d'IA, il sera impossible de déterminer si un droit – comme le droit à la non-discrimination – a effectivement été respecté. De la même manière, il sera difficile pour l'utilisateur de contester la décision. L'asymétrie de l'information qui existe entre les personnes qui subissent les incidences négatives des systèmes d'IA et celles qui conçoivent et utilisent ces systèmes fait par conséquent ressortir la nécessité de renforcer les mécanismes assurant la responsabilité, l'obligation de rendre des comptes et les *recours* et de veiller à la *traçabilité* et à *l'auditabilité* des systèmes d'IA[129]. Si ces insuffisances ne sont pas corrigées, par exemple en garantissant le respect de ces autres principes au moyen de droits et d'obligations concrètes, les personnes qui en subissent les effets négatifs – ainsi que d'autres parties prenantes, parmi lesquelles les autorités de réglementation et les représentants de la loi – ne seront pas en mesure d'établir l'existence d'atteintes aux droits de l'homme ou d'autres manquements.

86. Par ailleurs, les instruments actuels ne tiennent suffisamment compte des mesures que les concepteurs et les diffuseurs de systèmes d'IA doivent prendre pour garantir *l'efficacité* de ces systèmes chaque fois qu'ils peuvent avoir un impact sur les droits de l'homme, la démocratie ou l'État de droit[130], et pour s'assurer que les concepteurs et les diffuseurs d'IA ont *les compétences ou les qualifications professionnelles* nécessaires pour le faire. En outre, la *dimension sociétale* des risques liés à l'IA qui dépasse l'impact sur les individus, comme l'impact sur le processus électoral et les institutions démocratiques ou le système juridique, n'est pas encore suffisamment prise en compte. Si un certain nombre de mécanismes nationaux et internationaux permettent aux individus d'introduire un recours devant un tribunal en cas d'atteinte aux droits de l'homme dans le contexte de l'IA, il n'existe en revanche aucun dispositif en cas d'entrave à la démocratie ou à l'État de droit. Il est important de souligner que cette lacune appelle un contrôle public de la conception, du développement et de l'utilisation des systèmes d'IA dès lors qu'existent ces risques, en établissant pour cela des obligations ou des exigences claires.[131]

87. Ces lacunes peuvent également créer une insécurité juridique pour les acteurs concernés, notamment les développeurs, les déployeurs et les utilisateurs de systèmes d'IA, qui ne disposent pas d'un cadre juridique prévisible et solide applicable à la conception et à la mise en œuvre des systèmes d'IA. Cette insécurité risque de freiner les innovations bénéfiques dans le domaine de l'IA et, partant, d'empêcher les citoyens et la société au sens large d'exploiter pleinement le potentiel dont l'IA est porteuse. La mise en place d'un cadre juridique complet applicable aux systèmes d'IA, reposant sur une approche fondée sur les risques, peut aider à fixer les limites dans lesquelles l'innovation peut être stimulée et favorisée et le potentiel de l'IA optimisé, tout en garantissant – et en renforçant – le respect des droits de l'homme, de la démocratie et de l'État de droit au moyen de voies de recours effectives.

88. Enfin, les diverses lacunes des instruments juridiques existants - ainsi que l'approche fragmentée de l'application de ces instruments au contexte de l'IA dans toute l'Europe - suscitent également des incertitudes quant à la manière de traiter la nature transfrontalière de l'impact généré par le développement et l'utilisation des applications de l'IA. L'absence de normes communes au niveau international pourrait entraver le commerce transfrontalier des produits et services d'IA, car l'absence de normes communes et de règles équitables peut faire obstacle à la confiance mutuelle, ce qui pourrait aussi empêcher que les bénéfices des applications d'IA ne dépassent les frontières nationales.[132]

---

129. À cet égard, on peut noter que les administrations publiques exigent des niveaux de responsabilité encore plus élevés que le secteur privé. En même temps, il faut reconnaître que la distinction entre la participation du secteur public et du secteur privé est parfois floue, car les entités du secteur public dépendent souvent fortement des acteurs privés pour le développement, l'acquisition et l'utilisation des systèmes et des ensembles de données d'IA.
130. Dans les domaines qui ont un impact matériel sur la vie humaine, comme la médecine, la société s'appuie sur des instruments solides pour garantir que les technologies et les agents humains impliqués sont efficaces à la fois pour atteindre l'objectif visé (par exemple : guérir les maladies) et pour éviter les effets secondaires négatifs (par exemple : faire courir un risque excessif aux patients). Lorsque les systèmes d'IA peuvent avoir un impact sur les droits de l'homme, la démocratie ou l'État de droit (par exemple : dans des environnements juridiques, judiciaires ou répressifs), des instruments similaires sont nécessaires. Les systèmes de reconnaissance faciale, par exemple, s'ils sont utilisés dans les services répressifs, devraient être généralement efficaces pour identifier avec précision les individus (l'objectif visé dans une action répressive donnée), et avoir une précision raisonnablement uniforme d'une ethnie à l'autre (pour faire respecter les droits de l'homme et l'État de droit).
131. Des progrès techniques récents ont été réalisés dans ce domaine. Il existe plusieurs exemples de « modèles supplémentaires » et d'autres méthodes permettant de comprendre comment une décision a été prise. Les modèles supplémentaires sont de plus en plus courants, tout comme l'utilisation de systèmes d'IA plus interprétables, voir https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-ai/.
132. Cela est particulièrement important dans le cas des petits pays qui sont extrêmement dépendants de la réglementation de leurs voisins. Pour les petits pays dont les capacités de développement de l'IA sont limitées, une réglementation transfrontalière ou un socle commun de principes de réglementation serait particulièrement utile.

89. L'on peut conclure, à la lumière de ce qui précède, qu'une approche réglementaire de l'IA devrait s'employer à remédier à ces insuffisances. Au-delà des cadres juridiques existants, cette approche pourrait prévoir des dispositions contraignantes permettant d'assurer le respect des droits de l'homme, de la démocratie et de l'État de droit dans le contexte de l'IA, afin de garantir un niveau de protection plus complet quel que soit le secteur concerné et être complétées par des règles spécifiques au secteur, le cas échéant..[133] Il conviendrait, pour y parvenir, de clarifier et d'élargir le champ d'application des droits existants et d'imposer à cette fin le respect d'autres principes et exigences. Outre cette approche contraignante, on peut également envisager l'élaboration de dispositions sectorielles spécifiques et de lignes directrices pour les problématiques qui sont uniquement ou particulièrement pertinentes dans un domaine ou une application donnée[134]. À cet égard, on peut se référer au chapitre 8 de la présente étude de faisabilité, qui présente diverses options pour un cadre juridique du Conseil de l'Europe pour la conception, le développement et l'application de l'IA.

90. Comme indiqué dans le rapport d'état d'avancement du CAHAI, les travaux qu'entreprend ce comité permettent d'enrichir et de compléter d'autres initiatives internationales dans ce domaine (par exemple celles de l'OCDE, de l'Union européenne, de l'UNESCO et des Nations Unies en général, organisations avec lesquelles une coordination et des synergies sont en permanence recherchées[135]) en adoptant un instrument concret basé sur les normes du Conseil de l'Europe relatives aux droits de l'homme, à la démocratie et à l'État de droit, dans le cadre d'un mécanisme juridique mondial de régulation des technologies numériques. À cet égard, le CAHAI a souligné que le Conseil de l'Europe pouvait apporter une réelle valeur ajoutée lors de l'élaboration d'un instrument juridique sur l'IA, notamment parce que, outre la protection des droits de l'homme, il peut également traiter des enjeux sociétaux et environnementaux de l'IA pour la démocratie et l'État de droit[136]. L'élaboration d'un instrument juridiquement contraignant reposant sur les normes du Conseil de l'Europe – si telle est la solution retenue par le Comité des Ministres – contribuerait à faire de l'initiative du Conseil de l'Europe une initiative unique qui se distinguerait des autres initiatives internationales, qui soit s'attachent à l'élaboration d'un type différent d'instrument, soit diffèrent en termes de portée et de contexte. Il est important de garder à l'esprit la nature spécifique des normes régionales, et de faire appel à tout le spectre des compétences du Conseil de l'Europe dans l'accomplissement de ce travail.

5.5. Instruments juridiques internationaux, lignes directrices en matière d'éthique et acteurs du secteur privé

91. Les instruments du Conseil de l'Europe s'adressent généralement aux États membres plutôt qu'aux acteurs du secteur privé. Les droits accordés aux États membres en vertu de ces instruments et les obligations qui en découlent peuvent néanmoins concerner indirectement les acteurs privés. Les États ont l'obligation[137] de s'assurer que les acteurs du secteur privé agissent dans le respect des droits de l'Homme en les mettant en œuvre et en les appliquant dans leur législation nationale et en veillant à ce que des voies de recours effectives, judiciaires ou extra-judiciaires, soient disponibles au niveau national. De leur côté, les acteurs du secteur privé, conformément aux Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme, sont tenus, du fait de la responsabilité des entreprises, de respecter les droits de l'homme de leurs clients et de toutes les parties prenantes concernées[138].

92. Un certain nombre d'instruments internationaux mettent directement l'accent sur la nécessité pour les entreprises de veiller au respect des droits de l'homme et d'œuvrer en faveur d'une recherche et d'une

133. À cet égard, on peut noter que de nombreux systèmes d'IA peuvent être réutilisés pour être utilisés dans d'autres secteurs. Il peut donc être souhaitable d'adopter une approche qui prévoit certaines garanties dans tous les secteurs, éventuellement associées à des garanties complémentaires ou à des lignes directrices plus spécifiques à chaque secteur si nécessaire.
134. Le CAHAI reconnaît à cet égard la spécificité contextuelle de certains risques. Ainsi, l'utilisation à grande échelle de l'identification biométrique à distance fondée sur l'IA n'a pas la même incidence sur les droits de l'homme que l'utilisation d'un système basé sur l'IA pour recommander une chanson.
135. Lors de sa deuxième réunion plénière, le CAHAI a été informé des derniers travaux en cours de la FRA, de l'Union européenne, de l'OCDE, du Groupe de haut niveau des Nations Unies sur la coopération numérique et de l'UNESCO. Voir le rapport de la deuxième réunion plénière du CAHAI, paragraphes 78-84.
136. Notons que si le Livre blanc de la Commission européenne sur l'IA s'attache aux incidences de l'IA sur les droits fondamentaux, il ne traite en revanche pas explicitement de ses incidences sur la démocratie et l'État de droit.
137. Principes directeurs de l'ONU sur les entreprises et les droits de l'homme 2011 : Les États ont le devoir de protéger contre les violations des droits de l'homme sur leur territoire et/ou sous leur juridiction par des tiers, y compris les entreprises commerciales.
138. Voir Recommandation CM/Rec(2020)1 du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme, https://rm.coe.int/09000016809e1154. Voir également Recommandation CM/Rec(2016)3 du Comité des Ministres aux États membres sur les droits de l'homme et les entreprises https://rm.coe.int/droits-de-l-homme-etentreprises-recommandation-cm-rec-2016-3-du-comite/16806f2031.

innovation technologiques responsables[139]. Ces dernières années, les acteurs du secteur privé se sont montrés très désireux de promouvoir un développement et une utilisation responsables des systèmes d'IA, reconnaissant aussi bien les possibilités qu'ils offrent que les risques qu'ils présentent. Ils ont non seulement contribué à la multiplication de lignes directrices sur l'éthique en matière d'IA, mais certains se sont aussi clairement dit en faveur de la mise en place d'un cadre réglementaire pour assurer une plus grande sécurité juridique dans ce domaine[140].

93. Dans le cas où une approche réglementaire conjuguant instrument juridique et outils de droit souple serait préconisée, les acteurs du secteur privé, les organisations de la société civile, les milieux universitaires et les autres parties prenantes auraient un rôle important à jouer, non seulement en aidant les États à mettre en place un cadre juridique contraignant, mais aussi en contribuant à l'élaboration d'instruments sectoriels de droit souple mettant concrètement en œuvre les dispositions contraignantes en tenant compte du contexte (par exemple au moyen de lignes directrices sectorielles, de certifications et de normes techniques). L'efficacité d'un cadre réglementaire applicable aux systèmes d'IA reposera sur une étroite coopération entre toutes les parties prenantes, qu'il s'agisse des États et des entités publiques chargées d'assurer un contrôle public, des acteurs du secteur privé qui peuvent apporter leurs connaissances et garantir une innovation en matière d'IA qui soit profitable à la société et des organisations de la société civile qui peuvent représenter les intérêts du grand public, y compris les personnes sous-représentées ou issues de milieux défavorisés. Le CAHAI convient que le Conseil de l'Europe est le mieux placé pour mener cet effort et – en s'appuyant sur les cadres existants – pour veiller à la conformité des systèmes d'IA avec ses normes en matière de droits de l'homme, de démocratie et d'État de droit.

6. PRINCIPALES CONCLUSIONS DES CONSULTATIONS MULTIPARTITES

94. La consultation multipartite doit se tenir en 2021, sous l'égide du Groupe de consultation et de sensibilisation (CAHAI-COG), qui travaille actuellement en étroite collaboration avec le CAHAI-PDG pour définir le champ d'application, les groupes cibles et les modalités de la consultation, sur la base des indications fournies précédemment par le CAHAI. Le CAHAI prendra une décision sur ces questions lors de sa troisième réunion plénière en décembre 2020. Les résultats de la consultation, qui pourraient alimenter le travail d'élaboration des principaux éléments d'un cadre juridique que le CAHAI est mandaté pour développer, seront d'abord examinés par le CAHAI et ensuite présentés au Comité des Ministres dans le cadre du processus de rapport des activités du CAHAI.

---

139. En particulier, les Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme, notamment les articles 18 et 19. Voir également les Principes directeurs de l'OCDE à l'intention des entreprises multinationales et le Guide OCDE sur le devoir de diligence pour une conduite responsable des entreprises.

140. Outre les déclarations de certaines sociétés, comme Microsoft (https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/) ou IBM (https://www.ibm.com/blogs/policy/ai-precision-regulation/), les Recommandations en matière de politique et d'investissement du Groupe d'experts de haut niveau sur l'intelligence artificielle de la Commission européenne, qui compte des experts de plus de 20 sociétés, appellent à envisager l'adoption d'une nouvelle législation, recommandant par exemple page 40 : « *Pour les systèmes d'IA déployés par le secteur privé qui sont susceptibles d'avoir une incidence significative sur les vies humaines, par exemple en portant atteinte aux droits fondamentaux d'un individu à n'importe quel stade du cycle de vie d'un système d'IA, et pour les applications critiques pour la sécurité, envisager la nécessité d'introduire : une obligation de réaliser une évaluation fiable de l'IA (y compris une évaluation des incidences pour les droits fondamentaux couvrant également par exemple les droits de l'enfant, les droits des individus vis-à-vis de l'État et les droits des personnes handicapées) et une consultation des parties prenantes, notamment une consultation des autorités compétentes ; des exigences en matière de traçabilité, d'auditabilité et de contrôle ex ante ; et une obligation de garantir des procédures appropriées par défaut et dès la conception pour permettre d'obtenir une réparation effective et immédiate en cas d'erreurs, de préjudices et/ou de toute autre atteinte aux droits* ». Le document souligne également la nécessité de renforcer la sécurité juridique.

7. PRINCIPAUX ÉLÉMENTS D'UN CADRE JURIDIQUE POUR LA CONCEPTION, LE DÉVELOPPEMENT ET L'APPLICATION DE L'INTELLIGENCE ARTIFICIELLE

7.1. Valeurs, droits et principes clés découlant[141] dans une perspective ascendante – d'approches sectorielles et de lignes directrices éthiques, dans une perspective descendante – des exigences en matière de droits de l'homme, de démocratie et d'État de droit

95. Conformément au mandat du CAHAI, un cadre juridique applicable à l'IA devrait veiller à ce que le développement, la conception et l'application de cette technologie soient fondés sur les normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'État de droit. Il s'agit d'établir, suivant une approche fondée sur les risques, un cadre réglementaire qui tout en étant propice aux innovations bénéfiques de l'IA, tiendra compte des risques recensés au chapitre 3 et des lacunes juridiques de fond et de procédure identifiées au chapitre 5, de manière à garantir sa pertinence et son efficacité par rapport aux instruments existants.

96. L'un des moyens d'y parvenir serait dans un premier temps de formuler des principes fondamentaux devant être garanties dans le contexte de l'IA et dans un deuxième temps d'identifier des droits concrets que les individus peuvent invoquer (qu'il s'agisse de droits existants, de droits nouvellement adaptés aux défis et aux opportunités soulevés par l'IA, ou de clarifications supplémentaires des droits existants) ainsi que des exigences auxquelles les développeurs et les personnes qui déploient des systèmes d'IA devront satisfaire.[142] L'introduction éventuelle de nouveaux droits et obligations dans un futur instrument juridique doit se faire d'une manière qui soit nécessaire, utile et proportionnée à l'objectif à atteindre, à savoir la protection des effets négatifs potentiels du développement et de l'utilisation de systèmes d'IA sur les droits de l'homme, la démocratie et l'État de droit, et d'une manière qui tienne compte d'un équilibre entre les divers intérêts légitimes en jeu. En outre, le cas échéant, les exceptions aux droits existants et nouveaux devraient être conforme à la loi et nécessaire dans une société démocratique dans l'intérêt de la sécurité nationale, de la sécurité publique la sécurité ou d'autres intérêts publics légitimes.

97. La partie ci-après examine les grands principes[143] – y compris les droits et obligations qui s'y rapportent – considérés comme essentiels à respecter dans le contexte des systèmes d'IA et dont l'inclusion dans un futur cadre juridique du Conseil de l'Europe sur l'IA pourrait être potentiellement envisagé. Si, comme indiqué plus haut, ces principes, droits et exigences sont définis dans une perspective d'application horizontale, ils pourraient cependant être associés à une approche sectorielle énonçant des exigences contextuelles (plus détaillées) sous la forme d'instruments non contraignants, à l'instar de lignes directrices sectorielles ou de listes d'évaluation.

7.1.1. Dignité humaine

98. La dignité humaine est le fondement même de tous les droits de l'homme. Elle repose sur l'idée que tous les individus sont intrinsèquement dignes de respect du simple fait de leur statut d'être humain. La dignité humaine est un droit absolu[144], inviolable. Ainsi, même lorsqu'un droit fondamental est restreint – par exemple lorsqu'il est nécessaire de trouver un juste équilibre entre les droits et les intérêts – la dignité humaine doit toujours être protégée et ne peut être mise en balance. Dans le contexte de l'IA, cela signifie que la conception, le développement et l'utilisation de systèmes d'IA doivent respecter la dignité des êtres humains avec lesquels ils interagissent ou sur lesquels ils ont une incidence. Les êtres humains devraient être traités comme des sujets moraux et non comme de simples objets que l'on catégorise, note ou manipule.

---

141. En raison de la rapidité croissante de l'innovation et du développement technologique, les États membres, via leurs universités, leurs écoles d'ingénieurs ou par tout autre moyen, doivent promouvoir, former et encadrer les développeurs et les diffuseurs d'IA sur tous ces principes liés à l'éthique et aux principes de régulation de l'IA qui sont mentionnés, parmi beaucoup d'autres, dans ce document afin de suivre ce rythme rapide.
142. La liste des droits touchés par le développement et l'utilisation des systèmes d'IA telle que mentionnée dans ce chapitre ne doit en aucun cas être considérée comme une liste exhaustive.
143. Les principes énoncés dans ce chapitre sont issus des principes recensés dans le rapport de M. Ienca et E. Vayena (document CAHAI (2020)07-fin) et de discussions ultérieures au sein du CAHAI-PDG. Ils ne sont pas présentés dans un ordre précis.
144. Si le droit à la dignité humaine n'est pas explicitement inclus dans la Convention européenne des droits de l'homme, il a été reconnu comme implicitement inscrit dans celle-ci par la Cour européenne des droits de l'homme à de multiples reprises (voir également à cet égard Antoine Buyse, Le rôle de la dignité humaine dans la jurisprudence de la CEDH, octobre 2016, [en anglais uniquement] http://echrblog.blogspot.com/2016/10/the-role-of-human-dignity-in-echr-case.html.), ce droit est aussi explicitement inscrit à l'article 1 de la Charte des droits fondamentaux de l'Union européenne et est reconnu dans la Déclaration universelle des droits de l'homme.

99. Si les applications de l'IA sont parfois mises au service de la dignité humaine et de l'autonomisation des individus, il arrive que leur utilisation nuise à la dignité humaine ou y fasse (in-)volontairement obstacle. Pour garantir le respect de la dignité humaine, il est essentiel que les êtres humains sachent qu'ils interagissent avec un système d'IA et ne soient pas trompés à cet égard. Ils devraient en outre pouvoir choisir de ne pas interagir avec ce système et refuser d'être soumis à une décision prise par un système d'IA dès lors que cela est susceptible d'avoir d'importantes répercussions sur leur vie, surtout lorsque cela peut porter atteinte aux droits liés à leur dignité humaine. Par ailleurs, il pourrait être nécessaire de confier la réalisation de certaines tâches à des personnes humaines plutôt qu'à des machines compte tenu de leur incidence potentielle sur la dignité humaine. Plus généralement, les systèmes d'IA devraient être développés et utilisés de manière à protéger et servir l'intégrité physique et mentale des êtres humains.

Principaux droits matériels :

- Le droit à la dignité humaine, le droit à la vie (art 2 CEDH) et le droit à l'intégrité physique et mentale.
- Le droit de toute personne d'être informée du fait qu'elle interagit avec un système d'IA et non avec un être humain[145], en particulier lorsque le risque de confusion se présente et peut porter atteinte à la dignité humaine.
- Le droit de refuser l'interaction avec un système d'IA chaque fois que cela peut nuire à la dignité humaine.

Principales obligations :

- Lorsque des tâches réalisées par des machines plutôt que par des êtres humains risquent de nuire à la dignité humaine, les Etats membres devraient s'assurer qu'elles soient exclusivement confiées à des êtres humains.
- Les États membres devraient exiger que les personnes qui déploient l'IA informent les êtres humains du fait qu'elles interagissent avec un système d'IA plutôt qu'avec un être humain chaque fois qu'une confusion peut survenir.

7.1.2. Prévention des atteintes aux droits de l'homme, à la démocratie et à l'État de droit

100. Les systèmes d'IA peuvent être utilisés dans les systèmes de sécurité et de protection pour aider à minimiser le risque de dommages aux individus, à l'environnement et même à d'autres systèmes. Dans le même temps, les systèmes d'IA peuvent également être utilisés d'une manière qui nuit aux individus, aux sociétés et à l'environnement. La prévention des dommages est un principe fondamental qui doit être respecté, tant dans la dimension individuelle que collective, en particulier lorsque ces dommages concernent l'impact négatif sur les droits de l'homme, la démocratie et l'État de droit. L'intégrité physique et mentale des êtres humains doit être protégée de manière adéquate, avec des garanties supplémentaires pour les personnes et les groupes les plus vulnérables. Une attention particulière doit également être accordée aux situations dans lesquelles l'utilisation de systèmes d'IA peut causer ou exacerber des effets négatifs dus à des asymétries de pouvoir ou d'information, par exemple entre les employeurs et les employés, les entreprises et les consommateurs ou les gouvernements et les citoyens.

101. Il est important de noter qu'au-delà de l'impact des systèmes d'IA sur les individus, la prévention des dommages implique également de prendre en compte l'environnement naturel et tous les êtres vivants, ainsi que la manière dont les systèmes d'IA peuvent avoir un impact négatif sur ces derniers. Après tout, les individus dépendent d'un environnement naturel sûr et sain pour vivre. Il faut également prêter attention à la sûreté et à la sécurité des systèmes d'IA, y compris les garanties de leur robustesse technique, leur fiabilité et les mesures qui préviennent le risque d'attaques adverses ou d'utilisations malveillantes.

102. A la lumière de ce qui précède, les Etats membres doivent veiller à ce que des garanties adéquates soient mises en place pour minimiser et prévenir les dommages résultant du développement et de l'utilisation de l'IA, qu'il s'agisse de dommages physiques, psychologiques, économiques, environnementaux, sociaux ou juridiques. Les garanties susmentionnées sont particulièrement importantes dans le domaine des procédures de passation de marchés publics, ainsi que dans la conception des systèmes électroniques de passation de marchés publics. Lors de la mise en œuvre de mesures visant à prévenir les dommages, les

145. Cela a également été recommandé par les lignes directrices du Conseil de l'Europe sur l'IA et la protection des données, https://rm.coe.int/lignes-directrices-sur-l-intelligence-artificielle-et-la-protection-de/168091ff40

États membres devraient adopter une approche fondée sur les risques. En outre, lorsque les circonstances particulières le justifient, par exemple en cas de degré élevé d'incertitude associé à un niveau élevé de risque, une approche de précaution, y compris des interdictions potentielles, devrait être adoptée. Enfin, les États membres pourraient également envisager l'utilisation de garanties fondées sur l'IA pour minimiser et prévenir les dommages résultant des actions des êtres humains.

Principaux droits matériels :

- Le droit à la vie (art 2 CEDH) et le droit à l'intégrité physique et mentale.
- Le droit à la protection de l'environnement et le droit à une communauté et une biosphère durables.

Principales obligations :

- Les États membres devraient veiller à ce que les développeurs et déployeurs de systèmes d'IA prennent les mesures nécessaires pour réduire au minimum tout risque de préjudice physique ou mental pour les individus.
  - Cela pourrait, par exemple, être fait en s'assurant que les systèmes d'IA potentiellement nuisibles fonctionnent sur la base d'un modèle d'acceptation (opt-in) plutôt que d'un modèle de refus (opt-out). Lorsque cela n'est pas possible, des instructions claires doivent être fournies sur la manière dont les personnes peuvent refuser l'utilisation du système et sur les méthodes alternatives non fondées sur l'IA.
- Les États membres devraient s'assurer que les développeurs et déployeurs de systèmes d'IA satisfont aux exigences en matière de sûreté, de sécurité et de robustesse (dès la conception).
  - Ces exigences devraient inclure, entre autres, la résistance aux attaques, l'exactitude et la fiabilité, et la nécessité de garantir la qualité et l'intégrité des données. En outre, les systèmes d'IA doivent être dûment testés et vérifiés avant leur utilisation ainsi que tout au long de leur cycle de vie, notamment par des examens périodiques visant à réduire ces risques au minimum.
- Les États membres devraient veiller à ce que les systèmes d'IA soient développés et utilisés de manière durable, dans le plein respect des normes de protection environnementale.
- Le cas échéant, les États membres pourraient encourager l'utilisation de systèmes d'IA pour éviter et atténuer les dommages causés par les actions des êtres humains et d'autres systèmes technologiques, tout en préservant les normes en matière de droits de l'homme, de démocratie et d'État de droit.
- Les États membres pourraient également envisager de promouvoir des solutions d'IA qui protègent et soutiennent l'intégrité humaine et qui peuvent contribuer à résoudre les problèmes environnementaux. 6.0.3

7.1.3. Liberté et autonomie de la personne humaine

103. La liberté et l'autonomie humaines sont des valeurs fondamentales qui trouvent leur expression dans divers droits de l'homme consacrés par la CEDH. Dans le contexte de l'IA, ces valeurs renvoient à la capacité de chaque être humain d'agir de manière autodéterminée, en décidant de façon éclairée et autonome de l'utilisation d'un système d'IA et des conséquences que cette utilisation peut entraîner pour lui-même et pour autrui. Cela concerne également la décision de savoir si, quand et comment utiliser un système d'IA. Comme indiqué au chapitre 3, la liberté et l'autonomie humaines peuvent être affectées de diverses manières par l'IA : une surveillance (de masse) ou une manipulation ciblée par intelligence artificielle peuvent ainsi être exercées - que ce soit par des entités publiques ou privées – par exemple au moyen de systèmes de reconnaissance biométrique à distance ou de suivi en ligne.

104. En général, les systèmes d'IA ne doivent pas être utilisés pour subordonner, contraindre, tromper, manipuler ou conditionner les humains, mais plutôt pour compléter et augmenter leurs capacités. Des mécanismes de surveillance humaine doivent être mis en place, garantissant que l'intervention humaine est possible chaque fois que cela est nécessaire pour sauvegarder les droits de l'homme, la démocratie et l'État de droit. Comme indiqué au chapitre 5, la mise en place de mécanismes de contrôle humains adéquats n'est pas encore garantie par la loi. L'étendue et la fréquence de la surveillance doivent être adaptées au contexte

spécifique d'application de l'AI[146] et l'autonomie de ces interventions humaines doit être préservée[147]. Il faut cependant veiller à ce que, lorsqu'une intervention humaine est nécessaire, elle soit effectuée par une personne ayant la capacité réellement autonome de passer outre à la décision du système[148] (sans entraver l'automatisation ou le manque de temps pour l'examen).[149]

Principaux droits matériels :

- Le droit à la liberté et à la sécurité (art 5 CEDH)
- Le droit de la personne humaine à l'autonomie et à l'autodétermination. Le droit de ne pas être soumis à une décision reposant uniquement sur un traitement automatisé lorsque celui-ci produit des effets juridiques sur des personnes physiques ou affecte de manière similaire et significative.[150]
- Le droit de contester et d'attaquer effectivement les décisions informées et/ou prises par un système d'IA et d'exiger que cette décision soit revue par une personne (droit de retrait).
- Le droit de décider librement d'être exclu de la manipulation, du profilage individualisé et des prédictions basées sur l'IA, ainsi qu'en cas de traitement de données non personnelles.
- Le droit d'avoir la possibilité, lorsqu'il n'est pas exclu par des motifs impérieux légitimes concurrents., de choisir d'avoir un contact avec un être humain plutôt qu'avec un robot.

Principales obligations :

- Toute manipulation, profilage individualisé et prédiction utilisant l'IA et impliquant le traitement de données à caractère personnel doit respecter les obligations énoncées dans la Convention du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel. Les États membres devraient mettre en œuvre efficacement la version modernisée de la Convention ("Convention 108+") pour mieux traiter les questions liées à l'IA.
- Les États membres devraient exiger des développeurs et des déployeurs de systèmes d'IA qu'ils mettent en place un mécanisme approprié de surveillance humaine assurant le respect de l'autonomie humaine, d'une manière qui soit adaptée aux risques spécifiques découlant du contexte dans lequel le système d'IA est développé et utilisé :
  - Un niveau approprié d'intervention humaine devrait être assuré durant le fonctionnement du système d'IA, sur la base d'une évaluation contextuelle des risques tenant compte de l'impact du système sur les droits de l'homme, la démocratie et l'État de droit.
  - Chaque fois que nécessaire, sur la base d'une anticipation rigoureuse des risques, une personne humaine qualifiée devrait pouvoir désactiver le système d'IA ou en modifier les fonctionnalités.
  - Les personnes qui développent et exploitent des systèmes d'IA devraient être dotées des compétences nécessaires pour ce faire, afin qu'une surveillance puisse être dûment exercée pour permettre la protection des droits de l'homme, de la démocratie et de l'État de droit.
  - Afin de protéger l'intégrité physique et mentale des êtres humains, les personnes qui déploient l'IA doivent s'efforcer d'éviter l'utilisation de modèles d'« économie de l'attention » qui peuvent limiter l'autonomie humaine.

146. Voir à cet égard la distinction établie entre les approches dites « human-on-the-loop » (l'humain supervise le processus), « human-in-the-loop » (l'humain intervient dans le processus) et « human-in-command approach » (l'humain reste aux commandes) dans les Lignes directrices en matière d'éthique pour une IA digne de confiance, page 19, disponible à l'adresse suivante : https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419.
147. Par exemple, en veillant - lorsque cela est approprié et possible - à ce que la personne qui intervient ne connaisse pas la décision prise par la machine.
148. Concernant le recours excessif aux solutions fournies par les applications d'IA et la crainte de contester les décisions suggérées par les applications d'IA, qui risquent d'altérer l'autonomie de l'intervention humaine dans les processus décisionnels, voir également les lignes directrices du T-PD sur l'IA et la protection des données (T-PD(2019)01) où il est dit que « Le rôle de l'intervention humaine dans les processus décisionnels et la liberté des décideurs humains de ne pas se fier au résultat des recommandations fournies à l'aide de l'IA doivent donc être préservés ».
149. Il faut veiller à ce que l'« humain dans la boucle » ne devienne pas une « zone de fracture » morale ou juridique, qui peut être utilisée pour décrire comment la responsabilité d'une action peut être attribuée à tort à un acteur humain qui avait un contrôle limité sur le comportement d'un système automatisé ou autonome.
150. On peut noter qu'un droit similaire existe dans la Convention 108+, mais la protection qu'elle offre est moins complète (article 9(a) : « le droit de ne pas être soumis à une décision l'affectant de manière significative prise sur le seul fondement d'un traitement automatisé de données, sans que son avis soit pris en considération »). Par exemple, elle ne s'applique pas dans les situations qui ne relèvent pas du champ d'application de la Convention, comme lorsque les données à caractère personnel d'une personne n'ont pas été traitées, alors qu'un système d'IA peut également avoir une incidence sur les personnes sans traiter leurs données à caractère personnel.

► Les États membres devraient exiger des concepteurs et des diffuseurs d'IA qu'ils communiquent dûment et en temps utile les possibilités de recours.

7.1.4. Non-Discrimination, égalité de genre151, équité et diversité

105. Comme indiqué au chapitre 3, l'utilisation des systèmes d'IA peut avoir des effets négatifs sur le droit à la non-discrimination et le droit à l'égalité et à l'égalité de traitement. Plusieurs études ont montré que l'utilisation de ces systèmes pouvait perpétuer et exacerber des biais discriminatoires ou injustes et des stéréotypes nuisibles, et ainsi avoir une incidence négative non seulement sur les personnes soumises à cette technologie, mais aussi sur l'ensemble de la société[152]. Le recours à des systèmes d'IA entachés de biais injustes pourrait en effet accroître les inégalités et dès lors mettre en péril la cohésion sociale et l'équité dont la démocratie a besoin pour prospérer.

106. Bien que le droit à l'égalité et à la non-discrimination soit déjà inscrit dans de nombreux instruments juridiques internationaux, il importe, comme nous l'avons fait observer au chapitre 5, de l'adapter aux défis spécifiques que pose l'IA afin d'en assurer le respect. En particulier, l'importance accrue de la discrimination par procuration dans le contexte de l'apprentissage machine peut soulever des questions d'interprétation sur la distinction entre discrimination directe et indirecte ou, en fait, sur l'adéquation de cette distinction telle qu'elle est traditionnellement comprise. De même, il peut y avoir des questions d'interprétation sur la signification des normes traditionnelles de justification de la discrimination dans le contexte de l'apprentissage automatique. L'impact de l'utilisation des systèmes d'IA sur l'égalité de genre doit faire l'objet d'une attention particulière, dans la mesure où ces systèmes risquent de perpétuer (de manière non intentionnelle) la discrimination fondée sur le sexe, les stéréotypes sexistes et le sexisme. Il convient également de faire preuve de prudence quant à l'amplification potentielle de la discrimination à l'égard des personnes marginalisées et, plus généralement, des personnes en situation de vulnérabilité, y compris la discrimination fondée sur la race, l'origine ethnique ou culturelle et le racisme qui pourraient être perpétués par l'IA.[153] Le manque actuel de diversité parmi les personnes qui développent et prennent des décisions dans le secteur de l'IA est une source de préoccupation, et la représentation diversifiée dans les processus consultatifs concernant les applications du système d'IA dans des domaines sensibles devrait être encouragée. Cela contribuerait à prévenir et à atténuer les effets négatifs sur les droits de l'homme, notamment en ce qui concerne l'égalité et la non-discrimination. Il importe également de dûment tenir compte du risque de discrimination croisée que peut entraîner l'utilisation de systèmes d'IA[154] ainsi que le traitement fondé sur des motifs de différenciation ou des associations erronées qui pourraient ne pas être couverts par l'article 14 de la CEDH[155].

Principaux droits matériels :

► Le droit à la non-discrimination et le droit à l'égalité de traitement.
  – Le droit à la non-discrimination (sur la base des motifs protégés énoncés à l'article 14 de la CEDH et dans le protocole 12 à la CEDH), y compris la discrimination intersectionnelle.

151. Selon la définition du Conseil de l'Europe, « L'égalité entre les femmes et les hommes implique des droits égaux pour les femmes et les hommes, les filles et les garçons ainsi que la même visibilité, autonomisation, responsabilité et participation dans tous les domaines de la vie publique et privée. Elle implique également l'égalité des femmes et des hommes dans l'accès aux ressources et dans la distribution de celles-ci. »
152. Voir par exemple l'étude du CdE réalisée par F. Zuiderveen Borgesius, Discrimination, intelligence artificielle et décisions algorithmiques, 2018, disponible à l'adresse suivante : https://rm.coe.int/etude-sur-discrimination-intelligence-artificielle-et-decisions-algori/1680925d84 ; Joy Buolamwini, Timnit Gebru ; Actes de la 1e conférence sur l'équité, la responsabilité et la transparence, PMLR 81 :77-91, 2018. Voir également le 1er rapport de réunion du CAHAI-PDG, p.5.
153. Dans le cas des processeurs de langue naturelle et des assistants linguistiques basé sur des systèmes d'IA, cela est particulièrement important pour les langues des minorités qui peuvent être discriminées si seules les langues les plus courantes sont utilisées.
154. On parle de discrimination croisée lorsque plusieurs motifs ou caractéristiques personnels opèrent et interagissent simultanément au point de devenir indissociables. Les systèmes actuels d'intelligence artificielle ont particulièrement tendance à exercer ce genre de discrimination puisqu'ils se contentent de rechercher des corrélations entre différentes caractéristiques. Le cadre juridique du Conseil de l'Europe devrait plus particulièrement s'intéresser à cette question, la discrimination croisée étant rarement couverte par la législation nationale qui ne porte généralement que sur un motif de discrimination à la fois.
155. Voir l'exemple figurant dans l'étude précitée du Conseil de l'Europe sur l'IA et la discrimination, page 38 : "(…) *supposons qu'un système d'IA découvre une corrélation entre l'utilisation d'un certain navigateur et l'acceptation de prix supérieurs. Un magasin en ligne pourrait demander des prix plus élevés aux utilisateurs de ce navigateur. Cela ne relèverait pas de la législation anti-discrimination, le navigateur ne faisant pas partie des caractéristiques protégées* ».

- Les systèmes d'IA peuvent également donner lieu à un traitement injuste basé sur de nouveaux types de différenciation qui ne sont pas traditionnellement protégés.[156]
- Ce droit s'applique tout au long du cycle de vie d'un système d'IA (conception, développement, mise en œuvre et utilisation) et concerne également les choix humains qui entourent l'utilisation d'un système d'IA, que ce système soit utilisé dans le secteur public ou privé.

Principales obligations :

- Les États membres sont tenus de veiller à ce que les systèmes d'IA qu'ils déploient n'entraînent pas de discrimination illégale, de stéréotypes nuisibles (y compris, mais sans s'y limiter, les stéréotypes liés au genre) et d'inégalités sociales plus larges, et doivent donc appliquer le plus haut niveau de contrôle lorsqu'ils utilisent ou encouragent l'utilisation de systèmes d'IA dans des domaines sensibles de la politique publique, y compris, mais sans s'y limiter, l'application de la loi, la justice, l'asile et la migration, la santé, la sécurité sociale et l'emploi.
- Les États membres devraient inclure la non-discrimination et la promotion des exigences d'égalité dans les processus de passation de marchés publics pour les systèmes d'IA, et veiller à ce que les systèmes fassent l'objet d'un audit indépendant des effets discriminatoires avant leur déploiement. Les systèmes d'IA doivent être dûment testés et vérifiés avant leur utilisation ainsi que tout au long de leur cycle de vie, notamment par des audits et des examens périodiques.
- Les États membres devraient imposer des exigences visant à contrer efficacement les effets discriminatoires potentiels des systèmes d'IA déployés par les secteurs public et privé et à protéger les individus contre les conséquences négatives de ces systèmes. Ces exigences doivent être proportionnées aux risques encourus.
  - Ces exigences doivent couvrir l'ensemble du cycle de vie d'un système d'IA et doivent notamment viser à combler les lacunes existantes en matière de données sur le genre, la représentativité, la qualité et l'exactitude des ensembles de données[157], la conception et la fonction d'optimisation des algorithmes, l'utilisation du système et les processus de test et d'évaluation adéquats pour vérifier et atténuer le risque de discrimination.
  - La transparence et l'auditabilité des systèmes d'IA doivent être assurées pour permettre la détection de la discrimination tout au long du cycle de vie d'un système d'IA (voir ci-dessous).
- Les États membres devraient encourager la diversité et l'équilibre entre les sexes au sein de la main-d'œuvre d'IA et le retour d'information périodique de la part d'un éventail diversifié de parties prenantes. La prise de conscience du risque de discrimination, y compris des nouveaux types de différenciation, et des préjugés dans le contexte de l'IA doit être encouragée.
- Les États membres devraient encourager le déploiement de systèmes d'IA là où ils peuvent lutter efficacement contre les discriminations existantes dans la prise de décision basée sur l'homme et la machine.

7.1.5. Principe de transparence et d'explicabilité des systèmes d'IA

107. En l'absence d'informations transparentes permettant de savoir si un produit ou un service utilise un système d'IA et, le cas échéant, selon quels critères il fonctionne, il est souvent difficile, voire impossible, de déterminer si ce système a une incidence sur les droits de l'homme, la démocratie et l'État de droit. Sans ces informations, il n'est possible ni de dûment contester une décision ni d'améliorer ou de corriger le système lorsqu'il cause un préjudice. La transparence est donc essentielle pour garantir le respect d'autres principes et droits, notamment le droit à un recours effectif en cas de violation, ce qui inclut le droit de contester une décision prise en connaissance de cause par un système d'IA et de demander réparation. Par conséquent, les

156. Certains experts ont suggéré que, plutôt que d'étendre la liste des motifs de différenciation injuste avec de nouveaux motifs, une disposition « fourre-tout » pourrait être plus appropriée pour combler ce vide juridique spécifique. Voir J. Gerards et F. Zuiderveen Borgesius, "*Protected grounds and the system of non-discrimination law in the context of algorithmic decision-making and artificial intelligence*", Nov. 2020, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3723873.

157. Cela pourrait également englober le recours obligatoire à des ensembles croisés de données d'entraînement, la création d'indicateurs croisés et l'introduction d'audits croisés. Le respect de ces exigences peut être évalué à l'aune des résultats produits par le système d'IA ce qui signifie que l'accès aux procédures d'entraînement, de test et d'évaluation en tant que telles n'est pas toujours nécessaire. Cela nécessite toutefois des procédures appropriées pour permettre un examen significatif des résultats du système en termes, par exemple, de représentativité, d'exactitude et de qualité.

principes de transparence et d'explicabilité[158] sont donc indispensables dans le contexte de l'IA, en particulier lorsque le système est susceptible d'avoir une incidence sur les droits de l'homme, la démocratie et l'État de droit. Comme nous l'avons fait observer au chapitre 5, ces principes ne sont toutefois pas encore suffisamment protégés par les instruments juridiques existants.

108. La transparence suppose la mise en place de mesures permettant d'assurer la traçabilité des processus d'un système d'IA, par exemple des mécanismes de documentation ou de journalisation de ces processus, ainsi que la communication d'informations pertinentes sur les capacités, les limites et la finalité du système. Ces informations doivent être adaptées au contexte et au public auxquelles elles sont destinées. Les États membres devraient définir des procédures afin que les systèmes d'IA fassent l'objet d'audits indépendants et efficaces permettant de véritablement évaluer leur impact. Les personnes qui font l'objet d'une décision reposant exclusivement ou largement sur un système d'IA devraient en être avisées et recevoir rapidement les informations susmentionnées. Il convient par ailleurs que le processus de prise de décisions puisse leur être expliqué.[159] Il n'est pas toujours possible d'expliquer pour quelle raison un système a généré un résultat en particulier[160], d'où l'importance de garantir l'auditabilité du système[161]. Si les secrets des affaires et les droits de propriété intellectuelle doivent être respectés, ils devraient cependant être mis en balance avec d'autres intérêts légitimes. Les pouvoirs publics doivent pouvoir réaliser un audit des systèmes d'IA lorsqu'il existe des indications solides de non-conformité pour vérifier le respect de la législation en vigueur. Les charges techniques de transparence et d'explicabilité ne doivent pas restreindre de manière déraisonnable les possibilités du marché, en particulier lorsque les risques pour les droits de l'homme, la démocratie et l'État de droit sont moins importants. Il convient donc d'adopter une approche fondée sur le risque et de trouver un équilibre approprié pour prévenir ou réduire au minimum le risque d'enracinement des principaux acteurs du marché et/ou d'éviction et, ce faisant, de diminution de la recherche et du développement de produits innovants et socialement bénéfiques.

Principaux droits matériels :

- Le droit d'être rapidement informé qu'une décision produisant des effets juridiques ou d'autres effets d'importance comparable sur la vie d'un individu est générée par un système d'IA.[162]
- Le droit de recevoir une explication du fonctionnement de ce système d'IA, de la logique d'optimisation qu'il suit, du type de données qu'il utilise et de la manière dont cela affecte les intérêts de l'utilisateur, chaque fois que ce système génère des effets juridiques ou produit des effets d'importance comparable sur la vie de l'intéressé. L'explication doit être adaptée au contexte et fournie d'une manière qui soit utile et compréhensible pour un individu, permettant à celui-ci de protéger efficacement ses droits.
- Le droit d'un utilisateur d'un système d'IA d'être assisté par un être humain lorsqu'un système d'IA est utilisé pour interagir avec des individus, en particulier dans le cadre de services publics.

Principales obligations :

- Les États membres devraient exiger que les concepteurs et les utilisateurs de systèmes d'IA assurent une communication adéquate :
  - Les utilisateurs devraient être clairement informés de leur droit à être assistés par un être humain chaque fois qu'est utilisé un système d'IA pouvant avoir un impact sur leurs droits ou les affecter de manière significative, en particulier dans le contexte des services publics, et de la manière de demander cette assistance.

158. La mise en œuvre de ces principes doit se faire de manière à les mettre en balance avec d'autres intérêts légitimes, tels que la sécurité nationale et les droits de propriété intellectuelle.
159. Bien que différents types d'explications soient possibles, il est important de veiller à ce que l'explication soit adaptée au contexte et au public spécifiques. Cette explication doit au moins fournir les éléments nécessaires pour permettre à une personne de comprendre et de contester une décision qui a été prise ou informée par un système d'IA et qui affecte sa situation juridique ou sa vie de manière substantielle.
160. Notons que, dans certains cas, seule une diminution des performances et de la précision du système permettra d'obtenir un niveau plus élevé d'explicabilité.
161. Cela signifie que les auditeurs (indépendants) doivent être en mesure d'apprécier et d'évaluer les différentes étapes qui ont été prises pour concevoir, développer, former et vérifier le système, englobant à la fois les algorithmes du système et les ensembles de données qui ont été utilisés, afin de garantir l'alignement du système sur les droits de l'homme, la démocratie et l'État de droit. Les mécanismes de vérification les plus appropriés dépendront du contexte et de l'application.
162. Des exceptions à ce droit devraient être prévues par la loi, pour sauvegarder des intérêts publics légitimes tels que la sécurité nationale, lorsque cela est nécessaire dans une société démocratique et dans le respect du principe de proportionnalité.

- Chaque fois que l'utilisation de systèmes d'IA risque d'avoir un effet négatif sur les droits de l'homme, la démocratie et l'État de droit, les États membres devraient imposer des exigences aux concepteurs et aux déployeurs d'IA en ce qui concerne la traçabilité et la fourniture d'informations :
  - Les personnes qui ont un intérêt légitime (par ex. consommateurs, citoyens, autorités de contrôle notamment) devraient avoir facilement accès aux informations nécessaires et pertinentes sur les systèmes d'IA.
  - Ces informations devraient être compréhensibles, facilement accessibles et pourraient notamment englober le type de décisions ou de situations faisant l'objet d'un traitement automatisé, les critères intervenant dans une décision, les informations sur les données utilisées, une description de la méthode utilisée pour la collecte de données. Une description des potentiels effets juridiques, ou autres, devrait être accessible pour examen/audit par des organismes indépendants ayant les compétences nécessaires.[163]
  - Une attention particulière doit être accordée si des enfants ou d'autres groupes vulnérables sont soumis à une interaction avec des systèmes d'IA.
- Les États membres devraient imposer des exigences aux développeurs et aux déployeurs d'IA en matière de documentation :
  - La traçabilité et l'auditabilité des systèmes d'IA susceptibles d'avoir une incidence négative sur les droits de l'homme, la démocratie et l'État de droit devraient être assurées. Les ensembles de données et les processus permettant au système d'IA de rendre une décision, y compris les processus de collecte et d'étiquetage des données ainsi que les algorithmes utilisés, devraient être documentés, de manière à faciliter l'auditabilité ex post du système.
  - Des procédures de documentation qualitatives et efficaces devraient être mises en place.
- Les États membres devraient rendre publiques et accessibles toutes les informations pertinentes sur les systèmes d'IA (y compris leur fonctionnement, leur fonctionnement d'optimisation, la logique sous-jacente, le type de données utilisées) qui sont utilisés pour la fourniture de services publics, tout en préservant les intérêts légitimes tels que la sécurité publique ou les droits de propriété intellectuelle, tout en assurant le plein respect des droits de l'homme.

7.1.6. Protection des données et respect de la vie privée

109. Le droit à la vie privée fait partie du droit au respect de la vie privée et familiale au titre de l'article 8 de la CEDH et bénéficie d'une protection spécifique dans le cadre du traitement automatisé des données à caractère personnel dans la Convention 108. Il est également fondamental pour la jouissance d'autres droits de l'homme. Dès lors, le développement, l'entraînement, la phase d'essai et l'utilisation de systèmes d'IA qui reposent sur le traitement de données à caractère personnel doivent garantir pleinement le droit de toute personne au respect de la vie privée et familiale, y compris le « droit à l'autodétermination en matière d'information » en ce qui concerne ses données. Toute personne devrait pouvoir exercer un contrôle sur ses données. Le principe du consentement éclairé – bien que ce ne soit pas là le seul fondement juridique qui s'applique au traitement des données à caractère personnel – joue un rôle important à cet égard. Toutefois, pour être valable, le consentement doit être informé, spécifique, donné librement et sans ambiguïté (sinon « explicite » lorsque le traitement concerne des données sensibles). Les situations d'asymétrie de pouvoir ou d'information peuvent affecter l'exigence de consentement librement donné, ce qui implique certaines limitations à sa fonction de protection dans certaines situations et la nécessité d'une base juridique plus appropriée pour le traitement dans ces situations.

110. Les États membres devraient effectivement mettre en œuvre la Convention modernisée du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (« Convention 108+ ») ainsi que les autres instruments internationaux contraignants sur la protection des données et le respect de la vie privée qui sont juridiquement. Les systèmes d'IA ne traitent pas tous des données à caractère personnel. Mais, même lorsqu'ils ne sont pas conçus pour cela et qu'ils exploitent des données anonymisées, anonymes, ou non personnelles, la frontière entre données personnelles et données non personnelles est de plus en plus ténue. Il faut donc examiner plus avant l'interaction entre données personnelles et non personnelles afin de combler tout vide juridique éventuel en matière de protection. Les systèmes d'apprentissage automatique notamment peuvent déduire des informations personnelles sensibles sur des individus à partir de données anonymisées ou anonymes, voire à partir de données d'autres

163. Sans préjudice des droits et obligations existants à cet égard, inscrits dans la Convention 108.

personnes. À cet égard, une attention particulière doit être portée à la protection des personnes contre les données personnelles déduites[164].

111. Enfin, quels que soient les avantages que l'utilisation d'un système d'IA particulier pourrait apporter, toute ingérence dans l'exercice du droit à la vie privée, en particulier par une autorité publique, doit être conforme à la loi, notamment aux droits fondamentaux potentiellement en conflit, et nécessaire dans une société démocratique. Pour établir si une atteinte particulière à ce droit est « nécessaire dans une société démocratique », la Cour européenne des droits de l'homme a précisé que le terme « nécessaire » n'a pas la souplesse d'expressions telles que « utile », « raisonnable » ou « souhaitable », mais implique plutôt l'existence d'un « besoin social pressant » pour l'ingérence en question.[165] Il appartient aux autorités nationales de procéder à l'évaluation initiale du besoin social pressant auquel l'utilisation d'un système d'IA pourrait répondre dans chaque cas, sous réserve du contrôle de la Cour. Les autorités nationales sont encouragées à consulter un large éventail de parties prenantes dans le cadre de cette évaluation et à en assurer le réexamen périodique.

Principaux droits matériels et principales obligations :

Principal droit matériel :

- Le droit au respect de la vie privée et familiale et à la protection des données à caractère personnel (art 8 CEDH).
- Le droit à l'intégrité physique, psychologique et morale à la lumière du profilage basé sur l'IA et la reconnaissance de ses effets.
- Tous les droits inscrits dans la Convention 108 et dans sa version modernisée, et en particulier en ce qui concerne le profilage basé sur l'IA et la localisation.

Principales obligations :

- Les États membres devraient veiller à ce que le droit à la vie privée et à la protection des données soit garanti tout au long du cycle de vie des systèmes d'IA qu'ils déploient, ou qui sont déployés par des acteurs privés. Le traitement des données à caractère personnel à tout stade du cycle de vie d'un système d'IA, y compris les ensembles de données, doit être fondé sur les principes énoncés dans la Convention 108+ (notamment l'équité et la transparence, la proportionnalité, la licéité du traitement, la qualité des données, le droit de ne pas être soumis à des décisions purement automatisées et les autres droits de la personne concernée, la sécurité des données, la responsabilité, les évaluations d'impact et le respect de la vie privée dès la conception).
- Les États membres devraient prendre des mesures particulières pour protéger efficacement les personnes contre la surveillance de masse liée à l'IA, par exemple grâce à la technologie de reconnaissance biométrique à distance ou à d'autres technologies de suivi basées sur l'IA, car cela n'est pas compatible avec les normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'État de droit. A cet égard, comme mentionné au chapitre 3, lorsque cela est nécessaire et approprié pour protéger les droits de l'homme, les Etats devraient envisager l'introduction de mesures réglementaires supplémentaires ou d'autres restrictions pour l'utilisation exceptionnelle et contrôlée de l'application et, lorsque cela est essentiel, une interdiction ou un moratoire.
- Lors de l'acquisition ou de la mise en œuvre de systèmes d'IA, les États membres devraient évaluer et atténuer tout impact négatif de ces systèmes sur le droit à la vie privée et à la protection des données, ainsi que sur le droit plus large au respect de la vie privée et familiale, en considérant en particulier la proportionnalité du caractère invasif du système par rapport à l'objectif légitime qu'il devrait atteindre, ainsi que sa nécessité pour y parvenir.
- Les États membres devraient envisager le développement et l'utilisation d'applications d'IA qui peuvent exploiter l'utilisation bénéfique des données (personnelles) lorsqu'elles peuvent contribuer à la promotion et à la protection des droits de l'homme, tels que le droit à la vie (par exemple dans

164. Voir, par exemple, S. Wachter et B. Mittelstadt, A Right to Reasonable Inferences : Re-Thinking Data Protection Law in the Age of Big Data and AI, Columbia Business Law Review, 2019(2). Cela soulève également la question de savoir ce qui peut précisément être considéré comme les « données propres » d'une personne qui doivent être protégées, et dans quelle mesure cette protection doit englober les données brutes, les données analysées, ainsi que les conclusions qui peuvent être tirées sur la base des données (personnelles).

165. Cour européenne des droits de l'homme, «Guide sur l'article 8 de la Convention européenne des droits de l'homme» (2020), p12. disponible à l'adresse : https://www.echr.coe.int/Documents/Guide_Art_8_FRA.pdf

le contexte de la médecine expérimentale basée sur l'IA). Ce faisant, ils doivent garantir l'exécution de tous les droits de l'homme, et en particulier le droit à la vie privée et à la protection des données en assurant le plein respect de la Convention du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel et en mettant effectivement en œuvre la version modernisée de la Convention («Convention 108+»).

- Étant donné l'importance des données dans le contexte de l'IA, les États membres devraient mettre en place des garanties appropriées pour les flux de données transfrontaliers afin de s'assurer que les règles de protection des données ne sont pas contournées, conformément à la Convention 108 et à sa version modernisée.

7.1.7. Obligation de rendre des comptes (accountability) et responsabilité

112. Les personnes, y compris les organisations publiques et privées, qui conçoivent, développent, déploient ou évaluent des systèmes d'IA doivent assumer la responsabilité de ces systèmes et être tenus de rendre compte chaque fois que les normes juridiques basée sur les principes mentionnés ci-dessus ne sont pas respectés ou que l'utilisateur final ou toute autre personne subit un préjudice. Autrement dit, des mécanismes appropriés devraient être mis en place pour veiller à ce que les systèmes d'IA, tant avant qu'après leur développement, leur déploiement et leur utilisation, soient en adéquation avec les normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'État de droit. Les États membres doivent prendre les mesures appropriées pour garantir cela, par exemple en imposant une responsabilité civile ou pénale lorsque la conception, le développement et l'utilisation d'applications d'IA portent atteinte aux droits de l'homme ou affectent négativement le processus démocratique et l'État de droit. Il est indispensable que les effets négatifs potentiels des systèmes d'IA puissent être identifiés, évalués, documentés et atténués, et que ceux qui en rendent compte (les lanceurs d'alerte, par exemple) soient protégés. Sur la base d'une approche fondée sur le risque, un contrôle public efficace et des mécanismes de contrôle doivent être garanti, afin de veiller à ce que les développeurs et déployeurs de systèmes d'IA agissent conformément aux dispositions juridiques pertinentes, tout en prévoyant la possibilité d'interventions appropriés des pouvoirs publics quand ce n'est pas le cas.

113. Les Etats membres doivent s'asurer que les personnes susceptibles de subir les effets négatifs des systèmes d'IA disposent d'un recours effectif et accessible contre les concepteurs ou déployeurs de systèmes d'IA qui en sont responsables. L'accessibilité d'une voie de recours de cette nature doit leur être communiquée de façon claire, en accordant une attention particulière aux personnes ou groupes vulnérables et marginalisées. Des voies de recours effectives devraient permettre d'obtenir rapidement des réparations pour tout préjudice subi, et peuvent inclure des mesures civiles, administratives ou, le cas échéant, pénales. En outre, en raison du fait que les applications de l'IA sont multiples, tout recours doit être adapté aux applications visées. Il devrait inclure la cessation des comportements illicites ainsi que des garanties de non-répétition, de même que l'obligation de réparer le dommage causé, et le respect des règles générales concernant le partage et le renversement de la charge de la preuve dans la législation anti-discrimination.[166]

Principaux droits matériels

- Le droit à un recours effectif (art. 13 CEDH).
- Cela devrait également inclure le droit à des recours efficaces et accessibles chaque fois que le développement ou l'utilisation de systèmes d'IA par des entités privées ou publiques cause un préjudice injuste ou viole les droits légalement protégés d'un individu.

Principales obligations

- Les États membres devraient veiller à ce que des recours efficaces soient disponibles dans le cadre des juridictions nationales respectives, y compris en matière de responsabilité civile et pénale, et à ce que des mécanismes de recours accessibles soient mis en place pour les personnes dont les droits sont affectés par le développement ou l'utilisation d'applications d'IA.
- À cet égard, ils pourraient également envisager l'introduction de recours collectifs dans le contexte du préjudice causé par l'utilisation des systèmes d'IA et veiller à ce que les règles générales concernant le partage et le renversement de la charge de la preuve dans la législation anti-discrimination soient appliquées.

166. A cet égard, voir le § 11 de la recommandation de politique générale n°7 de l'ECRI et les §§ 29 et suivants de son exposé des motifs, accessibles à l'adresse : https://www.coe.int/fr/web/european-commission-against-racism-and-intolerance/recommendation-no.7

- Les États membres devraient mettre en place des mécanismes de contrôle public des systèmes d'IA susceptibles d'enfreindre les normes juridiques dans le domaine des droits de l'homme, de la démocratie ou de l'État de droit.
- Les États membres devraient veiller à ce que les concepteurs et les utilisateurs de systèmes d'IA :
  - identifient, documentent et rendent compte des impacts négatifs potentiels des systèmes d'IA sur les droits de l'homme, la démocratie et l'État de droit ;
  - mettent en place des mesures d'atténuation adéquates pour garantir la responsabilité et l'obligation de rendre compte de tout dommage causé.
- Les États membres devraient mettre en place des mesures visant à garantir que les autorités publiques soient toujours en mesure de contrôler les systèmes d'IA utilisés par des acteurs privés[167], afin d'évaluer leur conformité avec la législation en vigueur et de tenir les acteurs privés responsables.

7.1.8. Démocratie

114. Afin de faire face de manière adéquate aux risques qui menacent la démocratie, soulignés au chapitre 3, des mécanismes de contrôle démocratique efficaces, transparents et inclusifs sont nécessaires pour veiller à ce que le processus de décision démocratique - et les valeurs qui lui sont corrélées, de pluralisme, d'accès à l'information et d'autonomie – soient préservés, ainsi que les droits économiques et sociaux qui peuvent en être affectés négativement.

115. Lorsque cela est pertinent et raisonnablement possible, les États membres devraient s'appuyer sur une approche participative positive, des différentes parties prenantes (de la société civile, du secteur privé, des universités et des médias) dans les processus décisionnels concernant le déploiement de systèmes d'IA dans le secteur public, en accordant une attention particulière à l'inclusion des individus et des groupes sous-représentés et vulnérables, ce qui est essentiel pour garantir la confiance en la technologie et son acceptation par toutes les parties prenantes.

116. L'utilisation de systèmes d'IA peut également avoir une influence négative, notamment en renforçant le désordre de l'information (par exemple par la diffusion de désinformation et de contenu trompeur, ainsi que par un comportement inauthentique coordonné) qui peut affecter les principes d'élections libres et équitables et interférer dans l'égalité des chances et la liberté des électeurs de se former une opinion. Il est essentiel de veiller à ce que les processus électoraux soient conformes aux normes du Conseil de l'Europe et aux autres normes internationales applicables.

117. Le recours aux systèmes d'IA peut accroître l'efficacité des institutions publiques, mais au prix d'une moindre transparence et d'une diminution de l'intervention et de la surveillance humaines. En outre, pour se procurer des systèmes d'IA et les déployer, les pouvoirs publics dépendent souvent d'acteurs privés, ce qui risque d'éroder davantage cette confiance, car ce fait suscite des interrogations portant sur l'obligation de rendre des comptes, l'indépendance de la surveillance et le contrôle public, qui pourraient être amplifiées par l'utilisation de systèmes d'IA opaques. Un cadre de gouvernance approprié devrait donc permettre aux concepteurs et aux déployeurs d'IA d'agir de manière responsable et conformément aux dispositions juridiques pertinentes, tout en laissant ouvertes les voies de recours et d'intervention appropriées des pouvoirs publics quand ce n'est pas le cas.

118. Dans les processus de passation de marchés publics liés à l'IA, il est indispensable d'inclure des critères tels que l'égalité, l'équité, l'obligation de rendre des comptes et la transparence,[168] et l'adoption de garanties à cette fin peut servir deux objectifs. Premièrement, elle veillerait à ce que les pouvoirs publics n'utilisent que des systèmes compatibles avec les droits de l'homme, la démocratie et l'État de droit. Deuxièmement, elle inciterait économiquement le secteur privé à développer et utiliser des systèmes conformes à ces critères. L'utilisation de systèmes d'IA dans les services publics étant tenue de satisfaire à des normes de transparence plus élevées, les autorités publiques ne devraient par conséquent pas acquérir de systèmes d'IA auprès de tiers qui ne respectent pas les obligations légales d'information en ce qui concerne leurs systèmes d'IA, ou qui ne sont pas disposés à lever les restrictions à l'échange d'informations (confidentialité ou secrets

167. Comme il a déjà été noté ci-dessus, si les secrets d'affaires et les droits de propriété intellectuelle doivent être respectés, ils doivent être mis en balance avec d'autres intérêts légitimes.
168. Cela a également été recommandé par les lignes directrices du Conseil de l'Europe sur l'IA et la protection des données (Section III), https://rm.coe.int/lignes-directrices-sur-l-intelligence-artificielle-et-la-protection-de/168091ff40.

industriels, par exemple) lorsque ces restrictions entravent le processus (i) d'évaluation d'impact sur les droits de l'homme (y compris la recherche/l'examen externes[169]) et (ii) la mise à disposition de ces évaluation au public.

Principaux droits matériels

- Le droit à la liberté d'expression, la liberté d'association et de réunion (art 10 et 11 CEDH).
- Le droit de vote et d'éligibilité, le droit à des élections libres et équitables, et en particulier le suffrage universel, égal et libre, y compris l'égalité des chances et la liberté des électeurs de se forger une opinion. À cet égard, les individus ne doivent être soumis à aucune tromperie ou manipulation.
- Le droit à l'information (diversifiée), au libre discours et à l'accès à la pluralité des idées et des perspectives.
- Le droit à une bonne administration.

Principales obligations

- Les États membres devraient prendre des mesures adéquates pour lutter contre l'utilisation ou l'abus de systèmes d'IA ayant pour objectif de s'ingérer dans les processus électoraux, effectuer un ciblage politique personnalisé sans mécanismes de transparence, et de responsabilité adéquats et permettant de rendre des comptes, ou plus généralement pour modeler les comportements politiques des électeurs ou manipuler l'opinion publique d'une manière qui peut entraver les droits de l'homme, la démocratie et l'État de droit.
- Les États membres devraient adopter des stratégies et mettre en place des mesures pour lutter contre la désinformation et identifier les discours haineux en ligne afin de garantir une pluralité informationnelle équitable.
- Les États membres devraient soumettre les passations de marchés publics faites par des systèmes d'IA à des mécanismes de contrôle adéquats :
  - Les États membres devraient soumettre leurs procédures de passation de marchés publics à des exigences juridiquement contraignantes qui garantissent l'utilisation responsable de l'IA dans le secteur public en garantissant le respect des principes susmentionnés, notamment la transparence, l'équité, la responsabilité et l'obligation de rendre des comptes.
- Les États membres devraient soumettre l'utilisation des systèmes d'IA dans le secteur public à des mécanismes de contrôle adéquats :
  - Cela peut également inclure l'octroi de recours aux médiateurs et aux tribunaux.
  - Les États membres devraient également assurer le contrôle de l'utilisation des systèmes d'IA dans les différentes organisations du secteur public et intervenir et se coordonner le cas échéant pour garantir leur conformité avec les droits de l'homme, la démocratie et l'État de droit.
  - Les États membres devraient veiller à ce que, lorsque le secteur public utilise des systèmes d'IA, cela se fasse avec la participation de personnes possédant les compétences appropriées dans un large éventail de domaines, y compris l'administration publique et les sciences politiques, afin de garantir une compréhension approfondie des implications potentielles pour la gouvernance de l'État administratif et la relation entre le citoyen et l'État.
- Les États membres devraient rendre publiques et accessibles toutes les informations pertinentes sur les systèmes d'IA (y compris leur fonctionnement, leur manœuvre d'optimisation, la logique sous-jacente, le type de données utilisées) qui sont utilisés dans le cadre de la fourniture de services publics, tout en préservant les intérêts légitimes tels que la sécurité publique.
- Les États membres devraient mettre en place des mesures visant à accroître la culture et les compétences numériques dans tous les segments de la population. Leurs programmes d'enseignement devraient s'adapter pour promouvoir une culture d'innovations responsables qui respecte les droits de l'homme, la démocratie et l'État de droit.
- Les États membres devraient encourager l'utilisation de solutions d'IA et d'autres outils qui peuvent :
  - renforcer l'autonomie informationnelle des citoyens, améliorer la manière dont ils collectent des informations sur les processus politiques et les aider à y participer ;

169. Avec un accent particulier sur l'impact du système sur les communautés marginalisées.

- aider à lutter contre la corruption et la criminalité économique, et renforcer la légitimité et le fonctionnement des institutions démocratiques. Cela peut contribuer à l'impact positif des systèmes d'IA dans la sphère démocratique et renforcer la confiance ;
- aider à la fourniture de services publics.
- Ce faisant, ils devraient toujours garantir le respect des droits de l'homme, de la démocratie et de l'État de droit.

7.1.9. État de droit

119. L'utilisation de systèmes d'IA peut accroître l'efficacité des systèmes judiciaires, mais peut également – comme indiqué au chapitre 3 – présenter des défis majeurs à l'État de droit. Selon la Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement[170], lorsque des outils d'IA sont utilisés pour trancher un litige ou en tant qu'outils d'aide à la décision judiciaire ou d'orientation du justiciable, ils ne doivent pas porter atteinte aux garanties du droit d'accès au juge et du procès équitable.

120. Ce qui signifie en particulier que les principes d'égalité des armes et de respect du contradictoire doivent être garantis. En outre, l'utilisation de systèmes d'IA ne doit mettre en danger ni l'indépendance ni l'impartialité des juges. Pour ce faire, la CEPEJ a souligné l'importance que revêtent la garantie de la qualité et de la sécurité des décisions juridictionnelles et des données judiciaires, ainsi que la transparence, l'impartialité et l'équité dans les méthodologies de traitement des données. De plus, des garanties d'accessibilité et d'explicabilité des méthodologies de traitement des données, comportant la possibilité d'audits externes, doivent de même être adoptées. Les États membres devraient par conséquent soumettre à des contrôles approfondis l'utilisation de systèmes d'IA au sein du système judiciaire, et assurer leur conformité à tous les principes susmentionnés.

121. En cas de différend juridique survenant dans le cadre de l'utilisation de systèmes d'IA – tant dans la sphère privée que publique –, les personnes formant un recours contre les effets négatifs d'un système d'IA devraient avoir accès, pour fonder leur requête, aux informations pertinentes en possession d'un défendeur ou d'un tiers qui sont nécessaire pour permettre un recours efficace. L'accès aux informations pertinentes par les parties dans une procédure judiciaire est également essentiel lorsque les systèmes d'IA ont été utilisés pour soutenir la prise de décision judiciaire, car cela représente une condition importante pour préserver l'égalité des armes entre les parties. Il peut s'agir, le cas échéant, des données relatives à la formation et aux contrôles, d'informations sur la manière dont le système d'IA a été utilisé, d'informations significatives et compréhensibles sur la manière dont le système d'IA est parvenu à une recommandation, une décision ou une prédiction, et des détails sur la manière dont les résultats du système d'IA ont été interprétés et mis à exécution. À cet égard, il convient de rechercher un juste équilibre entre divers intérêts légitimes des parties concernées, qui peuvent inclure des considérations de sécurité nationale dans le cas d'un système d'IA utilisé publiquement, par exemple, ainsi que des droits de propriété intellectuelle et autres, tout en assurant la pleine protection des droits de l'homme.. En outre, les personnes qui forment un recours contre des violations présumées des droits de l'homme dans le cadre des systèmes d'IA ne devraient pas être tenues de satisfaire à des exigences plus élevées en matière de preuve.[171]

Principaux droits matériels

- Le droit à un procès équitable et une procédure régulière (art 6 CEDH), y compris la possibilité de prendre connaissance d'une décision fondée sur l'IA et de la contester dans le cadre de l'application de la loi ou de la justice, y compris le droit à un réexamen de cette décision par une personne humaine.
- Le droit à l'indépendance et à l'impartialité judiciaires, le droit d'accès à un avocat.
- Le droit à un recours effectif (art 13 CEDH) également en cas de préjudice injuste ou de violations des droits de l'homme d'un justiciable dans le contexte des systèmes d'IA.

170. L'analyse de la CEPEJ porte sur les défis soulevés par l'utilisation des systèmes d'IA également dans le domaine de la résolution de litiges et l'application de la loi en ligne.

171. Rappelons également que les règles générales relatives au partage et au renversement de la charge de la preuve dans la législation anti-discrimination devraient en principe s'appliquer dans de tels cas.

Principales obligations

- Les États membres devraient veiller à ce que les systèmes d'IA utilisés dans le domaine de la justice et de l'application de la loi soient conformes à la règle fondamentale du droit à un procès équitable. À cette fin, ils devraient tenir dûment compte de la nécessité de garantir la qualité et la sécurité des décisions de justice et des données judiciaires, ainsi que la transparence, l'impartialité et l'équité en matière de méthodologies de traitement des données. Des garanties d'accessibilité et d'explicabilité des méthodologies de traitement des données, y compris la possibilité d'audits externes, devraient être adoptées dans ce but.
- Les États membres devraient veiller à ce que des voies de recours appropriées soient disponibles et que des mécanismes de réparation accessibles soient mis en place pour les personnes qui pourraient subir un préjudice résultant du développement ou de l'utilisation de systèmes d'IA dans des contextes liés à l'État de droit.
- Les États membres devraient informer de manière appropriée les justiciables sur l'utilisation de systèmes d'IA dans le secteur public chaque fois que cela peut avoir un impact significatif sur la vie des individus. Ces informations devraient être fournies en particulier lorsque des systèmes d'IA sont utilisés dans le domaine de la justice et de la répression, tant en ce qui concerne le rôle des systèmes d'IA dans le cadre de la procédure judiciaire que la possibilité de contester les décisions prises de cette façon.
- Les États membres devraient veiller à ce que l'utilisation des systèmes d'IA n'interfère pas avec le pouvoir décisionnel des juges ou l'indépendance judiciaire et à ce que toute décision judiciaire soit soumise à un contrôle humain.

7.2. Rôle et responsabilités des États membres et des acteurs privés en matière de développement d'applications d'IA conformes à ces exigences

122. Les systèmes d'IA peuvent avoir des répercussions sur les droits de l'homme, la démocratie et l'État de droit à raison de leur développement et de leur utilisation par des acteurs privés et publics. Comme indiqué au chapitre 5, outre la responsabilité qui leur incombe de protéger les droits de l'homme dans la sphère publique, les États membres ont également l'obligation positive de veiller au respect par les acteurs privés des normes en matière de droits de l'homme. En outre, plusieurs cadres internationaux précisent également que les acteurs privés doivent également respecter les droits de l'homme (tels que les Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme).

123. Les obligations qui incombent aux États membres d'assurer la conformité avecles normes du Conseil de l'Europe relatives aux droits de l'homme, à la démocratie et à l'État de droit dans le cadre des systèmes d'IA ont déjà été soulignées dans la section ci-dessus. Plus généralement, les autorités nationales devraient procéder à des évaluations fondées sur des données probantes de leur législation nationale pour vérifier sa conformité avec les droits de l'homme – et sa capacité à dûment les protéger – et adopter de nouvelles législations pour remédier aux lacunes juridiques potentielles. En outre, elles devraient mettre en place des mécanismes de contrôle et des recours judiciaires effectifs pour assurer une réparation chaque fois que le développement et l'utilisation de l'IA conduisent à des violations du droit. À cette fin, des autorités nationales de contrôle devraient être en mesure d'auditer et évaluer le fonctionnement des systèmes d'IA (publics ou privés), en particulier lorsqu'il existe des indications de non-conformité. Cette évaluation devrait compléter les obligations de contrôle existantes dans le cadre de la législation en vigueur, y compris la législation sur la protection des données (principe d'obligation de rendre des comptes, étude d'impact[172], consultation préalable des autorités de contrôle, etc.) afin d'accroître la transparence. Il peut y avoir des circonstances limitées où, en raison de préoccupations concernant la vie privée ou la propriété intellectuelle, le maintien d'un certain degré de confidentialité serait nécessaire.

124. Il convient de noter que de nombreux acteurs publics dépendent d'acteurs privés pour s'équiper en systèmes d'IA, obtenir les données nécessaires à leur déploiement et accéder à l'infrastructure sous-jacente sur laquelle repose le fonctionnement de ces systèmes. En conséquence, étant donné le rôle essentiel des acteurs privés en la matière, la responsabilité éminente leur incombe de veiller à ce que leurs systèmes soient développés et utilisés conformément aux principes, droits et conditions exposés ci-dessus. Les motivations

172. Comme précisé au chapitre 9. Voir également à cet égard l'étude de la FRA qui souligne la nécessité d'évaluations d'impact sur les droits de l'homme, «*Getting the Future Right - Artificial Intelligence and Fundamental Rights in the EU*», 14 décembre 2020, https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights.

des acteurs privés, d'une part, et celles des particuliers et de la société, d'autre part, n'étant pas toujours en phase, une structure juridique qui contraindrait les acteurs privés à respecter des droits et conditions spécifiques dans le cadre de l'IA pourrait y contribuer, surtout lorsque le risque que les acteurs privés et les intérêts individuels soient divergents existe. En outre, une telle structure garantirait l'accès à la justice si les acteurs privés en cas de manquement à ces obligations[173].

125. Comme indiqué ci-dessus, lorsque les États membres prennent des mesures pour sauvegarder les principes, droits et exigences énumérés dans le contexte de l'IA, une approche fondée sur les risques - complétée par une approche de précaution si nécessaire - est recommandée. Cette approche reconnaît que tous les systèmes d'IA ne présentent pas le même niveau de risque élevé et que les mesures réglementaires doivent en tenir compte. En outre, elle exige que les risques que présentent les systèmes d'IA pour les droits de l'homme, la démocratie et l'État de droit soient évalués de manière systématique et que des mesures d'atténuation soient spécifiquement adaptées à ces risques.

126. Lorsqu'ils mettent en œuvre une approche fondée sur les risques et évaluent le type d'intervention réglementaire nécessaire pour atténuer les risques, les États membres peuvent être guidés par un certain nombre de facteurs qui sont couramment utilisés dans les évaluations de l'impact des risques. Ces facteurs de risque comprennent, par exemple, l'étendue potentielle des effets négatifs sur les droits de l'homme, la démocratie et l'État de droit ; la probabilité ou la possibilité qu'un impact négatif se produise ; l'ampleur et l'omniprésence de cet impact ; sa portée géographique ; son extension temporelle ; et la mesure dans laquelle les effets négatifs potentiels sont réversibles. En outre, un certain nombre de facteurs spécifiques à l'IA qui peuvent influencer le niveau de risque (tels que le niveau d'automatisation de l'application, la technique d'IA sous-jacente, la disponibilité de mécanismes de test, le niveau d'opacité) peuvent également être pris en compte.

7.3. Responsabilité en cas de dommages causés par l'intelligence artificielle

127. Le développement et l'utilisation de systèmes d'IA soulèvent des questions inédites sur le plan de la sécurité et de la responsabilité. Les opinions diffèrent toutefois sur la question de savoir si les régimes de responsabilité existants devraient s'appliquer ou si des régimes spécifiques devraient être élaborés dans le contexte de l'IA. Il convient néanmoins de noter que l'utilisation généralisée des systèmes d'IA est à l'origine de problèmes épineux s'agissant de l'interprétation et de la mise en œuvre de la législation en vigueur en matière de responsabilité. La Convention du Conseil de l'Europe sur la responsabilité du fait des produits (STE n° 91), en ce qui concerne les dommages corporels et les décès, par exemple, ne s'applique qu'aux systèmes d'IA considérés comme des produits meubles (matériel informatique) et non comme des logiciels, et ne s'applique qu'aux systèmes d'IA mis en circulation en tant que produit et non en tant que service[174]. Par conséquent, il pourrait être souhaitable de préciser que les logiciels autonomes peuvent être qualifiés de produits au sens de la législation existante en matière de responsabilité du fait des produits. L'opacité de certains systèmes d'IA, associée à l'asymétrie des informations entre les développeurs et les producteurs d'IA, d'une part, et les personnes qui peuvent être affectées négativement par les systèmes d'IA, d'autre part, peut dans certains cas rendre difficile pour ces dernières de satisfaire au niveau de preuve requis pour étayer une demande de dommages et intérêts. Toutefois, en général, l'attribution actuelle de la charge de la preuve peut apporter des solutions appropriées et raisonnables en ce qui concerne les systèmes d'IA.

128. Si le Comité des Ministres décide de traiter la question de la responsabilité au sein d'un futur cadre juridique au niveau du Conseil de l'Europe, le CAHAI recommande que soient promus les principes suivants :

- Un régime de responsabilité approprié et équilibré dans chaque secteur se révèle important pour les consommateurs comme pour les fabricants et peut contribuer à asseoir une sécurité juridique.
- Il est essentiel de garantir le même niveau de protection aux personnes ayant subi un préjudice du fait des systèmes d'IA qu'à celles ayant subi un préjudice causé par des technologies traditionnelles.
- Tout dommage injuste engage la responsabilité du système d'IA envisagé dans la totalité de son cycle de vie.

173. C. Muller, p. 16 ; cela signifie qu'il convient d'aller au-delà de la simple référence à 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 aux Principes directeurs des Nations unies relatifs aux entreprises et aux droits de l'homme).

174. Il en va de même pour la directive européenne en matière de responsabilité du fait des produits défectueux (Directive 85/374/CEE du Conseil). C'est l'une des raisons pour lesquelles le Livre blanc de la Commission européenne sur l'IA insiste sur la nécessité d'aborder les questions relatives à l'IA et à la responsabilité

- Il peut être nécessaire d'établir une distinction en ce qui concerne l'attribution de la responsabilité dans les contextes des relations d'entreprise à consommateur et d'entreprise à entreprise. La responsabilité des agents commerciaux pourrait, par exemple, être mieux traitée par des stipulations contractuelles plutôt que par l'adoption d'un régime de responsabilité spécifique.
- La question de la responsabilité transfrontalière doit être prise en compte. Cela est particulièrement pertinent lorsque, par exemple, une entreprise utilisant un système d'IA est enregistrée dans un État, le développeur de ce système dans un autre État et qu'un utilisateur qui subit un préjudice réside habituellement dans un État tiers.
- Les règles de responsabilité peuvent être complétées, dans certaines applications sectorielles spécifiques, par des codes de conduite éthiques (volontaires) de l'industrie qui serviraient à renforcer la confiance du public dans les domaines sensibles de l'IA.
- La mesure dans laquelle les acteurs privés assurent et investissent dans des mécanismes de due diligence peut être un facteur pertinent lorsqu'on examine la responsabilité des acteurs privés et la charge de la preuve.[175]

8. OPTIONS ENVISAGEABLES POUR L'ÉTABLISSEMENT D'UN CADRE JURIDIQUE DU CONSEIL DE L'EUROPE POUR LE DÉVELOPPEMENT ET L'APPLICATION DE L'INTELLIGENCE ARTIFICIELLE, FONDÉES SUR LES DROITS DE L'HOMME, LA DÉMOCRATIE ET DE L'ÉTAT DE DROIT.

129. Afin de combler les lacunes en matière de protection juridique constatées au chapitre 5, différentes options sont envisageables pour l'établissement d'un cadre juridique existent au sein du Conseil de l'Europe, y compris des instruments juridiques contraignants et non contraignants. Elles sont brièvement décrites, et leurs avantages et inconvénients soulignés, dans les paragraphes suivants. Alors que le chapitre précédent portait sur la *teneur* du cadre juridique, le présent chapitre se concentre sur sa *forme*.

8.1. Modernisation des instruments juridiques contraignants en vigueur

130. Un premier scénario envisageable consisterait à modifier les instruments juridiques contraignants en vigueur, pour les compléter et/ou les adapter en fonction des particularités des systèmes d'IA.

131. Un protocole additionnel à la CEDH pourrait être adopté pour consacrer de nouveaux droits de l'homme ou adapter les droits de l'homme reconnus dans le contexte des systèmes d'IA[176] . Ceux-ci pourraient être tirées du chapitre 7 ci-dessus.[177] Il n'est pas improbable que, eu égard à l'interprétation dynamique et évolutive adoptée par la Cour européenne des droits de l'homme, les droits en vigueur énoncés par la Convention européenne des droits de l'homme, tels que le droit au respect de la vie privée, de la liberté de pensée, de la liberté d'expression et le droit à la non-discrimination, puissent être interprétés de sorte à inclure les droits susmentionnés. Toutefois, l'intérêt que présenterait l'adoption d'un protocole additionnel serait que la reconnaissance de certains droits dans le contexte de l'IA ne dépende pas d'une décision de la Cour européenne des droits de l'homme, ce qui assurerait par conséquent davantage de clarté et de sécurité juridique (et éviterait que la Cour soit éventuellement critiquée pour interpréter les droits prévus par la Convention de manière trop extensive). Mais si l'adoption d'un protocole additionnel affirmerait, le plus vigoureusement possible, l'engagement des États membres à protéger les droits fondamentaux essentiels des citoyens, l'État de droit et la démocratie, ce ne serait pas un instrument approprié pour prescrire des exigences ou des obligations spécifiques dans le contexte des systèmes d'IA.[178] Il convient également de noter que les protocoles additionnels à la Convention européenne des droits de l'homme ne lient que les États les ayant ratifié, ce qui peut aboutir à ce que seuls certains États membres soient liés et que la Cour européenne des droits de l'homme

175. Compte tenu des travaux en cours dans l'Union européenne concernant un éventuel système communautaire obligatoire de due diligence pour les entreprises, il pourrait être utile de veiller à l'alignement sur ce système si ce point devait être envisagé. Voir l'étude de la Commission européenne sur les exigences de due diligence à travers la chaîne d'approvisionnement de janvier 2020, disponible à l'adresse suivante (uniquement en anglais) : https://op.europa.eu/en/publication-detail/-/publication/8ba0a8fd-4c83-11ea-b8b7-01aa75ed71a1/language-en
176. Cela se ferait en étroite collaboration avec les comités directeurs concernés et en particulier le Comité directeur pour les droits de l'homme (CDDH),
177. Voir CAHAI (2020)06-fin, paragraphe 77.
178. Tels que ceux mentionnés au chapitre 7 de la présente étude de faisabilité.

n'exerce qu'un contrôle fragmentaire. En outre, la Cour européenne des droits de l'homme est déjà submergée par sa charge de travail actuelle et ne devrait donc pas être chargée de questions supplémentaires, dont la décision nécessite des connaissances techniques qui n'y sont pas nécessairement disponibles.

132. La modernisation des instruments verticaux en vigueur, tels que la Convention de Budapest sur la cybercriminalité (STE n° 185) ou la « Convention 108+ », pourrait constituer un autre scénario plausible. Comparée à l'élaboration d'une nouvelle convention (voir ci-dessous), une telle approche aurait pour atout considérable de pouvoir mobiliser les réseaux existants de contrôle et de mise en œuvre (comme dans le cas de la Convention 108+, les autorités nationales indépendantes chargées de la protection des données, dont le champ d'interventions réglementaires pourrait être étendu à l'intelligence artificielle). Mais l'inconvénient de cette approche, outre la longueur et la complexité de l'adoption, réside dans la portée limitée de chacun des instruments verticaux ou sectoriels en vigueur, ce qui nécessiterait de réaliser de multiples interventions pour répondre aux diverses préoccupations exposées dans les chapitres précédents. La modernisation de la « Convention 108+ », par exemple, pourrait ne pas tenir compte de toutes les préoccupations liées aux systèmes d'IA, étant donné la place privilégiée (actuelle) qu'elle accorde à la protection des personnes à l'égard du traitement des données à caractère personnel ; dans le même temps, il convient de noter que bon nombre des principes de haut niveau identifiés jusqu'à présent pour relever les défis posés par les systèmes d'IA (par exemple l'obligation de rende des comptes, la transparence, les décisions automatisées) sont, dans une certaine mesure, déjà inclus dans la Convention 108+.[179] En outre, la « Convention 108+ » ayant été ouverte à la ratification en 2018, il pourrait se révéler difficile de la moderniser de nouveau à court terme[180].

133. On pourrait répondre aux deux préoccupations exprimées à l'occasion de chacune des options en combinant les deux idées, à savoir celle d'un protocole additionnel à la Convention européenne des droits de l'homme et celle de la modernisation de (certains) instruments verticaux comme la « Convention 108+ ». Alors que le premier prescrirait les principes et valeurs de haut niveau, la seconde pourrait préciser davantage les obligations des États et établir un réseau efficace d'autorités indépendantes compétentes pour assurer la mise en œuvre effective de ces garanties. Ces autorités pourraient traiter des actes ou omissions des États concernant les systèmes d'IA et, dans certaines circonstances, engager la responsabilité de l'État au titre de la Convention. La longue durée d'un processus combiné demeurerait toutefois problématique relativement au rythme accéléré de la mise en place des systèmes d'IA.

8.2. Adoption d'un nouvel instrument juridique contraignant : convention ou convention-cadre

134. Un deuxième scénario envisageable serait l'adoption d'un nouvel instrument juridique contraignant qui pourrait revêtir la forme d'une convention ou d'une convention-cadre. Il convient de noter que la convention et la convention-cadre sont toutes deux des traités multilatéraux, qu'elles sont de même nature juridique et toutes deux soumises aux règles usuelles s'appliquant aux traités internationaux tels que définies par la Convention de Vienne sur le droit des traités (1969). De plus, toutes deux peuvent être assorties d'un système de gouvernance (pour plus de détails, voir le chapitre 9.4) et peuvent être complétées par des protocoles additionnels. Ce qui les différencie, c'est le fait que la convention tend à réglementer une question ou un domaine spécifique de manière plus concrète, généralement en énonçant certains droits et obligations, tandis que la convention-cadre tend à établir des principes plus généraux et des domaines d'action plus larges qui ont été convenus entre les États Parties.

135. Une convention-cadre ne prévoit ordinairement qu'une obligation générale incombant aux États Parties d'entreprendre certaines actions, d'atteindre certains objectifs ou de reconnaître certains droits, sans directement attribuer ces droits à des personnes physiques ou morales. Par conséquent, la ratification par un État d'une convention-cadre serait insuffisante pour que des personnes physiques et morales puissent en inférer certains droits, des mesures législatives nationales supplémentaires devant être prises à ce titre. Les États disposent de ce fait d'une marge d'appréciation considérable s'agissant de la mise en œuvre des principes et objectifs plus généraux.

136. Une convention pourrait réglementer de manière plus exhaustive la conception, le développement et l'application de systèmes d'IA ou de prise de décision algorithmique, en s'inscrivant dans le prolongement

179. Il est nécessaire de prendre en compte les développements réglementaires dans d'autres forums internationaux, tels que l'UE, car les limites du règlement général de l'UE sur la protection des données (qui sont parallèles aux limites de la Convention 108+) dans le contexte des systèmes d'IA, ont conduit l'UE à proposer une nouvelle réglementation dans ce domaine. Une proposition de règlement est attendue au premier trimestre 2021.

180. Il convient de noter à cet égard que l'entrée en vigueur des amendements ou d'un protocole d'amendement nécessite normalement l'acceptation/la ratification par toutes les Parties à la Convention, ce qui est un long processus.

de l'étude de faisabilité et de la recommandation CM/Rec(2020)1.[181] Elle énumérerait certains droits et obligations susceptibles de contribuer à la sauvegarde et la protection des droits de l'homme, de la démocratie et de l'État de droit dans le cadre de systèmes d'IA, offrant de ce fait une protection juridique aux personnes physiques et morales dès son entrée en vigueur. Elle soulignerait l'importance d'une adhésion rapide du plus grand nombre de Parties afin de faciliter l'élaboration d'un régime juridique exhaustif en matière de systèmes d'IA tels que spécifiés dans la convention, et demanderait instamment aux États membres et autres Parties à la convention d'engager le processus prévu par leur législation nationale conduisant à la ratification, l'approbation ou l'acceptation de la convention. Il convient de noter que l'Assemblée Parlementaire du Conseil de l'Europe (APCE) a proposé en octobre 2020 que le « Comité des Ministres soutienne l'élaboration "d'un instrument juridiquement contraignant" gouvernant l'IA, éventuellement sous la forme d'une Convention ». L'APCE a en outre recommandé que le Comité des Ministres veille à ce que « cet instrument juridiquement contraignant soit fondé sur une approche globale, porte sur l'ensemble du cycle de vie des systèmes fondés sur l'intelligence artificielle, s'adresse à toutes les parties prenantes et comprenne des mécanismes visant à assurer » sa mise en œuvre.[182]

137. Sa valeur ajoutée résiderait dans la formation d'un instrument juridiquement contraignant concernant spécifiquement la conception, le développement et l'application de l'IA, fondé sur les normes du Conseil de l'Europe en matière de droits de l'homme, d'État de droit et de démocratie. Elle harmoniserait les règles et les obligations entre les États et renforcerait ainsi la confiance dans les produits et services transfrontaliers d'IA, à la lumière d'un accord sur la manière dont les systèmes d'IA devraient être conçus, développés et appliqués. La Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (STE n° 108) et la Convention de Budapest sur la cybercriminalité (STCE n° 185) sont deux exemples probants de cadres juridiques innovants élaborés dans des domaines connexes par le Conseil de l'Europe dans le passé.

138. Parallèlement, la rédaction d'un projet de convention prévoyant des obligations juridiques détaillées concernant les systèmes d'IA pourrait être considérée comme prématurée. Une approche excessivement normative et rigide pourrait entraîner un rejet de l'instrument et un manque de volonté de la part des États de signer, d'exprimer leur consentement à être liés par la convention et de la mettre effectivement en œuvre dans la pratique. À l'inverse, une large approbation d'une convention contenant des règles trop rigides pourrait entraver l'innovation et réduire la recherche sur le développement et le déploiement de nouvelles technologies et de solutions de pointe aux problèmes existants, dont beaucoup pourraient sauver des vies et profiter à la société dans son ensemble.

139. Toutefois, il est important de noter qu'un ensemble concret de règles internationales contraignantes apporterait une sécurité juridique aux acteurs étatiques et privés, tout en protégeant fortement les droits des individus et en établissant des principes clairs de responsabilité entre les acteurs impliqués dans l'utilisation des systèmes d'IA. En outre, il serait possible de répondre à cette dernière préoccupation en veillant à ce que les droits et obligations énoncés dans la convention ne soient ni trop prescriptives ni trop détaillées. On pourrait également y répondre en adoptant une convention-cadre sur l'IA, qui permettrait de consacrer les principes et valeurs fondamentaux devant être respectés en les consignant dans un instrument contraignant portant sur la conception, le développement et l'application de l'IA, conformément aux normes du Conseil de l'Europe en matière de droits de l'homme, d'État de droit et de démocratie, et qui dès lors laisserait une large marge d'appréciation aux États parties dans leur mise en œuvre respective. En application de l'approche dite de la « convention-cadre et de son protocole », les parties peuvent convenir d'un traité plus général – la convention-cadre – et de protocoles plus détaillés ou d'autres instruments pour édicter des dispositions spécifiques. Cette technique réglementaire, qui présente un certain nombre d'avantages par rapport aux seuls traités « fragmentaires » en droit international, pourrait être particulièrement appropriée dans le domaine de l'IA. Dans ce contexte, il convient d'examiner attentivement si une telle structure basée sur un traité plus général et l'élaboration éventuelle d'instruments spécifiques supplémentaires (tels que des protocoles) accroîtrait la complexité du cadre juridique résultant et rendrait le respect des dispositions plus difficile.[183]

---

181. Conseil de l'Europe, Comité des Ministres, Recommandation CM/Rec(2020)1 du Comité des Ministres aux Etats membres sur les impacts des systèmes algorithmiques sur les droits de l'homme, avril 2020.
182. Voir la recommandation 2181(2020) et la résolution 2341(2020), « Nécessité d'une gouvernance démocratique de l'intelligence artificielle », https://pace.coe.int/fr/news/8059/establishing-a-legally-binding-instrument-for-democratic-governance-of-ai.
183. À cet égard, il convient de considérer que tout autre instrument contraignant du Conseil de l'Europe, tel qu'un protocole à une convention, devra également être ratifié.

140. Une convention-cadre imposerait aux États qu'ils s'accordent sur la portée du cadre juridique et de la procédure à respecter pour offrir des garanties efficaces et fondées sur les normes du Conseil de l'Europe dans la conception, le développement et l'application des systèmes d'IA. Elle comprendrait les règles et principes fondamentaux acceptés d'un accord commun en matière de recherche, développement et mise en œuvre de l'IA, dans l'intérêt de la société humaine. Elle pourrait également contenir des dispositions spécifiques portant sur les garanties procédurales, les mesures préventives, les juridictions, la coopération internationale. Par exemple, un modèle d'accord pour l'échange d'informations pourrait être élaboré, ou pour la mobilisation d'un réseau d'autorités compétentes indépendantes déjà existantes, comme celles qui se consacrent à la protection des données ou au contrôle de la concurrence au niveau national. La convention-cadre pourrait également énoncer les règles et procédures nécessaires aux États pour mettre en œuvre la convention.

141. La Convention-cadre pour la protection des minorités nationales (FCMN) est un exemple de convention-cadre de ce type existant au niveau du Conseil de l'Europe. C'est un instrument de droit international juridiquement contraignant qui prévoit un système de suivi, mais le mot « cadre » souligne la possibilité pour les États membres de transposer les dispositions de la convention en fonction de leur situation spécifique par le biais de lois et de mesures nationales appropriées. Un autre exemple, bien que son titre officiel ne comporte pas le terme « cadre », est la Convention pour la protection des droits de l'homme et de la dignité de l'être humain à l'égard des applications de la biologie et de la médecine, titre résumé en Convention sur les droits de l'homme et la biomédecine (dite « Convention d'Oviedo »). Cette convention a été adoptée en 1997 pour répondre aux utilisations abusives potentielles des progrès scientifiques dans les domaines de la biologie et de la médecine. Elle s'inspire des principes affirmés par la Convention européenne des droits de l'homme et vise à protéger la dignité et l'identité de tous les êtres humains. Elle énonce des principes fondamentaux applicables à la pratique médicale quotidienne et concerne plus spécifiquement la recherche biomédicale, la génétique et la transplantation d'organes et de tissus. Elle est explicitée et complétée par des protocoles additionnels portant sur des sujets spécifiques, comme par exemple l'interdiction du clonage d'êtres humains.

142. Au-delà du choix d'une convention ou d'une convention-cadre, il convient de noter que les destinataires sont les États et des organismes étatiques, en particulier les États membres du Conseil de l'Europe. Toutefois, l'adhésion peut être ouverte à d'autres signataires que les États membres, ce qui peut contribuer significativement à la portée et à l'efficacité globales de l'instrument[184]. La force des traités tient à leur caractère formel et au fait qu'ils sont juridiquement contraignants pour les États qui les ont acceptés. En devenant partie à une convention, un État contracte des obligations légales qui sont exécutoires au titre du droit international.

143. Ce qui précède n'exclut pas le rôle essentiel que doivent jouer les acteurs privés dans la mise en œuvre de réglementations spécifiques appliquées à l'échelon national sur la base d'engagements internationaux de portée générale. En particulier, ils sont susceptibles de jouer un rôle de tout premier plan dans la conception de mécanismes de corégulation par lesquels les États, en interaction étroite avec les acteurs privés, approfondiraient leurs engagements internationaux.

144. Bien qu'il ne soit pas possible de tirer quelque conclusion générale à propos de la rapidité de préparation et d'entrée en vigueur d'un traité (les délais peuvent varier de quelques mois à plusieurs années, en fonction de la nature et de la complexité du sujet, mais aussi de la volonté politique des États membres), un inconvénient potentiel de cette option réside dans le processus d'entrée en vigueur d'un traité international. Il n'existe aucune obligation légale pour les États membres de ratifier, d'approuver ou d'accepter une nouvelle convention même s'ils ont voté en faveur de son adoption ou l'ont signée, et il n'existe aucun moyen d'obliger les États membres à accélérer leurs procédures internes de ratification, d'acceptation ou d'approbation. Il n'y a donc aucun moyen de garantir qu'un ou tous les États membres exprimeront leur consentement à être liés par une convention. En outre, sans une plus large adhésion des acteurs internationaux (par exemple, les États observateurs), même l'option d'une convention ou d'une convention-cadre risquerait de créer des divergences et, partant, la fragmentation des approches réglementaires internationales.

184. En général, les États non-membres du Conseil de l'Europe peuvent adhérer à une convention du CdE sur invitation du Comité des ministres et après avoir obtenu l'accord unanime des Parties à la convention. Certaines conventions du Conseil de l'Europe sont devenues des normes mondiales : la « Convention 108 » compte 55 États Parties, la « Convention de Budapest » compte 64 États Parties.

8.3. Instruments juridiques non contraignants

8.3.1. Conseil de l'Europe

145. Il convient d'opérer une distinction entre les instruments juridiques non contraignants (ou *soft law,* « droit souple ») d'une part au niveau du Conseil de l'Europe et, d'autre, part au niveau national. Les premiers sont déjà utilisés dans plusieurs secteurs (cf. chapitre 4), mais pourraient être complétés par un instrument général de droit souple, tel qu'une recommandation ou une déclaration, qui consolide les principes communs. Un tel instrument pourrait fonctionner comme un document autonome ou compléter un instrument contraignant pour rendre ses dispositions plus opérationnelles. D'autres options comprennent l'élaboration de documents d'orientation pour mieux faire comprendre la relation entre la protection des droits de l'homme, la démocratie et l'État de droit, et l'IA (par exemple en fournissant des informations sur la jurisprudence de la Cour européenne des droits de l'homme), et contribuer ainsi à renforcer la protection au niveau national. Des « manuels » ou « guides » de ce type pourraient être élaborés dans le cadre d'une vaste consultation multipartite avec les gouvernements, les entreprises privées, les organisations de la société civile et les représentants de la communauté technique et du monde universitaire. Ces documents devraient être évolutifs, mis à jour périodiquement et complétés de façon collaborative à la lumière des nouveaux développements. Parmi les précédents, l'on peut citer le Manuel sur les droits de l'homme et l'environnement et le Guide des droits de l'homme pour les utilisateurs d'internet.

8.3.2. Au niveau des États membres

146. Un cadre juridique instauré par le Conseil de l'Europe pourrait encourager l'approbation par les autorités nationales compétente de mécanismes de droit souple, afin de traduire ses dispositions en mesures opérationnelles plus concrètes et mettre en évidence sa conformité. Ces instruments de droit souple pourraient consister en documents approuvés, lignes directrices, codes de conduite, dispositifs d'attribution de labels, marques et sceaux, ainsi que mécanismes de certification. S'il est vrai que, précisément en raison de leur caractère non contraignant, les mesures de droit souple ne peuvent satisfaire l'objectif de garantir que les applications d'IA protègent et soutiennent les droits de l'homme, la démocratie et l'État de droit, elles peuvent toutefois y apporter d'importantes contributions. Une approche privilégiant le droit souple a notamment pour avantages la flexibilité, l'adaptabilité, l'immédiateté de la mise en œuvre, une force d'attraction plus large et la possibilité d'être révisée et modifiée rapidement.

147. Des acteurs privés, y compris les institutions universitaires et les organes de standardisation, peuvent contribuer à garantir que ces instruments de droit souple produisent en pratique des effets. Les organisations qui développent et déploient l'IA seraient susceptibles d'intégrer un instrument juridique de droit souple dans leurs structures de gouvernance, processus de passation de marchés, fonctionnement et pratiques en matière d'audit (comme elles le font déjà avec de nombreuses normes et certifications se rapportant à la sécurité, par exemple). De plus, des agences de notation pourraient potentiellement également jouer un rôle, par exemple en établissant un classement annuel, fondé sur des données tangibles, des organisations privées se conformant aux prescriptions de droit souple.

148. Il convient toutefois de souligner que, si l'autorégulation peut constituer une méthode complémentaire de mise en œuvre de certains principes et règles, elle ne saurait se substituer aux obligations positives qui incombent aux États membres au titre de la Convention européenne des droits de l'homme, de protéger et préserver de manière effective les droits de l'homme, la démocratie et l'État de droit en ce qui concerne l'IA. Les approches volontaires, autorégulatrices et fondées sur l'éthique manquent de mécanismes efficaces d'exécution, de responsabilité et assurant l'obligation de rendre des comptes, et il conviendrait par conséquent qu'elles ne soient pas considérées comme constituant à elles seules un moyen suffisant et efficace de réglementer l'IA qui a un impact sur les droits de l'homme, la démocratie ou l'État de droit. De surcroît, les mécanismes de certification ne sont pas à l'abri d'erreurs et de méprises. D'où la nécessité de réunir un certain nombre de conditions si l'on veut que ces mécanismes soient efficaces.

8.4. Autres formes de soutien apporté aux États membres, tel que l'identification des bonnes pratiques

149. Les façons d'identifier ou d'encourager les bonnes pratiques sont nombreuses (maintes d'entre elles étant déjà connues ou mises en œuvre par les États membres ou les acteurs économiques). Par exemple, parmis plusieurs options, un Institut européen d'analyse comparative pourrait constituer une source d'identification, de définition et de consensus très efficace, efficiente et digne de confiance portant sur les éléments

d'appréciation disponibles qui devraient orienter des bonnes pratiques éprouvées.[185] Ces éléments d'appréciation disponibles sont à leur tour susceptibles de servir de socle à un large éventail de bonnes pratiques qui peuvent être relayées de manière efficace et effective par des normes techniques et des certifications éprouvées. La valeur ajoutée d'un tel institut par rapport à d'autres organismes de normalisation tels que l'ISO et la CEI devrait néanmoins être soigneusement examinée. La coopération avec les organismes de normalisation peut être encouragée de manière plus générale.

150. Par ailleurs, un modèle uniforme ou un outil élaboré au niveau du Conseil de l'Europe en vue de la réalisation d'une étude d'impact relative aux droits de l'homme, à la démocratie et à l'État de droit pourrait se révéler extrêmement utile sur le plan de l'harmonisation de la mise en œuvre par les États membres de normes juridiques et de valeurs communes en matière de systèmes d'IA. En ce qui concerne les mécanismes pratiques qui pourraient contribuer à la fois à la mise en œuvre et à l'application du cadre juridique, nous renvoyons au chapitre 9 de la présente étude de faisabilité.

8.5. Éventuelle complémentarité entre les éléments horizontaux et transversaux susceptibles de faire partie d'un instrument de type convention, et le travail vertical et sectoriel qui pourrait donner lieu à des instruments spécifiques de nature différente.

151. Le travail sectoriel du Conseil de l'Europe sur les systèmes d'IA, qui devrait se développer dans les années à venir et qui alimente, de manière complémentaire, la dimension horizontale et transversale des travaux du CAHAI, a été décrit au chapitre 4. Les éléments horizontaux susceptibles de faire partie d'un instrument de type convention permettraient d'affiner un travail sectoriel et de donner une impulsion à l'élaboration d'instruments spécifiques dans les domaines où progresse l'analyse de l'impact des systèmes d'IA et des mesures correctrices politiques nécessaires. Un potentiel instrument juridiquement contraignant horizontal pourrait prévoir des références explicites aux instruments en vigueur ou à élaborer au sein des différents piliers du Conseil de l'Europe.

152. La création d'un dispositif/organisme de certification commun (volontaire ou obligatoire), comparable à celui en vigueur dans le secteur pharmaceutique (la Direction européenne de la qualité du médicaments et soins de santé (EDQM) et sa Pharmacopée), pourrait potentiellement être considéré comme un autre mécanisme visant à assurer la complémentarité. Ce dispositif/organisme de certification commun pourrait, par exemple, être chargé d'élaborer des lignes directrices plus détaillées concernant les études d'impact sur les droits de l'homme la démocratie et l'Etat de droit et les normes de qualité communes au niveau européen. De plus, il pourrait avoir pour mission d'épauler la mise en œuvre et le contrôle de l'application des normes de qualité pour les systèmes d'IA (qui adhèrent volontairement ou obligatoirement au système de certification), tout comme le fait l'EDQM en matière de sûreté des médicaments et de leur utilisation.

153. Pour conclure ce chapitre, étant donné la nature en constante évolution de l'IA et les défis qu'elle pose, un cadre juridique solide sera probablement composé d'une combinaison d'instruments juridiques contraignants et non contraignants, se complétant les uns les autres. Un instrument contraignant, convention ou convention-cadre, de caractère horizontal, pourrait consolider des principes généraux communs contextualisés pour s'appliquer aux risques inhérents à l'IA et prévoir des dispositions plus concrètes pour sauvegarder les droits, principes et obligations identifiés au chapitre 7. Elle pourrait servir de base à la législation nationale pertinente dans ce domaine, de manière harmonisée, et peut favoriser les meilleures pratiques pour la réglementation de l'IA de manière plus générale. Un tel instrument, qui pourrait inclure des mécanismes et des processus de suivi appropriés, pourrait être combiné à des instruments sectoriels supplémentaires du Conseil de l'Europe (contraignants ou non contraignants), établissant d'autres principes sectoriels spécifiques et prévoyant des dispositions détaillées sur la façon de faire face aux défis sectoriels spécifiques à l'IA. Cette combinaison fournirait le niveau d'orientation nécessaire aux acteurs privés souhaitant prendre des initiatives d'autorégulation.

154. Une telle approche permettrait également la flexibilité requise pour le développement technologique, car les révisions des instruments verticaux pourraient être entreprises avec relativement moins de formalités et de complexité.

185. On peut, par exemple, se référer aux travaux entrepris dans ce domaine par le *National Institute of Standards and Technology* (NIST) des États-Unis, par exemple à l'adresse suivante : https://www.nist.gov/speech-testimony/facial-recognition-technology-frt-0.

9. MÉCANISMES PRATIQUES ET DE SUIVIS POSSIBLES POUR ASSURER LA CONFORMITÉ ET L'EFFICACITÉ DU CADRE JURIDIQUE

9.1. Le rôle des mécanismes de conformité

155. L'efficacité recherchée d'un cadre juridique dépend de l'étendue de son adoption et du respect de ses dispositions. Des mécanismes concrets (tels qu'études d'impact, audit du cycle de vie, méthodes de suivi et de certification, bacs à sable) offrent un moyen d'assurer la conformité, en aidant les États membres à comprendre et contrôler le respect du cadre juridique. De tels mécanismes confèrent des avantages supplémentaires, en renforçant la transparence qui doit prévaloir en matière d'utilisation de l'IA et en créant un cadre commun propice au développement de la confiance, par exemple.

156. Tout cadre juridique du Conseil de l'Europe devrait énoncer comme condition requise théorique l'élaboration de mécanismes de conformité, à un niveau général, et identifier les principes devant être respectés en pratique par tout mécanisme ayant pour objet de garantir la conformité. Il appartient aux États membres de décider comment mettre à exécution ces principes au travers de leur législation, y compris quels mécanismes pratiques ils choisissent de rendre obligatoires, ou à quels acteurs ou institutions ils donnent compétence pour assurer un contrôle indépendant, qualifié et efficace. Cela permet une mise en œuvre qui tienne compte, au niveau local, du rôle des institutions existantes, de la culture réglementaire et des règles juridiques. De préférence à la prescription d'une solution unique, cette voie d'approche se prête également à la création d'un écosystème d'assurance de l'IA constituant un environnement potentiellement favorable à une participation diversifiée et à l'émergence d'approches inédites et innovantes en matière de conformité. Cela dit, il convient de considérer la collaboration entre les États membres comme essentielle pour prévenir le risque d'adoption d'approches divergentes et de fragmentation des marchés.

157. Des mécanismes de conformité sont susceptibles d'être utilisés pour évaluer la conception de systèmes reposant sur l'IA, de même que leurs processus opérationnels, leur mise en œuvre contextuelle et leurs exemples d'utilisation. Sur la question de savoir à quel moment les systèmes d'IA devraient être soumis à une telle évaluation, le CAHAI a convenu de l'importance capitale d'une évaluation *ex ante* et d'une évaluation continue effectuée à différentes étapes du cycle de vie du projet d'IA, y compris après le déploiement et l'utilisation initiaux. Les mécanismes de conformité devraient également évoluer dans le temps pour tenir compte de la nature évolutive du système. Pour que les analyses d'impact puissent être utilisées efficacement, il convient de veiller tout particulièrement à ce qu'elles soient compréhensibles et accessibles à tous les acteurs concernés. Les garanties juridiques doivent veiller à ce que les mécanismes de conformité ne soient pas utilisés par les organisations pour se protéger contre d'éventuelles actions en responsabilité liées à leur conduite.

158. La modalité de l'évaluation continue offre trois avantages essentiels. Premièrement, elle permet une meilleure compréhension des implications du projet d'IA (dans ses phases de conception, de développement et de déploiement). Deuxièmement, elle facilite la prise de décision en vue de réexaminer de futures utilisations imprévues du projet d'IA. Troisièmement, elle surveille en permanence les changements qui interviennent *ex post* dans le comportement du modèle (fonction plus particulièrement décisive dans les contextes d'apprentissage par renforcement et d'apprentissage dynamique, par exemple). L'acquisition de solutions d'IA préfabriquées et les développements techniques tels que les scénarios d'apprentissage par transfert présentent des difficultés qui doivent être envisagées.

9.2. Le rôle des différents acteurs

159. Comme il a été souligné ci-dessus, chaque État membre devrait veiller à la conformité de sa réglementation nationale avec tout cadre juridique futur. Les différents acteurs devraient concourir en complémentarité à l'émergence d'une nouvelle culture d'applications d'IA conformes aux principes du cadre juridique et aux réglementations locales, en vue de produire des mesures d'incitation pertinentes en matière de conformité et de surveillance, soit en tant qu'assureurs et développeurs, soit en tant qu'opérateurs et utilisateurs.

9.2.1. Assureurs de systèmes

160. Les États membres devraient également endosser la responsabilité de l'identification et de l'habilitation d'acteurs indépendants chargés d'assurer une surveillance. Ces acteurs indépendants devraient représenter des groupes de parties prenantes clairement identifiés et concernés par les applications pratiques de l'IA et être tenu de rendre des comptes devant eux. Il pourrait s'agir, selon les cas, d'un comité d'experts, des universitaires, des régulateurs sectoriels ou des auditeurs du secteur privé. Les États membres pourraient envisager

de mettre en place, s'ils n'existent pas déjà, des organes de contrôle indépendants dotés d'une expertise interdisciplinaire, de capacités et de ressources appropriées et suffisantes pour exercer leur fonction de contrôle. Ces organes pourraient être dotés de pouvoirs d'intervention et tenus de rendre compte, par exemple à un parlement national ou à d'autres organes publics, et publier régulièrement des rapports d'activités[186]. Ils pourraient également résoudre des litiges instruits au nom de citoyens ou de consommateurs. Par exemple, en complément de mécanismes plus contraignants, les États pourraient étendre le mandat des institutions des droits de l'homme, des organismes de promotion de l'égalité, des institutions de médiation ou d'autres organismes de contrôle existants, ou en créer de nouveaux en vue d'évaluer et donner suite aux plaintes ou recours. Comme il serait déraisonnable d'espérer qu'une telle entité soit en mesure de couvrir tous les produits et systèmes basés sur l'IA, il serait par conséquent important d'examiner son champ d'application. Si de nouvelles entités sont créés, leurs mandats ne doivent pas empiéter sur ceux d'entités existantes dont les fonctions de surveillance incluraient également les systèmes d'IA si leur utilisation spécifique fait partie de leur mandat, ni entrer en conflit avec eux. Il est également important de reconnaître le rôle essentiel des institutions (nationales) des droits de l'homme, des organismes de promotion de l'égalité[187] et des institutions de médiation existants, dont les structures resteront pertinentes pour assurer une surveillance efficace dans le cadre de leur mandat sur les questions liées à l'IA.

161. De multiples systèmes d'IA sont déployés dans de nombreuses juridictions. Pour assurer une surveillance adéquate, le partage d'informations entre les États membres est essentiel. Des mécanismes de partage et de communication d'informations et de rapports sur les systèmes d'IA pourrait être inclus dans le cadre réglementaire de chaque État (informations sur des systèmes d'IA certifiés, des applications d'IA interdites ou du statut actuel d'une application d'IA spécifique, par exemple). Des acteurs du secteur privé peuvent également jouer un rôle dans la garantie des systèmes.

162. En plus des services de contrôle (volontaires ou obligatoires), des mécanismes de certification peuvent soutenir un cadre juridique et promouvoir l'attribution au secteur privé d'un rôle actif dans la prévention et la gestion des risques associés aux systèmes d'IA en ce qui concerne leurs effets négatifs sur les droits de l'homme. En effet, plus généralement, les mécanismes de certification, très polyvalents, peuvent fournir des instruments fondés sur des éléments factuels, à partir desquels il est envisageable d'élaborer des régimes de gouvernance flexibles pour répondre aux besoins de différents domaines et aux aspects dont il doit être tenu compte dans les régimes réglementaires nationaux. Des normes et des certifications peuvent être élaborées à chaque étape du développement et des opérations d'IA, et impliquer tous les agents concernés pour mettre en œuvre certaines obligations. Les pratiques de passation de marchés auxquelles ont recours les organisations intergouvernementales et les entités nationales du secteur public peuvent contribuer à leur adoption. Lorsqu'ils sont dûment mis en œuvre, ils peuvent contribuer à doter les citoyens ordinaires de moyens d'action en faisant office de « monnaie de confiance » à laquelle peuvent se référer les spécialistes comme les non-spécialistes (de manière analogue aux étiquettes nutritionnelles ou aux tests de collision pour l'amélioration de la sécurité automobile). Les éléments d'appréciation sous-jacents recherchés par ces normes et certifications peuvent également être utilisés pour encourager, accélérer et récompenser l'innovation au moyen d'initiatives d'évaluation comparative, ouvertes et récurrentes, de l'innovation en matière d'IA.

163. Dans le cadre des mécanismes de certification, le cadre juridique pourrait figurer dans les programmes de formation professionnelle. D'une manière générale, les milieux universitaires et la société civile pourraient prendre part à la politique éducative afin de diffuser le cadre juridique et les évolutions techniques de l'IA, mener des travaux de recherches en la matière et l'enseigner. Cette approche se traduirait également par des avantages supplémentaires dans une économie de marché mondiale.

164. En outre, la certification professionnelle au niveau des développeurs et des systèmes peut constituer une autre stratégie pour garantir que l'IA est utilisée conformément aux normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'État de droit. Ce mécanisme de certification pourrait être similaire aux mécanismes de certification déjà existants dans divers pays pour certaines professions.

---

186. Voir la Recommandation de la Commissaire aux droits de l'homme du Conseil de l'Europe, « Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme ».

187. Dans sa Recommandation de politique générale n°2 révisée, l'ECRI demande aux Etats membres de doter les organismes de promotion de l'égalité des pouvoirs et des ressources nécessaires pour traiter efficacement les questions d'égalité et de non-discrimination. Cela s'étend également à la discrimination résultant de l'utilisation de l'IA, en insistant sur le rôle des organismes de promotion de l'égalité dans les enquêtes sur des affaires spécifiques, le conseil aux victimes, la sensibilisation des organisations publiques et privées utilisant l'IA et du grand public aux risques potentiels liés à l'utilisation des systèmes d'IA. En outre, le nouveau rapport 2020 d'Equinet souligne le rôle important et le potentiel des organismes (nationaux) de promotion de l'égalité dans le contexte de l'IA, accessible à l'adresse suivante (uniquement en anglais) : https://equineteurope.org/2020/equinet-reportregulating-for-an-equal-ai-a-new-role-for-equality-bodies/

9.2.2. Développeurs de systèmes

165. Les acteurs chargés de la mise en place de systèmes reposant sur l'IA (tant dans le secteur privé que public) devraient envisager les mesures qu'ils peuvent prendre pour renforcer la conformité de ces systèmes à un futur cadre juridique. Par exemple, des politiques peuvent être adoptées pour accroître la visibilité des sites où ces technologies sont déployées (en particulier par la publication des contrats du secteur public, la constitution de registres publics[188] ou la mise en place de systèmes de notification), ou bien en définissant des normes et en développant des outils standardisés pour la réalisation d'audits internes et d'autocertifications (sans ignorer les limites de cette approche). Des considérations tenant à la responsabilité devraient également être prises en compte.

9.2.3. Opérateurs et utilisateurs de systèmes

166. Les opérateurs et utilisateurs d'IA sont susceptibles de créer une demande portant sur des applications d'IA qui soient conformes au futur cadre juridique. Cela vaut en particulier pour le secteur public et son relatif pouvoir d'achat. La promotion des supports de confiance, comme les mecanismes ou labels de certification sur le cycle de vie des systèmes d'IA, et les audits et présentations de rapports périodiques, sont des réponses du marché motivées par les préférences et les attentes des opérateurs et des utilisateurs de systèmes d'IA. Dès lors que les opérateurs et les utilisateurs de systèmes d'IA sont mieux informés de leurs droits et des mécanismes de réparation des préjudices, le coût de transaction de la surveillance s'en trouve significativement réduit.

9.3. Exemples de types de mécanismes de conformité

167. Dans de nombreux contextes, comme par exemple les services financiers et les soins de santé, les organisations sont déjà tenues de satisfaire aux normes ou aux réglementations. Chacune d'entre elle a évolué en écosystèmes de services qui leur permettent d'attester, vis-à-vis d'elles-mêmes, de leurs clients et des régulateurs, qu'elles satisfont à une norme nécessaire. Des mécanismes distincts fonctionneront mieux dans différents contextes, en fonction des infrastructures, des mécanismes sectoriels et des institutions existantes. Il convient également d'examiner, par une approche fondée sur le risque, quels composants d'un système d'IA peuvent être soumis au critère de conformité, par exemple, les données d'apprentissage utilisées, la construction d'algorithmes, la pondération des données d'entrée ou la précision des données de sortie. Des processus participatifs inclusifs devraient être engagés pour mettre en place dans chaque cas les mécanismes de réglementation et d'application pertinents.

168. Un futur cadre juridique pourrait prévoir que des mécanismes pratiques respectent un ensemble de principes renforçant les valeurs fondamentales dudit cadre. Ces principes pourraient être les suivants :

- **Dynamique (non statique) :** évaluation *ex ante* et à différents moments du cycle de vie du projet d'IA, pour rendre compte des choix effectués au cours des processus de conception, de développement et de déploiement et pour tout changement dans le comportement d'application des modèles d'apprentissage dynamique.
- **Technologie adaptative :** pour renforcer la pérennité de tout futur **mécanisme de conformité.**
- **Accessibilité différentielle :** compréhensible par les spécialistes comme par les non-spécialistes, et simplifiant à son tour les éventuelles procédures de recours et d'appel.
- **Indépendant :** mené ou supervisé par une tierce partie indépendante.
- **Fondé sur des données probantes :** s'appuyant sur des éléments d'appréciation produits par des normes et certifications techniques. Intégrant par exemple des données collectées à la faveur de bonnes pratiques telles que le transfrontière, la normalisation ou les paramètres déterminants développés, par exemple, au moyen d'évaluations comparatives de performance (*benchmark*).

169. Tout mécanisme doit pouvoir être mis en œuvre en pratique, compte tenu de l'infrastructure de gouvernance en vigueur et des limites techniques existantes. Il convient par conséquent de considérer les mécanismes concrets décrits ci-dessous comme constituant une boîte à outils offrant maintes possibilités d'innovation et d'amélioration en matière de réglementation :

188. De tels registres existent déjà aux Pays-Bas et au Royaume-Uni : https://algoritmeregister.amsterdam.nl/ ; https://ai.hel.fi/en/ai-register/.

(1) Due diligence en matière de droits de l'homme, y compris les études d'impact sur les droits de l'homme [189] – Les entreprises sont tenues de faire preuve de due diligence en matière de droits de l'homme afin d'assumer leur responsabilité de respecter les droits de l'homme tels qu'ils sont définis dans les principes directeurs des Nations unies sur les entreprises et les droits de l'homme (PDNU). Les entreprises doivent identifier, prévenir et atténuer les effets négatifs de leurs activités sur les droits de l'homme et en rendre compte. La due diligence en matière de droits de l'homme devrait inclure l'évaluation des impacts réels et potentiels sur les droits de l'homme, l'intégration et l'action des résultats, le suivi des réponses et la communication de la manière dont les impacts sont traités.[190] Les évaluations d'impact sur les droits de l'homme devraient faire partie du processus plus large de due diligence en matière de droits de l'homme, dans le cadre duquel les risques et les impacts identifiés sont atténués et traités de manière efficace, et devrait faire partie d'un processus d'évaluation continu plutôt que d'être un exercice statique. En outre, la recommandation du Conseil de l'Europe sur l'impact des systèmes algorithmiques sur les droits de l'homme a recommandé que les organisations publiques et privées réalisent une évaluation d'impact sur les droits de l'homme. Ces évaluations pourraient explicitement valider la conformité avec les principes énoncés dans un futur cadre juridique. Dans des contextes spécifiques, des « **évaluations d'impact intégrées** » pourraient être jugées plus appropriées pour réduire la charge administrative des équipes de développement (en regroupant, par exemple, les considérations relatives aux droits de l'homme, à la protection des données, à la transparence, à la responsabilité, aux compétences et à l'égalité). Lors de la réalisation d'une évaluation d'impact sur les droits de l'homme, il convient d'adopter une approche globale, dans laquelle tous les droits civils, politiques, sociaux, culturels et économiques pertinents sont pris en compte. Un modèle uniforme et des orientations élaborés au niveau du Conseil de l'Europe pour une évaluation d'impact sur les droits de l'homme, la démocratie et l'État de droit, ou une évaluation d'impact intégrée, pourraient être utiles pour valider la conformité avec les principes énoncés dans un futur cadre juridique du Conseil de l'Europe.

(2) Certification et labellisation de la qualité – Des obligations *ex ante*, prises en charge par des organismes reconnus et contrôlées de manière indépendante, contribueraient à instaurer la confiance. Une distinction pourrait être faite entre les normes et les certifications qui peuvent s'appliquer (i) aux produits / systèmes d'IA ou (ii) aux organisations qui développent ou utilisent des systèmes d'IA. Une date d'expiration serait la garantie d'un réexamen régulier des systèmes. Ce type de mécanisme pourrait être facultatif (par exemple pour les systèmes à faible risque) ou obligatoire (par exemple pour les systèmes à risque plus élevé), en fonction de la maturité de l'écosystème. Des garanties juridiques doivent veiller à ce que les certifications ne soient pas utilisées par les entreprises pour se protéger contre d'éventuelles actions en responsabilité fondées sur leur conduite, ou pour obtenir un avantage concurrentiel déloyal. La certification devrait faire l'objet d'une réglementation portant sur la qualification des personnes chargées du contrôle, les normes adoptées et la gestion des conflits d'intérêts. Elle devrait également viser l'amélioration permanente et tenir compte des critiques.[191] Des travaux d'élaboration de normes multipartites en cours viendraient appuyer ce processus mené par des organismes de normalisation.

(3) Audits – Les évaluations ou audits indépendants des systèmes d'IA régulièrement réalisés par des experts ou des groupes accrédités forment également un mécanisme qui devrait être utilisé tout au long du cycle de vie de tout système reposant sur des systèmes d'IA pouvant affecter négativement les droits de l'Homme, la démocratie et l'Etat de droit, afin de vérifier son intégrité, son impact, sa robustesse et son absence de biais. Les audits favorisent la transition vers une utilisation plus transparente et responsable des systèmes d'IA. Ils pourraient certifier des organisations dans leur ensemble, sans se limiter à des cas d'utilisation spécifiques.

---

189. Voir également Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, Lignes directrices sur l'intelligence artificielle et la protection des données, janvier 2019, et Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, Lignes directrices sur la protection des personnes à l'égard du traitement des données à caractère personnel dans un monde de données volumineuses, janvier 2017. Voir également à cet égard la nouvelle étude de la FRA, mentionnée dans la note de bas de page 114 ci-dessus, qui développe la nécessité d'introduire des les évaluations d'impact sur les droits dans le contexte des systèmes d'IA. Les travaux en cours au niveau de l'Union européenne sur un éventuel système communautaire obligatoire de due diligence en matière de droits de l'homme et d'environnement sont également pertinents. Voir à cet égard l'étude de la Commission européenne sur les exigences de due diligence à travers la chaîne d'approvisionnement de janvier 2020, disponible à l'adresse (uniquement en anglais) : https://op.europa.eu/en/publication-detail/-/publication/8ba0a8fd-4c83-11ea-b8b7-01aa75ed71a1/language-en.
190. Voir UNGP 17, https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_fr.pdf ; OHCHR B-Tech «*Key characteristics of Business Respect for Human Rights*», https://www.ohchr.org/Documents/Issues/Business/B-Tech/key-characteristics-business-respect.pdf.
191. Lorsque de nouveaux mécanismes de certification sont créés, il est important de tenir compte des initiatives existantes dans ce domaine, comme les travaux en cours de la CEPEJ concernant une certification spécifique pour les systèmes d'IA dans le secteur juridique, ainsi que les différents types de certification et labels établis au sein de l'UE, par exemple.

(4) Bacs à sable réglementaires[192] – Les bacs à sable réglementaires, en particulier ceux autorisant un soutien réglementaire plus appuyé, offrent une voie d'approche dynamique et sûre pour tester de nouvelles technologies et sont le cas échéant utilisables pour renforcer les capacités d'innovation dans le domaine de l'IA.[193] Ils pourraient être particulièrement utiles lorsqu'une introduction sur le marché, effectuée à point nommé et éventuellement limitée, semble justifiée pour des raisons d'intérêt général, par exemple en cas de crises extraordinaires telles qu'une pandémie, ou si des cadres juridiques actuels qui n'auraient pas été mis à l'épreuve dans la pratique risquent d'aboutir à une entrave à l'innovation. Il est important que l'établissement de bacs à sable réglementaires se fasse dans un cadre juridique approprié qui protège les droits de l'homme. Les bacs à sable interjuridictionnels offrent d'autres possibilités de collaboration, sur le modèle du *Global Financial Innovation Network*[194].

(5) Surveillance continue et automatisée – Des systèmes automatisés peuvent être déployés parallèlement aux systèmes d'IA pour surveiller et évaluer en permanence leur fonctionnement afin de garantir le respect des normes établies, par exemple lorsque les systèmes d'IA présentent un risque important.

170. Il conviendrait de considérer l'obligation d'adopter des mécanismes pratiques pour assurer le respect de la conformité comme n'étant qu'un élément d'un ensemble plus large d'initiatives nécessaires pour entraîner un changement. Les États membres pourraient renforcer les mécanismes de mise en conformité en prenant un certain nombre d'initiatives. Par exemple, en investissant dans la culture numérique, l'acquisition et le perfectionnement des compétences et des capacités des développeurs, des décideurs et de la société dans son ensemble, pour comprendre les implications des systèmes d'IA en matière de droits de l'homme ; en suscitant l'adoption généralisée de normes telles que le libre accès au code source ; en établissant, à différents stades du développement, des rapports directs avec les organisations de la société civile de défense des droits de l'homme, considérées comme principales parties prenantes[195].

171. Ce travail plus exhaustif visant à élaborer des bonnes pratiques et des normes dans le cadre des régimes juridiques et réglementaires en vigueur devrait s'accompagner, au niveau national et international, d'échanges permanents entre les acteurs, d'une collaboration durable et d'un partage des bonnes pratiques. Les centres d'expertise seraient bien placés pour faciliter la collaboration sur des solutions innovantes apportées aux projets de réglementation intersectorielles[196].

9.4. Mécanismes de suivi

172. Outre les mécanismes pratiques mentionnés ci-dessus, le CAHAI a pris note de la diversité des mécanismes et processus de suivi, ainsi que des mesures de coopération internationale qui sont prévues dans le cadre des instruments juridiques du Conseil de l'Europe, et dont les caractéristiques varient en fonction du type et du contenu desdits instruments.

173. Pour ce qui est des mécanismes et processus de suivi, le CAHAI a noté qu'ils peuvent inclure, par exemple, la désignation d'une ou plusieurs entités – telles que groupes d'experts indépendants, comités créés en vertu de conventions, commissions permanentes, comités consultatifs et comités de parties[197] – qui peuvent être chargées de missions telles qu'effectuer le suivi de la mise en œuvre d'une convention donnée, faciliter l'utilisation et la mise en œuvre effectives d'une convention, et échanger des informations et bonnes pratiques relatives aux évolutions significatives, juridiques, politiques ou technologiques, dans un domaine donné. En outre, un observatoire pourrait être créé pour suivre la mise en œuvre et l'impact d'un éventuel cadre juridique du Conseil de l'Europe sur l'IA. Il pourrait également suivre les conséquences sociétales de l'adoption de systèmes d'IA sur les droits de l'homme, la démocratie et l'État de droit, et garder une vue d'ensemble des contributions apportées par d'autres parties prenantes dans ce domaine.

192. Les bacs à sable s'entendent comme des cadres concrets qui, en fournissant un contexte structuré pour l'expérimentation, permettent de tester dans un environnement réel des technologies, produits, services ou approches innovants, en particulier dans le contexte de la digitalisation, pendant un temps limité et généralement dans une partie limitée d'un secteur ou d'un domaine, sous le contrôle réglementaire de l'autorité concernée, en veillant à ce que des garanties appropriées soient mises en place.
193. Dans le même temps, il convient de veiller à ce que la protection des droits de l'homme - et plus encore lorsqu'il s'agit de droits de l'homme absolus - reste assurée.
194. https://www.fca.org.uk/firms/innovation/global-financial-innovation-network
195. CAHAI(2020)21 rev PDG contributions p. 45-46.
196. CAHAI(2020)21 rev PDG contributions p. 32-33.
197. Le comité des parties pourrait être chargé de la collecte et du partage d'informations sur la législation et les meilleures pratiques dans un domaine donné.

174. Quant aux mesures potentielles de coopération internationale, elles pourraient inclure la désignation de points de contact ou la création de réseaux entre les États Parties visant à stimuler l'assistance mutuelle et la coopération en matière pénale ou civile.

175. Bien que l'identification de solutions précises soit trop prématurée à ce stade, et en gardant à l'esprit que les caractéristiques concrètes des mécanismes et processus de suivi dépendront de la nature et des éléments de fond de l'instrument (ou des instruments) juridique(s) choisi(s), le CAHAI recommande de veiller à ce qu'un futur cadre juridique portant sur l'IA comporte des mécanismes et processus de suivi appropriés, ainsi que des mesures de coopération internationale, dans le droit fil des normes et pratiques juridiques du Conseil de l'Europe. Cette intégration est d'une importance capitale pour garantir au niveau international l'efficacité des principes, droits et obligations fondamentaux énoncés au chapitre 7, et pour compléter les mesures pratiques et de surveillance décrites plus haut dans le présent chapitre, qui peuvent être mises en œuvre au niveau national.

10. CONSIDÉRATIONS FINALES

176. La présente étude a confirmé que les systèmes d'IA sont susceptibles d'offrir de remarquables possibilités de développement individuel et sociétal, ainsi que pour les droits de l'Homme, la démocratie et l'Etat de droit. Dans le même temps, elle a également confirmé que les systèmes d'IA peuvent avoir un impact négatif sur plusieurs droits de l'homme protégés par la Convention européenne des droits de l'homme et d'autres instruments du Conseil de l'Europe, ainsi que sur la démocratie et l'État de droit. L'étude a constaté qu'il n'existe aucun instrument juridique international spécifiquement adapté aux défis posés par l'IA et que le niveau actuel de protection assuré par les instruments internationaux et nationaux en vigueur présente des lacunes. L'étude a identifié les principes, droits et obligations qui pourraient constituer les principaux éléments d'un futur cadre juridique applicable à la conception, au développement et à l'application de l'IA, fondé sur les normes du Conseil de l'Europe et dont l'élaboration a été confiée au CAHAI.

177. Un cadre juridique approprié sera probablement composé d'une combinaison d'instruments juridiques contraignants et non contraignants se complétant les uns les autres. Un instrument contraignant, convention ou convention-cadre, de caractère horizontal pourrait consolider des principes généraux communs – contextualisés pour s'appliquer à l'environnement d'IA et utilisant une approche fondée sur les risques – et prévoir des dispositions plus granulaires conformes aux droits, principes et obligations identifiés dans la présente étude de faisabilité. Tout document contraignant, quelle que soit sa forme, ne devrait pas être trop prescriptif afin de garantir son caractère évolutif. En outre, il devrait garantir que l'innovation en matière d'IA socialement bénéfique puisse se développer, tout en abordant de manière adéquate les risques spécifiques posés par la conception, le développement et l'application des systèmes d'IA.

178. Cet instrument pourrait être associé à d'autres instruments sectoriels contraignants ou non contraignants du Conseil de l'Europe pour relever les défis posés par les systèmes d'IA dans des secteurs spécifiques. Cette association renforcerait également la sécurité juridique apportée aux parties prenantes de l'IA, et fournirait le niveau d'orientation juridique nécessaire aux acteurs privés souhaitant prendre des initiatives d'autorégulation. En outre, en établissant des normes communes au niveau international, la confiance transfrontalière dans les produits et services d'IA serait assurée, garantissant ainsi que les bénéfices engendrés par les systèmes d'IA peuvent franchir les frontières nationales. Il est essentiel que ce cadre juridique comporte des mécanismes concrets ayant pour objet d'atténuer les risques suscités par les systèmes d'IA, ainsi que des mécanismes et processus de suivi appropriés et des mesures de coopération internationale.

179. Le Comité des Ministres est invité à prendre note de cette étude de faisabilité et charger le CAHAI de concentrer ses travaux sur l'élaboration des éléments spécifiques d'un cadre juridique approprié. Celui-ci pourrait comprendre un instrument juridique contraignant, ainsi que des instruments non contraignants le cas échéant, parallèlement aux progrès qui peuvent être réalisés sur des instruments sectoriels.

Avec cette étude de faisabilité, le CAHAI a exploré les raisons pour lesquelles un cadre juridique sur le développement, la conception et l'application de l'IA, fondé sur les normes du Conseil de l'Europe sur les droits de l'homme, la démocratie et l'état de droit, est nécessaire. Un certain nombre de lacunes substantielles et procédurales ont été identifiées, qui montrent la nécessité d'un cadre de gouvernance plus complet et d'une réponse juridique internationale efficace. L'étude conclut qu'un cadre juridique complet combinant des instruments juridiques contraignants et non contraignants, qui se complètent mutuellement, est la voie à suivre.

www.coe.int/cahai

PREMS 053621

FRA

www.coe.int

Le Conseil de l'Europe est la principale organisation de défense des droits de l'homme du continent. Il comprend 47 États membres, dont l'ensemble des membres de l'Union européenne. Tous les États membres du Conseil de l'Europe ont signé la Convention européenne des droits de l'homme, un traité visant à protéger les droits de l'homme, la démocratie et l'État de droit. La Cour européenne des droits de l'homme contrôle la mise en œuvre de la Convention dans les États membres.

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE