



COMMISSION NATIONALE  
CONSULTATIVE  
DES DROITS DE L'HOMME

RÉPUBLIQUE FRANÇAISE

AVIS

A - 2022 - 6

# AVIS RELATIF À L'IMPACT DE L'INTELLIGENCE ARTIFICIELLE SUR LES DROITS FONDAMENTAUX

7 AVRIL 2022



*Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux  
a été adopté lors de l'Assemblée plénière du 7 avril 2022.  
(Adoption à l'unanimité)*

**RÉSUMÉ.**

Alors que les recherches sur l'intelligence artificielle (IA) et la mise en œuvre de ses applications pratiques se développent, la réglementation actuelle demeure lacunaire pour endiguer les atteintes possiblement majeures aux droits fondamentaux. Dans le contexte de l'adoption prochaine de la proposition de règlement de l'UE sur le sujet, et des travaux en cours au sein du Conseil de l'Europe, la CNCDH invite les pouvoirs publics à promouvoir un encadrement juridique ambitieux en la matière.

Elle recommande, d'une part, d'interdire certains usages de l'IA jugés trop attentatoires aux droits fondamentaux, tels que le scoring social ou l'identification biométrique à distance des personnes dans l'espace public et les lieux accessibles au public. D'autre part, elle recommande de faire peser sur les utilisateurs d'un système d'IA des exigences en mesure de garantir le respect des droits fondamentaux : une étude d'impact, une consultation des parties prenantes, une supervision du système tout au long de son cycle de vie. La CNCDH appelle enfin à reconnaître des droits aux personnes ayant fait l'objet d'une décision impliquant un algorithme, notamment le droit à une intervention humaine dans le processus de décision, ou encore un droit au paramétrage des critères de fonctionnement du système d'IA.

## TABLE DES MATIÈRES.

<b>Introduction.</b>	<b>4</b>
<b>1. Les lignes rouges.</b>	<b>9</b>
<b>1.1. Les apports et les limites des interdictions posées par la proposition de règlement UE.</b>	<b>9</b>
<b>1.2. La nécessité d'étendre l'interdiction à d'autres domaines.</b>	<b>12</b>
1.2.1. L'IA en justice.	13
1.2.2. L'IA et la reconnaissance des émotions.	14
<b>2. Un encadrement garant du respect des droits fondamentaux.</b>	<b>15</b>
<b>2.1. Un contrôle à tous les stades de développement du système IA.</b>	<b>15</b>
2.1.1. Une étude d'impact sur les droits fondamentaux : un préalable nécessaire.	16
2.1.2. Une supervision du système tout au long du cycle de vie.	19
2.1.3. Une formation et une sensibilisation aux enjeux de l'IA.	22
<b>2.2. Des garanties du respect des droits fondamentaux à l'égard des décisions individuelles.</b>	<b>23</b>
2.2.1. Une intervention humaine garante de la prise en compte des spécificités individuelles.	23
2.2.2. Une information garante de la dignité de la personne humaine.	26
<b>Synthèse des recommandations.</b>	<b>30</b>
<b>Liste des personnes auditionnées.</b>	<b>33</b>

## Introduction.

1. Depuis quelque temps déjà, ce qu'il est désormais courant d'appeler « intelligence artificielle » ou « IA » agit au cœur du quotidien de chacun : propositions de contenu sur les réseaux sociaux ou les plateformes en ligne, accès à des applications ou à des lieux par authentification biométrique, diagnostics médicaux automatisés, etc. Elle fait cependant l'objet d'un engouement particulier depuis plusieurs années, en profitant d'investissements publics et privés massifs. Si certains la placent en effet au cœur d'une « nouvelle révolution industrielle », d'autres s'inquiètent d'une nouvelle vague d'automatisation des activités jusque-là réservées à l'être humain et, plus largement, des dérives d'une nouvelle gouvernance par les données, ainsi que, plus largement, d'atteintes possiblement majeures aux droits fondamentaux, sans compter son impact croissant sur l'environnement.

2. À titre liminaire, la Commission nationale consultative des droits de l'homme (CNCDDH) tient à exprimer ses réserves à l'égard de la terminologie usitée en la matière. Elle observe en effet un excès d'anthropomorphisation dans les termes employés, à commencer par celui de l'« intelligence artificielle », mais également lorsqu'il est question de « réseaux de neurones », d'« apprentissage profond », etc. Cela engendre des confusions sur les possibilités réelles offertes par des systèmes de traitement de données, qui reposent sur des procédures codées dans des systèmes informatiques : il s'agit avant tout de mathématiques. L'ensemble des acteurs, tant du secteur public que du secteur privé, devrait donc s'affranchir de cette expression en raison de son impact psychologique, source de réticences ou au contraire de confiance et d'acceptation exagérées. C'est pourquoi la CNCDDH recommande que les institutions publiques et les médias privilégient des expressions plus neutres, telles que des « systèmes algorithmiques d'aide à la décision » (SAAD). Néanmoins, pour des raisons de commodité rédactionnelle, et parce qu'il s'agit de l'usage actuellement consacré, la CNCDDH fera dans cet avis référence à l'« IA ».

**Recommandation n°1** : La CNCDDH recommande de privilégier, dans la communication institutionnelle, une terminologie plus neutre et objective que l'expression « intelligence artificielle », telle celle de « système algorithmique d'aide à la décision ».

3. Ce vocable recouvre, plus précisément, des technologies informatiques qui reposent sur des logiques de fonctionnement différentes : on distingue principalement l'IA symbolique (ou cognitiviste) de l'IA connexionniste. La première suppose de programmer une série d'instructions explicites et univoques – autrement dit un algorithme – appelées à donner un résultat, prévisible parce qu'il se présente comme le traitement logique des données rentrées dans le système. La seconde, d'apparition plus récente, repose sur un autre type d'algorithme, non plus axé sur une approche logique

du traitement de l'information mais sur une approche probabiliste : les programmeurs conçoivent un algorithme d'apprentissage et soumettent à l'ordinateur une série de données (*data set*) à partir desquelles il va « apprendre » ou, plus exactement, inférer des règles. Cet apprentissage peut être supervisé ou non supervisé : dans le premier cas, les données utilisées pour l'apprentissage sont étiquetées, tandis que dans le second cas elles sont « brutes ». Dans cette dernière hypothèse, l'apprentissage machine établit des corrélations entre les informations qui ont alimenté le système.

4. Ce dernier type d'apprentissage, l'apprentissage machine, soulève des enjeux inédits par rapport à l'IA symbolique. Tandis que les instructions codées dans un logiciel « classique » peuvent aisément être communiquées (encore que la compréhension du système puisse être rendue malaisée lorsque ces instructions sont nombreuses et les données traitées volumineuses, comme c'est le cas par exemple de la plateforme en ligne d'admission dans l'enseignement supérieur, Parcoursup), le modèle auquel parvient la machine au terme de son apprentissage peut plus ou moins aisément faire l'objet d'une information, étant donné que les concepteurs du système sont dans certains cas extrêmes (en matière de *deep learning* notamment) bien incapables de connaître le modèle de fonctionnement auquel est parvenue la machine pour obtenir ses résultats.

5. Si l'IA pourrait permettre, selon certains, d'« activer nos droits fondamentaux »<sup>1</sup>, elle présente toutefois des risques incontestables pour ces derniers. Au niveau national, la Commission nationale de l'informatique et des libertés (CNIL) a publié en 2017 un rapport sur les algorithmes, fruit d'une vaste consultation auprès d'acteurs du secteur et de citoyens, face à la nécessité de « permettre à l'homme de garder la main »<sup>2</sup>, et a également axé sa réflexion sur certaines applications particulières de l'IA<sup>3</sup>. En 2017, encore, le Défenseur des droits alertait sur les risques de discrimination engendrés par l'utilisation d'algorithmes dans le cadre de la lutte contre la fraude aux prestations sociales, en ciblant des catégories de personnes à contrôler en priorité<sup>4</sup>. Depuis, le Défenseur des droits mène une réflexion plus large sur l'IA et les discriminations<sup>5</sup>. Au niveau international, de nombreuses instances ont également alerté sur l'impact de

---

1 Selon la formule présente dans le rapport Villani : « Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne », 28 mars 2018.

2 CNIL, « Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle », décembre 2017.

3 CNIL, « Chatbots : des humains comme les autres », LINC, 3 février 2017 ; « Reconnaissance faciale : pour un débat à la hauteur des enjeux », 2019. Plus récemment, la CNIL a également mis en ligne un ensemble de ressources pour le grand public et les professionnels ([www.cnil.fr/fr/intelligence-artificielle/la-cnil-publie-ressources-grand-public-professionnels](http://www.cnil.fr/fr/intelligence-artificielle/la-cnil-publie-ressources-grand-public-professionnels)), ainsi que plusieurs articles, notamment sur la sécurité des systèmes d'IA ou sur l'apprentissage fédéré (<https://linc.cnil.fr/fr/dossier-intelligence-artificielle>).

4 Défenseur des droits, « Lutte contre la fraude aux prestations sociales : à quel prix pour les droits des usagers », septembre 2017.

5 Défenseur des droits, en partenariat avec la CNIL, « Algorithmes : prévenir l'automatisation des discriminations », 2020. Plus récemment, le DDD a publié un rapport sur les technologies biométriques : « Technologies biométriques : l'impératif respect des droits fondamentaux », 2021.

l'IA sur les droits fondamentaux<sup>6</sup>. En particulier, l'Agence des droits fondamentaux de l'Union européenne (FRA), et le Comité ad hoc sur l'intelligence artificielle (CAHAI), organe du Conseil de l'Europe chargé d'examiner les possibilités de mettre en place un cadre juridique relatif à l'IA, ont dressé un inventaire des droits fondamentaux susceptibles d'être remis en cause par l'IA : notamment le respect de la dignité humaine, le respect de la vie privée et la protection des données, l'égalité et la non-discrimination, l'accès à la justice, l'accès aux droits sociaux, etc.

6. Le déploiement de l'IA inquiète d'autant plus qu'il n'existe pas, à l'heure actuelle, de cadre juridique global, à l'échelon tant national qu'international, pour endiguer ses débordements. Les réglementations en vigueur fournissent des références seulement partielles, qu'il s'agisse de la protection des données personnelles – avec en particulier, au sein de l'Union européenne, le Règlement général sur la protection des données (RGPD) – ou de la non-discrimination. Cela demeure toutefois insuffisant dès lors qu'un grand nombre de systèmes d'IA fonctionnent à partir de données non identifiantes et peuvent avoir des conséquences sur les droits fondamentaux excédant la protection des données personnelles et la non-discrimination, sans compter les formes de discriminations susceptibles de viser des groupes non couverts par les critères de discrimination prohibés par le droit<sup>7</sup>.

7. Depuis plusieurs années, des initiatives émanent du secteur privé. Conscients de la nécessité de proposer des solutions d'IA dignes de confiance pour en assurer le succès commercial, des professionnels proposent des guides d'éthique à destination des concepteurs et des développeurs<sup>8</sup>. Par ailleurs, des institutions internationales adressent aux Etats des recommandations allant dans le même sens, par exemple celles adoptées par l'UNESCO le 24 novembre 2021 afin de « *mettre les systèmes d'IA au service de l'humanité, des individus, des sociétés, de l'environnement et des écosystèmes, ainsi que de prévenir les préjudices* »<sup>9</sup>.

8. Les préoccupations exprimées à travers ces textes, souvent formulées à partir d'une référence à des « principes éthiques », coïncident en grande partie avec les

---

6 Voir not. : Agence des droits fondamentaux de l'UE, « Getting the future right : Artificial intelligence and fundamental rights », 14 décembre 2020 ; UNESCO, Recommandation sur l'éthique de l'intelligence artificielle, novembre 2021 ; CAHAI, « Étude de faisabilité », 17 décembre 2020 ; OCDE, Recommandation du Conseil sur l'intelligence artificielle, 22 mai 2019.

7 CNIL, « Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle », décembre 2017, p. 49. Certains algorithmes sont appelés à établir des corrélations entre différentes caractéristiques individuelles, à partir desquelles ils constituent des groupes et font des prédictions sur le comportement au niveau d'un groupe, par exemple le groupe des « propriétaires de chiens résidant en région parisienne, âgés de 35 à 40 ans, et qui pratiquent au moins deux fois par semaine une activité sportive ». Le fait d'être identifié comme membre de ce groupe peut dès lors entraîner des prises de décisions automatisées ayant des effets néfastes ou bénéfiques pour les membres individuels, comme un tarif différencié pour l'assurance maladie par exemple.

8 Voir notamment le guide pratique « Ethical IA », lancé en septembre 2021 par Numeum, premier syndicat professionnel des entreprises du numérique en France.

9 Dernier texte en date : UNESCO, Recommandation sur l'éthique de l'intelligence artificielle, novembre 2021.

droits de l'Homme, en particulier lorsqu'il est question de l'autonomie ou de la liberté, du respect de la dignité de la personne humaine, ou encore de la non-discrimination. Elles disposent cependant d'une portée limitée, s'en remettant aux vertus de l'auto-régulation des acteurs, au bon vouloir des industriels et des entreprises, et n'enjoignent aucune obligation aux États.

9. Étant donné l'impact important de l'IA sur les droits fondamentaux, cette approche n'apparaît pas suffisante. C'est pourquoi la CNCDH s'est auto-saisie de la question. L'ampleur des enjeux soulevés à l'égard des droits fondamentaux par la conception, le déploiement et l'utilisation des systèmes d'IA plaide pour la mise en place d'un cadre juridique contraignant en mesure de garantir le respect de ces droits. La CNCDH a suivi avec attention les réflexions et les travaux actuellement menés en ce sens au sein du Conseil de l'Europe par le CAHAI<sup>10</sup>, qui pourraient aboutir à l'adoption d'un cadre juridique pour « *le développement, la conception et l'application de l'intelligence artificielle, fondés sur les normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et de droits de l'homme* »<sup>11</sup>. La CNCDH appelle de ses vœux l'adoption d'une « Convention 108+ de l'IA »<sup>12</sup>. Quant à la proposition de règlement sur l'IA de l'Union européenne (dénommée dans la suite « proposition de règlement »), qui fixe un premier cadre juridique régional afin de « *favoriser le développement, l'utilisation et l'adoption de l'intelligence artificielle dans le marché intérieur, tout en respectant un niveau élevé de protection des intérêts publics, tels que la santé et la sécurité et la protection des droits fondamentaux, tels que reconnus et protégés par le droit de l'Union* »<sup>13</sup>, elle est pour la CNCDH un préalable nécessaire à la prise en compte des droits fondamentaux. Néanmoins, la Commission relève l'insuffisance des garanties propres à assurer un respect effectif de ces derniers. Dans la mesure où la proposition de règlement doit être à la hauteur des enjeux de protection de ces droits dans le recours à des systèmes d'IA, la Commission recommande que ce texte assure, à cette fin, la mise en place d'un cadre juridique contraignant.

## **Recommandation n°2 : La Commission recommande de renforcer, au sein de la proposition de règlement de l'Union européenne relatif à l'IA, les dispositions propres à as-**

<sup>10</sup> Le CAHAI ayant rempli son mandat (2019-2021), il a été remplacé par le Comité sur l'intelligence artificielle (CAI) pour l'élaboration d'un cadre juridique approprié sur le développement, la conception et l'application de l'intelligence artificielle, sur la base des normes du Conseil de l'Europe.

<sup>11</sup> Voir not. sa dernière publication : CAHAI, « *Éléments potentiels d'un cadre juridique sur l'intelligence artificielle, fondés sur les normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'Etat de droit* », 2 décembre 2021.

<sup>12</sup> « Convention 108+ » : Convention pour la protection des personnes à l'égard du traitement des données à caractère personnel, disponible en ligne : <https://rm.coe.int/convention-108-convention-pour-la-protection-des-personnes-a-l-egard-d/16808b3726>. Ouverte à la signature le 28 janvier 1981, au sein du Conseil de l'Europe, la Convention 108 fut le premier instrument international juridique contraignant dans le domaine de la protection des données. Afin de répondre aux défis soulevés en la matière par le numérique, un protocole d'amendement été ouvert, le 10 octobre 2018, à la signature des Etats partis à la Convention. On parle désormais de « Convention 108+ » pour désigner cette nouvelle version.

<sup>13</sup> Proposition de règlement IA, cons. 5. Disponible en ligne : <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:52021PC0206>.

**sur la mise en place d'un cadre juridique contraignant, garant du respect effectif des droits fondamentaux. De surcroît, la CNCDH recommande l'adoption, dans le cadre du Conseil de l'Europe, d'une « Convention 108+ de l'IA ».**

10. Ainsi, c'est en s'inspirant des différentes démarches déjà esquissées par des instances nationales comme le Défenseur des droits ou la CNIL et les instances européennes et internationales, que la CNCDH souhaite définir les grandes lignes d'un cadre général, respectueux des droits fondamentaux, pour les systèmes d'IA. Ce faisant, son avis pourra contribuer à nourrir les nécessaires modifications du règlement IA de l'UE. D'autres avis suivront à l'avenir afin d'identifier les risques pour les droits de l'homme, propres à l'utilisation de l'IA dans certains secteurs, ainsi que les garanties susceptibles d'y remédier. D'ailleurs, la CNCDH s'est déjà exprimée dans son avis sur la lutte contre la haine en ligne<sup>14</sup> à propos de l'utilisation des algorithmes de modération des contenus sur les réseaux sociaux.

11. La CNCDH précise que la valorisation d'un cadre juridique contraignant n'exclut évidemment pas des dispositifs de droit souple, avec des certifications, labels, sous le contrôle de l'autorité de régulation, afin de favoriser le développement de bonnes pratiques susceptibles d'accompagner la mise en œuvre de cette réglementation.

12. Soucieuse de promouvoir une approche fondée sur les droits de l'Homme<sup>15</sup>, la CNCDH s'attachera ici à la fois à mettre en avant la nécessité d'inclure dans la réflexion et la supervision à l'égard des systèmes d'IA les segments les plus marginalisés de la population et, au-delà, à insister sur l'importance de mettre en place un cadre juridique permettant de garantir le respect des droits fondamentaux. L'approche fondée sur les droits de l'Homme implique de placer la personne humaine en position de définir ses besoins et donc de soutenir le développement d'une IA au service de l'humain et de son autonomie. Cette approche devrait davantage irriguer les réformes en cours, dès lors qu'elles entendent garantir le respect des droits fondamentaux.

**Recommandation n°3 : La CNCDH recommande la prise en compte d'une approche fondée sur les droits de l'Homme dans les réformes en cours, dès lors qu'elles entendent garantir le respect des droits fondamentaux.**

13. Les observations et recommandations de la CNCDH traiteront des deux composantes d'un encadrement de l'IA respectueux des droits fondamentaux : une définition des « lignes rouges », autrement dit les usages de l'IA à proscrire (1); des garanties à promouvoir pour assurer un encadrement des systèmes d'IA respectueux des droits fondamentaux (2).

---

<sup>14</sup> CNCDH, *Avis relatif à la lutte contre la haine en ligne*, Assemblée plénière du 8 juillet 2021, JORF n°0170 du 24 juillet 2021, texte n° 79.

<sup>15</sup> CNCDH, *Avis « Pour une approche fondée sur les droits de l'Homme »*, Assemblée plénière du 3 juillet 2018, JORF n°0161 du 14 juillet 2018, texte n° 104.

14. Dans le cadre de cet avis, et en s'inspirant de la terminologie consacrée par les instances de l'Union européenne<sup>16</sup>, la CNCDH désignera par « concepteur » ou « fournisseur » la personne physique ou morale qui développe un système d'IA ; par « utilisateur », toute personne physique ou morale, autorité publique, agence ou autre organisme, y compris de droit privé, utilisant sous sa propre autorité un système d'IA ; et par « personne concernée » ou « personne visée par un système d'IA », toute personne physique exposée à un système d'IA ou impactée par celui-ci.

## 1. Les lignes rouges.

15. Certains usages de l'IA portent une atteinte trop grave aux droits fondamentaux pour être admis. Il revient aux pouvoirs publics d'en prohiber la mise en place. La proposition de règlement de l'UE interdit quelques cas d'utilisation de l'IA considérés, à juste titre, comme « *particulièrement néfastes* »<sup>17</sup>. La CNCDH relève toutefois certaines limites dans la définition des usages proscrits. En outre, d'autres usages paraissent tout aussi dangereux pour les droits fondamentaux et la dignité humaine et mériteraient, en tant que tels, d'être interdits également.

### 1.1. Les apports et les limites des interdictions posées par la proposition de règlement UE.

16. La proposition de règlement de l'UE relative à l'IA dresse une liste des utilisations « *interdites en raison de leur caractère contraire aux valeurs de l'Union* »<sup>18</sup>. Cette technologie peut en effet être « *utilisée à mauvais escient et fournir des outils nouveaux et puissants à l'appui de pratiques de manipulation, d'exploitation et de contrôle social* »<sup>19</sup>. La CNCDH se rallie à l'idée que certains usages de l'IA devraient faire l'objet d'une prohibition pure et simple, au regard de l'ampleur de leur impact sur les droits et libertés fondamentaux.

17. Les systèmes particulièrement problématiques de ce point de vue, identifiés par la proposition de règlement, représentent de graves menaces pour la protection des droits et libertés fondamentaux :

- les systèmes reposant sur des composants subliminaux que les personnes ne peuvent pas percevoir, ou exploitant les fragilités des enfants et des personnes vulnérables en raison de leur âge ou de leurs handicaps physiques ou mentaux, et qui, en altérant leur comportement, peuvent leur causer un préjudice, physique ou psychologique ;
- les systèmes d'IA permettant la notation sociale des personnes physiques, en

---

<sup>16</sup> Voir le RGPD et la proposition de règlement sur l'IA.

<sup>17</sup> Proposition de règlement IA, « Exposé des motifs ».

<sup>18</sup> Proposition de règlement IA, « Exposé des motifs ».

<sup>19</sup> *Ibid.*, cons. 15.

fonction de leur comportement ou de leurs caractéristiques personnelles, par les autorités publiques ou pour le compte de celles-ci, à des fins de traitement préjudiciable ou défavorable de certaines personnes physiques ou de groupes de personnes ;

- l'identification biométrique à distance «en temps réel », à des fins répressives, de personnes physiques dans des espaces accessibles au public.

18. La CNCDH s'interroge sur la définition et la portée des interdictions ainsi imposées à l'utilisation des systèmes d'IA.

19. D'abord, s'agissant du premier cas de figure, les termes employés pour définir des applications « malveillantes » pourraient couvrir un grand nombre de situations, d'interfaces et de services en ligne, actuellement très en vogue : les dispositifs de « nudge » ou de « sludge »<sup>20</sup>, mis en place par les réseaux sociaux ou les sites d'achat en ligne, exploitent les biais cognitifs pour orienter le comportement des utilisateurs, afin de capter leur attention, les inciter à acheter un produit, etc. Des termes tels que « subliminal » ou « substantiel » relèvent d'une appréciation particulièrement complexe, et rendent incertaine la portée de cette première limite à l'utilisation de l'IA. Cette interdiction présente toutefois le mérite de susciter des interrogations sur les risques admissibles de manipulation par l'IA, notamment en ce qui concerne les processus d'automatisation du traitement des informations issues de recherches en neuromarketing<sup>21</sup>. Elle est d'autant plus essentielle, de surcroît, que se développent des « *augmented dark patterns* » permettant, par un traitement algorithmique, d'agir dynamiquement sur les stimuli des utilisateurs pour exploiter leurs vulnérabilités. À cet égard, la CNCDH soutient la volonté des parlementaires européens, exprimée par ailleurs<sup>22</sup>, d'interdire aux très grandes plateformes en ligne l'utilisation de techniques de manipulation qui visent à inciter les utilisateurs à faire certains choix.

**Recommandation n°4 : La CNCDH recommande l'interdiction du recours aux interfaces de choix dès lors qu'elles ont pour objet ou pour effet de manipuler, à leur détriment, les utilisateurs en exploitant leurs vulnérabilités.**

20. Ensuite, et de manière plus inquiétante, la proposition du texte de l'UE prohibe la notation sociale (« *social scoring* »), qui consiste à évaluer les personnes en fonction de leur comportement social ou de caractéristiques personnelles ou de personnalité

---

20 Les « nudges » correspondent à des techniques pour orienter l'utilisateur d'un service en ligne à faire tel ou tel choix. Par exemple, indiquer sur un site qu'un article ou un produit est « populaire » incitera l'internaute à le consulter ou à l'acheter. Voir not. : « La forme des choix », CNIL, 2019. Les « sludges » ont, au contraire, une finalité dissuasive. Par exemple, s'agissant de la gestion des cookies sur un site, certains sites simplifient l'option « tout accepter » et complexifient l'option « paramétrer les cookies », de telle sorte que l'utilisateur aura tendance à privilégier la première.

21 Le neuromarketing étudie le fonctionnement du cerveau, notamment à l'aide de l'imagerie cérébrale (IRM), pour mieux comprendre comment le consommateur réagit à la publicité et aux dispositifs de vente.

22 Voir les débats relatifs à la proposition de règlement du Parlement européen et du Conseil relatif à un marché intérieur des services numériques (DSA).

connues ou prédites, uniquement lorsqu'elle est pratiquée par les administrations ou pour le compte de celles-ci. Or, les entreprises privées, par exemple les réseaux sociaux, peuvent aussi traiter de vastes quantités de données personnelles et effectuer du *scoring*. Par conséquent, la CNCDH se rallie à la position des autorités européennes de protection des données, favorable à une interdiction de tout type de notation sociale, quelle que soit la nature, publique ou privée, de l'entité qui la met place<sup>23</sup>.

**Recommandation n°5 : La CNCDH recommande d'interdire tout type de notation sociale (« social scoring ») mis en place par les administrations ou par toute entreprise, publique ou privée.**

21. Enfin, la prohibition de l'identification biométrique à distance « *en temps réel* » de personnes physiques – autrement dit l'utilisation de l'IA en vue d'une reconnaissance automatisée de caractéristiques humaines telles que le visage, la voix, la démarche, etc. – dans des espaces accessibles au public et à des fins répressives soulève, en tant que telle, des interrogations sur son champ d'application, mais également des inquiétudes au regard de l'étendue des dérogations admises par la proposition de règlement.

22. La CNCDH rejoint la Commission européenne lorsqu'elle met en avant non seulement le risque pour le respect de la vie privée engendré par un tel dispositif, mais aussi le « *sentiment de surveillance constante* » susceptible d'être généré par cette technologie et le risque de « *dissuader indirectement l'exercice de la liberté de réunion et d'autres droits fondamentaux* »<sup>24</sup>, à commencer par la liberté de circulation. La CNCDH souligne toutefois que cet effet dissuasif vaut tout autant pour un système d'identification biométrique a posteriori, c'est-à-dire en temps différé par rapport au moment où les images ont été collectées<sup>25</sup>. Par ailleurs, à défaut d'une interdiction expresse, l'utilisation de cette technologie à des fins préventives est autorisée ; mais les réserves exprimées plus haut, à l'égard de son utilisation dans un cadre répressif, valent également, si ce n'est plus, lorsqu'elle intervient pour prévenir des atteintes à l'ordre public.

23. Par ailleurs, la proposition de règlement prévoit trois exceptions à l'interdiction de l'identification biométrique en temps réel des personnes : la recherche ciblée de victimes potentielles, notamment d'enfants disparus ; la prévention d'une menace spécifique, substantielle et imminente pour la vie ou la sécurité des personnes ou la

23 Avis conjoint 05/2021 de l'EDPB et du CEPD sur la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle. Disponible en ligne : [https://edpb.europa.eu/system/files/2021-10/edpb-edps\\_joint\\_opinion\\_ai\\_regulation\\_fr.pdf](https://edpb.europa.eu/system/files/2021-10/edpb-edps_joint_opinion_ai_regulation_fr.pdf).

24 Proposition de règlement IA, Cons. 18. La CNCDH pointait aussi ce risque de « *chilling effect* » associé à l'utilisation de drones équipés d'un logiciel de reconnaissance faciale : CNCDH, *Avis sur la proposition de loi relative à la sécurité globale*, Assemblée plénière du 26 novembre 2020, JORF n°0290 du 1<sup>er</sup> décembre 2020, texte n° 83.

25 La proposition de règlement admet la possibilité d'utiliser cette technologie dans ce cas-là, en la rangeant néanmoins dans la catégorie des systèmes d'IA à haut risque.

prévention d'une attaque terroriste ; la détection, la localisation, l'identification ou les poursuites à l'encontre de l'auteur, ou de la personne soupçonnée, d'une infraction pénale visée à l'article 2, paragraphe 2, de la décision-cadre 2002/584/JAI du Conseil<sup>26</sup>. Cette troisième exception est particulièrement préoccupante puisqu'en admettant cette technologie pour plus d'une trentaine d'infractions, elle vide d'une grande partie de son effectivité l'interdiction de principe.

24. En définitive, la CNCDH recommande donc l'interdiction de l'identification biométrique à distance des personnes dans l'espace public et les lieux accessibles au public, en raison des risques d'atteinte grave aux droits et libertés fondamentaux liés à une remise en cause, réelle ou supposée, de l'anonymat dans l'espace public<sup>27</sup>. Elle admet toutefois la légitimité des deux premiers types d'exceptions envisagées par la proposition de règlement, tout en insistant sur la nécessité d'en assurer, le cas échéant, un strict encadrement. L'exception devrait ainsi notamment se limiter à la recherche ciblée de victimes potentielles ou à la prévention d'une menace grave et imminente pour la vie ou la sécurité des personnes et celle des ouvrages, installations et établissements d'importance vitale.

**Recommandation n°6 : La CNCDH recommande d'interdire l'identification biométrique à distance des personnes dans l'espace public et les lieux accessibles au public, en admettant par exception son utilisation, dès lors que celle-ci est strictement nécessaire, adaptée et proportionnée pour la prévention d'une menace grave et imminente pour la vie ou la sécurité des personnes et celle des ouvrages, installations et établissements d'importance vitale.**

25. Après avoir passé en revue les utilisations de l'IA jugées inacceptables par la proposition de règlement, la CNCDH s'interroge sur la méthode suivie par la Commission européenne pour parvenir à cette identification, à défaut d'indications suffisantes à ce sujet. En s'inspirant du « triple test » issu de la jurisprudence de la Cour européenne des droits de l'homme, entériné par les jurisprudences du Conseil constitutionnel et du Conseil d'État<sup>28</sup>, la CNCDH estime que, pour être considérée comme légitime, l'atteinte apportée par un système d'IA à une liberté doit être « *adaptée, nécessaire et proportionnée* » : adaptée, c'est-à-dire pertinente par rapport à l'objectif légitime

---

26 Notamment : participation à une organisation criminelle, trafic illicite d'armes, de munitions et d'explosifs, corruption, crimes contre l'environnement, y compris le trafic illicite d'espèces animales menacées et le trafic illicite d'espèces et d'essences végétales menacées, aide à l'entrée et au séjour irréguliers, racisme et xénophobie, vols organisés ou avec arme, trafic illicite de biens culturels, y compris antiquités et œuvres d'art, escroquerie, racket et extorsion de fonds, falsification de documents administratifs et trafic de faux, trafic illicite de substances hormonales et autres facteurs de croissance, trafic illicite de matières nucléaires et radioactives, trafic de véhicules volés.

27 Dans le même sens : EDPB et CEPD, Avis conjoint 05/2021 sur la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle, 18 juin 2021).

28 CC, Décision n° 2008-562 DC du 21 février 2008, *Loi relative à la rétention de sûreté et à la déclaration d'irresponsabilité pénale pour cause de trouble mental*.

poursuivi ; nécessaire, dès lors qu'elle ne doit pas excéder ce qu'exige la réalisation de cet objectif et que d'autres moyens moins attentatoires à la liberté n'étaient pas envisageables ; proportionnée, en ce qu'elle ne doit pas, par les charges qu'elle crée, être hors de proportion avec le résultat recherché<sup>29</sup>.

## 1.2. La nécessité d'étendre l'interdiction à d'autres domaines.

26. La CNCDH a opté dans le présent avis pour une approche globale et transversale des systèmes d'IA. Elle n'a donc pas procédé à un examen détaillé des différentes applications susceptibles de porter une atteinte excessive aux droits fondamentaux justifiant leur interdiction. Ce type d'analyse pourra être mené dans des avis futurs consacrés à de nouvelles technologies ou à des secteurs d'activité particuliers. D'ores et déjà toutefois, en complément des observations relatives aux cas d'usage prohibés par la proposition de règlement UE sur l'IA, la CNCDH souhaiterait mettre en avant deux types d'utilisation particulièrement préoccupants pour le respect des droits de l'Homme : la justice prédictive et la reconnaissance des émotions à l'appui d'un processus de sélection.

### 1.2.1. L'IA en justice.

27. Des applications telles que celles utilisées aux États-Unis pour évaluer le risque de récidive de personnes condamnées<sup>30</sup> font peser une grave menace sur les droits fondamentaux de ces dernières. Le manque de transparence des logiciels utilisés (conçus par des sociétés privées) remet en cause leur compatibilité avec les droits fondamentaux des personnes et la garantie des droits de la défense. En effet, sous couvert du droit au secret des affaires et de la propriété intellectuelle, les concepteurs de ces logiciels ne sont pas tenus de partager le code source de leurs algorithmes, d'où résulte leur mode d'emploi. Il est donc impossible à la fois pour le juge et les parties au procès de comprendre précisément la méthodologie utilisée par l'algorithme pour produire ses résultats<sup>31</sup>.

28. D'ores et déjà ce type de recours à l'IA est interdit en France puisqu'« aucune décision de justice impliquant une appréciation sur le comportement d'une personne ne peut avoir pour fondement un traitement automatisé de données à caractère personnel

---

29 Voir not. : M. Guyomar, « Le passeport biométrique au contrôle : empreintes et cliché », *Actualité Juridique Droit Administratif*, 2012, p. 35 ; J.-M. Sauvé, « Le principe de proportionnalité, protecteur des libertés », Institut Portalis, Aix-en-Provence, 17 mars 2017.

30 Voir not. COMPAS, l'algorithme d'évaluation et d'appréciation des risques de récidive utilisé dans plusieurs États des États-Unis.

31 Voir : Cour suprême du Wisconsin, 13 juillet 2016, l'affaire *State v. Loomis*.

destiné à évaluer certains aspects de la personnalité de cette personne »<sup>32</sup>. La formule ainsi retenue n'écarte cependant pas toute possibilité de fournir aux magistrats une application d'IA répondant à d'autres finalités, par exemple pour automatiser le calcul de l'indemnisation d'un préjudice.

29. Au regard de la charge de travail qui pèse actuellement sur les magistrats, les garanties qui pourraient être prévues afin de préserver l'impartialité du juge (explicabilité/intelligibilité du fonctionnement de l'algorithme, renforcement des moyens de remédier aux biais d'automatisation) ne paraissent pas suffisants pour endiguer le risque d'une reprise quasi systématique des résultats de la machine. En outre, l'intervention d'un système d'IA pour apporter une assistance au juge pourra éveiller des doutes chez le justiciable à l'égard de son impartialité. Or, le juge ne doit pas seulement être indépendant et impartial, il doit également en donner l'apparence<sup>33</sup>. C'est donc le droit à un procès équitable qui est ainsi remis en cause par ce type de logiciel.

30. Par ailleurs, si l'on peut admettre que le juge dispose d'un outil informatique pour faciliter son évaluation de l'indemnisation des victimes d'un préjudice<sup>34</sup>, celui-ci ne saurait reposer sur de l'apprentissage machine, trop opaque pour satisfaire à l'exigence d'explicabilité que les justiciables sont en droit d'attendre<sup>35</sup>. La CNCDH relève par ailleurs que le ministère de la Justice a mis fin prématurément, en janvier dernier, à son expérimentation d'un tel logiciel, en réalisant la multiplicité des critères à prendre en compte pour caractériser l'étendue d'un préjudice corporel et l'importance excessive des moyens à mobiliser pour étudier et prévenir les biais algorithmiques afin d'atteindre un niveau de performance satisfaisant<sup>36</sup>.

**Recommandation n°7 : La CNCDH recommande de poursuivre et d'approfondir la réflexion afin d'identifier les apports et les limites d'une utilisation de l'IA dans le cadre des procédures juridictionnelles.**

### 1.2.2. L'IA et la reconnaissance des émotions.

32 Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, art. 47 modifié par l'ordonnance n°2018-1125 du 12 décembre 2018. Sur ce point, la loi française va plus loin que le RGPD puisque ce dernier prohibe seulement les décisions, notamment de justice, fondées « exclusivement » sur un traitement automatisé.

33 D'après la théorie des apparences, consacrée par la Cour européenne des droits de l'homme, « *justice must not only be done, it must also be seen to be done* » (il faut non seulement que justice soit faite, mais aussi qu'elle le soit au vu et au su de tous).

34 L'analyse automatisée des décisions de justice repose, techniquement, le plus souvent sur le traitement du langage naturel et l'apprentissage machine : ils peuvent représenter un outil d'information utile pour les professionnels du droit.

35 Sur ce point, la loi n° 2016-1321 du 7 octobre 2016 dite « République numérique » impose la transparence des algorithmes publics, offrant ainsi une garantie contre un éventuel phénomène de « boîte noire » en matière d'usages judiciaires de l'intelligence artificielle. Voir la réponse du ministre de la Justice publiée dans le JO Sénat du 01/10/2020, page 4462, disponible en ligne : <https://www.senat.fr/questions/base/2020/qSEQ200616942.html>.

36 E. Marzolf, « Le ministère de la Justice renonce à son algorithme DataJust », *Acteurspublics*, 14 janvier 2022.

31. Les technologies de reconnaissance des émotions reposent sur un postulat dont la scientificité fait défaut, à savoir que les émotions sont détectables par des expressions du visage ou, plus largement, par des manifestations corporelles. Or, comme le rappelle le CAHAI, « 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 »<sup>37</sup>. Outre la performance du système d'IA, à laquelle des progrès dans la conception pourraient remédier, son caractère peut être inapproprié. Plusieurs instances, nationales<sup>38</sup> et internationales<sup>39</sup>, ont déjà exprimé leur préoccupation à cet égard, les instances européennes de protection des données ayant même préconisé récemment d'interdire la déduction d'émotions par l'utilisation de l'IA, sauf dans certains cas bien précis « notamment à des fins de santé et de recherche »<sup>40</sup>.

32. Partageant les mêmes craintes, la CNCDH recommande donc d'appliquer en la matière un principe d'interdiction, à moins de pouvoir démontrer que cette technologie biométrique est en mesure de renforcer l'autonomie des personnes, ou plus largement l'effectivité de leurs droits fondamentaux. À cette fin et malgré ses approximations, cette technologie d'IA peut par exemple favoriser des activités d'apprentissage auprès de personnes en situation de handicap ou s'avérer utile pour d'autres interactions humain-machine (des robots de compagnie pour les personnes âgées par exemple).

**Recommandation n°8 : La CNCDH recommande d'interdire les technologies de reconnaissance des émotions, en admettant par exception leur utilisation dès lors qu'elles visent à renforcer l'autonomie des personnes, ou plus largement l'effectivité de leurs droits fondamentaux.**

## 2. Un encadrement garant du respect des droits fondamentaux.

33. Même si beaucoup d'applications de l'IA ne portent pas une atteinte disproportionnée aux droits et libertés fondamentaux justifiant leur interdiction, la CNCDH appelle cependant les pouvoirs publics à faire respecter certaines garanties par les acteurs publics et privés, aux stades de la conception, du développement et de l'utilisation de l'IA. Cela doit passer essentiellement par un contrôle et une supervision du système d'IA, à tous les stades de son cycle de vie, au regard de son impact sur les droits fondamentaux. Outre cette vigilance à exercer à l'égard du système envisagé globalement (2.1), les décisions issues de sa mise en œuvre doivent être assorties de garanties en mesure de protéger les individus (2.2).

37 CAHAI, Étude de faisabilité, p. 8.

38 DDD, « Technologies biométriques : l'impératif respect des droits fondamentaux », 2021 ; CNIL, « Reconnaissance faciale : pour un débat à la hauteur des enjeux », 2019.

39 CEPD, EDPB, Avis conjoint 05/2021 de l'EDPB et du CEPD sur la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle, 18 juin 2021.

40 CEPD, EDPB, *ibid.*, p.14

## 2.1. Un contrôle à tous les stades de développement du système IA.

34. En fonction des domaines concernés (organisation du travail, calcul des prestations sociales, des aides aux entreprises, etc.), opter pour une solution IA peut avoir des répercussions sur les salariés ou sur les personnes visées par les décisions du logiciel. Une approche de l'IA fondée sur les droits de l'Homme suppose l'implication des personnes concernées par son déploiement, en particulier les plus vulnérables<sup>41</sup> : 1. en amont, la réalisation d'une étude d'impact sur les droits fondamentaux représente une étape essentielle, dont le résultat conditionnera le niveau de participation des parties prenantes à ce stade ; 2. ensuite, la phase de supervision de l'exécution du système d'IA, doit les inclure afin d'évaluer si son fonctionnement conduit à des atteintes aux droits fondamentaux ; 3. plus généralement, l'étendue de ce contrôle humain suppose de renforcer l'offre de formation des professionnels et de sensibilisation des particuliers.

### 2.1.1. Une étude d'impact sur les droits fondamentaux : un préalable nécessaire.

35. À côté de certains usages de l'IA à prohiber en raison de la gravité des risques qu'ils font peser sur les droits et libertés fondamentaux<sup>42</sup>, d'autres usages peuvent avoir un impact plus ou moins négatif sur les droits fondamentaux. Ces effets négatifs sont évoqués le plus souvent à partir d'une référence aux secteurs « sensibles » dans lesquels le système est déployé, tels que la police, la justice, ou la santé, ou bien encore à partir d'un inventaire des droits et libertés fondamentaux susceptibles d'être mis en cause par telle ou telle technologie d'IA<sup>43</sup>.

### *L'approche du règlement IA de l'UE : une définition a priori et centralisée des risques pour les droits fondamentaux.*

36. De son côté, la proposition de règlement réserve un traitement particulier aux systèmes d'IA qualifiés « à haut risque », identifiés comme tels par la Commission européenne en raison de l'« ampleur » de leur incidence négative sur la santé, la sécurité ou sur les droits fondamentaux. Il s'agit de systèmes d'IA qui interviennent dans des domaines « sensibles » : identification biométrique, gestion et exploitation des infrastructures critiques, éducation et formation, emploi, accès aux services privés essentiels, aux services publics et aux prestations sociales, police, justice, gestion de la migration. L'évaluation de l'impact des systèmes d'IA sur les droits de l'Homme est par conséquent centralisée et réalisée *a priori* : les systèmes à haut risque figurent dans une liste<sup>44</sup> que la Commission a la possibilité de compléter à certaines conditions et

---

41 Voir l'avis de la CNCDH, *Avis « Pour une approche fondée sur les droits de l'Homme »*, Assemblée plénière du 3 juillet 2018, JORF n°0161 du 14 juillet 2018, texte n° 104.

42 Voir plus haut.

43 Voir not. l'étude de faisabilité du CAHAJ.

44 Annexe 3 de la proposition de règlement.

selon une procédure susceptible de prendre du temps<sup>45</sup>. Demeure bien évidemment pour les responsables de traitements de données personnelles l'obligation de procéder, conformément aux exigences du RGPD, à une analyse d'impact relative à la protection des données et, plus largement, aux droits et libertés des personnes physiques<sup>46</sup>.

37. La CNCDH émet des réserves sur ce dispositif pour plusieurs raisons. D'abord, l'innovation technologique évolue toujours plus vite que la réglementation, et la liste des systèmes d'IA à haut risque retenue par les institutions européennes pourrait ne pas prendre en compte des utilisations présentes et à venir particulièrement préoccupantes pour les droits fondamentaux. Ensuite, l'IA peut présenter la particularité, notamment s'agissant de l'apprentissage machine dans sa version la plus automatisée (*deep learning*), d'évoluer de manière autonome par rapport aux intentions de ses concepteurs : c'est pourquoi des risques ignorés au moment de la conception du système peuvent apparaître au stade du développement de l'algorithme puis de son utilisation<sup>47</sup>. Surtout, l'inscription dans cette liste engendre une série d'obligations, principalement à la charge des fournisseurs, relatives à la gestion des risques, l'intégrité des données, le contrôle et le suivi du système, etc. tandis que peu d'obligations pèsent sur l'utilisateur d'un système d'IA à haut risque<sup>48</sup>.

### ***Une étude d'impact sur les droits fondamentaux à la charge des utilisateurs.***

38. La CNCDH considère qu'il faut aller plus loin en exigeant également de la part de l'utilisateur du système d'IA qu'il réalise une étude de son impact sur les droits fondamentaux. Pour plusieurs motifs : d'abord, parce cette étude responsabilise l'organisme public ou privé qui souhaite recourir à cette option informatique ; ensuite, parce qu'elle pourra éventuellement permettre de nourrir un dialogue entre toutes les parties prenantes (salariés, clients, usagers d'un service public, etc.) sur l'opportunité de son utilisation ; enfin, parce qu'elle constituera une source d'information pour la supervision du système d'IA tout au long de son cycle de vie, voire aussi pour la personne visée *in fine* par une décision automatisée.

39. S'il n'est pas question dans le cadre de cet avis à portée générale de détailler les éléments à prendre en compte pour mener cette analyse, la CNCDH souhaiterait

---

45 La Commission dispose en effet du pouvoir de modifier la liste de l'annexe 3 par des « actes délégués » dont la procédure d'adoption figure dans l'article 73 de la proposition de règlement : 6 mois peuvent passer entre la notification de la proposition de modification et son entrée en vigueur, sachant que le Parlement européen et le Conseil peuvent s'y opposer.

46 Conformément à l'article 35 du RGPD. La proposition de règlement prend d'ailleurs le soin de le rappeler à l'article 29.

47 Par exemple, l'agent conversationnel (« chatbot ») conçu par Microsoft, et destiné en 2016 à participer à des échanges sur les réseaux sociaux, a rapidement tenu des propos injurieux et racistes, à rebours de l'intention de ses programmeurs. Cette dérive est le fruit de l'interaction du système d'IA avec des personnes mal intentionnées : M. Tual, « A peine lancée, une intelligence artificielle de Microsoft dérape sur Twitter », *Le Monde*, 24 mars 2016.

48 Voir l'article 29 de la proposition de règlement IA et l'article 52 pour certains systèmes d'IA plus particuliers.

formuler un certain nombre de recommandations relatives aux grandes lignes qui pourraient orienter sa réalisation.

40. Elle recommande ainsi qu'il revienne à l'utilisateur d'évaluer l'impact du recours au système d'IA sur les droits fondamentaux et, en cas de risques identifiés, de procéder à leur évaluation en tenant compte de la probabilité et de la gravité de ces derniers. Cela concernerait par exemple des algorithmes de détection des fraudes à la réglementation fiscale, ou aux prestations sociales, les systèmes d'aide à la gestion du personnel, les logiciels de commande vocale dans les plateformes de distribution, etc.

41. Cette analyse inclura éventuellement la communication d'éléments fournis par le concepteur, lorsqu'ils ont une incidence sur les droits fondamentaux : les jeux de données utilisés pour l'apprentissage de la machine, par exemple pour les logiciels de détection de comportements suspects destinés à équiper des caméras de surveillance ; les différents types de paramétrage possible, par exemple s'agissant des outils de modération automatisée de contenu, etc. L'utilisateur devra par conséquent indiquer et justifier ses choix de paramétrage en mettant en perspective la finalité recherchée et les risques d'atteinte aux droits fondamentaux.

### ***Le contenu de l'étude d'impact.***

42. S'agissant de son contenu, toute analyse d'impact devrait, au minimum, consister dans un premier temps à mentionner les finalités attachées à l'utilisation du système d'IA envisagé, et à identifier les droits fondamentaux susceptibles d'être affectés.

43. Dans un second temps, cette étude devrait faire apparaître les réponses apportées par l'utilisateur aux questions qui alimentent habituellement le régime d'encadrement des droits et libertés fondamentaux :

- Est-il nécessaire de recourir à un système d'IA pour la tâche considérée ? Dans quelle mesure apporte-t-il une plus-value au fonctionnement antérieur ?
- Le système d'IA est-il adapté à la finalité poursuivie ? Dans quelle mesure parviendra-t-il à accomplir la tâche considérée ?
- Le système d'IA est-il proportionné ? Dans quelle mesure l'atteinte susceptible d'être portée aux droits de l'Homme, y compris l'impact environnemental, est-elle justifiée par rapport aux bénéfices attendus pour réaliser l'objectif légitime assigné au système ?

44. Enfin, et sur un plan plus technique, l'étude d'impact devrait faire figurer les procédures mises en place pour assurer le suivi de l'application, ainsi que les mesures

d'atténuation des risques encourus<sup>49</sup>.

***Les modalités de consultation des parties prenantes conditionnées par les conclusions de l'étude d'impact.***

45. En fonction du niveau de risque pour les droits fondamentaux identifié au terme de l'étude d'impact, une consultation avec les parties prenantes devrait éventuellement être prévue afin d'échanger sur le choix de recourir ou non à la solution d'IA envisagée par la direction. Trois niveaux de risques pourraient être retenus afin de déterminer les modalités de cette consultation :

- un niveau élevé (pour tous les traitements ayant une incidence sur les droits des personnes) : une consultation de l'ensemble des parties prenantes, associant notamment les représentants du personnel, les associations d'usagers des services publics ou de consommateurs, en veillant à inclure des associations de personnes défavorisées ;
- un niveau modéré (pour les traitements n'ayant pas directement une incidence sur les droits des personnes, par exemple un logiciel de gestion des congés) : la consultation évoquée en cas de niveau élevé serait facultative, et pourrait être par exemple sollicitée par les personnes concernées ou leurs représentants ;
- un niveau faible : l'étude d'impact serait communiquée à la demande des personnes concernées par le système d'IA ou de leurs représentants.

46. Une étude d'impact à la charge de l'utilisateur est d'autant plus nécessaire qu'elle facilitera la supervision du système d'IA une fois qu'il sera mis en place.

**Recommandation n°9 :** La CNCDH recommande que l'utilisateur d'un système d'IA évalue l'impact du recours à ce système sur les droits fondamentaux et, en cas de risques identifiés, procède à leur évaluation en tenant compte de la probabilité et de la gravité de ces derniers. L'étude d'impact devrait inclure au minimum :

- une mention de la ou des finalités attachées à l'utilisation du système d'IA envisagé ;
- une identification des droits fondamentaux susceptibles d'être affectés par le système ;
- un examen du système d'IA envisagé, à partir d'une évaluation de sa nécessité, de son caractère adapté, et de la proportionnalité des atteintes portées aux droits fondamentaux par rapport au but escompté ;
- les procédures mises en place pour assurer le suivi de l'application, et les mesures d'atténuation au regard des risques encourus.

---

<sup>49</sup> Le groupe d'experts indépendants de haut niveau sur l'IA (UE) renvoie à ce type d'analyse, en amont de l'auto-évaluation qu'elle préconise et qu'elle fait reposer sur une check-list multi-critères pour une « IA digne de confiance » : High Level Expert Group on Artificial Intelligence, The Assessment List for Trustworthy Artificial Intelligence (ALTAI), juin 2020.

**Recommandation n°10 :** En fonction des risques engendrés par un système d'IA sur les droits fondamentaux dans un contexte d'usage particulier, la CNCDH recommande d'assurer préalablement à la décision d'y recourir, une consultation des parties prenantes, selon des modalités adaptées, en incluant par exemple les représentants du personnel et, plus largement, les personnes visées par le système d'IA.

### 2.1.2. Une supervision du système tout au long du cycle de vie.

47. La qualité et la pertinence des données sélectionnées pour concevoir les algorithmes et le caractère proportionné des éventuelles atteintes portées aux droits fondamentaux par le système d'IA peuvent faire l'objet d'un contrôle en amont de l'utilisation du système d'IA. Toutefois, des atteintes aux droits fondamentaux peuvent survenir après sa prise en main par l'utilisateur. C'est pourquoi une vigilance continue à l'égard du fonctionnement du système d'IA doit être assurée. Cette supervision doit être organisée, à l'instar de ce qui a été exposé plus haut s'agissant de la consultation des parties prenantes à l'issue de l'étude d'impact, selon des modalités plus ou moins exigeantes pour l'utilisateur en fonction des risques pour les droits fondamentaux identifiés par cette étude.

#### ***Une vigilance en continue à l'égard des effets du système d'IA sur les droits fondamentaux.***

48. Un contrôle périodique doit être assuré aux différentes étapes de l'utilisation d'un système d'IA. Il doit se fonder sur l'analyse d'impact à l'égard des droits de l'Homme en s'assurant que les risques identifiés en amont ne se sont pas concrétisés ou, le cas échéant, identifier les mesures à prendre pour neutraliser ces risques. Des atteintes aux droits fondamentaux non identifiées par l'étude d'impact peuvent aussi être relevées à cette occasion.

49. À cet égard, une attention particulière pourrait être portée aux risques de discrimination engendrés par les systèmes d'IA. Ces risques ont déjà été largement documentés<sup>50</sup>. Alors que l'automatisation des processus décisionnels pourrait laisser penser au premier abord qu'ils s'affranchissent des préjugés inhérents à la subjectivité humaine, les algorithmes peuvent reproduire, renforcer ou générer des biais, notamment systémiques, susceptibles d'aggraver des situations discriminatoires. Comme le souligne le Défenseur des droits à la suite de la Commissaire aux droits de l'Homme du Conseil de l'Europe, il existe alors un risque majeur d'« essentialisation » et de renforcement des « stéréotypes » car le caractère prédictif de l'algorithme est fondé sur le comportement ou les caractéristiques homogénéisées de groupes.

50. Ces biais peuvent être la conséquence des données utilisées pour alimenter

---

<sup>50</sup> Voir not. les rapports du Défenseur des droits: cf *supra*, nbp n°4.

la machine. Les jeux de données utilisés pour entraîner le modèle algorithmique peuvent en effet inclure des discriminations, par exemple si l'on attend d'une IA qu'elle sélectionne les meilleures candidatures pour un poste donné à partir des seuls dossiers de personnes recrutées précédemment, le modèle pourra reproduire des biais discriminatoires dès lors qu'ils caractérisaient ces recrutements (raciste ou sexiste par exemple). Ces jeux de données peuvent aussi ne pas être suffisamment représentatifs de la diversité de la population, par exemple si des logiciels de reconnaissance faciale sont alimentés principalement par des photos de personnes de type caucasien ; ceux-ci auront tendance à commettre des erreurs d'identification sur les personnes de couleur noire, à l'origine d'interpellations injustifiées.

51. Les classifications algorithmiques risquent donc de générer des discriminations envers les personnes, en raison de leur appartenance à un groupe, qui peut recouper indirectement un groupe protégé par le droit de la non-discrimination (par un proxy, c'est-à-dire une variable liée à un critère de discrimination prohibé – par exemple une habitude alimentaire qui témoignerait des convictions religieuses).

52. En outre, les effets discriminatoires de l'algorithme sont peu perceptibles à l'échelle individuelle, compte tenu de l'opacité du fonctionnement des dispositifs algorithmiques mais aussi du fait que ces discriminations s'observent beaucoup mieux à l'échelle de groupes qu'à l'échelle de l'individu<sup>51</sup>. C'est pourquoi, une vigilance en continu doit s'exercer sur les systèmes d'IA et la classification algorithmique lorsque ces systèmes produisent des résultats ayant des effets, même indirects, sur les droits et libertés des personnes<sup>52</sup>.

### ***Les modalités de la supervision par l'utilisateur.***

53. La proposition de règlement relatif à l'IA contraint les fournisseurs de systèmes d'IA à haut risque à instaurer un système de gestion de la qualité, censé documenter l'ensemble des procédures et instructions qu'ils mettent en place pour se conformer aux exigences du règlement<sup>53</sup>. Ces instructions doivent notamment inclure les modalités de la surveillance établie par le fournisseur après la commercialisation d'un système d'IA, afin qu'il respecte durablement la réglementation<sup>54</sup>. Le contrôle sur la manière dont le fournisseur assure cette surveillance est confié par la proposition de règlement à des organismes d'évaluation de la conformité homologués<sup>55</sup>, voire relève d'une évaluation interne du fournisseur<sup>56</sup>.

---

51 DDD, en partenariat avec la CNIL, « Algorithmes : prévenir l'automatisation des discriminations », 2020, p. 6.

52 Dans le même sens, le DDD et la CNIL recommandent un contrôle régulier des effets des algorithmes après leur déploiement, sur le modèle du contrôle des effets indésirables des médicaments : DDD et CNIL, « Algorithmes : prévenir l'automatisation des discriminations », p. 10.

53 Proposition de règlement IA, art. 17.

54 Proposition de règlement IA, art. 61.

55 Les « organismes notifiés ».

56 Proposition de règlement IA, art. 43.

54. La CNCDH estime cette approche nécessaire mais insuffisante lorsque sont concernés des systèmes d'IA pour lesquels l'étude d'impact révèle un risque important d'atteinte aux droits fondamentaux. Dans ce cas-là, la supervision devrait reposer en complément sur des tests et des sondages mis en œuvre par les parties prenantes, l'utilisateur et les personnes concernées, selon une périodicité à déterminer en fonction du contexte d'usage. De ce point de vue, le « contrôle humain » prévu par l'article 14 de la proposition de règlement de l'UE, en particulier pour « *appréhender totalement les capacités et les limites du système d'IA* », mériterait d'être précisé et complété par une référence à la collégialité du processus et à l'ouverture de ce collège aux différentes parties prenantes.

55. Le contrôle relatif à l'impact d'un système d'IA de recrutement ou de management au sein d'une entreprise ou d'une administration doit, par exemple, inclure une participation des représentants du personnel. Ces derniers sont en effet des acteurs de premier plan, notamment au sein du comité social et économique (CSE) de l'entreprise, pour évaluer les risques psycho sociaux engendrés par un système d'IA au sein de leur organisme mais également son impact sur l'organisation du travail.

56. Si l'étude d'impact ne révèle pas de risques importants pour les droits fondamentaux, la supervision du système d'IA pourra relever uniquement de l'utilisateur. La CNCDH insiste en effet sur la nécessité de mettre en avant son rôle dans la vigilance à maintenir tout au long du déploiement du système, au-delà de la simple communication de « *données pertinentes* » au fournisseur pour satisfaire à l'obligation de surveillance qui incombe à ce dernier, conformément à la proposition de règlement<sup>57</sup>.

**Recommandation n°11 :** La CNCDH recommande de mettre en place une supervision du système d'IA, selon une procédure susceptible de varier selon les risques d'atteintes aux droits fondamentaux tels qu'identifiés par l'étude d'impact, afin de maintenir une vigilance en continu de la part de l'utilisateur à l'égard des effets du système, notamment ses effets discriminatoires.

### 2.1.3. Une formation et une sensibilisation aux enjeux de l'IA.

57. Vu la place à accorder aux parties prenantes dans le contrôle et la supervision des systèmes d'IA, notamment les salariés et les personnes concernées par les décisions du système, la CNCDH recommande la mise en place de formations et de campagnes de sensibilisation aux technologies d'IA. Des modules de formation pourraient être largement diffusés auprès des salariés et de tout un chacun, par exemple sous forme de MOOC. La CNCDH recommande donc des investissements publics dans la conception d'outils de formation et d'information accessibles au plus grand nombre.

**Recommandation n°12 :** La CNCDH recommande de favoriser les investissements pu-

---

<sup>57</sup> Selon les termes de l'article 61 de la proposition de règlement IA.

## **blics dans la conception d'outils de formation et d'information accessibles au plus grand nombre.**

58. En outre, les pouvoirs publics devraient organiser des débats publics sur la question. Sur le modèle des États généraux de la bioéthique organisés par le Comité consultatif national d'éthique (CCNE), ces consultations auraient une double vocation : d'une part, informer les citoyens du fonctionnement de ces systèmes, de leur finalité, et d'autre part, les mettre en capacité de se positionner à l'égard des orientations nationales en la matière. Ce faisant, il s'agit de favoriser l'expression d'une diversité de points de vue sur un certain nombre d'usages de l'IA. Une attention particulière devrait être accordée aux personnes les plus démunies, afin de s'assurer qu'elles soient en mesure d'y participer.

**Recommandation n°13** : La CNCDH recommande l'organisation de consultations nationales sur le modèle des États généraux de la bioéthique organisés par le Comité consultatif national d'éthique.

59. La part prise par les algorithmes dans la vie quotidienne et le fonctionnement de la société appelle à l'acquisition d'une véritable culture informatique dès le plus jeune âge. La CNCDH souligne ainsi la nécessité pour l'Éducation nationale de renforcer la formation des élèves aux enjeux techniques, politiques et sociétaux de l'intelligence artificielle et de proposer, à cette fin, des supports pédagogiques à destination des enseignants.

**Recommandation n°14** : La CNCDH recommande à l'Éducation nationale de renforcer la formation des élèves aux enjeux techniques, politiques et sociétaux de l'intelligence artificielle et de proposer, à cette fin, des supports pédagogiques à destination des enseignants.

## **2.2. Des garanties du respect des droits fondamentaux à l'égard des décisions individuelles.**

60. La part accordée à l'algorithme dans la prise de décision varie d'un usage à l'autre : une intervention humaine peut être prévue, voire parfois requise, mais dans certains cas la décision peut aussi être totalement automatisée. La nécessité et les modalités de cette intervention dépendent du niveau de risque d'atteinte à des droits fondamentaux. Par ailleurs, les personnes visées par un système d'IA doivent en être informées, et disposer également d'informations intelligibles tant sur le fonctionnement de l'algorithme que sur la part prise par cet algorithme dans la décision individuelle.

### **2.2.1. Une intervention humaine garante de la prise en compte des spécificités individuelles.**

61. Pour « permettre à l'homme de garder la main »<sup>58</sup> sur l'IA, la CNCDH appelle à réintroduire de l'humain au terme du processus de décision automatisée : soit au niveau de l'utilisateur, chargé d'assurer un contrôle sur le résultat produit par l'algorithme, soit au niveau de la personne concernée.

***Une intervention humaine au niveau de l'utilisateur : un contrôle du résultat algorithmique.***

62. Dans certains cas, la supervision humaine du fonctionnement général du système d'IA, évoquée plus haut, requiert en complément une intervention humaine à l'égard des décisions individuelles fondées sur les résultats de ce système. Le RGPD a consacré un droit « *de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire* »<sup>59</sup>. Le législateur a pris soin de l'inscrire dans la loi de 1978 relative à l'informatique et aux libertés en précisant, vu le champ d'application du RGPD, que la prohibition concerne tout « *traitement automatisé de données à caractère personnel* »<sup>60</sup>.

63. La réglementation actuelle présente cependant une double limite : d'une part, elle ne couvre pas les traitements algorithmiques alimentés par des données anonymisées et, d'autre part, elle envisage de nombreuses dérogations à son interdiction de principe, à commencer par la faculté laissée aux États membres de l'UE d'autoriser ce type de traitement, à condition toutefois de prévoir des « *mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée* ».

64. La nécessité d'une intervention humaine a déjà été soulignée par le Conseil constitutionnel, pour les décisions individuelles adoptées à partir d'un algorithme dont les modalités de fonctionnement ne peuvent pas être communiquées (parce que leur communication porterait atteinte au secret défense, à la sûreté de l'État, etc)<sup>61</sup>. Plus récemment, s'agissant du dispositif français de détection des connexions susceptibles de révéler une menace terroriste, la CJUE a fait valoir que « *tout résultat positif obtenu à la suite d'un traitement automatisé doit être soumis à un réexamen individuel par des moyens non automatisés avant l'adoption d'une mesure individuelle produisant des effets préjudiciables à l'égard des personnes concernées* »<sup>62</sup>.

65. La CNCDH considère que la nécessité d'une intervention humaine devrait s'imposer de manière plus générale, à des degrés divers, selon le domaine considéré, pour l'ensemble des traitements algorithmiques produisant des effets sur les droits des

---

58 Selon la formule consacrée par la CNIL dans son rapport de 2017. Cf supra.

59 Proposition de règlement IA, art. 22.

60 Loi 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, art. 47, al. 2.

61 CC, Décision n° 2018-765 DC, *Loi relative à la protection des données personnelles*, § 70.

62 CJUE, 6 octobre 2020, aff. C 511/18, *La Quadrature du Net*, § 182.

personnes.

66. Afin d'assurer un contrôle humain effectif dans le contexte d'usage du système d'IA, les modalités de l'intervention humaine pourront varier en fonction de l'impact du système d'IA sur les droits fondamentaux, s'agissant notamment :

- de la composition de l'instance en charge du contrôle (individu ou collègue) ;
- de l'étendue des informations mises à la disposition de l'intervenant, sachant que, dans certains cas, il faudra lui communiquer des données complémentaires à celles traitées par le système ;
- du type de formation dont doivent bénéficier les personnes chargées d'intervenir ;
- du moment propice pour intervenir (à l'issue du résultat obtenu par la machine, donc en amont de la décision, ou bien ex post à la demande de la personne).

67. Assurer une intervention humaine effective suppose d'informer l'intervenant sur les caractéristiques de l'algorithme utilisé : la technologie à l'origine de sa conception, le type de données utilisées pour sa modélisation, les paramètres de fonctionnement et la pondération des critères retenus par le concepteur de l'algorithme, la fiabilité, etc. Cette information relative aux « rouages de la machine » s'impose pour favoriser une prise de distance vis-à-vis du système d'IA utilisé, et ainsi réduire le biais cognitif d'automatisation qui consiste à accorder une confiance excessive aux processus de décision automatisés. C'est pourquoi la CNCDH rappelle la nécessité pour les professionnels assistés par une application d'IA dans l'exercice de leurs fonctions (médecins, magistrats, agents administratifs, recruteurs, etc.) de bénéficier d'une information claire, complète, et compréhensible sur ces aspects.

68. Afin de neutraliser le biais d'automatisation, la CNCDH recommande également de former tout intervenant appelé à assurer un contrôle sur les résultats individuels produits par le système d'IA, qui insistera notamment sur ses limites (les biais issus des données, la nature probabiliste des résultats obtenus, etc.). En outre, la Commission recommande de ne pas faire peser de contrainte particulière, par un surcroît de formalités par exemple, sur ces personnes lorsqu'elles s'écartent de l'indication algorithmique.

69. Enfin, et s'agissant plus particulièrement des administrés, l'automatisation des décisions ne peut qu'aggraver la mise à l'écart ressentie, face à la dématérialisation croissante des services publics et à la difficulté de faire valoir leurs droits<sup>63</sup>. La CNCDH souscrit donc à la recommandation du Défenseur des droits lorsqu'il préconise « *le maintien systématique d'un accès alternatif et la possibilité d'un accompagnement suffisamment proche, compétent et disponible* »<sup>64</sup>.

---

63 Défenseur des droits, « Dématérialisation des services publics : trois ans après, où en est-on ? », Rapport, 2022.

64 *Ibid.*, p. 5.

**Recommandation n°15** : La CNCDH recommande de :

- garantir une intervention humaine pour le contrôle des décisions individuelles issues d'un système d'IA selon des modalités correspondant au niveau de risque de ce dernier ;
- en assurer l'effectivité par une formation appropriée et une information de l'intervenant sur les caractéristiques du système, sans lui imposer de contrainte particulière lorsqu'il s'écarte de la préconisation issue du système d'IA ;
- assurer aux usagers du service public le maintien systématique d'un accès alternatif à un agent humain.

***Une intervention humaine au niveau de la personne concernée : le droit au paramétrage.***

70. Dans certains cas, la personne concernée par le fonctionnement de l'IA est directement exposée aux résultats d'un algorithme conçu, selon le discours marketing qui l'accompagne, pour « répondre à ses besoins ». Sous cet angle, l'algorithme se présente comme un outil façonné par un opérateur et utilisé par les personnes qui recourent à ses services. À cet égard, la CNCDH s'inquiète particulièrement de l'absence de maîtrise par l'utilisateur des paramètres de fonctionnement du système.

71. Comme la CNCDH l'a déjà relevé dans son avis sur la lutte contre la haine en ligne, une telle maîtrise paraît notamment nécessaire s'agissant des algorithmes de sélection des contenus sur les réseaux sociaux. La liberté de conscience exige en effet de renforcer l'autonomie des utilisateurs et d'accroître leur maîtrise sur les contenus qui leur sont proposés. La CNCDH renouvelle, dès lors, sa recommandation de leur reconnaître un droit au paramétrage des critères permettant de déterminer les contenus reçus, s'agissant tant de leur sélection que de leur présentation<sup>65</sup>.

72. Plus généralement, ce droit d'agir sur les paramètres de l'algorithme traduirait une nouvelle manifestation du rôle qu'il convient de reconnaître à l'utilisateur dès lors que l'on prône une véritable « IA centrée sur l'humain », intention partagée par les autorités françaises, européennes et internationales. L'utilisateur pourrait ainsi se voir conférer un droit d'agir sur les paramètres des systèmes d'IA dès lors que ceux-ci se nourrissent des interactions interpersonnelles ou humain-machine qui sont notamment appelées à se développer avec l'essor des agents conversationnels.

**Recommandation n°16** : La CNCDH recommande de reconnaître au bénéficiaire des utilisateurs des systèmes d'IA un droit au paramétrage de leurs critères, notamment afin de déterminer la sélection et la présentation des contenus reçus, et plus généralement dans l'hypothèse des interactions humain-machine.

---

<sup>65</sup> Pour une présentation plus détaillée de ce droit et de ses implications, voir : CNCDH, *Avis relatif à la lutte contre la haine en ligne*, Assemblée plénière du 8 juillet 2021, JORF n°0170 du 24 juillet 2021, texte n° 79.

### 2.2.2. Une information garante de la dignité de la personne humaine.

73. L'information relative aux caractéristiques du système d'IA utilisé s'impose pour assurer le contrôle de la décision individuelle. Elle doit également permettre à la personne concernée d'en comprendre les motifs et, éventuellement, de la contester. Pour ce faire, la personne doit être informée que la décision dont elle fait l'objet repose, en partie ou totalement, sur un processus automatisé. Elle doit, ensuite, disposer d'éléments lui permettant de comprendre le fonctionnement de l'algorithme utilisé.

#### *L'information relative à l'intervention d'un système d'IA dans la décision.*

74. La personne exposée à un système d'IA doit en être informée. La CNCDH estime que c'est un préalable qui ne doit souffrir aucune exception.

75. La réglementation actuelle reconnaît d'ores et déjà un droit, pour les usagers de l'administration concernés par l'utilisation d'un système d'IA, d'être informés qu'une décision administrative individuelle a été prise sur le fondement d'un algorithme<sup>66</sup>. Ce droit connaît toutefois un grand nombre d'exceptions. En effet, la loi prévoit d'exclure cette information au cas où cela aurait une incidence sur : le secret des délibérations du Gouvernement et des autorités responsables relevant du pouvoir exécutif ; la défense nationale ; la conduite de la politique extérieure de la France ; la sûreté de l'État, la sécurité publique, la sécurité des personnes ou la sécurité des systèmes d'information des administrations ; la monnaie et le crédit public ; le déroulement des procédures engagées devant les juridictions ou d'opérations préliminaires à de telles procédures, sauf autorisation donnée par l'autorité compétente ; la recherche et la prévention, par les services compétents, d'infractions de toute nature ; les autres secrets protégés au titre de l'article L. 124-4 du code de l'environnement<sup>67</sup>.

76. La CNCDH regrette l'étendue de ces motifs, d'autant plus qu'ils ont été prévus initialement pour justifier l'exclusion d'une communication ou d'une consultation d'un document administratif. Il faut en effet distinguer deux types d'information : d'une part, l'information sur la nature du processus à l'origine de la décision, en l'occurrence l'utilisation d'un processus de décision automatisé et, d'autre part, l'information sur le fonctionnement de l'algorithme utilisé. Si l'on peut admettre que les règles logiques composant l'algorithme puissent échapper à une communication en raison d'un certain nombre d'impératifs publics, cela ne saurait justifier l'absence de mention que la décision individuelle repose sur l'utilisation d'un algorithme, alors même que cet élément d'information lui sera utile à l'intéressé en cas de recours juridictionnel.

77. La CNCDH estime que cette obligation d'information doit être étendue aux personnes privées, tout en étant conçue de manière plus large à l'égard des personnes

<sup>66</sup> Art. L 311-3-1 du code des relations entre le public et l'administration.

<sup>67</sup> Art. L 311-5 du code des relations entre le public et l'administration.

publiques.

78. Si la proposition de règlement de l'UE mentionne une « obligation de transparence », elle le fait de manière restrictive en la réservant à certains systèmes d'IA. Elle fait d'abord peser une obligation sur les fournisseurs de systèmes d'IA destinés à interagir avec l'être humain, en leur enjoignant de les concevoir de manière à ce que les personnes concernées en soient averties. Ensuite, elle fait peser une obligation sur les utilisateurs de systèmes de reconnaissance des émotions, ou d'un système de catégorisation biométrique, d'informer les personnes qui y sont exposées. Enfin, les manipulations d'image ou de contenus audio ou vidéo, tels les « hypertrucages » (ou *deepfake*), doivent être accompagnées d'un message d'avertissement.

**Recommandation n°17 :** La CNCDH recommande d'informer systématiquement les personnes lorsqu'elles sont exposées ou amenées à interagir avec un système d'IA et, lorsqu'elles font l'objet d'une décision, que cette dernière se fonde, le cas échéant, en partie ou totalement sur un traitement algorithmique.

79. Les personnes visées par une décision issue d'un système d'IA doivent non seulement disposer d'un droit d'en être informées, mais également disposer d'un droit de contester cette décision auprès d'un être humain. Ce dernier devra être en mesure de réexaminer le dossier de la personne. Afin de rendre effectif ce droit à un réexamen, des canaux facilement accessibles doivent être mis à la disposition des personnes concernées.

**Recommandation n°18 :** La CNCDH recommande de garantir à la personne concernée un droit au réexamen, par un être humain, de toute décision individuelle fondée totalement, ou même en partie, sur un traitement algorithmique, dès lors qu'elle emporte des conséquences significatives pour elle.

*L'information sur les modalités de la prise de décision automatisée.*

80. L'exigence d'explicabilité des systèmes d'IA figure dans la plupart des textes de référence sur l'encadrement de l'IA<sup>68</sup>. Parfois assimilée à la « transparence », en tout cas susceptible d'acceptions diverses, l'explicabilité renvoie en substance à la nécessité pour le concepteur et/ou pour l'utilisateur d'un système d'IA d'être en mesure de fournir aux personnes concernées des éléments d'information compréhensibles sur le fonctionnement de l'algorithme. Il convient de garantir un droit de disposer des moyens de comprendre les raisons pour lesquelles l'IA est parvenue à son résultat. Ce droit à l'information doit s'étendre à la communication des modalités de l'éventuelle intervention humaine dans le processus de décision.

---

68 Voir not. : OCDE, Recommandation du Conseil sur l'IA, 1.3 ; UNESCO, Recommandation sur l'éthique de l'IA, §§ 37 et s.

81. À l'heure actuelle, la réglementation n'impose cette exigence qu'à l'égard de l'administration. Celle-ci doit en effet communiquer « *sous une forme intelligible* » aux personnes ayant fait l'objet d'une décision individuelle prise sur le fondement d'un traitement algorithmique, si elles en font la demande, les informations suivantes<sup>69</sup> :

- 1° Le degré et le mode de contribution du traitement algorithmique à la prise de décision ;
- 2° Les données traitées et leurs sources ;
- 3° Les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé ;
- 4° Les opérations effectuées par le traitement.

82. La CNCDH recommande de mener une réflexion s'agissant de l'extension de cette obligation aux organismes privés dont l'activité peut affecter les droits des personnes (réseaux sociaux, banques, assurances, etc.). A cet égard, la protection de la propriété intellectuelle ne saurait constituer un obstacle dirimant : il ne s'agirait pas de rendre public le code source d'un logiciel, mais de communiquer à la personne qui en fait la demande des informations sur les éléments pris en compte par la machine (notamment les critères saillants relatifs à sa situation individuelle), dans un langage facile à comprendre, afin d'expliquer le processus ayant conduit à la décision.

83. Dans des avis ultérieurs, la CNCDH examinera la proportionnalité des restrictions apportées à l'exigence de communication de ces informations par l'administration et les organismes privés, lorsqu'elles sont susceptibles de porter atteinte à un secret protégé par la loi<sup>70</sup>, en fonction des domaines considérés.

84. L'exigence d'explicabilité vaut plus largement pour garantir l'effectivité du droit au recours à l'encontre des décisions individuelles fondées sur un algorithme.

85. La complète information de la personne qui fait l'objet d'une décision, par l'administration ou par un organisme privé, un établissement bancaire par exemple, suppose de lui communiquer des explications sur le fonctionnement de l'algorithme appliqué à sa situation personnelle et, plus encore, de lui restituer, le cas échéant, les modalités d'une intervention humaine.

**Recommandation n°19 :** La CNCDH recommande aux administrations de communiquer sous une forme intelligible les informations sur le fonctionnement de l'algorithme, ainsi que sur la part éventuellement prise par une intervention humaine dans le processus de décision. Elle recommande par ailleurs de mener une réflexion s'agissant de l'extension de cette obligation aux organismes privés.

86. Compte tenu de l'importance grandissante du déploiement des systèmes

---

69 R 311-3-1-2 du code des relations entre le public et l'administration.

70 Notamment le « secret-défense » et le secret des affaires.

d'intelligence artificielle, des enjeux majeurs qui en résultent pour le respect des droits fondamentaux, ainsi que pour la préservation de l'État de droit et de la démocratie, sans oublier celle de l'environnement, la Commission nationale consultative des droits de l'homme entend poursuivre à l'avenir ses travaux sur l'intelligence artificielle, tout particulièrement pour examiner ses impacts, notamment dans les domaines de la santé, de l'éducation, de l'emploi et de la justice.

## Synthèse des recommandations.

**Recommandation n°1** : La CNCDH recommande de privilégier, dans la communication institutionnelle, une terminologie plus neutre et objective que l'expression « intelligence artificielle » telle que « systèmes algorithmiques d'aide à la décision » (SAAD).

**Recommandation n°2** : La Commission recommande de renforcer, au sein de la proposition de règlement l'Union européenne relatif à l'IA, les dispositions propres à assurer la mise en place d'un cadre juridique contraignant, garant du respect effectif des droits fondamentaux. De surcroît, la CNCDH recommande l'adoption, dans le cadre du Conseil de l'Europe, d'une « Convention 108+ de l'IA ».

**Recommandation n°3** : La CNCDH recommande la prise en compte d'une approche fondée sur les droits de l'Homme dans les réformes en cours, dès lors qu'elles entendent garantir le respect des droits fondamentaux.

**Recommandation n°4** : La CNCDH recommande l'interdiction du recours aux interfaces de choix dès lors qu'elles ont pour objet ou pour effet de manipuler, à leur détriment, les utilisateurs en exploitant leurs vulnérabilités.

**Recommandation n°5** : La CNCDH recommande d'interdire tout type de notation sociale (« *social scoring* ») mis en place par les administrations ou par toute entreprise, publique ou privée.

**Recommandation n°6** : La CNCDH recommande d'interdire l'identification biométrique à distance des personnes dans l'espace public et les lieux accessibles au public, en admettant par exception son utilisation, dès lors que celle-ci est strictement nécessaire, adaptée et proportionnée pour la prévention d'une menace grave et imminente pour la vie ou la sécurité des personnes et celle des ouvrages, installations et établissements d'importance vitale.

**Recommandation n°7** : La CNCDH recommande de poursuivre et d'approfondir la réflexion afin d'identifier les apports et les limites d'une utilisation de l'IA dans le cadre des procédures juridictionnelles.

**Recommandation n°8** : La CNCDH recommande d'interdire les technologies de reconnaissance des émotions, en admettant par exception leur utilisation dès lors qu'elles visent à renforcer l'autonomie des personnes, ou plus largement l'effectivité de leurs droits fondamentaux.

**Recommandation n°9** : La CNCDH recommande que l'utilisateur d'un système d'IA évalue l'impact du recours à ce système sur les droits fondamentaux et, en cas de risques

identifiés, procède à leur évaluation en tenant compte de la probabilité et de la gravité de ces derniers. L'étude d'impact devrait inclure au minimum :

- une mention de la ou des finalités attachées à l'utilisation du système d'IA envisagé ;
- une identification des droits fondamentaux susceptibles d'être affectés par le système ;
- un examen du système d'IA envisagé, à partir d'une évaluation de sa nécessité, de son caractère adapté, et de la proportionnalité des atteintes portées aux droits fondamentaux par rapport au but escompté ;
- les procédures mises en place pour assurer le suivi de l'application, et les mesures d'atténuation au regard des risques encourus.

**Recommandation n°10** : En fonction des risques engendrés par un système d'IA sur les droits fondamentaux dans un contexte d'usage particulier, la CNCDH recommande d'assurer préalablement à la décision d'y recourir, une consultation des parties prenantes, selon des modalités adaptées, en incluant par exemple les représentants du personnel et, plus largement, les personnes visées par le système d'IA.

**Recommandation n°11** : La CNCDH recommande de mettre en place une supervision du système d'IA, selon une procédure susceptible de varier selon les risques d'atteintes aux droits fondamentaux tels qu'identifiés par l'étude d'impact, afin de maintenir une vigilance en continu de la part de l'utilisateur à l'égard des effets du système, notamment ses effets discriminatoires.

**Recommandation n°12** : La CNCDH recommande de favoriser les investissements publics dans la conception d'outils de formation et d'information accessibles au plus grand nombre.

**Recommandation n°13** : La CNCDH recommande l'organisation de consultations nationales sur le modèle des États généraux de la bioéthique organisés par le Comité consultatif national d'éthique.

**Recommandation n°14** : La CNCDH recommande à l'Éducation nationale de renforcer la formation des élèves aux enjeux techniques, politiques et sociétaux de l'intelligence artificielle et de proposer, à cette fin, des supports pédagogiques à destination des enseignants.

**Recommandation n°15** : La CNCDH recommande de :

- garantir une intervention humaine pour le contrôle des décisions individuelles issues d'un système d'IA selon des modalités correspondant au niveau de risque de ce dernier ;
- en assurer l'effectivité par une formation appropriée et une information de l'intervenant sur les caractéristiques du système, sans lui imposer de contrainte particulière lorsqu'il s'écarte de la préconisation issue du système d'IA ;
- assurer aux usagers du service public le maintien systématique d'un accès

alternatif à un agent humain.

**Recommandation n°16** : La CNCDH recommande de reconnaître au bénéfice des utilisateurs des systèmes d'IA un droit au paramétrage de leurs critères, notamment afin de déterminer la sélection et la présentation des contenus reçus, et plus généralement dans l'hypothèse des interactions humain-machine.

**Recommandation n°17** : La CNCDH recommande d'informer systématiquement les personnes lorsqu'elles sont exposées ou amenées à interagir avec un système d'IA et, lorsqu'elles font l'objet d'une décision, que cette dernière se fonde, le cas échéant, en partie ou totalement sur un traitement algorithmique.

**Recommandation n°18** : La CNCDH recommande de garantir à la personne concernée un droit au réexamen, par un être humain, de toute décision individuelle fondée totalement, ou même en partie, sur un traitement algorithmique, dès lors qu'elle emporte des conséquences significatives pour elle.

**Recommandation n°19** : La CNCDH recommande aux administrations de communiquer sous une forme intelligible les informations sur le fonctionnement de l'algorithme, ainsi que sur la part éventuellement prise par une intervention humaine dans le processus de décision. Elle recommande par ailleurs de mener une réflexion s'agissant de l'extension de cette obligation aux organismes privés.

## Liste des personnes auditionnées.

**Celine Castets-Renard**, Professeure de droit privé, Chaire de recherche de l'Université d'Ottawa sur l'intelligence artificielle responsable à l'échelle mondiale.

Pour la CGT-UGICT : **Sylvain Delaitre** (CGT Thalès), **Jean-Luc Molins** (secrétaire national), **Matthieu Trubert** (co-animateur du Collectif Numérique UGICT-CGT).

**Mona Caroline Chammas**, Attorney & Integrity Director GOVERN&LAW.

**Régis Chatellier**, Directeur de projet au Pôle innovation, études et prospective de la Commission nationale de l'informatique et des libertés (CNIL).

**Raja Chatilla**, Directeur de recherche INRIA, Institut des Systèmes Intelligents et de Robotique, Sorbonne Université.

Défenseur des droits, représenté par **Sarah Bénichou**, adjointe au directeur de la « promotion de l'égalité et de l'accès aux droits » et **Gaëtan Goldberg**, Chargé de mission Numérique, Droits et Libertés.

**Sonia Desmoulin-Canselier**, Chargée de recherche CNRS au laboratoire Droit et Changement Social de l'Université de Nantes.

**Laurence Devillers**, Professeure d'informatique appliquée aux sciences sociales à l'Université Paris IV et chercheuse en Informatique au sein du Laboratoire Interdisciplinaire des Sciences du Numérique du CNRS à Saclay.

**Tim Engelhardt**, Responsable des droits de l'homme au sein de la Section État de droit et démocratie (Haut-Commissariat des Nations unies aux droits de l'homme).

**Jean-Gabriel Ganascia**, Professeur d'informatique à la Faculté des sciences de Sorbonne Université et chercheur en intelligence artificielle au LIP6.

**Alexei Grinbaum**, Directeur de recherche au laboratoire de philosophie des sciences (LARSIM) du CEA-Saclay.

**David Gruson**, Membre du comité de direction de la Chaire Santé de Sciences Po Paris.

**Joanne Kirkham**, Chercheuse à l'Université Paris II Panthéon-Assas.

**Claude Kirchner**, Directeur du Comité national pilote d'éthique du numérique.

**Michel Lansard**, ATD Quart Monde.

**Karine Lefeuvre**, Vice-présidente du Comité consultatif national d'éthique.

**Daniel Le Métayer**, Directeur de recherche à l'Institut national de recherche en sciences et technologies du numérique (INRIA).

**Grégoire Loiseau**, Professeur de droit privé, Université Paris 1 Panthéon-Sorbonne

**Frédéric Marty**, Chargé de recherche CNRS au GREDEG, Université Côte d'Azur.

**Winston Maxwell**, Directeur d'Etudes Droit et Numérique à Télécom Paris, Institut polytechnique de Paris.

**Yannick Meneceur**, Magistrat détaché au Conseil de l'Europe, conseiller en transformation numérique et en intelligence artificielle et chercheur associé à l'IHEJ.

**Arthur Messaud**, La Quadrature du Net.

Pour Numeum : **Valentin Hueber**, Délégué ESN/ICT ; **Anissa Kemiche**, Chargée des affaires européennes ; **Katya Lainé**, Présidente de la commission IA de Numeum ; **François Lhemery**, Directeur délégué aux affaires publiques et à la communication.

**Louis Perez**, Chercheur à l'Université Paris II Panthéon-Assas.

**Patrick Perrot**, Officier de gendarmerie, Coordonnateur pour l'IA au sein de la Gendar-

merie.

**Fabien Tarissan**, Chercheur en informatique au CNRS (section 06), professeur attaché à l'ENS Paris-Saclay.

**Romain Tinière**, Professeur de droit public, Université Grenoble-Alpes.

**Renaud Vedel**, Préfet, Coordonnateur de la Stratégie nationale pour l'Intelligence artificielle.

**Cédric Villani**, Député.

**Serena Villata**, Chargée de recherche au CNRS, Laboratoire d'informatique, Signaux et Systèmes (I3S), Université Sophia Antipolis.



COMMISSION NATIONALE  
CONSULTATIVE  
DES DROITS DE L'HOMME

RÉPUBLIQUE FRANÇAISE

Créée en 1947 sous l'impulsion de René Cassin, la **Commission nationale consultative des droits de l'homme (CNC DH) est l'institution nationale de promotion et de protection des droits de l'homme française, accréditée de statut A par les Nations Unies.**

L'action de la CNC DH s'inscrit dans une quadruple mission :

Conseiller les pouvoirs publics en matière de droits de l'Homme ;

Contrôler l'effectivité des engagements de la France en matière de droits de l'Homme et de droit international humanitaire ;

Assurer un suivi de la mise en oeuvre par la France des recommandations

formulées par les comités de suivi internationaux et régionaux ;

Sensibiliser et éduquer aux droits de l'Homme.

L'indépendance de la CNC DH est consacrée par la loi. Son fonctionnement s'appuie sur le principe du pluralisme des idées.

Ainsi, seule institution assurant un dialogue continue entre la société civile et les experts français en matière de droits de l'homme, elle est composée de 64 personnalités qualifiées et représentants d'organisations non gouvernementales issues de la société civile.

La CNC DH est le rapporteur national indépendant sur la lutte contre toutes les formes de racisme depuis 1990, sur la lutte contre la traite et l'exploitation des êtres humains depuis 2014, sur la mise en oeuvre des Principes directeurs des Nations Unies sur les entreprises et les droits de l'Homme depuis 2017, sur la lutte contre la haine et les discriminations anti-LGBTI depuis avril 2018 et sur l'effectivité des droits des personnes handicapées depuis 2020.

La CNC DH est en outre la Commission française de mise en oeuvre du droit international humanitaire » au sens du Comité international de la Croix-Rouge (CICR).

20 Avenue Ségur - TSA 40 720 - 75334 PARIS Cedex 07

Tel : 01.42.75.77.09

Mail : [cncdh@cncdh.fr](mailto:cncdh@cncdh.fr)

[www.cncdh.fr](http://www.cncdh.fr)



@CNC DH



@cncdh.france