COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 15 juin 2020 CAHAI(2020)08-fin

COMITE AD HOC SUR L'INTELLIGENCE ARTIFICIELLE (CAHAI)

Elaboration de l'étude de faisabilité

Analyse des instruments internationaux juridiquement contraignants

Rapport final

Document préparé par Alessandro Mantelero*

www.coe.int/cahai

* Maître de conférences en droit privé et en éthique et protection des données, université polytechnique de Turin (Politecnico di Torino). Les opinions exprimées dans cette analyse ne reflètent pas nécessairement la position du CAHAI ou du Conseil de l'Europe

Table des matières

RESUME

Les développements les plus récents de l'intelligence artificielle (IA) ont une action transformatrice de plus en plus marquée sur la société et amènent de nouvelles interrogations dans des domaines aussi divers que la médecine prédictive, la modération de contenus, le « quantified self » (automesure connectée) et les systèmes judiciaires, sans oublier les questions relatives à leur impact environnemental.

Une analyse des instruments juridiques internationaux contraignants est donc le point de départ obligé pour définir le cadre juridique existant, en recenser les valeurs fondamentales et vérifier si ce cadre et ses principes répondent bien à l'ensemble des problématiques soulevées par l'IA.

Afin de continuer à assurer l'harmonisation du cadre juridique actuel dans le domaine des droits de l'homme, de la démocratie et de l'État de droit, la présente étude a pour objectif de contribuer à l'élaboration de la future réglementation de l'IA en s'appuyant sur les instruments contraignants existants, en contextualisant leurs principes et en fournissant les principales orientations réglementaires d'un **futur cadre juridique.**

Cette approche a pour fondement théorique l'idée que les principes généraux fournis par les instruments internationaux relatifs aux droits de l'homme devraient sous-tendre l'ensemble des activités humaines, y compris l'innovation basée sur l'intelligence artificielle. Par ailleurs, seul le cadre des droits de l'homme peut offrir **une référence universelle à la réglementation de l'IA**, tandis que les autres sphères (par exemple l'éthique) n'ont pas la même dimension internationale, sont davantage liées au contexte et se caractérisent par une diversité d'approches théoriques.

L'analyse des instruments juridiques contraignants actuels contenue dans ce document ne se limite pas à une étude d'harmonisation, qui consiste à extraire des valeurs et des principes communs d'un ensemble donné de règles relatives à l'IA. Elle comprend également un examen plus structuré en plusieurs étapes.

Après une première analyse sectorielle visant à recenser les grands principes directeurs dans quatre domaines essentiels (protection des données, santé, démocratie et justice), ces principes sont contextualisés compte tenu des mutations opérées par l'IA au sein de la société. Ce faisant, nous tirons profit des instruments non contraignants existants qui appliquent avec un plus grand niveau de détail les principes contenus dans les instruments juridiques contraignants et donnent également, dans certains cas, des orientations spécifiques sur l'IA.

Cette **contextualisation des principes directeurs et des valeurs juridiques** permet de les formuler de manière plus fine, en tenant compte de la nature spécifique des produits et services d'IA. En outre, elle permet de mieux répondre aux défis soulevés par l'IA. Par conséquent, il est possible de définir **un premier ensemble de dispositions pour la future réglementation de l'IA**, en mettant l'accent sur les points les plus problématiques de chaque secteur examiné.

Compte tenu du nombre important de documents adoptés par divers organismes internationaux et intergouvernementaux et de l'étude parallèle sur les instruments d'éthique actuellement menée par le CAHAI, ce document se limite aux instruments juridiquement contraignants et aux instruments non contraignants adoptés pour leur mise en œuvre.

L'étude comporte deux parties. La première identifie la portée et la méthodologie de cette analyse, tandis que la seconde présente les résultats de l'analyse sectorielle des principes directeurs.

Dans l'analyse sectorielle, les deux premiers domaines essentiels examinés sont la **santé** et la **protection des données**. L'intersection entre ces deux sphères est intéressante pour la présente étude, compte tenu du grand nombre d'applications de l'IA relatives aux données de santé et de leurs dénominateurs communs que l'on retrouve dans plusieurs dispositions de la Convention d'Oviedo, de la Convention 108+ et des instruments non contraignants. L'autre question qui occupe une place centrale aussi bien dans le domaine de la protection des données que dans celui de la biomédecine est celle de l'autodétermination de l'individu ; les problématiques liées à l'IA présentent donc un intérêt tout particulier à cet égard et suscitent des préoccupations communes du point de vue de la complexité et de l'opacité des opérations de traitement.

Les quatrième et cinquième sous-parties s'intéressent à la **démocratie** et à la **justice**. Ici, le champ d'investigation est plus large et il n'existe pas d'instruments juridiques complets pouvant offrir des principes sectoriels spécifiques comme la Convention 108+ ou la Convention d'Oviedo. Par conséquent, l'analyse se concentre davantage sur les principes essentiels et leur contextualisation ; en comparaison avec les sous-parties précédentes moins de dispositions clés sont élaborées.

La dernière sous-partie donne une vue d'ensemble des principes directeurs recensés et suggère un cadre d'harmonisation qui met en avant les corrélations existantes et les dénominateurs communs entre ces principes, tout en soulignant les contributions uniques de chaque secteur à la future réglementation de l'IA.

L'objectif principal de cette étude n'est pas d'ajouter une nouvelle liste de principes directeurs à ceux déjà fournis par diverses instances et entités, mais d'atteindre un résultat différent sur le plan méthodologique et matériel.

Premièrement, **l'analyse effectuée et la solution proposée trouvent leurs racines dans les droits de l'homme et les libertés**, adoptant une approche concrète centrée sur les instruments juridiques internationaux existants. Les autres études sont souvent sectorielles et s'appuient sur différents ensembles de références normatives (nationales ou régionales) ou adoptent une approche théorique énonçant des principes ou se référant aux droits de l'homme de manière générale et abstraite. Bien que ces travaux fassent avancer le débat éthique et juridique sur l'IA, leur impact en termes de contribution au cadre réglementaire est limité et n'est pas spécialement contextualisé dans le cadre des normes du Conseil de l'Europe relatives aux droits de l'homme, à la démocratie et à l'État de droit.

Deuxièmement, le résultat de cette analyse des instruments juridiquement contraignants, y compris les instruments non contraignants adoptés pour les mettre en œuvre, n'est pas simplement une liste de principes aussi précis soient-ils. **Il est important de recenser les principes directeurs communs, mais cela ne suffit pas à constituer une feuille de route pour la future réglementation de l'IA**. La transparence, l'obligation de rendre compte, le contrôle humain et de nombreux autres principes déjà énumérés dans plusieurs chartes sur l'IA sont des concepts abstraits qui ne sont pas bien contextualisés.

Tel est l'objectif principal de cette étude, fournir précisément cette **contextualisation eu égard au cadre juridique et aux problématiques de l'IA**. Si ce document parvient à suggérer des **moyens concrets et efficaces de formuler et de codifier les principes directeurs de l'IA** et à concrètement **ancrer les normes du Conseil de**

l'Europe relatives aux droits de l'homme, à la démocratie et à l'État de droit dans le projet de réglementation de l'IA, il aura atteint son objectif d'aider à structurer la relation entre les humains et l'IA d'un point de vue juridique.

PARTIE I – METHODOLOGIE

1. Introduction

À l'instar d'Internet, de l'électricité ou de la propulsion à vapeur, l'intelligence artificielle (IA) englobe une multitude de technologies – dont les interactions humain-robot – qui ont des répercussions importantes sur diverses activités humaines et sur la société.

Dans ce contexte, de nombreuses questions se posent. Par exemple, quand un système d'IA devrait-il prendre une décision ? Quels critères le système devrait-il appliquer ? Qui est responsable des décisions susceptibles d'avoir des effets négatifs sur les personnes et la société ? À la lumière de ces questions et de nombreuses questions émergentes, les réglementations existantes devraient être réexaminées.

Une analyse des instruments juridiques internationaux contraignants est donc le point de départ obligé pour définir le cadre juridique existant, en recenser les valeurs fondamentales et vérifier si ce cadre et ces valeurs répondent bien à l'ensemble des problématiques soulevées par l'IA, afin de continuer à assurer l'harmonisation du cadre juridique actuel dans le domaine des droits de l'homme, de la démocratie et de l'État de droit.

Le but n'est pas de créer un tout nouveau cadre de référence complet, la réglementation devant mettre l'accent sur les changements que l'IA apportera à la société plutôt que de repenser tous les domaines dans lesquels elle peut s'appliquer[1]. Cette approche ciblée est rendue possible en utilisant les instruments contraignants existants et en contextualisant leurs principes directeurs, puis en définissant les principales règles d'un futur cadre juridique relatif à l'IA qui pourra couvrir les domaines non encore réglementés par les instruments contraignants existants.

Il est important de souligner à cet égard la différence entre les instruments juridiquement contraignants existants et d'autres documents comme les instruments non contraignants ou les chartes d'éthique relatives à l'IA. Les instruments juridiquement contraignants existaient avant l'essor actuel de l'IA. Ils n'ont pas été rédigés en pensant à l'intelligence artificielle et ne fixent pas un ensemble spécifique de règles applicables à ce domaine, tandis que les instruments non contraignants et les documents d'éthique relatifs à l'IA portent sur des points précis, mais les abordent sous des angles différents.

L'analyse des instruments juridiques contraignants existants ne se limite donc pas à une étude d'harmonisation (qui consisterait à extraire des valeurs et principes communs d'un ensemble donné de règles relatives à l'IA), mais nécessite un processus plus structuré qui englobe l'harmonisation sans s'y limiter. Ce processus, décrit dans les parties suivantes, peut être divisé en trois étapes distinctes : (i) inventaire et une identification des concepts clé; (ii) une contextualisation ; et (iii) une harmonisation

[1] voir par exemple l'approche de l'UE en matière de réglementation du commerce électronique

I.1 Le scénario

Les développements les plus récents de l'IA ont une action transformatrice de plus en plus marquée sur la société et amènent de nouvelles interrogations dans des domaines aussi divers que la médecine prédictive, la modération de contenus, le *quantified self* (automesure connectée) et les systèmes judiciaires, sans oublier les questions relatives à leur impact environnemental.

L'évolution rapide de l'IA appliquée ces dernières années est incompatible avec l'adoption d'une réponse juridique spécifique sous la forme d'instruments internationaux juridiquement contraignants centrés sur l'IA. C'est pourquoi deux stratégies opérationnelles différentes ont été mises en œuvre pour traiter ces questions : (i) un travail considérable d'interprétation du cadre juridique existant à la lumière des problématiques liées à l'IA (voir par exemple le débat en cours sur les dispositions du RGPD relatives à transparence et aux décisions automatisées) ; (ii) le recours à des instruments non contraignants pour contextualiser les principes contenus dans les instruments contraignants existants (par exemple Lignes directrices T-PD(2019)01 sur l'intelligence artificielle et la protection des données ; CEPEJ. 2019. Charte éthique européenne d'utilisation de l'intelligence artificielle (IA) dans les systèmes judiciaires et leur environnement[2]).

La future réglementation de l'IA devrait par conséquent prendre appui sur ces initiatives, en mettant l'accent d'une part sur les principes directeurs et valeurs tirés des instruments contraignants existants et d'autre part, sur leurs instruments de mise en œuvre non contraignants, qui envisagent déjà dans certains cas le nouveau scénario de l'IA.

I.2 Axe de recherche et méthodologie

Cette étude vise principalement à définir les principes de base de la future réglementation de l'intelligence artificielle en analysant le cadre juridique existant. La méthodologie sera donc nécessairement déductive et consistera à extraire les principes fondamentaux des diverses réglementations applicables aux domaines dans lesquels des solutions d'intelligence artificielle pourraient être adoptées.

Cette approche a pour fondement théorique l'idée que les principes généraux fournis par les instruments internationaux relatifs aux droits de l'homme devraient sous-tendre l'ensemble des activités humaines, y compris l'innovation basée sur l'intelligence artificielle[3]. Par ailleurs, seul le cadre des droits de l'homme peut offrir une référence universelle pour la réglementation de l'IA, tandis que les autres sphères (par exemple l'éthique) n'ont pas la même dimension internationale, sont davantage liées au contexte et se caractérisent par une diversité d'approches théoriques.

Dans ce contexte, de nombreuses questions se posent. Par exemple, quand un système d'IA devrait-il prendre une décision ? Quels critères le système devrait-il appliquer ? Qui est responsable des décisions susceptibles d'avoir des effets négatifs sur les personnes et la société ? À la lumière de ces questions et de nombreuses questions émergentes, les réglementations existantes devraient être réexaminées.

Afin de fournir un cadre réglementaire harmonisé pour relever les défis de l'IA, des orientations communes de haut niveau sur les principes et les valeurs à consacrer

[2] Commission européenne pour l'efficacité de la justice (CEPEJ). 2018. Charte éthique européenne sur l'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement

[3] Voir Comité directeur sur les médias et la société de l'information (CDMSI). 2019. Projet de Recommandation du Comité des Ministres aux États membres concernant les impacts des systèmes algorithmiques sur les droits de l'homme (préparé et finalisé par le Comité d'experts sur la dimension droits de l'homme des traitements automatisés de données et différentes formes d'intelligence artificielle (MSI-AUT) ; Conseil de l'Europe - Comité d'experts sur les intermédiaires d'Internet (MSI-NET). 2018.

devraient être tirées des chartes internationales des droits de l'homme (par exemple, la Déclaration universelle des droits de l'homme, le Pacte international relatif aux droits civils et politiques, la Convention de sauvegarde des droits de l'homme et des libertés fondamentales, la Charte des droits fondamentaux de l'Union européenne).

Les principes directeurs doivent être envisagés dans le cadre d'un scénario de transformation basé sur l'IA, qui, dans de nombreux cas, nécessitera leur adaptation. Ces principes restent valables, mais leur fonctionnement doit être reconsidéré à la lumière des changements sociaux et techniques induits par l'IA (par exemple, la liberté de choix en cas de "boîtes noires"). Cela permettra une application plus contextualisée et plus granulaire des principes afin qu'ils puissent apporter une contribution concrète à la forme de la future réglementation sur l'IA.

Pour mener cette étude, il faut commencer par définir les principaux sujets d'étude, en considérant à la fois les effets potentiels de l'IA et les domaines d'action du Conseil de l'Europe. À cet égard, quatre domaines essentiels ont été sélectionnés : les données, la santé, la démocratie et la justice.

I.3 Analyse et résultats escomptés

L'étude adopte une approche descendante en vue de contribuer au futur cadre réglementaire relatif à l'IA qui sera mis en œuvre au moyen d'instruments contraignants complétés par des instruments non contraignants, sur le modèle de ce qui a été fait dans le domaine de la biomédecine. Il devrait en ressortir un ensemble de dispositions relatives aux domaines étudiés et les principes directeurs communs essentiels, reposant sur une analyse globale de tous les instruments, contraignants ou non, adoptés.

Première étape : recensement des principes essentiels. Des principes directeurs seront recensés dans les domaines étudiés. L'analyse commence par les différentes thématiques, car les instruments contraignants sont spécifiques à des secteurs et non basés sur des droits. Les deux tableaux suivants donnent un premier aperçu de cet exercice de recensement, qui part d'une vue d'ensemble des sphères de la protection des données et de la justice afin d'identifier les principes directeurs de la future réglementation de l'IA.

Figure 1 : Protection des données

Instruments contraignants	Convention 108+ Convention sur la cybercriminalité
Zones touchées	Systèmes de prise de décision Vie privée et dimension collective du groupe Profilage
Instruments non contraignants connexes	CdE. 2019. Lignes directrices sur les implications de l'intelligence artificielle en matière de protection des données

	CdE. 2017. Lignes directrices sur la protection des personnes physiques à l'égard du traitement des données à caractère personnel dans un monde de "Big Data CdE. 2010. Recommandation sur la protection des personnes physiques à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage [en cours de révision]. UNESCO. 2019. Étude préliminaire sur un éventuel instrument normatif sur l'éthique de l'intelligence artificielle OCDE. 2019. Recommandation du Conseil sur l'intelligence artificielle 40e Conférence internationale des commissaires à la protection des données et de la vie privée. 2018
Principes directeurs et valeurs juridiques	Responsabilité Une approche fondée sur les risques Le principe de précaution Qualité et sécurité des données Transparence Équité Approche contextuelle Rôle des experts Participation/Inclusion Liberté de choix/Autonomie Contrôle/supervision humain Sensibilisation Alphabétisation Innovation responsable Coopération entre les autorités de contrôle

Figure 2 : Justice

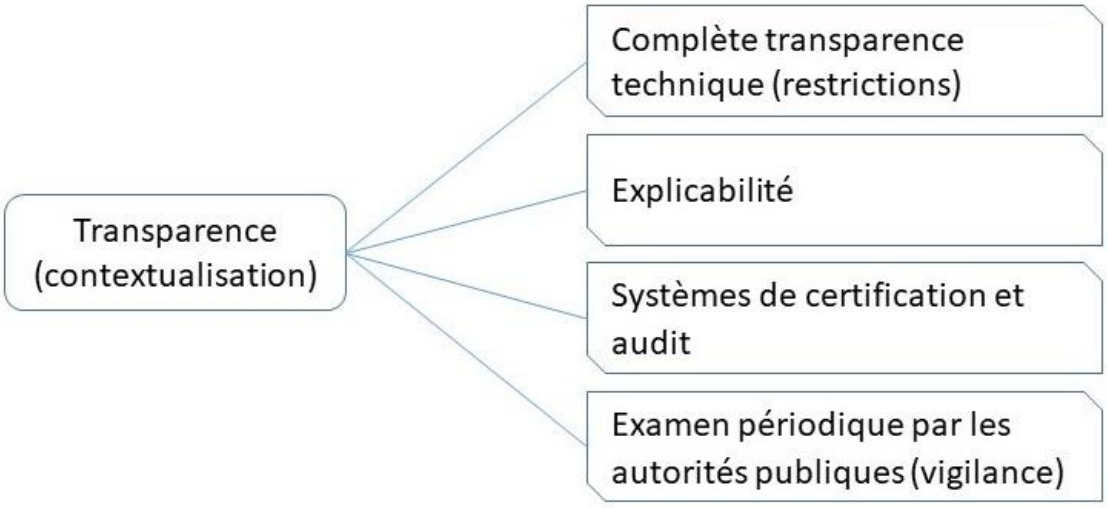
Instruments contraignants	Déclaration universelle des droits de l'homme Pacte international relatif aux droits civils et politiques Convention internationale sur l'élimination de toutes les formes de discrimination raciale Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes Convention de sauvegarde des droits de l'homme et des libertés fondamentales Charte des droits fondamentaux de l'Union européenne

Zones touchées	Traitement des décisions judiciaires et des données La police prédictive
Instruments non contraignants connexes	CEPEJ. 2019. Charte éthique européenne sur l'utilisation de l'intelligence artificielle (IA) dans les systèmes judiciaires et leur environnement .
Principes directeurs et valeurs juridiques	Non-discrimination Qualité et sécurité des données Transparence Impartialité Équité Approche contextuelle Liberté de choix/ Indépendance des juges (processus décisionnel) Contrôle/supervision humain Garanties du droit à un procès équitable

Deuxième étape : Contextualisation. Les valeurs directrices identifiées dans l'exercice de cartographie doivent être contextualisées à la lumière des changements de société produits par l'IA. Cette phase bénéficiera des instruments non contraignants existants qui fournissent des applications plus granulaires des principes inscrits dans les instruments contraignants, en fournissant également dans certains cas des orientations spécifiques sur l'IA.

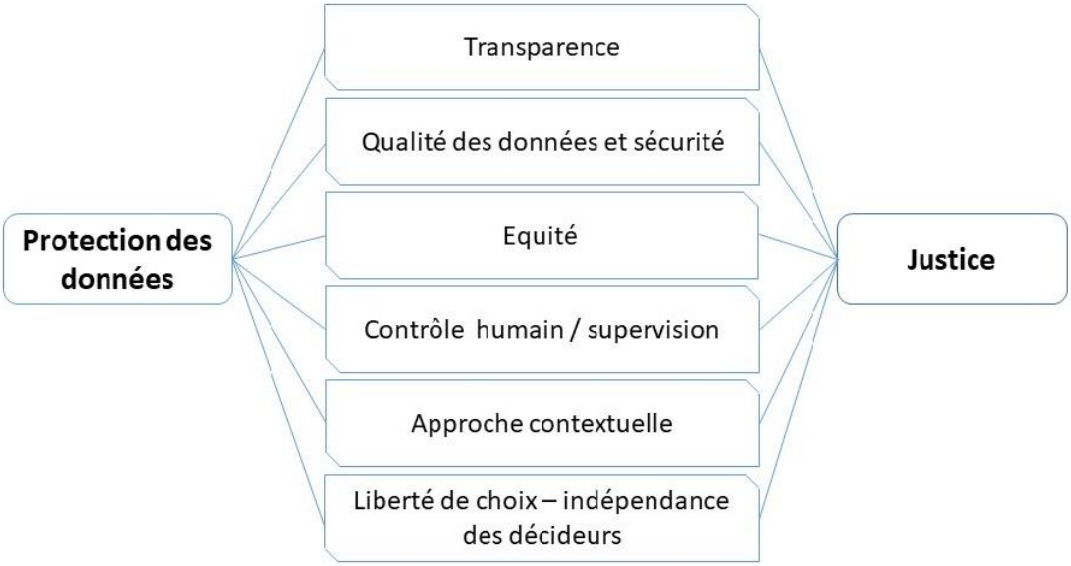
Cette contextualisation des principes directeurs et des valeurs juridiques permettra de les formuler de manière plus précise et plus élaborée, compte tenu de la nature spécifique des produits et services de l'IA. À ce stade, il sera donc possible de formuler une première série de dispositions pour la future réglementation sur l'IA en se concentrant sur les questions les plus difficiles dans chaque secteur.

Figure 3 : Mise en œuvre adaptée au contexte du principe de transparence



Troisième étape : harmonisation (analyse intersectorielle). À partir de l'analyse sectorielle menée dans cette étude, une liste de principes directeurs essentiels communs aux différentes sphères sera établie dans la dernière sous-partie (figure 4). Ces principes constitueront le fondement commun des futures dispositions relatives à l'IA.

Figure 4 : Valeurs fondamentales communes aux domaines de la protection des données et de la justice



PARTIE II – ANALYSE

Compte tenu du nombre important de documents adoptés par divers organismes internationaux et intergouvernementaux et de l'étude parallèle sur les instruments d'éthique actuellement menée par le CAHAI, cette partie se limite aux instruments juridiquement contraignants.

Les codes d'éthique ne seront donc pas examinés à ce stade et les documents relatifs aux futures stratégies réglementaires (par ex. livres blancs) ne seront pris en compte qu'à titre d'informations contextuelles.

Cette partie se subdivise en six sections, suivies de conclusions provisoires. La première sous-partie contient une vue d'ensemble des instruments existants adoptés par le Conseil de l'Europe et des principes/valeurs sur lesquels ils reposent. Cet inventaire aidera à définir les principes fondamentaux d'une future réglementation de l'IA et à assurer sa cohérence avec le cadre existant.

Les deuxième et troisième sous-parties mettent l'accent sur deux domaines essentiels allant de pair : la santé et la protection des données. L'intersection entre ces deux sphères est intéressante pour la présente étude sur les fondements de la future réglementation de l'IA, compte tenu du grand nombre d'applications de l'IA relatives aux données de santé et de leurs dénominateurs communs que l'on retrouve dans plusieurs dispositions de la Convention d'Oviedo, de la Convention 108+ et d'autres instruments non contraignants[4]. L'autre question qui occupe une place centrale aussi bien dans le domaine de la protection des données que dans celui de la biomédecine est celle de l'autodétermination de l'individu ; les problématiques liées à l'IA présentent donc un intérêt tout particulier à cet égard et suscitent des préoccupations communes du point de vue de la complexité et de l'opacité des opérations de traitement.

Les quatrième et cinquième sous-parties s'intéressent à la démocratie et à la justice. Ici, le champ d'investigation est plus large et il n'existe pas d'instruments juridiques généraux pouvant offrir des principes sectoriels, comme la Convention 108+ ou la Convention d'Oviedo. Par conséquent, l'analyse se focalise sur les principes essentiels et leur contextualisation ; en comparaison avec les sous-parties précédentes moins de dispositions clés sont élaborées.

La sixième sous-partie donne une vue d'ensemble des principes directeurs recensés et suggère un cadre d'harmonisation qui met en avant à la fois les corrélations existantes entre ces principes et la contribution unique de chaque secteur à la future réglementation de l'IA.

Comme il a été souligné dans les commentaires reçus lors de l'exercice de veille décrit dans la prochaine sous-partie, les technologies d'IA ont un impact sur de très nombreux secteurs et posent des questions qui mettent en jeu un vaste ensemble d'instruments réglementaires.[5] Cette première étude constitue donc un point de départ axé sur les quatre domaines essentiels cités. Malgré les limites de l'analyse, les résultats valident la méthodologie proposée et donnent un certain nombre d'indications pour les futures dispositions de la réglementation de l'IA.

[4] Voir la Recommandation Rec(2019)2 du Comité des Ministres aux Etats membres en matière de protection des données relatives à la santé.

[5] Voir Annexe 1.

II.1. Vue d'ensemble

L'IA s'appliquant à de nombreuses situations[6] couvertes par différents instruments contraignants qui portent sur divers domaines, il convient d'analyser l'ensemble des données disponibles pour définir les principes fondamentaux et les valeurs communes à prendre en considération dans la future réglementation.

Un travail de veille a été mené en ce sens entre le 12 et le 28 février 2020 auprès des différents secteurs du Conseil de l'Europe pour tirer parti de l'expertise spécifique des services qui travaillent depuis de nombreuses années sur divers sujets liés aux droits de l'homme, à la démocratie et à l'Etat de droit.

Au moyen d'un questionnaire ouvert, il leur a été demandé de donner des informations sur les points suivants : (i) instruments contraignants, (ii) domaines concernés (applications), (iii) instruments non contraignants connexes, (iv) principes directeurs et valeurs juridiques et (v) principes/questions non abordés. Une grande diversité d'informations ont pu être recueillies grâce à la participation active des différents secteurs.

D'un point de vue méthodologique, la structure de cette enquête préliminaire, basée sur des questions ouvertes, a nécessairement un impact sur les résultats de l'analyse quantitative. Les principales limites rencontrées sont dues à l'utilisation de catégories générales qui se chevauchent parfois et au niveau de précision des réponses.

Néanmoins, par agrégation au niveau macro et prise en compte des similarités (fréquence) entre les principes et valeurs recensés, il a été possible de mettre en perspective les résultats et d'obtenir une cartographie plus détaillée des instruments non contraignants existants adoptés par le Conseil de l'Europe, pouvant servir à établir le cadre juridique d'une future réglementation (voir annexe 1).

S'agissant des domaines concernés (voir annexe 2), l'exercice suggère de bâtir la future réglementation de l'IA autour de deux grands axes : l'utilisation de l'IA et le développement de l'IA. Dans les deux cas, différents droits de l'homme et libertés fondamentales pourraient être menacés ou au contraire, jouer un rôle important dans les futurs scénarii de l'IA[7].

Pour ce qui est de l'utilisation de l'IA, on recense quatre grands domaines d'application et donc de réglementation : l'analyse prédictive et les systèmes d'aide à la décision, l'analyse prédictive et les systèmes de prise de décision automatisée, la collecte de preuves/informatique judiciaire et la production de contenus.

Les deux premiers domaines sont bien connus et étudiés, car ils recouvrent un ensemble extrêmement vaste d'applications (voir annexe 2). Cela étant, la distinction entre systèmes d'aide à la décision et systèmes de décision automatisée est essentielle du point de vue de la conception centrée sur les valeurs et du rôle de l'humain dans les processus décisionnels : ces deux types de systèmes diffèrent dans leur nature et devront donc faire l'objet de garanties procédurales et matérielles distinctes dans la réglementation de l'IA.

Les deux autres domaines d'application, bien qu'ils relèvent de secteurs spécifiques, devraient être examinés séparément car ils ne concernent pas directement le processus décisionnel mais fournissent les éléments qui en constituent le fondement (collecte de preuves et informatique judiciaire) ou ont une influence sur les processus de création (production de contenus). Dans ce cas, l'enjeu principal semble résider davantage dans

[6] Voir également UNESCO, « Étude préliminaire concernant un éventuel instrument normatif sur l'éthique de l'intelligence artificielle », bibliothèque numérique de l'Unesco, consultée le 21 novembre 2019, https://unesdoc.unesco.org/ark:/48223/pf0000369455.
[7] Voir Comité d'experts du Conseil de l'Europe sur les intermédiaires d'Internet (MSI-NET), 2018.

les aspects procéduraux et leur cohérence avec les approches traditionnelles (non fondées sur l'IA).

Bien que la plupart des travaux et recommandations existants se concentrent sur les systèmes d'IA et leurs conséquences potentielles, deux autres aspects - le développement de l'IA et la prestation de services d'IA – ont également un impact important sur les droits de l'homme et les libertés fondamentales. C'est pourquoi la future réglementation devrait également prendre en considération les questions liées aux conditions de travail des personnes impliquées dans l'ensemble de la chaîne d'approvisionnement de produits et services d'IA[8].

Le deuxième ensemble d'informations fourni par l'exercice de veille concerne les principes directeurs et valeurs juridiques qui devraient sous-tendre le développement et les usages futurs de l'IA (voir annexe 3). La diversité des notions employées par les services interrogés plaide en faveur d'une agrégation des principes et des valeurs. Le résultat de cette analyse a permis de regrouper les principes directeurs et valeurs fondamentales autour d'un certain nombre d'éléments dont la répartition était la suivante en termes de fréquence :

Non-discrimination (15)
Diversité, inclusion et pluralisme (13)
Protection des données et de la vie privée (11)

Transparence (9)
Égalité (8)
Accès à la justice, droit à un procès équitable (7)
Contrôle humain (7)

Impartialité (6)
Accès à l'information (5)
Sécurité (5)
Traitement équitable (5)
Participation (5)
Liberté de choix (5)
Liberté d'expression et de création (5)

Obligation de rendre compte (3)
Compétence et capacité (2)
Indépendance (3)
Autonomie individuelle (3)
Coopération culturelle (2)
Durabilité (2)

Malgré les limites de l'analyse mentionnée, il apparaît clairement que les trois premiers principes sont considérés comme des éléments clés de la future réglementation de l'IA et qu'ils en constitueront donc les axes principaux. Ce constat est confirmé par le deuxième ensemble de principes/valeurs, qui est étroitement lié au premier : la transparence et le contrôle humain sont des facteurs importants dans la lutte contre les discriminations et la protection des données, tandis que l'accès à la justice est essentiel pour réagir aux éventuels manquements à ces valeurs. De la même manière, bien que plus substantiellement, l'égalité est à de nombreux égards liée aux trois premiers grands principes/valeurs. Il existe de plus grandes différences entre les autres valeurs/principes, qui se rapportent à des questions spécifiques liées à la mise en œuvre de l'IA.

[8] Voir également Crawford et Joler, 2018.

Cet exercice a permis de recenser les grands principes directeurs de la réglementation de l'IA, déjà codifiés dans des instruments juridiques contraignants et non contraignants, mais nécessitant une adaptation au contexte de l'IA. Lors de l'analyse sectorielle, cette contextualisation sera réalisée sur la base d'une analyse approfondie des instruments internationaux juridiquement contraignants, en faisant le point sur les éventuelles lacunes dans le cadre réglementaire existant, secteur par secteur.

L'IA étant une technologie intersectorielle, cette analyse devrait laisser entrevoir des possibilités d'interventions réglementaires du même type dans d'autres domaines, comme indiqué dans la partie consacrée à la méthodologie[9]. À l'issue de l'analyse sectorielle, ces interventions potentielles seront classées pour éviter tout doublon, puis réunies dans un cadre cohérent reposant sur des valeurs fondamentales qui devraient concorder avec les résultats de l'exercice de veille.

II.2 Protection des données

Ces dix dernières années, le cadre réglementaire international dans le domaine de la protection des données a subi une vaste refonte. Les instruments juridiques établis sur la base de principes définis dans les années 1970 et 1980[10] n'étaient plus adaptés au nouveau paysage social et technique apparu avec l'augmentation de la largeur de bande disponible pour le transfert de données, la conservation de données et les ressources informatiques (informatique en nuage), la mise en données progressive de parts importantes de nos vies et de notre environnement (Internet des objets), l'analyse de données à grande échelle et l'analyse prédictive reposant sur les mégadonnées et l'apprentissage-machine.

En Europe, les principales réponses à ce changement sont la version modernisée de la Convention 108 (Convention 108+) et le RGPD. Une redéfinition du cadre réglementaire est également en cours ou a été menée dans d'autres contextes internationaux comme l'OCDE[11] – ou par certains pays.

Cela dit, la dernière vague de développement de l'IA a été si rapide que certaines problématiques spécifiques à l'IA n'ont pas pu être traitées directement par ces nouveaux instruments contraignants ; plusieurs instruments non contraignants ont donc été adoptés pour combler ces lacunes et les futures stratégies réglementaires sont en cours d'examen[12].

L'analyse menée aux fins de la présente étude a porté sur les instruments juridiques non contraignants suivants[13] : T-PD(2019)01, Lignes directrices sur l'intelligence artificielle et la protection des données (GAI) ; T-PD(2017)1, Lignes directrices sur la protection des personnes à l'égard du traitement des données à caractère personnel à l'ère des mégadonnées ; Recommandation CM/Rec(2019)2 du Comité des Ministres du Conseil de l'Europe aux États membres en matière de protection des données relatives à la

[9] Voir ci-dessus partie I.3.

[10] Voir also Mayer-Schönberger, 1997; González Fuster, 2014.

[11] Voir OCDE. 2013. Recommandation du Conseil de l'OCDE relative aux Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel, C(80)58/FINAL, telle que modifiée le 11 juillet 2013 par la C(2013)79.

[12] Commission européenne. 2020. Rapport sur les conséquences de l'intelligence artificielle, de l'internet des objets et de la robotique sur la sécurité et la responsabilité, COM(2020) 64 final ; Commission européenne. 2020. Livre blanc sur l'intelligence artificielle – une approche européenne axée sur l'excellence et la confiance, COM(2020) 65 final. Voir également Commission européenne. 2020. Une stratégie européenne pour les données, COM(2020) 66 final.

[13] Voir annexe 4

santé[14] ; Recommandation CM/Rec(2010)13 du Comité des Ministres du Conseil de l'Europe aux Etats membres sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage ; [UNESCO]. 2019. Étude préliminaire concernant un éventuel instrument normatif sur l'éthique de l'intelligence artificielle[15] ; [OCDE]. 2019. Recommandation du Conseil sur l'intelligence artificielle ; 40e Conférence internationale des commissaires à la protection des données et de la vie privée. 2018 [ICDPPC]. Déclaration sur l'éthique et la protection des données dans le secteur de l'intelligence artificielle.

Ces instruments sont de nature différente : tandis que les instruments du Conseil de l'Europe contiennent des exigences et dispositions particulières, les autres établissent un certain nombre de principes mais ne fournissent pas, ou en partie seulement, d'indications plus détaillées quant à des exigences spécifiques. Les paragraphes suivants illustrent les principes essentiels tirés de ces différents instruments et la façon dont ils peuvent être contextualisés dans le cadre de l'IA.

Parmi les principes classés dans le domaine de la protection des données à caractère personnel (par ex. qualité des données), certains peuvent être étendus aux données non personnelles, principalement pour ce qui est de l'impact de l'utilisation de données non personnelles (par ex. données agrégées) sur les personnes et les groupes dans le cadre des processus décisionnels (par ex. données sur la mobilité ou sur la consommation d'énergie).

i) Primauté de l'être humain

Les systèmes d'IA devraient être conçus pour servir l'humanité ; toute création, évolution et utilisation de systèmes d'IA doit respecter pleinement les droits de l'homme, la démocratie et l'Etat de droit.[16]

ii) Contrôle humain

Les applications de l'IA devraient permettre l'exercice d'un contrôle humain significatif de leurs effets sur les individus et la société.[17]

iii) Transparence et extensibilité

Toute personne qui interagit directement avec un système d'IA a le droit de recevoir des informations adéquates et facilement compréhensibles sur l'objectif et les effets de ce dernier, y compris l'existence de décisions automatisées, pour vérifier qu'il reste conforme aux attentes des personnes concernées, pour assurer un contrôle humain

[14] Cette recommandation remplace la Recommandation Rec(97)5 du Comité des Ministres aux États membres sur la protection des données médicales. Voir également Rec(2016)8 sur le traitement des données à caractère personnel relatives à la santé à des fins d'assurance, y compris les données résultant de tests génétiques et son exposé des motifs.

[15] Bien que le titre ne mentionne que l'éthique, l'objet de cette étude est présenté comme suit : « Le présent document contient l'étude préliminaire sur les aspects techniques et juridiques liés à l'opportunité d'un instrument normatif sur l'éthique de l'intelligence artificielle, ainsi que les commentaires et observations du Conseil exécutif à ce sujet ».

[16] Voir CM/Rec(2019)2, ICDPPC, GAI II.1, UNESCO ; voir également considérant n° 4 RGPD.

[17] Voir GAI I.6

global sur les systèmes d'IA et pour permettre aux personnes qui subissent les effets néfastes d'un système d'IA de contester son résultat.[18]

Toute personne a le droit d'obtenir, à sa demande, connaissance du raisonnement qui sous-tend un processus décisionnel reposant sur l'IA lorsque les résultats de ce processus lui sont appliqués.[19]

Les États devraient soutenir la recherche scientifique sur l'intelligence artificielle explicable et les meilleures pratiques en matière de transparence et de vérifiabilité des systèmes d'IA.[20]

iv) Approche de précaution

Lorsque les risques potentiels des applications d'IA sont inconnus ou incertains, le développement de l'IA doit reposer sur le principe de précaution.[21]

v) Gestion des risques

Les développeurs, fabricants et prestataires de service en IA devraient évaluer et répertorier les éventuelles conséquences négatives des applications d'IA sur les droits de l'homme et les libertés fondamentales et adopter des mesures appropriées de prévention et de réduction des risques dès la conception (approche centrée sur les droits de l'homme dès la conception) et pendant toute leur durée de vie.[22]

Les conséquences préjudiciables incluent celles qui résultent de l'utilisation de données décontextualisées et de modèles algorithmiques décontextualisés.[23]

Les développeurs, fabricants et prestataires de service en IA devraient consulter les autorités de contrôle compétentes dès lors que les applications de l'IA peuvent avoir un impact significatif sur les droits de l'homme et les libertés fondamentales des individus.[24]

vi) Risque de réidentification

Des mesures appropriées devraient être mises en place pour éviter que des données anonymes et agrégées ne puissent déboucher sur une réidentification des personnes concernées[25]

vii) Qualité et minimisation des données

[18] Voir ICDPPC, CM/Rec(2019)2, OCDE, UNESCO. Voir également Comité d'experts sur la dimension « droits de l'homme » des traitements automatisés de données et des différentes formes d'intelligence artificielle (MSI-AUT). 2019.
[19] Voir Convention 108+, GAI II.11
[20] Voir ICDPPC
[21] Voir GAI II.2
[22] Voir GAI II.2 et II.3, ICDPPC, OCDE, UNESCO Voir également Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems.
[23] Voir GAI II.5
[24] Voir GAI III.5
[25] Voir CM/Rec(2010)13

Les développeurs d'IA devraient évaluer de manière critique la qualité, la nature, l'origine et la quantité des données à caractère personnel utilisées, en limitant les données inutiles, redondantes ou marginales lors des phases de conception et d'apprentissage de l'IA puis en vérifiant l'exactitude du modèle à mesure qu'il est alimenté par de nouvelles données. Le recours à des données synthétiques peut être considéré comme une solution possible pour minimiser la quantité de données personnelles traitées par des applications de l'IA. [26]

viii) Rôle des experts

Les développeurs, fabricants et prestataires de service en IA sont encouragés à créer et à consulter des comités indépendants composés d'experts issus de différents domaines ainsi qu'à collaborer avec des institutions universitaires indépendantes pouvant contribuer à concevoir des applications de l'IA fondées sur les droits de l'homme et orientées de façon éthique et sociale, ainsi qu'à détecter des biais potentiels. Ces comités peuvent jouer un rôle particulièrement important dans les secteurs où la transparence et la mobilisation des parties prenantes peuvent être plus difficiles à assurer en raison d'intérêts et de droits concurrents, par exemple dans les domaines de la justice prédictive, de la prévention du crime et de la détection des infractions.[27]

Des mécanismes adéquats devraient être mis en place pour assurer l'indépendance de ces comités d'experts.[28]

ix) Participation et contrôle démocratique du développement de l'IA

Des formes participatives d'évaluation des risques, reposant sur l'engagement actif des personnes et groupes potentiellement concernés par les applications de l'IA, devront être mises en place.

Les personnes, les groupes et les autres parties prenantes devraient être informés et impliqués de façon active dans le débat sur le rôle que l'IA devrait jouer dans la formation des dynamiques sociales et dans les processus décisionnels qui les concernent.[29]

Des dérogations peuvent être prévues dans l'intérêt général, lorsqu'elles constituent des mesures proportionnées dans une société démocratique et sont assorties de garanties adéquates.

x) Contrôle humain

Les produits et services de l'IA doivent être conçus de manière à garantir le droit des personnes de ne pas être soumises à des décisions qui les affectent de manière significative, prises uniquement sur le fondement d'un traitement automatisé de données, sans que leur point de vue soit pris en compte. Les produits et services de l'IA doivent permettre un contrôle humain global.[30]

[26] Voir GAI II.4, OCDE . Voir également CM/Rec(2020)1
[27] Voir Section II.3.
[28] Voir GAI II.6 et II.7, ICDPPC ; voir également article 11, UNESCO. Déclaration universelle sur le génome humain et les droits de l'homme (11 novembre 1997).
[29] Voir GAI II.7 et III.8, ICDPPC. Voir également CM/Rec(2020)1
[30] Voir Convention 108+, GAI II.8, ICDPPC, UNESCO

Le rôle de l'intervention humaine dans les processus décisionnels reposant sur l'IA et la liberté des décideurs humains de ne pas suivre les résultats de recommandations fondées sur l'utilisation de l'IA devraient être préservés.[31]

xi) Vigilance sur les algorithmes

Les développeurs, fabricants et prestataires de service en IA devraient adopter diverses formes de vigilance sur les algorithmes contribuant à la responsabilisation de toutes les parties prenantes par l'évaluation et la description des impacts attendus sur les individus et la société tout au long du cycle de vie du système d'IA, afin d'assurer le respect des droits de l'homme, de l'État de droit et de la démocratie.[32]

Les gouvernements devraient présenter des rapports réguliers sur leur utilisation de l'IA dans la police, le renseignement et la sécurité.[33]

xii) Liberté de choix

Afin d'accroître la confiance des utilisateurs, les développeurs, fabricants et prestataires de services en IA sont encouragés à concevoir leurs produits et services de manière à préserver la liberté de choix des utilisateurs concernant l'usage de l'IA, en proposant des alternatives réalistes aux applications de l'IA.[34]

xiii) Droit d'opposition

Le droit d'opposition quant aux systèmes d'IA basés sur des technologies qui influencent les opinions et le développement personnel des individus devrait être garanti.[35]

xiv) Interopérabilité

L'interopérabilité entre les systèmes d'IA doit être mise en œuvre conformément aux principes de licéité, de nécessité et de proportionnalité, en mettant en place des garanties adéquates concernant les droits de l'homme, la démocratie et l'État de droit.[36]

xv) Coopération

Une coopération doit être encouragée entre les autorités de contrôle ayant des compétences liées à l'IA.[37]

xvi) Éducation, éducation au numérique et formation professionnelle

[31] Voir GAI III. 4
[32] Voir GAI II.10, OCDE, ICDPPC . Voir également CM/Rec(2020)1
[33] Voir UNESCO
[34] Voir GAI II.9
[35] Voir GAI II.12 Voir également Section II.4
[36] Voir CM/Rec(2019)2
[37] Voir ICDPPC, GAI III.6

Les responsables politiques devraient investir des ressources dans l'éducation au numérique et l'éducation afin que les personnes concernées connaissent et comprennent mieux les applications de l'IA et leurs effets. Ils devraient également encourager la formation professionnelle des développeurs en IA pour les sensibiliser aux effets potentiels de l'IA sur les personnes et la société. Ils devraient soutenir la recherche sur l'IA orientée vers les droits de l'homme.[38]

xvii) Intégrité de la recherche scientifique

Lorsqu'une personne décide de se retirer d'un projet de recherche scientifique, le retrait du consentement ne compromet pas la licéité du traitement fondé sur le consentement effectué avant ce retrait ; ses données à caractère personnel devraient être détruites ou anonymisées de manière à ne pas compromettre la validité scientifique de la recherche et la personne concernée devrait en être informée.[39]

II.3 Santé

Le cadre réglementaire européen en matière de santé se compose des quelques instruments contraignants du Conseil de l'Europe et d'un certain nombre d'instruments sectoriels adoptés au niveau de l'UE, conformément à la nature, aux domaines d'activité et aux compétences réglementaires respectives de ces deux entités.

La Convention européenne des droits de l'homme, la Convention 108+ et la Charte sociale européenne contiennent plusieurs dispositions générales sur la protection de la santé et les droits connexes. Ces dispositions et principes, déjà énoncés dans d'autres instruments généraux au niveau international[40], sont toutefois contextualisés plus largement dans la Convention d'Oviedo.

La Convention d'Oviedo – seul instrument contraignant multilatéral entièrement consacré à la biomédecine – et ses protocoles additionnels constituent donc la principale référence pour définir les principes fondamentaux applicables dans ce domaine[41], principes qui devront être précisés et si nécessaire étoffés pour réglementer les applications de l'IA. La Convention est complétée par deux instruments non contraignants, la Recommandation sur les données relatives à la santé[42] et la Recommandation sur la recherche utilisant du matériel biologique d'origine humaine[43]. La première de ces deux recommandations illustre bien le lien étroit, examiné ci-après, entre la biomédecine (et plus généralement les soins de santé) et le traitement de données.

La réglementation en vigueur dans le domaine de la santé concerne principalement les traitements médicaux, la recherche (y compris les essais médicaux) et les dispositifs/produits médicaux. L'IA peut avoir un impact sur tous ces domaines, compte

[38] Voir ICDPPC, OCDE, GAI III.9, UNESCO. Voir également CM/Rec(2020)1

[39] Voir Convention 108+, CM/Rec(2019)2

[40] Certains principes généraux consacrés par cette convention étaient déjà présents dans de précédents instruments internationaux de droits de l'homme, tels que le Pacte international relatif aux droits civils et politiques, le Pacte international relatif aux droits économiques, sociaux et culturels et la Convention internationale des droits de l'enfant du 20 novembre 1989.

[41] Andorno, 2005 ; Seatzu, 2015.

[42] Voir la Recommandation Rec(2019)2 du Comité des Ministres aux États membres en matière de protection des données relatives à la santé.

[43] Voir la Recommandation Rec(2016)6 du Comité des Ministres aux États membres sur la recherche utilisant du matériel biologique d'origine humaine.

tenu de ses applications dans la médecine de précision[44], le diagnostic et les dispositifs et services médicaux.

Bien que la Convention d'Oviedo et les instruments non contraignants connexes aient été adoptés avant l'avènement de l'IA, ils apportent des garanties spécifiques concernant l'autodétermination, le traitement du génome humain et la recherche impliquant des êtres humains, non modifiées par l'application de l'IA à ce secteur.

Les problématiques qui se posent quant à l'autodétermination dans le domaine de la biomédecine sont les mêmes que celles qui ont déjà été évoquées dans le domaine du traitement des données. Malgré les différences de nature entre le consentement au traitement médical et le consentement au traitement de données, le haut niveau de complexité – et souvent, une certaine obscurité – des applications de l'IA peuvent compromettre dans les deux cas l'exercice effectif de l'autonomie individuelle[45].

À cet égard, la principale contribution de la Convention d'Oviedo à la future réglementation de l'IA ne provient pas des garanties qu'elle apporte dans un secteur donné, mais réside dans l'ensemble important de principes et de valeurs généraux qui peuvent en être tirés pour servir de fondement à cette réglementation.

Les principaux apports du Conseil de l'Europe dans le domaine de la médecine sont au nombre de huit et concernent la dignité humaine, la primauté de l'être humain, les normes professionnelles, la règle générale sur le consentement éclairé, la vie privée et le droit à l'information, la non-discrimination, la protection des personnes se prêtant à une recherche et le débat public. La contribution de la Convention d'Oviedo au débat sur la future réglementation de l'IA va au-delà de la biomédecine car plusieurs de ses dispositions, centrées sur le juste équilibre entre technologie et droits de l'homme, peuvent être étendues de manière générale au domaine de l'IA, comme expliqué dans les paragraphes suivants[46].

i) Primauté de l'être humain

Dans un contexte géopolitique et économique caractérisé par une course au développement de l'IA, la primauté de l'être humain doit être affirmée de façon générale comme un élément essentiel de l'approche européenne : l'amélioration des performances des systèmes reposant sur l'IA et leur rentabilité ne doivent pas l'emporter sur les intérêts et le bien-être des personnes.

Ce principe devrait s'appliquer aussi bien à la conception des systèmes d'IA (par ex. mise au point de systèmes qui violent les droits de l'homme et les libertés fondamentales) qu'à leurs usages.[47]

ii) Accès équitable aux soins de santé

Le principe de l'accès équitable peut être étendu à l'accès aux effets bénéfiques de l'IA. Cela suppose l'adoption de mesures adéquates pour traiter les risques liés à la fracture numérique, à la discrimination, à la marginalisation des personnes vulnérables ou des minorités culturelles et aux restrictions d'accès à l'information.[48]

---

[44] Voir Azencott, 2018 ; Ferryman et Pitcan, 2018.

[45] Voir ci-dessus, section 2.1 de cette partie.

[46] La dignité humaine et le consentement éclairé ne figurent pas dans le tableau car la première est une valeur commune aux instruments adoptés par le Conseil de l'Europe dans le domaine des droits de l'homme, de la démocratie et de l'Etat de droit (voir paragraphe 1, partie II) et le second est un principe également applicable dans le cadre du traitement de données.

[47] Voir Oviedo Convention, Article 2.

[48] Voir Oviedo Convention, Article 3

iii) Normes professionnelles

Le développement de l'IA recouvre une multitude de domaines d'expertise et lorsque la mise au point de systèmes d'IA peut avoir un impact sur les personnes et sur la société, elle doit être effectuée conformément aux obligations et aux normes professionnelles applicables dans chaque domaine d'expertise concerné.

Les normes et compétences professionnelles requises doivent se fonder sur l'état actuel des connaissances.[49]

Les États devraient encourager la formation des professionnels pour améliorer leur connaissance et leur compréhension de l'IA et de ses effets potentiels sur les personnes et sur la société. Ils devraient soutenir la recherche relative à l'IA axée sur les droits de l'homme.

Les États devraient coopérer en vue de l'élaboration de programmes de formation communs et de normes communes pour les professionnels concernés par l'IA et la société.

Lorsque l'IA est mise en œuvre dans le secteur médical, une attention particulière doit être portée à la confiance du patient envers son médecin et à la confiance mutuelle, qui ne doit pas être compromise par le recours à l'IA.

La décision de recourir à l'IA dans les processus décisionnels concernant les personnes et la société doit reposer sur le principe de proportionnalité, en prenant en compte les intérêts et le bien-être des personnes.

iv) Protection des personnes n'ayant pas la capacité de consentir et qui n'ont pas la capacité de consentir à une recherche

Le respect du principe du bénéfice doit être considéré comme une obligation lorsque le consentement individuel est soumis à des restrictions compte tenu de la complexité ou de l'opacité des traitements reposant sur l'IA et ne peut constituer le fondement exclusif du traitement.[50]

v) Vie privée et droit à l'information

Conformément à l'article 10 de la Convention d'Oviedo, les applications d'IA en matière de santé doivent garantir le droit à l'information et respecter la volonté des personnes de ne pas être informées, hormis lorsque le respect de cette volonté constitue un risque grave pour la santé d'autrui[51].

vi) Non-discrimination

Le principe de non-discrimination dans le domaine de la santé doit être complété en interdisant toute forme de discrimination contre une personne ou un groupe fondée sur des prédictions de problèmes de santé futurs[52].

vii) Rôle des experts

[49] Voir Oviedo Convention, Article 4; Recommendation CM/Rec(2019)2 on the protection of health-related data.

[50] Voir Oviedo Convention, Articles 6 and 17.
[51] Voir Oviedo Convention, Article 10
[52] Voir article 5, CM/Rec(2016)6. Voir également Oviedo Convention, Article 11.

L'expérience des comités d'éthique dans le domaine de la biomédecine doit être prise en compte en associant des comités d'experts multidisciplinaires à l'évaluation des applications de l'IA.[53]

viii) Débat public

Les évolutions de l'IA posent des questions fondamentales qui doivent être soumises à un débat public justifié au regard de leurs incidences sur le plan médical, social, économique, éthique et juridique, et leurs applications éventuelles doivent faire l'objet de consultations appropriées.[54]

Ces considérations montrent que le cadre juridique en vigueur dans le domaine de la biomédecine contient des principes et des éléments importants qui peuvent être étendus à la future réglementation de l'IA, au-delà du secteur de la santé. Cela dit, vu l'impact de l'IA, il reste encore un certain nombre de lacunes à combler dans les domaines suivants.

a) Systèmes de prise de décision [approche contextuelle, équité, qualité des données, surveillance/contrôle humains]

Ces dernières années, un nombre croissant d'applications de l'IA ont été développées et sont utilisées dans le secteur médical à des fins de diagnostic, grâce à des solutions d'analytique et d'apprentissage machine. De vastes ensembles de données sont ainsi créés et l'analyse prédictive est utilisée pour tenter de trouver des solutions à des cas cliniques en s'appuyant sur les connaissances et les pratiques existantes. De la même manière, en ce qui concerne la reconnaissance d'images, les applications d'apprentissage machine sembleraient pouvoir améliorer les capacités de dépistage du cancer. Pour ce qui est de la médecine de précision, la collecte et l'analyse à grande échelle de sources de données multiples (données médicales mais également données non médicales, comme la qualité de l'air et la qualité du logement) servent à obtenir des indications individualisées sur la santé et la maladie.

L'utilisation des données cliniques, des connaissances et pratiques médicales mais aussi de données non médicales n'est pas nouveau en soi dans le secteur de la médecine et des études de santé publique. Cela dit, l'ampleur de la collecte de données, le niveau de détail de l'information recueillie, la complexité (et dans certains cas l'opacité) du traitement de données et la nature prédictive des résultats de l'analyse font s'interroger sur les systèmes de prise de décisions et leurs éventuelles faiblesses.

La plupart de ces questions ne se limitent pas au secteur de la santé car les biais potentiels (notamment le manque de diversité et l'exclusion des valeurs aberrantes et des populations de plus petite taille), la qualité des données, la décontextualisation, la nature contextuelle de l'étiquetage des données et la réutilisation des données[55] sont des caractéristiques communes à de nombreuses applications de l'IA et concernent les

[53] Voir Oviedo Convention, Article 16.

[54] Voir Oviedo Convention, Article 28.

[55] Ferryman et Pitcan, 2018, 19-20 (« les désignations des maladies, par exemple la notion d'infection, n'étant pas clairement définies, elles peuvent décrire des réalités cliniques très différentes » et « ces enregistrements n'ont pas été conçus à des fins de recherche mais de facturation, ce qui peut être une source d'erreurs et de biais systématiques »)

données en général[56]. En cohérence avec la méthodologie adoptée[57], les lignes directrices existantes dans le domaine de la protection des données[58] peuvent aussi s'appliquer dans ce cas et les aspects relatifs à la qualité des données peuvent être étendus aux données non personnelles.

b) Autodétermination [liberté de choix/autonomie, sensibilisation]

L'opacité des applications d'IA et l'utilisation novatrice des données dans des analyses à grande échelle bouleversent la notion traditionnelle de consentement, que ce soit pour le traitement de données[59] ou les traitements médicaux, ce qui suggère l'adoption de nouvelles formes de consentement comme le consentement large[60] ou dynamique, lesquelles ne contribueraient toutefois qu'en partie à résoudre le problème.

c) Relation médecin-patient

Plusieurs aspects du diagnostic par IA, comme la perte des informations qui ne peuvent être encodées en données[61], le recours excessif à l'IA dans les décisions médicales, les effets des pratiques locales sur les ensembles de données d'entraînement et le risque de déqualification dans le secteur médical[62] peuvent avoir une incidence sur la relation entre les soignants et le patient[63] et devraient être évalués lorsque l'IA est adoptée dans ce domaine.

d) Gestion des risques [approche axée sur les risques, obligation de rendre compte]

Les dispositifs médicaux[64] constituent un exemple intéressant sur le plan de la gestion des risques car leur utilisation peut avoir d'importantes conséquences sur les personnes. L'Union européenne a adopté en la matière un modèle de classification basé sur les

[56] Voir sous-partie II.2 ci-dessus.

[57] Voir par ex. ci-dessus figure 3 : Valeurs fondamentales communes aux domaines de la protection des données et de la justice.

[58] Voir Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatique des données à caractère personnel 2017. Lignes directrices sur la protection des personnes à l'égard du traitement des données à caractère personnel à l'ère des mégadonnées. T-PD(2017)1 ; Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatique des données à caractère personnel 2019. Lignes directrices sur l'intelligence artificielle et la protection des données. T-PD(2019)01. Voir également les études préliminaires connexes : Mantelero, A. 2019 ; Rouvroy, A. 2016..

[59] Voir également la Recommandation Rec(2019)2 du Comité des Ministres du Conseil de l'Europe aux Etats membres en matière de protection des données relatives à la santé.

[60] Voir également le rapport explicatif de la Convention 108+, par. 43 (« En matière de recherche scientifique, il arrive fréquemment qu'il ne soit pas possible de cerner entièrement la finalité du traitement des données à caractère personnel à des fins de recherche scientifique au moment de la collecte des données. Par conséquent, les personnes concernées devraient pouvoir donner leur consentement en ce qui concerne certains domaines de la recherche scientifique, dans le respect des normes éthiques reconnues en matière de recherche scientifique. Les personnes concernées devraient pouvoir donner leur consentement uniquement pour ce qui est de certains domaines de la recherche ou de certaines parties de projets de recherche, dans la mesure où la finalité visée le permet ») et la Recommandation CM/Rec(2019)2 du Comité des Ministres du Conseil de l'Europe aux Etats membres en matière de protection des données relatives à la santé, 15.6 (« Dans la mesure où il n'est pas toujours possible de définir de façon préalable les finalités des différents projets de recherche au moment de la collecte des données, les personnes concernées devraient pouvoir donner un consentement uniquement pour certains domaines de recherche ou certaines parties de projets de recherche, dans la mesure où la finalité visée le permet et en tenant compte des normes éthiques reconnues »).

[61] Caruana et autres, 2015.

[62] Cabitza, Rasoini et Gensini, 2017.

[63] Voir aussi Déclaration d'Helsinki de l'AMM – Principes éthiques applicables à la recherche médicale impliquant des êtres humains, 9 juillet 2018, https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/.

[64] Commission européenne, 2014.

risques[65] qui prévoit des garanties progressives selon la classe de risque de chaque dispositif (procédures d'évaluation de conformité placées sous la seule responsabilité du fabricant ou avec l'intervention d'un organisme notifié, inspection par un organisme notifié ou, pour les risques les plus élevés, mise sur le marché soumise à autorisation préalable).

Un modèle reposant sur de telles garanties progressives pourrait être généralisé à la future réglementation de l'IA et adopté également en dehors du domaine des dispositifs médicaux, en prêtant attention à l'impact sur les droits de l'homme et les libertés fondamentales. Cela dit, il est plus difficile de procéder à une classification des produits/services de l'IA en raison de leur diversité et de leurs différents domaines d'application : il faudrait alors établir plusieurs classifications sectorielles, ou adopter des critères généraux sur la base de procédures d'évaluation des risques.

Des dispositions spécifiques sur la vigilance dans le domaine de l'IA et l'adoption du principe de précaution pour le développement de l'IA, dont il a été question précédemment[66], peuvent contribuer à résoudre ces problèmes.

II.4 Démocratie

La démocratie couvre un très large éventail de défis juridiques et sociétaux[67] dont la plupart sont susceptibles d'être relevés avec le soutien des TIC[68]. En effet, l'IA peut jouer un rôle important dans le développement présent et futur de la démocratie numérique au sein d'une société de l'information.

En comparaison avec les autres domaines examinés (protection des données et santé), ce sujet est si vaste qu'il est difficile de se référer à un seul instrument juridique contraignant spécifique au secteur. Plusieurs instruments internationaux traitent de la démocratie et de ses différents aspects, à commencer par la Déclaration des droits de l'homme des Nations Unies et le Pacte international relatif aux droits civils et politiques. De la même manière, dans le contexte européen, les principes essentiels de la démocratie sont présents dans plusieurs sources internationales.

Sur la base de l'article 25 du PIDCP, nous pouvons identifier deux principaux domaines d'intervention : i) la participation[69] et la bonne gouvernance et ii) les élections. Il est sans conteste difficile, voire impossible de tracer une ligne rouge entre ces domaines du fait qu'ils sont interconnectés de diverses manières. L'IA peut avoir un impact sur tous : la participation (par ex. l'engagement des citoyens, les plateformes de participation), la bonne gouvernance (par ex. l'e-gouvernance, les processus décisionnels, les villes intelligentes), la phase préélectorale (par ex. le financement, le

[65] Directive 93/42/CEE

[66] Voir ci-dessus paragraphe 2.1 de cette partie.

[67] Voir par ex. Conseil de l'Europe. Direction générale de la démocratie - Comité européen sur la démocratie et la gouvernance. 2016. Recueil des textes les plus pertinents du Conseil de l'Europe dans le domaine de la démocratie.

[68] Voir par ex. Direction générale de la démocratie et des affaires politiques - Direction des institutions démocratiques. 2009. Projet « Bonne gouvernance dans la société de l'information », CM(2009)9 Addendum 3. Guides indicatifs et Glossaire relatifs à la Recommandation Rec(2009) 1 du Comité des Ministres aux États membres sur la démocratie électronique (e-démocratie), élaborés par le Comité ad hoc pour la démocratie électronique (CAHDE) du Conseil de l'Europe ; Protocole additionnel à la Charte européenne de l'autonomie locale sur le droit de participer aux affaires des collectivités locales, 2009, article 2.2.iii.

[69] Pour une analyse plus détaillée, voir Faye Jacobsen, 2013. Voir également Maisley, 2017.

ciblage et le profilage, la propagande), les élections (par ex. la prévision des résultats électoraux, le vote électronique) et la période postélectorale (par ex. le règlement des litiges électoraux).

Comme dans toute classification, cette distinction se caractérise par une marge directionnelle. Il convient de souligner ici qu'il s'agit d'une classification fonctionnelle basée sur les différents impacts de l'IA, sans aucune intention de donner une représentation juridique ou politique de la démocratie et de ses divers éléments clés. La relation entre la participation, la bonne gouvernance et les élections peut ainsi être considérée sous différents angles et façonnée de plusieurs façons, uniformisant certains domaines ou les divisant davantage.

La participation s'exprime tant par la participation au débat démocratique que par le processus électoral. Cependant, dans ces deux cas, les outils d'IA interagissent différemment avec la participation et il existe des instruments juridiques internationaux distincts spécifiques au processus électoral.

II.4.1 Participation et bonne gouvernance

Le droit de participer aux affaires publiques (article 25 du Pacte), étroitement lié à la liberté d'expression, de réunion et d'association,[70] repose sur un vaste concept d'« affaires publiques »[71] qui englobe à la fois le débat public et le dialogue entre les citoyens et leurs représentants. À cet égard, l'IA est pertinente à deux points de vue : comme moyen de participation et comme objet des décisions participatives.

Si l'on considère l'IA comme un moyen, les obstacles techniques et éducatifs peuvent compromettre l'exercice du droit de participer. Les outils de participation fondés sur l'IA devraient donc tenir compte des risques de sous-représentation et de manque de transparence dans les processus participatifs (par ex. les plateformes de rédaction des projets de loi. Parallèlement, l'IA fait également l'objet de décisions participatives étant donné qu'elles incluent des décisions portant sur le développement de l'IA en général et sur son utilisation dans les affaires publiques.

Les plateformes participatives basées sur l'IA (Consul,[72] Citizenlab,[73] Decidim[74]) peuvent contribuer de manière significative au processus démocratique, en facilitant les échanges entre citoyens, la définition des objectifs prioritaires et les approches collaboratives en matière de prise de décisions[75] sur des sujets d'intérêt général à

[70] Voir Haut-Commissariat des Nations Unies aux droits de l'homme. 1996. Observation générale n° 25 : Le droit de participer aux affaires publiques, le droit de vote et le droit d'accéder, dans des conditions d'égalité, aux fonctions publiques (article 25). CCPR/C/21/Rev.1/Add.7.

[71] Voir également Haut-Commissariat des Nations Unies aux droits de l'homme. 1981. CESCR, Observation générale n° 1 : Rapports des États parties, paragraphe 5 (« de faciliter l'évaluation, par l'opinion publique, des politiques nationales en matière de droits économiques, sociaux et culturels, et d'encourager la participation des divers secteurs économiques, sociaux et culturels de la société à la formulation de ces politiques, à leur mise en œuvre et à leur réexamen »).

[72] Voir <https://consulproject.org/en/>, consulté le 29.12.2019.

[73] Voir < https://www.citizenlab.co/>, consulté le 29.12.2019.

[74] Voir < https://decidim.org/>, consulté le 29.12.2019.

[75] Voir également Conseil de l'Europe. Lignes directrices relatives à la participation civile aux décisions politiques CM(2017)83-final. Adoptées par le Comité des Ministres le 27 septembre 2017 lors de la 1295e réunion des Délégués des Ministres.

différents niveaux (quartier, municipalité, métropole, région, pays).[76] Ces plateformes étant utilisées dans un environnement social et dans le cadre de la collecte d'informations, les mêmes aspects déjà examinés concernant la protection des données, notamment la sécurité, peuvent être rappelés ici en étendant à ces applications les lignes directrices sur les données abordées dans la sous-partie précédente.

Cependant, d'autres questions plus spécifiques se posent en ce qui concerne les outils d'IA pour la participation démocratique (notamment ceux pour la prévention et la lutte contre la corruption[77]) ; elles concernent les quatre principaux domaines suivants : la transparence, l'**obligation de rendre compte**, l'**inclusivité**, et l'**ouverture**. À cet égard, les principes généraux énoncés dans les instruments internationaux contraignants trouvent une application importante dans la Recommandation CM/Rec(2009)1 du Comité des Ministres aux États membres sur la démocratie électronique (e-démocratie), qui jette les bases d'une élaboration plus précise des principes directeurs en matière de démocratie dans le domaine de l'IA.

La **transparence** est une exigence pour l'utilisation d'applications technologiques à des fins démocratiques.[78] Ce principe est commun aux sujets analysés ci-dessus, les données et la santé. Il s'agit, toutefois, d'une notion fondée sur le contexte. Si dans ces domaines la transparence est étroitement liée à l'autodétermination, elle prend ici une signification plus large. Dans un processus démocratique, la transparence n'est pas seulement une condition de l'autodétermination des citoyens relativement à un outil technique, c'est aussi un élément du processus participatif démocratique.[79] La transparence n'a plus de dimension individuelle, mais revêt une dimension collective comme garantie du processus démocratique.

Dans ce contexte, les solutions de démocratie électronique fondées sur l'IA doivent être utilisées en toute transparence eu égard à leur logique et à leur fonctionnement (par ex. la sélection du contenu des plateformes participatives), grâce à la communication d'informations claires, facilement accessibles, intelligibles et actualisées sur les outils d'IA concernés.[80]

De plus, la mise en œuvre de cette notion de transparence devrait aussi tenir compte de la grande diversité d'utilisateurs de ces outils. L'adoption d'une approche **accessible**[81] dès les premiers stades de la conception des outils d'IA permettrait de

[76] Voir également la Recommandation CM/Rec(2009)2 sur l'évaluation, l'audit et le suivi de la participation et des politiques de la participation aux niveaux local et régional.

[77] Voir l'article 13 de la Convention des Nations Unies contre la corruption, 2003.

[78] Voir le paragraphe 6 de la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie).

[79] Voir également les Lignes directrices relatives à la participation civile aux décisions politiques. CM(2017)83-final, IV.

[80] Voir le paragraphe 6 de la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie) (« faciliter et améliorer l'accès, l'accessibilité […] à l'aide, si possible, de moyens transparents […] ») et le paragraphe P.57 de l'annexe à la Recommandation CM/Rec(2009)1. Voir également les paragraphes 2.1.3 et 3.2 de l'Annexe à la Recommandation CM/Rec(2016)5 sur la liberté d'internet.

[81] Voir également le paragraphe B.IV de l'annexe à la Recommandation CM/Rec(2018)4 relative à la participation des citoyens à la vie publique locale.

garantir une transparence effective à l'égard des groupes vulnérables et des personnes handicapées, apportant une valeur ajoutée à l'accessibilité dans ce contexte.

La transparence et l'accessibilité sont étroitement liées à la nature de l'architecture utilisée pour élaborer les systèmes d'IA. Les solutions **à code source ouvert et les normes ouvertes**[82] peuvent ainsi contribuer au contrôle démocratique des applications d'IA les plus critiques.[83] Dans certains cas, l'ouverture est compromise par des restrictions dues à la nature de l'application d'IA spécifique (par ex. la prévention de la criminalité). Dans de tels cas, les audits et les mécanismes de certification jouent un rôle encore plus important qu'ils ne le font déjà pour les systèmes d'IA en général.[84]

Dans le contexte des applications d'IA visant à favoriser la participation démocratique, l'**interopérabilité**[85] peut également jouer un rôle majeur en facilitant l'intégration entre les différents services/plateformes de démocratie électronique à différents niveaux géographiques. Déjà pertinent pour la démocratie électronique en général,[86] cet aspect devrait être étendu à la conception de systèmes fondés sur l'IA.

Un autre principe clé de la démocratie électronique, comme dans les secteurs des données et de la santé, est l'**obligation de rendre des comptes**. Contrairement aux principes examinés ci-dessus, l'obligation de rendre des comptes garde ici la même signification. Par conséquent, elle ne semble pas nécessiter une application sectorielle dans le contexte de l'IA, autre que son application générale.

Enfin, étant donné le rôle des médias dans le contexte de la participation démocratique, et conformément à la Recommandation CM/Rec(2016)4 du Comité des Ministres du Conseil de l'Europe,[87] les applications d'IA ne doivent pas compromettre la confidentialité ou la sécurité des communications ni la protection des sources journalistiques et des lanceurs d'alerte.[88]

Lorsque l'on aborde les différents aspects liés au développement des solutions d'IA pour la participation démocratique, la première idée qui vient à l'esprit est qu'une approche démocratique est incompatible avec une approche techno-déterministe. Les solutions d'IA visant à traiter les problèmes sociétaux devraient, par conséquent, être

[82] Voir également le paragraphe 6 de la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie) et le paragraphe P.54 de l'annexe.

[83] Voir également le paragraphe G.58 de l'annexe à la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie).

[84] Il convient de souligner que les audits et les mécanismes de certification jouent également un rôle important dans le cas d'une architecture d'IA à code source ouvert, étant donné que cette nature ne suppose pas en soi l'absence de partialité ni de tout autre défaut. Voir également les paragraphes P.55 et G.57 de l'annexe à la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie), (« les logiciels de la démocratie électronique devraient être à code source ouvert et pouvoir être inspectés ou, alternativement, être homologués par un organisme indépendant »).

[85] Voir également les paragraphes P.56, G.56, 59 et 60 de l'annexe à la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie).

[86] Voir également le paragraphe 6 de la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie)

[87] Voir le paragraphe 2 de l'annexe à la Recommandation CM/Rec(2016)4 sur la protection du journalisme et la sécurité des journalistes et autres acteurs des médias, Conseil de l'Europe, Assemblée parlementaire. 2019. Résolution 2254 (2019) La liberté des médias en tant que condition pour des élections démocratiques.

[88] Voir également la Résolution 2300 (2019)1 de l'Assemblée parlementaire, Améliorer la protection des lanceurs d'alerte partout en Europe ; Recommandation CM/Rec(2014)7 sur la protection des lanceurs d'alerte.

le résultat d'un processus inclusif. Ainsi, les valeurs telles que la protection des minorités, le pluralisme et la diversité devraient absolument être examinées dans le cadre du développement de ces solutions.

D'un point de vue démocratique, la première question que nous devrions poser est la suivante : devons-nous vraiment privilégier une solution fondée sur l'IA plutôt qu'une autre possibilité pour résoudre un problème donné,[89] étant donné l'impact potentiel de l'IA sur les droits et les libertés ? Dans l'affirmative, l'étape suivante consiste à examiner **l'intégration des valeurs** dans le développement de l'IA.[90]

Les solutions d'IA proposées doivent être conçues sous l'angle des droits de l'homme. Elles doivent garantir le plein respect des droits de l'homme et des libertés fondamentales, et notamment adopter des **outils et des procédures d'évaluation** à cette fin.[91] Dans le cas des applications d'IA ayant un fort impact sur les droits de l'homme et les libertés, telles que les processus électoraux, il convient **préalablement d'évaluer** la conformité avec la législation. En outre, les systèmes d'IA prévus pour les tâches publiques devraient **pouvoir faire l'objet d'un audit** et, lorsque cela n'est pas exclu par des intérêts premiers concurrents, les audits devraient être publiés.

Un autre aspect important à prendre en considération est le **partenariat public-privé** qui caractérise souvent les services d'IA aux citoyens ; il convient d'évaluer le meilleur choix entre les solutions internes et tierces, y compris les nombreuses combinaisons de ces deux extrêmes. À cet égard, lorsque des solutions d'IA sont totalement ou partiellement développées par des sociétés privées, la **transparence des contrats** et la **clarté des règles sur l'accès aux données des citoyens et leur utilisation** ont une valeur essentielle en termes de contrôle démocratique.

Il est pertinent de restreindre l'accès aux données des citoyens et leur utilisation non seulement du point de vue de la protection des données (principes de la minimisation des données et de la limitation des finalités), mais plus généralement eu égard au volume de données générées par une communauté, parmi lesquelles on trouve également des données à caractère non personnel et des données agrégées. Dans l'environnement numérique, cette question devrait être considérée comme faisant partie de la démocratie. En outre, la **dimension collective** des ressources numériques générées par une communauté devrait comporter des formes de surveillance et de contrôle par les citoyens, comme c'est le cas pour les autres ressources d'un territoire/d'une communauté (par exemple, l'environnement).

Les considérations exprimées ci-dessus sur le caractère essentiel de l'ouverture dans les outils de participation démocratique devraient être rappelées ici, étant donné leur impact sur la conception des systèmes d'IA. De plus, la conception, le développement et le déploiement de ces systèmes devraient également prendre en compte l'adoption d'une stratégie écologique et durable.[92]

---

[89] Voir également le paragraphe 5.7 de l'annexe à la Recommandation CM/Rec(2020)1.

[90] Voir également la Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques (adoptée par le Comité des Ministres le 13 février 2019 lors de la 1337e réunion des Délégués des Ministres), paragraphe 7.

[91] Voir la Recommandation CM/Rec(2009)1, paragraphes 5 et 6, et Annexe à la Recommandation CM/Rec(2009)1, paragraphe G.67. Voir également la sous-partie II.2 sur les données et le rôle des comités d'experts et A Mantelero, 'AI *and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*' (2018) 34 Computer Law & Security Review 754.

[92] Voir également la Recommandation CM/Rec(2009)1, Annexe, paragraphe P. 58.

Enfin, il convient de noter que si la conception de l'IA est un élément clé de ces systèmes, elle n'est pas neutre. Des valeurs peuvent être intégrées dans les composants technologiques,[93] notamment les systèmes d'IA. Ces valeurs peuvent être choisies de manière intentionnelle ; dans le contexte de la démocratie électronique, ce choix doit reposer sur un processus démocratique. Cependant, certaines valeurs peuvent aussi être intégrées involontairement dans les solutions d'IA, du fait de la composition culturelle, sociale et hommes-femmes des équipes de développeurs d'IA. D'où la valeur ajoutée du **caractère inclusif** ici, à la fois en termes d'inclusion et de diversité[94] dans le développement de l'IA.

En ce qui concerne la bonne gouvernance,[95] les principes examinés pour la démocratie électronique peuvent être repris ici.[96] Tel est le cas avec les villes intelligentes et la gestion environnementale par capteurs, dans lesquels les processus décisionnels ouverts, transparents et inclusifs jouent un rôle majeur.[97] De même, l'utilisation de l'IA pour superviser les activités des collectivités locales,[98] à des fins de contrôle et de lutte contre la corruption,[99] devrait être fondée sur l'**ouverture** (logiciels à code source ouvert), la **transparence** et l'**auditabilité**.

Plus généralement, l'IA peut être utilisée dans les échanges entre les gouvernements et les citoyens afin d'automatiser les requêtes et les demandes d'informations des citoyens.[100] Cependant, dans ce cas, il est important de garantir le droit de savoir que nous interagissons avec une machine[101] et d'avoir un interlocuteur humain. En outre, l'accès aux services publics ne doit pas dépendre de la communication de données inutiles et inappropriées.

Il convient d'accorder une attention particulière à l'utilisation potentielle de l'IA dans l'interaction homme-machine afin de mettre en œuvre des stratégies d'incitation.[102] Ici, en raison de la complexité et de l'opacité des solutions techniques adoptées, l'IA peut accroître le rôle passif des citoyens et avoir des répercussions négatives sur le processus décisionnel démocratique. Il convient, au contraire, de privilégier une

[93] Voir également P-P Verbeek, 2011, 41-65.

[94] Voir également la Recommandation CM/Rec(2020)1, Annexe, paragraphe 3.5.

[95] Voir la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie), paragraphe P.4 de l'annexe (« […] bonne gouvernance, qui se caractérise par un exercice démocratique du pouvoir, efficient, efficace, participatif, transparent et responsable, s'appuyant sur des moyens électroniques, et inclut débat politique informel et intervention d'acteurs non gouvernementaux »).

[96] Voir également la Recommandation Rec(2004)15 sur la gouvernance électronique (« e-gouvernance ») ; Conseil de l'Europe. 2008. Les 12 Principes de bonne gouvernance ancrés dans la Stratégie sur l'innovation et la bonne gouvernance au niveau local, adoptée par une décision du Comité des Ministres du Conseil de l'Europe en 2008.

[97] Voir également Privacy International, 2017.

[98] Voir également la Recommandation CM/Rec(2019)3 sur le contrôle des actes des collectivités locales, Annexe, Lignes directrices concernant l'amélioration des systèmes de contrôle de l'action des collectivités locales, paragraphes 4 et 9.

[99] Voir aussi Savaget, Chiarini et Evans, 2019, examen de l'affaire brésilienne « Operação Serenata de Amor » (OSA).

[100] Voir Mehr. 2017.

[101] Voir également GAI 2.11.

[102] Voir, *ex multis*, Sunstein, 2015a ; Sunstein, 2015a ; Sunstein et Thaler, 2003 ; Thaler et Sunstein, 2008.

approche dynamique fondée sur une participation consciente et active aux objectifs de la communauté, qui sera mieux gérée par les outils de participation basés sur l'IA. Lorsqu'elles sont adoptées, les stratégies d'incitation devraient toutefois suivre une approche fondée sur des données probantes.

Enfin, l'utilisation des systèmes d'IA dans les tâches de gouvernance soulève des questions épineuses sur la relation entre les décideurs humains et le rôle de l'IA dans le processus décisionnel.[103] Ces questions concernent davantage les fonctions qui ont un fort impact sur les droits et les libertés individuels, comme c'est le cas des décisions juridictionnelles. Pour cette raison, les préoccupations relatives à la transparence (notamment l'explicabilité) du raisonnement de l'IA et à la relation entre l'utilisation de l'IA et la liberté des décideurs seront analysées dans la cinquième sous-partie.

II.4.2 Élections

Comme dans d'autres domaines, l'impact de l'IA sur les processus électoraux est vaste et touche la phase préélectorale, les élections et la phase postélectorale de diverses manières. Cependant, il ne suffit pas d'analyser les étapes du processus électoral pour mettre en évidence les différentes interactions avec les solutions d'IA.

L'influence de l'IA est mieux représentée par la distinction suivante : l'IA pour le processus électoral d'une part (vote électronique, prévision des résultats et règlement des litiges électoraux) et l'IA pour les campagnes électorales (micro-ciblage ou profilage, propagande et fausses nouvelles). Si dans le premier domaine l'IA consiste principalement en une amélioration technologique d'un processus existant, dans le domaine des campagnes électorales le profilage et la propagande basés sur l'IA suscitent de nouvelles préoccupations qui ne sont pas totalement prises en compte par le cadre juridique existant. Par ailleurs, plusieurs documents ont mis en avant le rôle actif des États dans la création d'un environnement propice à la liberté d'expression.[104]

En ce qui concerne la mise en œuvre technologique de la démocratie électronique (vote électronique, prévision des résultats et règlement des litiges électoraux), plusieurs principes essentiels mentionnés eu égard à la participation démocratique sont également pertinents ici. **L'accessibilité,**[105] **la transparence,**[106] **l'ouverture,**[107] **la**

[103] Voir également Calo et Citron, 2020, à venir.

[104] Voir également la Recommandation CM/Rec(2018)1 sur le pluralisme des médias et la transparence de leur propriété ; la Déclaration conjointe sur la liberté d'expression et les fausses nouvelles (« fake news »), la désinformation et la propagande, le Rapporteur spécial des Nations Unies (UN) sur la liberté d'expression, la Représentante de l'Organisation pour la sécurité et la coopération en Europe (OSCE) pour la liberté des médias, le Rapporteur spécial de l'Organisation des États américains (OEA) sur la liberté d'expression et le Rapporteur spécial de la Commission africaine des droits de l'homme et des peuples (CADHP) sur la liberté d'expression et à l'accès à l'information (3 mars 2017). Voir également la Recommandation CM/Rec(2016)5 sur la liberté d'internet, Annexe, paragraphes 1.5, 2.1 et 3 ; Commission européenne pour la démocratie par le droit (Commission de Venise). 2019. Rapport conjoint de la Commission de Venise et de la Direction de la société de l'information et de la lutte contre la criminalité de la Direction Générale des droits de l'homme et de l'État de droit (DGI) sur les technologies numériques et les élections, paragraphe 151.E; OSCE, 2020. Voir également Bychawska-Siniarska, 2017.

[105] Voir la Recommandation CM/Rec(2017)5 sur les normes relatives au vote électronique, Annexe I, Normes relatives au vote électronique, paragraphes 1 et 2.

[106] Voir la Recommandation CM/Rec(2017)5, Annexe I, paragraphe 32. Voir également Conseil de l'Europe. Direction générale de la démocratie et des affaires politiques - Direction des institutions démocratiques. 2011. Lignes directrices relatives à la transparence des élections par voie électronique.

[107] Voir la Recommandation CM/Rec(2017)5, Annexe I, paragraphe 35.

gestion des risques et l'obligation de rendre des comptes (y compris l'adoption de procédures de certification et d'audit)[108] sont des éléments fondamentaux des solutions technologiques adoptées à ces stades du processus électoral.

En ce qui concerne l'utilisation de l'IA dans les campagnes électorales (micro-ciblage et profilage, propagande et fausses nouvelles), plusieurs questions soulevées concernent le traitement des données à caractère personnel en général. Les principes énoncés dans la Convention 108+ peuvent donc être appliqués et correctement contextualisés.[109]

Dans le cas de la propagande et de la désinformation, il convient d'apporter des réponses nouvelles et plus spécifiques.[110] Ici, les instruments contraignants et non contraignants existants ne définissent pas de conditions précises. En effet, cette désinformation repose sur de nouvelles formes de communication, telles que les réseaux sociaux, qui diffèrent des médias traditionnels[111] et contournent souvent la médiation professionnelle des journalistes.

Cependant, les principes généraux, tels que le **principe de non-ingérence** par les pouvoirs publics dans les activités médiatiques visant à influencer les élections,[112] peuvent être étendus à ces nouvelles formes de propagande et de désinformation. L'IA étant utilisée pour automatiser la propagande, sa future réglementation devrait élargir le champ d'application des principes généraux de la non-ingérence aux systèmes d'IA utilisés pour diffuser des informations fausses, trompeuses et préjudiciables. En outre, afin d'éviter une telle ingérence, les États[113] et les fournisseurs de médias sociaux devraient adopter une **approche dès la conception** pour accroître leur résilience à la désinformation et à la propagande.

De même, en vertu de l'obligation de couvrir les campagnes électorales de manière **équitable, équilibrée et impartiale**[114], les médias et les opérateurs de médias sociaux devraient assumer les obligations relatives à la transparence de la logique des

[108] Voir la Recommandation CM/Rec(2017)5, Annexe I, paragraphes 36, 37, 38, 39 et 40.

[109] Recommandation CM/Rec(2010)13 sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage et son analyse en cours. Voir Conseil de l'Europe, Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel. 2019. Profilage et la Convention 108+ : Pistes pour une actualisation. T-PD(2019)07BISrev.

[110] Voir Manheim et Kaplan, 2019 ; Commission européenne - Direction générale des réseaux de communication, du contenu et des technologies, '*A Multi-Dimensional Approach to Disinformation Report of the Independent High Level Group on Fake News and Online Disinformation*' (2018). Voir également l'affaire *Stoll c. Suisse* [GC], Requête n° 69698/01, § 104.

[111] Voir également la Recommandation CM/Rec(2011)7 sur une nouvelle conception des médias.

[112] Voir la Recommandation CM/Rec(2007)15 sur des mesures concernant la couverture des campagnes électorales par les médias, paragraphe I.1.

[113] Voir également la Déclaration conjointe sur la liberté d'expression et les fausses nouvelles (« fake news »), la désinformation et la propagande, paragraphe 2.c.

[114] Voir également le paragraphe II.1 de la Recommandation CM/Rec(2007)15.

algorithmes utilisés pour la sélection des contenus,[115] garantissant ainsi le pluralisme et la diversité des voix,[116] notamment les voix critiques.[117]

De plus, les États et les intermédiaires devraient promouvoir et faciliter l'accès aux outils permettant de détecter la désinformation et les agents non humains, et soutenir les recherches indépendantes sur l'impact de la désinformation et les projets proposant aux utilisateurs des services de vérification des informations (« fast-checking »).[118]

Étant donné le rôle majeur de la publicité dans la désinformation et la propagande, les critères utilisés par les solutions basées sur l'IA pour la publicité politique devraient être **transparents**,[119] **pouvoir faire l'objet d'audits** et offrir des **conditions équitables** à tous les partis et candidats politiques.[120] En outre, les intermédiaires devraient revoir leurs modèles publicitaires pour s'assurer qu'ils ne nuisent pas à la **diversité des opinions et des idées**.[121]

II.5 Justice

Comme indiqué dans la sous-partie précédente, le domaine de la justice est vaste. Il serait trop ambitieux d'envisager l'analyse de l'ensemble des conséquences de l'IA sur la justice et de leurs effets sur la démocratie. Conformément à la portée de cette étude, cette sous-partie décrit les principaux enjeux liés à l'utilisation de l'IA et les principes qui, sur la base des instruments internationaux juridiquement contraignants, peuvent contribuer à sa réglementation future.

En l'absence d'instruments contraignants spécifiques et dédiés, tels que la Convention 108+ et la Convention d'Oviedo, la justice se distingue des domaines de la protection des données et de la santé. Par conséquent, cette analyse est davantage

[115] Voir également la Déclaration conjointe sur la liberté d'expression et les fausses nouvelles (« fake news »), la désinformation et la propagande ; la Recommandation CM/Rec(2018)2 sur les rôles et les responsabilités des intermédiaires d'internet, Annexe, paragraphes 2.1.3 et 2.3.5 (« Du fait de la capacité actuellement limitée des moyens automatisés à évaluer le contexte, les intermédiaires devraient mesurer minutieusement les incidences qu'une gestion automatisée des contenus peut avoir sur le plan des droits de l'homme et procéder à un contrôle humain si nécessaire. Ils devraient tenir compte du risque de blocages insuffisants ou au contraire excessifs dus à des systèmes d'algorithmes inexacts, et des effets potentiels de ces algorithmes sur les services fournis sur le débat public »).

[116] Voir également le Code de bonnes pratiques contre la désinformation de l'UE, 2018.

[117] Voir aussi le paragraphe 15 de l'annexe à la Recommandation CM/Rec(2016)4.

[118] Voir également la Déclaration conjointe sur les fausses nouvelles (« fake news »), la désinformation et la propagande, paragraphe 4.E ; Commission européenne pour la démocratie par le droit. 2019, paragraphe 151.D.

[119] Voir également Conseil de l'Europe. Assemblée parlementaire. Résolution 2254 (2019) La liberté des médias en tant que condition pour des élections démocratiques, paragraphes 9.2 et 11.1 ; Commission européenne pour la démocratie par le droit (Commission de Venise). 2019. Rapport sur les technologies numériques et les élections, établi conjointement par la Commission de Venise et la Direction de la société de l'information et de la lutte contre la criminalité (Direction générale Droits de l'homme et État de droit, DGI), paragraphes 151.A et 151.B.

[120] Voir également la Recommandation CM/Rec(2007)15, paragraphe II.5.

[121] Voir également la Déclaration conjointe sur les fausses nouvelles (« fake news »), la désinformation et la propagande, paragraphe 4.e.

centrée sur la contextualisation des principes directeurs généraux que sur des instruments juridiques précis.

Cet exercice est facilité par la Charte éthique européenne d'utilisation de l'intelligence artificielle (IA) dans les systèmes judiciaires et leur environnement, adoptée par la CEPEJ en 2019, qui traite directement de la relation entre justice et IA. Bien que cet instrument non contraignant soit considéré comme une charte éthique, dans une large mesure il concerne les principes juridiques ancrés dans les instruments internationaux.

Les principes directeurs du développement de l'IA dans le domaine de la justice peuvent être tirés des instruments contraignants suivants : la Déclaration universelle des droits de l'homme, le Pacte international relatif aux droits civils et politiques, la Convention pour la protection des droits de l'homme et des libertés fondamentales, la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, la Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes et la Convention pour la protection des droits de l'homme et des libertés fondamentales.[122]

Compte tenu de la diversité des types et des objectifs des opérations dans ce domaine et de la variété des professionnels et des procédures, cette sous-partie établit une distinction fonctionnelle entre deux thèmes : i) les décisions judiciaires et les modes alternatifs de règlement des conflits et ii) la prévention/prédiction de la criminalité. Avant d'analyser et de contextualiser les principes clés relatifs à ces deux domaines, il convient de formuler quelques observations générales, qui peuvent également s'appliquer à l'action de l'administration publique dans son ensemble.[123]

Premièrement, il convient de noter que – en comparaison avec les décisions humaines, et plus spécifiquement les décisions judiciaires – la logique qui sous-tend les systèmes d'IA ne ressemble pas à un raisonnement juridique. Ces systèmes exécutent simplement des codes basés sur une approche mathématique/statistique centrée sur les données.

Deuxièmement, les taux d'erreur de l'IA sont proches voire inférieurs à ceux du cerveau humain dans des domaines tels que l'étiquetage d'images, mais les processus décisionnels plus complexes affichent des taux d'erreur supérieurs. C'est le cas du raisonnement juridique en matière de résolution de problèmes.[124] Cela étant, si la classification erronée de l'image d'un chat peut avoir des effets négatifs limités, un taux d'erreur dans les décisions juridiques aura une incidence importante sur les droits et les libertés individuels.

---

[122]Voir également, pour ce qui est de la zone UE, la Charte des droits fondamentaux de l'Union européenne.

[123] Voir ci-dessus sous-partie II.4.

[124] Voir Dupont et al., 2018, 148 (« L'apprentissage profond n'a aucun moyen naturel de gérer la structure hiérarchique, ce qui signifie que toutes les variables disponibles sont considérées au même niveau, comme « plates » ou non hiérarchiques. Cela constitue un obstacle majeur lorsque les décisions ont un poids moral ou juridique important qui doit l'emporter sur d'autres éléments »). Voir également Osoba et Welser, 2017, 18 (« Si l'on examine le problème sous un autre angle, on constate que les jugements dans l'espace du comportement social sont souvent flous et non des critères binaires bien définis […]. Nous sommes capables d'apprendre à nous orienter dans des relations complexes et floues, comme celles avec les gouvernements et les lois, en nous fiant souvent à des évaluations subjectives pour ce faire. Les systèmes qui reposent sur un raisonnement quantifié (comme la plupart des agents artificiels) peuvent reproduire l'effet mais exigent souvent une conception soignée. Pour saisir cette nuance, il faudra peut-être plus que des informaticiens et des spécialistes des données. »). Voir également Cummings et al., 2018, 13.

Il convient de signaler que la différence entre les erreurs de décisions humaines et automatisées a une conséquence importante en termes d'échelle : si l'erreur humaine ne se répercute que sur des cas individuels, la mauvaise conception et la partialité des solutions d'IA produisent inévitablement des effets sur toutes les personnes qui se trouvent dans des circonstances identiques ou similaires puisque les outils d'IA sont appliqués à toute une série de cas. Cela peut créer une discrimination de groupe portant atteinte à des personnes appartenant à différentes catégories.

Au vu de la nature textuelle des documents juridiques, le traitement du langage naturel joue un rôle important dans les applications d'IA de la sphère juridique. Cela pose plusieurs questions cruciales autour des solutions commerciales développées spécifiquement pour le marché anglophone, moins efficaces dans un environnement juridique qui utilise une autre langue que l'anglais.[125] De plus, les décisions juridiques se caractérisent souvent par un raisonnement implicite non exprimé, qui peut être perçu par des systèmes experts, mais pas par des outils d'apprentissage automatique des langues. Enfin, la présence de clauses générales exige de bien comprendre l'interprétation juridique pertinente et les mises à jour continues qui ne peuvent être tirées de l'analyse de textes.

Toutes ces contraintes suggèrent d'adopter l'IA de manière prudente et plus critique dans le domaine de la justice par rapport aux autres domaines. Par ailleurs, en ce qui concerne les décisions de justice et les modes alternatifs de règlement des litiges, il convient de faire la distinction entre les affaires caractérisées par des évaluations factuelles régulières et les affaires caractérisées par une importante marge de raisonnement juridique et de discrétion.[126]

II.5.1 Décisions de justice et modes alternatifs de règlement des litiges

Plusieurs produits d'IA dits technico-juridiques n'ont pas d'impact direct sur les processus décisionnels au sein des tribunaux ou sur les modes alternatifs de règlement des litiges. Ils facilitent plutôt la gestion des connaissances et des contenus, la gestion organisationnelle et l'évaluation des performances.[127] Parmi ces applications figurent, par exemple, des outils pour la catégorisation des contrats, la détection des clauses contractuelles divergentes ou incompatibles, la preuve électronique, l'aide à la rédaction, l'extraction des dispositions juridiques et l'examen de conformité assisté. En outre, certaines applications peuvent offrir des fonctions basiques de résolution des problèmes fondées sur des questions et des situations types (par exemple des chatbots juridiques).

Bien que l'IA ait, dans de tels cas, un impact sur la pratique et les connaissances juridiques qui soulève plusieurs questions éthiques,[128] les éventuelles conséquences négatives sur les droits de l'homme, la démocratie et l'État de droit sont limitées. Elles sont principalement liées à des défauts et des lacunes de ces systèmes.

[125] Voir Conseil des barreaux européens, 2020, 29.

[126] Voir la sous-partie suivante sur la distinction entre justice codifiée et justice équitable.

[127] Voir Commission européenne pour l'efficacité de la justice (CEPEJ). 2018. Charte éthique européenne sur l'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, Annexe II.

[128] Voir également Nunez, 2017.

Dans le cas de la gestion de contenus et de connaissances, y compris les recherches et l'analyse de documents, ces défauts peuvent engendrer des représentations incomplètes ou inexactes des faits ou des situations. Néanmoins, cela concerne les méta-produits, les résultats d'un outil de recherche qui doivent être interprétés et dûment justifiés en cas d'utilisation au sein d'un tribunal. Les règles de la responsabilité, dans le cas de la responsabilité du fait des produits par exemple, peuvent résoudre ces problèmes.

Par ailleurs, il est possible de neutraliser la partialité (mauvaise sélection des affaires, mauvaise classification, etc.) des outils de recherche assistée par ordinateur fondés sur des textes standards utilisés pour l'analyse de la législation, de la jurisprudence et des documents,[129] grâce à l'**éducation et** à **la formation** adéquates des professionnels du droit. La **transparence** des systèmes d'IA (à savoir la description de leur logique, de leurs biais potentiels et de leurs limites) peut également réduire les conséquences négatives.

L'utilisation de l'IA par les tribunaux, pour les recherches juridiques et l'analyse de documents, devrait également être définie par la transparence. Les juges doivent être transparents en ce qui concerne les décisions qui dépendent de l'IA et la façon dont les résultats fournis par l'IA sont utilisés afin de contribuer aux motifs, conformément aux **principes de procès équitable et d'égalité des armes**.[130]

Enfin, la transparence peut jouer un rôle important eu égard aux chatbots juridiques fondés sur l'IA, en faisant connaître aux utilisateurs leur logique et les ressources utilisées (la liste des affaires analysées, par exemple). Une transparence totale devrait également inclure les sources utilisées pour entraîner ces algorithmes et accéder à la base de données utilisée pour fournir les réponses. Lorsque ces bases de données sont privées, des **audits** de tierces parties devraient être disponibles afin d'évaluer la qualité des ensembles de données et la façon dont les biais potentiels ont été traités, notamment le risque de sous- ou de sur- représentation de certaines catégories (**non-discrimination**).

D'autres problèmes majeurs touchent les applications d'IA conçues pour automatiser les modes alternatifs de règlement des litiges ou pour soutenir les décisions de justice. Ici, la distinction entre justice codifiée et justice équitable[131] suggère que l'IA devrait être limitée, à des fins de prise de décisions, aux affaires marquées par des évaluations factuelles régulières. D'où l'importance de mener des recherches supplémentaires sur la classification des différents types de processus décisionnels afin d'identifier les applications de raisonnement juridique systématiques qui peuvent être demandées à

[129] Voir la notion de justice électronique dans la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie), Annexe, paragraphe 38.

[130] Voir également Commission européenne pour l'efficacité de la justice (CEPEJ). 2018. Charte éthique européenne sur l'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, Annexe II.

[131] Voir Re et Solow-Niederman, 2019, 252-254 (« La justice équitable englobe à la fois la réflexion sur les valeurs mises en place par le système juridique et l'application raisonnée de ces valeurs, dans le contexte […] La justice codifiée désigne l'application routinière de procédures normalisées à un ensemble de faits […] En résumé, la justice codifiée voit les vices de la discrétion, tandis que la justice équitable voit ses vertus »).

l'IA, en gardant dans tous les cas la surveillance humaine qui garantit également la créativité juridique des décideurs.[132]

En ce qui concerne la justice équitable, comme l'indiquent les ouvrages,[133] sa logique est plus compliquée que la simple conclusion d'affaires individuelles. Les valeurs et les considérations exprimées et non exprimées, à la fois juridiques et non juridiques, caractérisent le raisonnement des tribunaux et ne peuvent être reproduites par la logique de l'IA. Les systèmes d'apprentissage automatique ne sont pas capables de produire un raisonnement juridique. Ils tirent des conclusions en identifiant des modèles dans les ensembles de données juridiques, ce qui n'est pas la même chose que d'élaborer un raisonnement juridique.

Étant donné le vaste contexte du rôle social des tribunaux, la jurisprudence est un système évolutif, ouvert aux nouveaux enjeux sociétaux et politiques. Les outils d'IA dépendants du parcours suivi pourraient donc faire obstacle à ce processus évolutif : la nature déductive et dépendante du parcours suivi de certaines solutions d'apprentissage automatique fondées sur l'IA peut remettre en question le rôle important des décideurs humains dans l'évolution du droit dans la pratique et le raisonnement juridique.

De plus, au niveau individuel, la dépendance au parcours suivi peut générer le risque d'« analyses déterministes »,[134] provoquant la résurgence de doctrines déterministes au détriment des doctrines favorisant l'individualisation de la sanction et portant atteinte au principe de réhabilitation et d'individualisation de la peine.

Par ailleurs, dans plusieurs cas, notamment les modes alternatifs de règlement des litiges, la médiation entre les exigences des parties mais aussi l'analyse de la composante psychologique des actions humaines (faute, intention) nécessitent une intelligence émotionnelle que les systèmes d'IA n'ont pas.

Le cadre juridique existant, fourni par les instruments juridiques internationaux, illustre ces préoccupations. La Déclaration universelle des droits de l'homme (articles 7 et 10), le PIDCP (article 14), la Convention pour la protection des droits de l'homme et des libertés fondamentales (article 6) et la Charte des droits fondamentaux de l'Union européenne (article 47) insistent sur les principales exigences liées à l'exercice du pouvoir judiciaire : le traitement équitable devant la loi, l'impartialité, l'indépendance et la compétence. Les outils d'IA ne possèdent pas ces qualités, ce qui limite leur contribution au processus décisionnel tel qu'il est suivi par les tribunaux.

Comme indiqué par la Commission européenne pour l'efficacité de la justice, « la neutralité des algorithmes est un mythe, leurs créateurs transférant, de manière consciente ou non, leurs propres systèmes de valeurs. » De nombreux cas de partialité relatifs à des applications d'IA confirment que ces systèmes donnent trop souvent - bien que cela ne soit pas conscient dans bien des cas - une représentation partielle de la société et des situations personnelles, ce qui n'est pas compatible avec les principes

[132] Voir également Clay, 2019. À cet égard, par exemple, un système juridique qui prévoit des indemnités en cas de blessures corporelles, sur la base des dommages patrimoniaux effectifs, pourrait être automatisé. Il ne sera, toutefois, pas capable de réévaluer le fondement du raisonnement juridique et d'étendre les indemnités aux dommages non personnels et existentiels.

[133] Voir Re et Solow-Niederman, 2019.

[134] Voir Commission européenne pour l'efficacité de la justice (CEPEJ). 2018. Charte éthique européenne sur l'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement.

du **traitement équitable devant la loi** et de la **non-discrimination**.[135] La **qualité des données** et d'autres formes d'**évaluation** de la qualité (évaluation de l'impact, audits, etc.) peuvent réduire ce risque.[136] Cependant, compte tenu de l'importance des intérêts potentiellement concernés en cas de décisions biaisées, les risques restent élevés dans le cas d'une justice équitable et semblent disproportionnés par rapport aux avantages, principalement en termes d'efficacité du système judiciaire.[137]

Les **principes du procès équitable et de l'égalité des armes**[138] suscitent d'autres préoccupations lorsque les décisions de justice reposent sur les résultats d'algorithmes propriétaires dont les données et la structure de formation ne sont pas accessibles au public.[139] Une notion de **transparence** plus large pourrait y remédier en ce qui concerne l'utilisation de l'IA dans les décisions judiciaires, mais la transparence de l'IA - un objectif ambitieux en soi - ne peut répondre aux autres objections structurelles et fonctionnelles précitées.

Par ailleurs, les spécialistes des données peuvent façonner les outils d'IA de différentes façons au cours des phases de conception et de formation, de sorte que si les outils d'IA devaient nécessairement faire partie du processus décisionnel, les gouvernements qui sélectionnent les outils à utiliser par les tribunaux risqueraient d'interférer indirectement avec l'**indépendance** des juges.[140]

Ce risque subsiste même si le juge reste libre d'ignorer les décisions d'IA en invoquant un motif spécifique. Bien que le **contrôle humain** constitue un élément important,[141] son impact effectif peut être diminué par la propension psychologique ou utilitaire (économique) des décideurs humains à tirer partie de la solution proposée par l'IA.[142]

II.5.2 Prévention de la criminalité

La complexité de la détection et de la prévention de la criminalité a stimulé la recherche sur les applications d'IA afin de faciliter les activités humaines. Ces dernières années, plusieurs solutions[143] et un nombre croissant d'ouvrages ont été élaborés dans le domaine de la police prédictive, une approche proactive de la prévention de la

[135] Voir également CEPEJ, 2018.

[136] Voir également CEPEJ, 2018.

[137] Voir également la Recommandation CM/Rec(2020)1, Annexe, paragraphe 11.

[138] Voir également CEPEJ, 2018, Annexe I, paragraphe 138.

[139] Voir également CEPEJ, Annexe I, paragraphe 131 (« l'opacité des processus de fonctionnement des algorithmes par les entreprises privées (qui revendiquent leur propriété intellectuelle) a été une autre source d'inquiétude. Si l'on tient compte du fait qu'elles tiennent leurs données sources des autorités étatiques elles-mêmes, leur absence d'esprit de responsabilité vis-à-vis des citoyens pose un problème démocratique majeur [...] par exemple, lorsque ProPublica a révélé les failles de l'algorithme COMPAS après le refus de la société propriétaire de le partager »).

[140] Voir également CEPEJ, 2018.

[141] Voir également CEPEJ, 2018.

[142] Voir également Mantelero, 2019 (« En effet, la nature soi-disant fiable des solutions d'IA fondées sur les mathématiques peut pousser les personnes qui s'appuient sur des algorithmes pour prendre leurs décisions à faire confiance au tableau des individus et de la société suggéré par les procédés analytiques »).

[143] Voir Završnik, 2019 ; Agence des droits fondamentaux de l'Union européenne, 2018, 98-100; Osoba et Welser, 2017.

criminalité fondée sur les données. Les solutions disponibles poursuivent essentiellement deux objectifs distincts : prédire où et quand des crimes pourraient être commis ou prédire qui pourrait commettre un crime.[144]

Ces deux objectifs n'ont pas le même impact potentiel sur les droits de l'homme et la liberté : cet impact est plus marqué lorsque l'IA est utilisée pour prédire les auteurs de crimes. Cependant, dans les deux cas, nous pouvons reprendre ici les considérations concernant les enjeux généraux de l'IA (opacité, droits de propriété intellectuelle, collecte de données à grande échelle[145], etc.) examinés dans les sous-parties précédentes et partiellement traités par la transparence, la **qualité des données**, la **protection des données**, les **audits** et les autres mesures. Il convient de noter que le rôle de la **transparence**[146] dans le contexte judiciaire pourrait être limité afin de ne pas entraver l'effet dissuasif de ces outils. La transparence totale pourrait donc être remplacée par des audits et un contrôle assurés par des organismes indépendants.

Abstraction faite des aspects organisationnels relatifs aux limites de l'autodétermination des policiers dans le cadre de leurs fonctions, les principales questions concernant l'utilisation de l'IA pour prévoir la criminalité sur une base géographique et temporelle concernent l'impact de ces outils sur le **droit à la non-discrimination**.[147] Les préjugés qui se confirment d'eux-mêmes, les préjugés communautaires[148] et les préjugés historiques[149] peuvent créer des formes de stigmatisation à l'égard de certains groupes et des lieux où ils résident habituellement.

Lorsque l'analyse des données est utilisée pour classer les crimes et déduire des preuves sur des réseaux criminels, les solutions propriétaires posent problème en termes de respect des **principes de procès équitable et d'égalité des armes** dans le cadre de la collecte et de l'utilisation des preuves. De plus, si les opérations quotidiennes des forces de police sont guidées par des logiciels prédictifs, cela pose le problème de la **responsabilité** des stratégies adoptées, étant donné qu'elles sont en partie déterminées par les logiciels, et donc par les sociétés qui les développent, et non par la police.

Les outils de police prédictive, qui utilisent le profilage pour soutenir les prévisions concernant les auteurs de crimes, sont encore plus en contradiction avec les droits de l'homme. Indépendamment de la question du traitement des données et du profilage,[150]

---

[144] Pour une taxonomie des méthodes prédictives, voir Perry et al., 2013, qui recense les quatre catégories suivantes : les méthodes de prévision des crimes (axées sur les lieux et les horaires des crimes), la méthode de prévision des auteurs de crimes (axée sur les personnes), les méthodes de prévision des identités des auteurs de crimes (axées sur les personnes), et les méthodes de prévision des victimes de crimes (axées sur les groupes et, dans certains cas, les personnes).

[145] Voir également la Recommandation Rec(2001)1 sur le Code européen d'éthique de la police, Annexe, paragraphe 42.

[146] Voir également Barrett, 2017, 361-62.

[147] Voir Agence des droits fondamentaux de l'Union européenne, 2018, 10.

[148] Voir également Barrett, 2017, 358-59 (« Pour certains, l'objectif de la sécurité collective mérite un sacrifice unilatéral d'un certain degré des droits individuels dans ce contexte particulier. Cependant, ce calcul peut changer si le sacrifice n'est pas collectif mais limité à des groupes minoritaires, ou devient fondamentalement arbitraire en vertu d'un degré d'erreur inacceptable. »)

[149] Voir Bennett Moses et Chan, 2018.

[150] Voir ci-dessus sous-partie II.2.

ces solutions peuvent également nuire au principe de la **présomption d'innocence**,[151] de l'**équité** procédurale et au droit à la **non-discrimination**.[152]

Bien que les questions de non-discrimination puissent être partiellement traitées, il semble plus difficile de résoudre les autres conflits. Du point de vue des droits de l'homme et en termes de proportionnalité (notamment le droit au respect de la vie privée et familiale[153]), le risque de porter atteinte à ces principes semble élevé et n'est pas suffisamment compensé par la preuve que ces solutions apporteront des bénéfices en matière de droits et de libertés individuels et collectifs.[154] À la lumière de la future réglementation de l'IA, cela devrait inciter à examiner attentivement ces questions, en établissant une distinction entre les possibilités techniques des solutions d'IA et leurs avantages concrets pour la protection et la consolidation des droits de l'homme et des libertés.

Enfin, dans une perspective plus large et plus globale des droits de l'homme, le fait que les outils d'IA basés sur les données mettent l'accent sur la criminalité conduit à une approche factuelle à court terme. Celle-ci sous-estime les problèmes sociaux, qui sont souvent liés à la criminalité et nécessitent des stratégies sociales à long terme impliquant le renforcement effectif des droits et des libertés individuels et sociaux.[155]

II.6 Harmonisation des principes identifiés

Les sous-parties précédentes ont identifié **plusieurs principes directeurs de la future réglementation de l'IA**. Ces principes ont été **contextualisés eu égard aux enjeux associés à l'IA** dans les différents domaines examinés, mais il est intéressant d'étudier le degré d'harmonisation qui existe entre ces principes.

Les conclusions de la présente étude indiquent que dans un certain nombre de cas **il existe des principes communs** (la primauté de l'être humain, l'autodétermination, la non-discrimination, le contrôle humain). Cela s'explique par plusieurs facteurs.

Premièrement, **certains principes sont sectoriels**. C'est le cas, par exemple, de l'indépendance des juges ou des principes du procès équitable et de l'égalité des armes, qui concernent uniquement la justice.[156]

Deuxièmement, certains principes directeurs sont les mêmes dans différents domaines, mais avec **différentes nuances** selon le contexte. C'est le cas de la

[151] Voir également la Recommandation Rec(2001)1 sur le Code européen d'éthique de la police, Annexe, paragraphe 47.

[152] Voir également la Recommandation Rec(2001)1 sur le Code européen d'éthique de la police, Annexe, paragraphe 49.

[153] Voir van Brakel et De Hert, 2011, 183. Voir également *Szabó et Vissy c Hongrie* [2016] Cour européenne des droits de l'homme quatrième section. Requête n° 37138/14.

[154] Voir Meijer et Wessels, 2019.

[155] Voir également Rosenbaum, 2006, 245–266.

[156] Voir également les principes de l'accès équitable et de la bienfaisance dans le secteur de la santé, ou les principes de la non-ingérence par les pouvoirs publics dans les activités médiatiques visant à influencer des élections et de l'obligation d'offrir des conditions de publicité équitables à tous les partis et candidats politiques.

transparence, souvent considérée comme cruciale dans la réglementation de l'IA, mais qui revêt diverses significations en fonction des cadres réglementaires.

Dans les domaines de la santé et des données à caractère personnel, la transparence a trait aux informations communiquées aux personnes sur le traitement de ces données ; elle met particulièrement l'accent sur les processus et les risques associés et revêt une forte connotation d'autodétermination. Cependant, la transparence est également pertinente en matière de protection des données afin de contrôler l'exercice du pouvoir sur les données aux mains des entités publiques et privées. Cette autre facette de la transparence est examinée eu égard aux applications d'IA pour la participation démocratique et la bonne gouvernance. Encore une fois, dans le contexte de la justice, la transparence a une importance plus complexe. En effet, elle est essentielle pour protéger les droits et les libertés fondamentaux (par exemple lorsque l'IA est utilisée dans les tribunaux), mais elle exige aussi des limites afin d'éviter de nuire à des intérêts concurrents (par exemple, dans le cas de la détection et de la prévention de la criminalité en matière de police prédictive).

Nous pouvons donc conclure que la transparence est une valeur fondamentale. Néanmoins, nous devons aller au-delà d'une simple exigence de transparence en tant que principe clé de la réglementation de l'IA. Comme pour les autres principes essentiels (la participation, l'inclusion, le contrôle démocratique et l'ouverture), il convient d'adopter des dispositions qui tiennent compte des différents contextes dans lesquels elles sont appliquées et donc de procéder à une contextualisation en bonne et due forme.

Troisièmement, certains principes sont différents mais appartiennent au **même domaine conceptuel** ; ils prennent différentes nuances selon le contexte. C'est le cas de l'obligation de rendre compte et des principes directeurs en matière de gestion des risques en général. Ici, le niveau de détail et les exigences correspondantes peuvent être plus ou moins élaborés. Par exemple, dans le domaine de la protection des données, plusieurs dispositions mettent en œuvre ces principes avec un degré de détail important, tandis que dans le cas de la démocratie et de la justice, ces principes sont moins développés eu égard aux applications à usage intensif de données telles que l'IA.

Enfin, la stratégie de réglementation de l'IA comprend certains éléments qui ne sont pas des principes mais des **approches et des solutions opérationnelles ;** celles-ci sont communes aux différents domaines mais nécessitent un développement fondé sur le contexte. C'est le cas du rôle important joué par l'éducation et la formation, l'interopérabilité et les comités d'experts.

Ces considérations suggèrent que seule une harmonisation partielle est possible. L'approche réglementaire de l'IA devrait donc reposer sur un **instrument juridiquement contraignant qui prévoit à la fois des dispositions générales** - axées sur des principes communs et des solutions opérationnelles - et **des dispositions sectorielles plus spécifiques**, couvrant les principes uniquement pertinents dans un domaine donné ou les cas où le même principe est contextualisé différemment selon les domaines.

II.7 Conclusions

Cette analyse a confirmé la validité de l'approche méthodologique adoptée, axée sur la **contextualisation** des principes directeurs tirés des instruments internationaux juridiquement contraignants ou non. Dans le même temps, elle a également mis en lumière la difficulté à classer les dispositions d'un vaste ensemble d'instruments, qui

diffèrent non seulement du point de vue de leur caractère contraignant mais également en termes d'objectifs, d'approche et de structure.

Les résultats ont également confirmé que le cadre existant fondé sur les droits de l'homme, la démocratie et l'État de droit pouvait fournir un contexte commun et approprié pour l'élaboration d'un **instrument contraignant plus spécifique visant à réglementer l'IA conformément aux principes et aux valeurs ancrés dans les instruments juridiques internationaux, capable de répondre de manière plus efficace aux questions soulevées par l'IA**.

Ce cadre international nous amène nécessairement à réaffirmer le rôle central de la dignité humaine dans le contexte de l'IA, où les solutions automatisées ne sauraient être autorisées à déshumaniser les personnes. En outre, il pourrait être prévu de limiter spécifiquement les technologies IA lorsqu'elles sont conçues ou utilisées d'une manière qui ne respecterait pas la dignité humaine, [157] les droits de l'homme, la démocratie et l'État de droit.

En vue d'une future réglementation de l'IA, ce résultat méthodologique et concret positif n'exclut pas l'existence de certaines lacunes. Elles concernent principalement de vastes domaines, tels que la démocratie et la justice, où il existe différentes possibilités et interprétations selon la vision politique et sociétale de la future relation entre les humains et les machines.

Des recherches supplémentaires dans le domaine des droits de l'homme et de l'IA, ainsi que le débat continu à l'échelle internationale et régionale, contribueront à combler ces lacunes. Toutefois, étant donné la nature évolutive de l'IA, il est souhaitable d'adopter une **approche de corégulation**.

Un instrument contraignant établissant le cadre juridique de l'IA, notamment les principes communs généraux mais aussi les dispositions précises répondant à des questions spécifiques, pourrait donc compléter les règles détaillées énoncées dans des **instruments sectoriels non contraignants supplémentaires**. Ce modèle offrirait à la fois un cadre réglementaire clair et la souplesse nécessaire pour prendre en compte le développement technologique.

[157] Voir également UNESCO. 1997. Déclaration sur le génome humain et les droits de l'homme, article 11.

Bibliographie

Andorno, R. 2005. The Oviedo Convention: A European Legal Framework at the Intersection of Human Rights and Health Law. Journal of International Biotechnology Law, January 2005. https://doi.org/10.1515/jibl.2005.2.4.133, consulté le 20.02.2020.

Azencott, C.-A. 2018. Machine Learning and Genomics: Precision Medicine versus Patient Privacy. Phil. Trans. R. Soc. A 376, n° 2128 (13 septembre 2018) : 20170350, https://doi.org/10.1098/rsta.2017.0350 consulté le 14.01.2020

Barrett, L. 2017. *Reasoably Suspicious Algorithms: Predictive Policing at the United States Border*. 41 N.Y.U. Rev. Law & Social Change 327.

Bennett Moses, L., Chan, J. 2018. *Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability*. 28 Policing and Society 806.

Bychawska-Siniarska, D. 2017. *Protecting the Right to Freedom of Expression under the European Convention on Human Rights*. Conseil de l'Europe.

Cabitza, F., Rasoini, R. et Gensini, G.F. 2017. Unintended Consequences of Machine Learning in Medicine. *JAMA* 318, no. 6 (8 août 2017): 517 https://doi.org/10.1001/jama.2017.7797. consulté le 18.12.2019.

Calo, R., Citron, D.K. 2020. *The Automated Administrative State: A Crisis of Legitimacy*. Emory Law Journal, à venir. https://ssrn.com/abstract=3553590, consulté le 20.04.2020.

Caruana, R. et autres 2015. Intelligible models for healthcare : predicting pneumonia risk and hospital 30-day readmission, dans : Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Cham, Switzerland : Springer International Publishing AG) 721-1730.

Clay, T. (ed). 2019. L'arbitrage en ligne. Rapport du Club des Juristes. Paris. 58 https://www.leclubdesjuristes.com/les-commissions/larbitrage-en-ligne/, consulté le 30.05.2020.

Comité des Ministres. 1999. Recommandation n° R (99) 5 sur la protection de la vie privée sur Internet. Adoptée par le Comité des Ministres le 23 février 1999 lors de la 660e réunion des Délégués des Ministres.

Comité des Ministres. 2001. Recommandation Rec(2001)10 du Comité des Ministres aux États membres sur le Code européen d'éthique de la police. Adoptée par le Comité des Ministres le 19 septembre 2001 lors de la 765e réunion des Délégués des Ministres.

Comité des Ministres. 2003. Recommandation CM/Rec(2003)4 sur les règles communes contre la corruption dans le financement des partis politiques et des campagnes électorales. Adoptée par le Comité des Ministres le 8 avril 2003 lors de la 835e réunion des Délégués des Ministres.

Comité des Ministres. 2004. Recommandation CM/Rec(2004)15 sur la gouvernance électronique (« e-gouvernance »). Adoptée par le Comité des Ministres le 15 décembre 2004, lors de la 909e réunion des Délégués des Ministres.

Comité des Ministres. 2007. Recommandation CM/Rec(2007)15 du Comité des Ministres aux États membres relative à des mesures concernant la couverture des campagnes électorales par les médias. Adoptée par le Comité des Ministres le 7 novembre 2010 lors de la 1010e réunion des Délégués des Ministres.

Comité des Ministres. 2009. Recommandation CM/Rec(2009)1 du Comité des Ministres aux États membres sur la démocratie électronique (e-démocratie). Adoptée par le Comité des Ministres le 18 février 2009, lors de la 1049e réunion des Délégués des Ministres.

Comité des Ministres. 2009. Recommandation CM/Rec(2009)2 du Comité des Ministres aux États membres sur l'évaluation, l'audit et le suivi de la participation et des politiques de la participation aux niveaux local et régional. Adoptée par le Comité des Ministres le 11 mars 2009 lors de la 1050e réunion des Délégués des Ministres.

Comité des Ministres. 2010. Recommandation CM/Rec(2010)13 sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage. Adoptée par le Comité des Ministres le 23 novembre 2010 lors de la 1099e réunion des Délégués des Ministres.

Comité des Ministres. 2011. Recommandation CM/Rec(2011)7 du Comité des Ministres aux Etats membres sur une nouvelle conception des médias. Adoptée par le Comité des Ministres le 21 septembre 2011, lors de la 1121e réunion des Délégués des Ministres.

Comité des Ministres. 2012. Recommandation CM/Rec(2012)3 sur la protection des droits de l'homme dans le contexte des moteurs de recherche. Adoptée par le Comité des Ministres le 4 avril 2012 lors de la 1139e réunion des Délégués des Ministres.

Comité des Ministres. 2012. Recommandation CM/Rec(2012)4 sur la protection des droits de l'homme dans le cadre des services de réseaux sociaux. Adoptée par le Comité des Ministres le 4 avril 2012 lors de la 1139e réunion des Délégués des Ministres.

Comité des Ministres. 2014. Recommandation CM/Rec(2014)7 sur la protection des lanceurs d'alerte. Adoptée par le Comité des Ministres le 30 avril 2014 lors de la 1198e réunion des Délégués des Ministres.

Comité des Ministres. 2016. Recommandation CM/Rec (2016)1 sur la protection et la promotion du droit à la liberté d'expression et du droit à la vie privée en lien avec la neutralité du réseau. Adoptée par le Comité des Ministres le 13 janvier 2020, lors de la 1244e réunion des Délégués des Ministres.

Comité des Ministres. 2016. Recommandation CM/Rec(2016)4 du Comité des Ministres aux États membres sur la protection du journalisme et la sécurité des journalistes et autres acteurs des médias. Adoptée par le Comité des Ministres le 13 avril 2016 lors de la 1253e réunion des Délégués des Ministres.

Comité des Ministres. 2016. Recommandation CM/Rec(2016)5 du Comité des Ministres aux États membres sur la liberté d'internet. Adoptée par le Comité des Ministres le 13 avril 2016 lors de la 1253e réunion des Délégués des Ministres.

Comité des Ministres. 2017. Recommandation CM/Rec(2017)5 du Comité des Ministres aux États membres sur les normes relatives au vote électronique. Adoptée par le Comité des Ministres le 14 juin 2017, lors de la 1289e réunion des Délégués des Ministres.

Comité des Ministres. 2018. Recommandation CM/Rec(2018)1 du Comité des Ministres aux États membres sur le pluralisme des médias et la transparence de leur propriété. Adoptée par le Comité des Ministres le 7 mars 2008, lors de la 1309e réunion des Délégués des Ministres.

Comité des Ministres. 2018. Recommandation CM/Rec(2018)2 du Comité des Ministres aux États membres sur les rôles et les responsabilités des intermédiaires d'internet. Adoptée par le Comité des Ministres le 7 mars 2008, lors de la 1309e réunion des Délégués des Ministres.

Comité des Ministres. 2018. Recommandation CM/Rec(2018)4 du Comité des Ministres aux États membres sur la participation des citoyens à la vie publique au niveau local. Adoptée par le Comité des Ministres le 21 mars 2008, lors de la 1311e réunion des Délégués des Ministres.

Comité des Ministres. 2019. Recommandation CM/Rec(2019)2 du Comité des Ministres aux États membres en matière de protection des données relatives à la santé. Adoptée par le Comité des Ministres le 27 mars 2019, lors de la 1342e réunion des Délégués des Ministres.

Comité des Ministres. 2019. Recommandation CM/Rec(2019)3 du Comité des Ministres aux États membres sur le contrôle des actes des collectivités locales. Adoptée par le Comité des Ministres le 4 avril 2019 lors de la 1343e réunion des Délégués des Ministres.

Comité des Ministres. 2020. Recommandation CM/Rec(2020)1 du Comité des Ministres aux Etats membres sur les impacts des systèmes algorithmiques sur les droits de l'homme. Adoptée par le Comité des Ministres le 8 avril 2020 lors de la 1373e réunion des Délégués des Ministres.

Conseil des barreaux européens. 2020. Considérations du CCBE sur les aspects juridiques de l'intelligence artificielle. Bruxelles.

Conseil de l'Europe - Commission de Venise, OSCE/BIDDH. 2011. Lignes directrices sur la réglementation des partis politiques https://www.osce.org/odihr/77812, consulté le 20.12.2019.

Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel du Conseil de l'Europe. 2019. Lignes directrices sur l'intelligence artificielle et la protection des données. T-PD(2019)01. https://rm.coe.int/2018-lignes-directrices-sur-l-intelligence-artificielle-et-la-protecti/168098e1b7, consulté le 16.11.2019.

Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel du Conseil de l'Europe. 2017. Lignes directrices sur la protection des personnes à l'égard du traitement des données à caractère personnel à l'ère des mégadonnées T-PD(2017)1. https://rm.coe.int/lignes-directrices-sur-la-protection-des-personnes-a-l-egard-du-traite/16806f06d1, consulté le 16.11.2019

Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel du Conseil de l'Europe. 2019. Profilage et la Convention 108+ : Pistes pour une actualisation. T-PD(2019)07BISrev.

Conseil de l'Europe, Direction générale de la démocratie - Comité européen sur la démocratie et la gouvernance. 2016. Recueil des textes les plus pertinents du Conseil de l'Europe dans le domaine de la démocratie https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806b6140, consulté le 18.11.2019.

Conseil de l'Europe, Direction générale de la démocratie et des affaires politiques - Direction des institutions démocratiques. 2011. Lignes directrices relatives à la transparence des élections par voie électronique.

Conseil de l'Europe, Direction générale de la démocratie et des affaires politiques - Direction des institutions démocratiques. 2009. Projet « Bonne gouvernance dans la société de l'information », CM(2009)9.

Conseil de l'Europe, Assemblée parlementaire. 2019. Résolution 2254 (2019) La liberté des médias en tant que condition pour des élections démocratiques.

Conseil de l'Europe. 2008. Les 12 Principes de bonne gouvernance https://www.coe.int/fr/web/good-governance/12-principles, consulté le 16.03.2020.

Conseil de l'Europe. 2009. Protocole additionnel à la Charte européenne de l'autonomie locale sur le droit de participer aux affaires des collectivités locales, Utrecht, 16.XI.2009.

Conseil de l'Europe. 2017. Lignes directrices relatives à la participation civile aux décisions politiques (CM(2017)83-final). Adoptées par le Comité des Ministres le 27 septembre 2017, lors de la 1295e réunion des Délégués des Ministres.

Conseil de l'Europe. 2019. Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques. Adoptée par le Comité des Ministres le 13 février 2019, lors de la 1337e réunion des Délégués des Ministres.

Conseil de l'Europe. 2019. Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques (Adoptée par le Comité des Ministres le 13 février 2019 lors de la 1337e réunion des Délégués des Ministres).

Conseil de l'Europe. Lignes directrices relatives à la participation civile aux décisions politiques CM(2017)83-final. https://www.coe.int/fr/web/civil-society/guidelines, consulté le 15.03.2020.

Comité d'experts du Conseil de l'Europe sur les intermédiaires d'internet (MSI-NET). 2018. Algorithmes et droits humains. Étude sur les dimensions des droits humains dans les techniques de traitement automatisé des données (en particulier les algorithmes) et éventuelles implications réglementaires. https://rm.coe.int/etude-sur-les-algorithmes-version-finale/1680770cc2 , consulté le 28.11.2019.

Cummings, M. L., Roff H. M., Cukier K., Parakilas J. et Bryce H. 2018. *Chatham House Report. Artificial Intelligence and International Affairs Disruption Anticipated.* London: Chatham House. The Royal Institute of International Affairs. https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial-intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf, consulté le 21.03.2020.

Dupont, B. et al. 2018. *L'intelligence artificielle dans le contexte de la criminalité et de la justice pénale*. Institut coréen de criminologie 2018. https://www.cyberjustice.ca/publications/lintelligence-artificielle-dans-le-contexte-de-la-criminalite-et-de-la-justice-penale/, consulté le 30.05.2020.

Code européen de bonnes pratiques contre la désinformation en ligne, 2018 https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation, 23.03.2020.

Commission européenne pour la démocratie par le droit (Commission de Venise). 2002. Code de bonne conduite en matière électorale. Lignes directrices et rapport explicatif. Adoptés par la Commission lors de ses 51e et 52e sessions plénières (Venise, 5-6 juillet et 18-19 octobre 2002).

Commission européenne pour la démocratie par le droit (Commission de Venise). 2019. Rapport sur les technologies numériques et les élections, établi conjointement par la Commission de Venise et la Direction de la société de l'information et de la lutte contre la criminalité (Direction générale Droits de l'homme et État de droit, DGI).

Commission européenne pour l'efficacité de la justice (CEPEJ). 2018. Charte éthique européenne sur l'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, Annexe II. Adoptée par la CEPEJ lors de sa 31e réunion plénière (Strasbourg, 3-4 décembre 2018) https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b, consulté le 04.12.2018.

Commission européenne, Direction générale des réseaux de communications, du contenu et des technologies. 2018. *A Multi-Dimensional Approach to Disinformation Report of the Independent High Level Group on Fake News and Online Disinformation*.

https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation, consulté le 22.03.2018.

Commission européenne. 2014. Livre vert sur la santé mobile. https://eur-lex.europa.eu/resource.html?uri=cellar:0de99b25-c0af-11e3-86f9-01aa75ed71a1.0002.01/DOC_1&format=PDF, consulté le 12.01.2020.

Commission européenne. 2020. Une stratégie européenne pour les données, COM(2020) 66 final. https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_fr.pdf, consulté le 20.02.2020.

Commission européenne. 2020. Rapport sur les conséquences de l'intelligence artificielle, de l'internet des objets et de la robotique sur la sécurité et la responsabilité, COM(2020) 64 final. https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_fr.pdf, consulté le 20.02.2020.

Commission européenne. 2020. Livre blanc sur l'intelligence artificielle - une approche européenne axée sur l'excellence et la confiance, COM(2020) 65 final. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_fr.pdf, consulté le 20.02.2020.

Agence des droits fondamentaux de l'Union européenne. 2018. *#BigData: Discrimination in Data-Supported Decision Making*. https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making, consulté le 20.05.2020.

Agence des droits fondamentaux de l'Union européenne. 2018. *Guide pour la prévention du profilage illicite aujourd'hui et demain*. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-preventing-unlawful-profiling-guide_fr.pdf , consulté le 20.05.2020.

Faye Jacobsen, A. 2013. *The Right to Public Participation. A Human Rights Law Update*. Document de réflexion. Institut danois pour les droits de l'homme.

Ferryman, K. et Pitcan, M. 2018. Fairness in Precision Medicine. Data & Society, February, https://datasociety.net/wp-content/uploads/2018/02/Data.Society.Fairness.In_.Precision.Medicine.Feb2018.FINAL-2.26.18.pdf

González Fuster, G. 2014. *The Emergence of Personal Data Protection as a Fundamental Right of the EU*. Cham-New York: Springer International Publishing.

Maisley, N. 2017. *The International Right of Rights? Article 25(a) of the ICCPR as a Human Right to Take Part in International Law-Making*. 28 European Journal of International Law 89.

Manheim, K., Kaplan, L. 2019. *Artificial Intelligence: Risks to Privacy and Democracy*. 21 Yale J.L. & Tech. 106.

Mantelero, A. 2018. *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*. 34 Computer Law & Security Review 754.

Mantelero, A. 2019. Intelligence artificielle et protection des données : Enjeux et solutions possibles. Rapport sur l'intelligence artificielle (Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel : Strasbourg), T-PD(2018)09Rev, https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6. consulté le 20.02.2020

Mayer-Schönberger, V. 1997. *Generational development of data protection in Europe?* In Agre, P.E., Rotenberg, M. (eds). *Technology and privacy: The new landscape*. Cambridge, MA: MIT Press.

Mehr, H. 2017. *Artificial Intelligence for Citizen Services and Government*. Harvard Kennedy School. Ash Center for Democracy and Innovation.

Meijer, A., Wessels, M. 2019. *Predictive Policing: Review of Benefits and Drawbacks*. 42 International Journal of Public Administration 1031.

Nunez, C. 2017. *Artificial Intelligence and Legal Ethics: Whether AI Lawyers Can Make Ethical Decisions*. 20 Tul. & Tech. & Intell. Prop. 189-204.

OCDE. 2013. Recommandation du Conseil de l'OCDE concernant les Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel, C(80)58/FINAL, telle que révisée le 11 juillet 2013 par le document C(2013)79.

OSCE. 2020. *Non-Paper on the Impact of Artificial Intelligence on Freedom of Expression*. https://www.osce.org/representative-on-freedom-of-media/447829, consulté le 11.06.2020.

Osoba, O.A., Welser, W. 2017. *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. RAND Corporation. https://www.rand.org/pubs/research_reports/RR1744.html, consulté le 20.05.2020.

Assemblée parlementaire. 2019. Résolution 2254 (2019) La liberté des médias en tant que condition pour des élections démocratiques.

Assemblée parlementaire. 2019. Résolution 2300 (2019)1 Améliorer la protection des lanceurs d'alerte partout en Europe

Perry, W.L. et al. 2013. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Corporation 2013. https://www.rand.org/pubs/research_reports/RR233.html, consulté le 30.03.2020.

Privacy International. 2017. *Smart Cities: Utopian Vision, Dystopian Reality*. https://privacyinternational.org/report/638/smart-cities-utopian-vision-dystopian-reality, consulté le 21.03.2020.

Re, R.M., Solow-Niederman, A. 2019. *Developing Artificially Intelligent Justice*. 22 Stan. Tech. L. Rev. 242.

Rosenbaum, D., 2006. *The limits of hot spots policing*. In: D. Weisburd et A. Braga, eds. *Police innovation: contrasting perspectives*. New York, NY: Cambridge University Press, 245–266.

Rouvroy, A. 2016. « Des données et des hommes » – Droits et libertés fondamentaux dans un monde de données massives (Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel : Strasbourg), T-PD-BUR(2015)09Rev, http://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a6020 consulté le 04.11.2019.

Savaget, P., Chiarini, T., Evans, S. 2019. *Empowering Political Participation through Artificial Intelligence*. 46 Science and Public Policy 369.

Seatzu, F. 2015. The Experience of the European Court of Human Rights with the European Convention on Human Rights and Biomedicine. 31(81) Utrecht Journal of International and European Law 5, DOI: http://dx.doi.org/10.5334/ujiel.da.

Sunstein, C.R. 2015a. *The Ethics of Nudging*. Yale Journal on Regulation 32: 413

Sunstein, C.R. 2015b. *Why Nudge? The Politics of Libertarian Paternalism*. New Haven: Yale University Press.

Sunstein, C.R., Thaler, R. 2003. *Libertarian Paternalism in Not an Oxymoron*. University of Chicago Law Review 70 (4): 1159.

Thaler, R., Sunstein, C.R. 2008. *Nudge*. New Haven, CT: Yale University Press.

Le Rapporteur spécial des Nations Unies (ONU) sur la promotion et la protection du droit à la liberté d'opinion et d'expression, la Représentante de l'Organisation pour la sécurité et la coopération en Europe (OSCE) sur la liberté des médias, le Rapporteur spécial pour la liberté d'expression de l'Organisation des Etats Américains (OEA) et le Rapporteur spécial sur la liberté d'expression et l'accès à l'information de la Commission africaine des droits de l'homme et des peuples (CADHP). 2017. Déclaration conjointe sur la liberté d'expression et les fausses nouvelles (« fake news »), la désinformation et la propagande. https://www.law-democracy.org/live/wp-content/uploads/2018/11/mandates.decl .2017.French.pdf, consulté le 02.02.2020.

T-PD(2019)07BISrev. Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel. 2019. Profilage et la Convention 108+ : Pistes pour une actualisation

Bureau du Haut-Commissaire des Nations Unies aux droits de l'homme 1981. CESCR Observation générale n° 1 : Rapports des États parties. Adoptée lors de la treizième session du Comité des droits économiques, sociaux et culturels, le 27 juillet 1981 (Contenue dans le document E/1989/22).

Bureau du Haut-Commissaire des Nations Unies aux droits de l'homme 1996. Observation générale n° 25 : Le droit de participer aux affaires publiques, le droit de vote et le droit d'accéder, dans des conditions d'égalité, aux fonctions publiques (article 25). CCPR/C/21/Rev.1/Add.7.

UNESCO. 2019. Étude préliminaire concernant un éventuel instrument normatif sur l'éthique de l'intelligence artificielle. https://unesdoc.unesco.org/ark:/48223/pf0000369455, consulté le 21.11.2019.

UNESCO. Déclaration universelle sur le génome humain et les droits de l'homme (11 novembre 1997).

Convention des Nations Unies contre la corruption, 2003.

van Brakel, R., De Hert, P. 2011. *Policing, Surveillance and Law in a Pre-Crime Society: Understanding the Consequences of Technology Based Strategies*. 3 Cahiers Politiestudies, Jaargang 163.

Verbeek, P-P. 2011. *Understanding and Designing the Morality of Things*. Chicago-London: The University of Chicago Press.

Završnik, A. 2019. *Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings*. European Journal of Criminology, 1-20, https://doi.org/10.1177/1477370819876762, consulté le 20.02.2020.

Annexe 1 – Instruments juridiques

Instruments contraignants	Instruments non contraignants connexes
Biomédecine	
Convention sur les droits de l'homme et la biomédecine (Convention d'Oviedo) Protocole additionnel, relatif aux tests génétiques à des fins médicales Protocole additionnel, relatif à la recherche biomédicale	Rec(2016)8 sur le traitement des données à caractère personnel relatives à la santé à des fins d'assurance, y compris les données résultant de tests génétiques et son exposé des motifs Recommandation CM/Rec(2016)6 du Comité des Ministres aux États membres sur la recherche utilisant du matériel biologique d'origine humaine Plan d'action stratégique sur les droits de l'homme et les technologies en biomédecine 2020-2025
Anti-discrimination	
- Déclaration universelle des droits de l'homme - Pacte international relatif aux droits civils et politiques - Pacte international relatif aux droits économiques, sociaux et culturels - Convention internationale sur l'élimination de toutes les formes de discrimination raciale - Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes (CEDAW) - Convention des Nations Unies relative aux droits des personnes handicapées - Convention européenne des droits de l'homme (CEDH) et ses protocoles (n° 12 en particulier) - Charte sociale européenne - Convention sur la cybercriminalité et son Protocole additionnel relatif à l'incrimination d'actes de nature raciste et xénophobe commis par le biais de systèmes informatiques - Convention sur la prévention et la lutte contre la violence à l'égard des femmes et la violence domestique (Convention d'Istanbul) - Charte des droits fondamentaux de l'Union européenne	Recommandations de politique générale de l'ECRI, et en particulier les recommandations n°2 (sur les organismes de promotion de l'égalité), 11 (sur la lutte contre le racisme et la discrimination raciale dans les activités de la police) et 15 (sur la lutte contre le discours de haine). Recommandation 2098 (2017) de l'APCE Mettre fin à la cyberdiscrimination et aux propos haineux en ligne Recommandation (2019)1 du Comité des Ministres sur la prévention et la lutte contre le sexisme
Cybercriminalité et preuves électroniques	
Convention sur la cybercriminalité	Notes d'orientation du Comité de la Convention sur la cybercriminalité sur les attaques DDOS, les attaques visant les infrastructures d'information critiques, les logiciels malveillants, les spams, l'usurpation d'identité, etc.
Justice	
- Déclaration universelle des droits de l'homme - Pacte international relatif aux droits civils et politiques	CEPEJ 2019. Charte éthique européenne d'utilisation de l'intelligence artificielle (IA) dans les systèmes judiciaires et leur environnement

- Convention internationale sur l'élimination de toutes les formes de discrimination raciale - Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes - Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH) - Charte des droits fondamentaux de l'Union européenne	
Congrès des pouvoirs locaux et régionaux	
Charte européenne de l'autonomie locale	Résolution 435 (2018) et Recommandation 424 (2018) du Congrès : « Transparence et gouvernement ouvert » Résolution 417 (2017) et Recommandation 398 (2017) du Congrès : « Le libre accès aux données = amélioration des services publics » Résolution 394 (2015) du Congrès : « Médias électroniques : une nouvelle donne pour les responsables politiques locaux et régionaux » Résolution 290 (2009) du Congrès : « La démocratie électronique : perspectives et risques pour les collectivités locales ».
Démocratie et participation	
- Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH) - Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (STE n°108) de 1981 et Protocole de 2018 modernisant la Convention	- Recommandation Rec(2003)4 du Comité des Ministres sur les règles communes contre la corruption dans le financement des partis politiques et des campagnes électorales - Code de bonne conduite en matière électorale (Commission de Venise) - Lignes directrices conjointes sur la réglementation des partis politiques (Commission de Venise et BIDDH/OSCE) - Recommandation CM/Rec(2007)15 du Comité des Ministres aux États membres sur des mesures concernant la couverture des campagnes électorales par les médias - voir également la Recommandation CM/Rec(2018)1 sur le pluralisme des médias et la transparence de leur propriété, la Recommandation CM/Rec(2018)2 sur les rôles et les responsabilités des intermédiaires d'Internet, la Recommandation CM/Rec(2016)1 sur la protection et la promotion du droit à la liberté d'expression et du droit à la vie privée en lien avec la neutralité du réseau - Recommandation n° R (99) 5 du Comité des Ministres sur la protection de la vie privée sur Internet Recommandation CM/Rec(2010)13 sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage, Recommandation CM/Rec(2012)3 sur la protection des droits de l'homme dans le contexte des moteurs de recherche

	Recommandation CM/Rec(2012)4 sur la protection des droits de l'homme dans le cadre des services de réseaux sociaux
Liberté d'expression	
- Convention européenne des droits de l'homme - Pacte international relatif aux droits civils et politiques - Charte des droits fondamentaux de l'Union européenne	Déclaration universelle des droits de l'homme CM/Rec(2018)2 sur les rôles et les responsabilités des intermédiaires d'Internet CM/Rec(2020)x sur les impacts des systèmes algorithmiques sur les droits de l'homme Déclaration (13/02/2019) sur les capacités de manipulation des processus algorithmiques CM/Rec(2018)1 sur le pluralisme des médias et la transparence de leur propriété CM/Rec(2020)x sur la promotion d'un environnement favorable à un journalisme de qualité à l'ère numérique
Elections	
Déclaration universelle des droits de l'homme Pacte international relatif aux droits civils et politiques Convention des Nations Unies sur l'élimination de toutes les formes de discrimination raciale Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes Convention des Nations Unies relative aux droits des personnes handicapées Convention des Nations Unies contre la corruption Convention de sauvegarde des droits de l'homme et des libertés fondamentales (STE n° 5) Protocole additionnel à la Convention de sauvegarde des droits de l'homme et des libertés fondamentales (STE n° 9) Charte européenne de l'autonomie locale (STE n° 122) Charte européenne des langues régionales ou minoritaires (STE n° 148) Convention sur la cybercriminalité (STE n° 185) Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (STE n° 108)	Code de bonne conduite en matière électorale, adopté par le Conseil des élections démocratiques du Conseil de l'Europe et la Commission européenne pour la démocratie par le droit (Commission de Venise) Recommandation Rec(2003)3 du Comité des Ministres aux États membres sur la participation équilibrée des femmes et des hommes à la prise de décision politique et publique Convention sur les normes en matière d'élections démocratiques et les droits et libertés électoraux dans les États membres de la Communauté des États indépendants (CDL-EL(2006)031rev) Recommandation Rec(99)5 du Comité des Ministres sur la protection de la vie privée sur Internet Recommandation Rec(2004)15 du Comité des Ministres aux États membres sur la gouvernance électronique (e-gouvernance) Recommandation CM/Rec(2007)15 du Comité des Ministres aux États membres relative à des mesures concernant la couverture des campagnes électorales par les médias Recommandation CM/Rec(2009)1 du Comité des Ministres aux États membres sur la démocratie électronique Recommandation CM/Rec(2017)5 du Comité des Ministres aux États membres sur les normes relatives au vote électronique

Protocole additionnel à la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, concernant les autorités de contrôle et les flux transfrontières de données (STE n° 181) Charte des droits fondamentaux de l'Union européenne Convention-cadre pour la protection des minorités nationales et rapport explicatif	Document de la réunion de Copenhague de la conférence sur la dimension humaine de l'OSCE Rapport sur l'abus de ressources administratives pendant les processus électoraux, adopté par le Conseil des élections démocratiques et par la Commission de Venise (CDL-AD(2013)033) Rapport sur les règles électorales et les actions positives en faveur de la participation des minorités nationales aux processus de décision dans les pays européens, adopté par le Conseil des élections démocratiques et par la Commission de Venise (CDL-AD(2005)009) Code de bonne conduite en matière référendaire, adopté par le Conseil des élections démocratiques et par la Commission de Venise Stratégie du Conseil de l'Europe sur le handicap 2017-2023 Résolution 1897 (2012) de l'APCE, Garantir des élections plus démocratiques Code de bonne conduite en matière de partis politiques et rapport explicatif, adoptés par la Commission de Venise (CDL-AD(2009)021)
Démocratie (hors questions relatives aux élections et au cycle électoral)	
- Déclaration universelle des droits de l'homme - Pacte international relatif aux droits civils et politiques - Convention internationale sur l'élimination de toutes les formes de discrimination raciale - Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes - Charte des droits fondamentaux de l'Union européenne - Convention 108+ - Convention de sauvegarde des droits de l'homme et des libertés fondamentales et ses protocoles - Charte européenne de l'autonomie locale - Convention-cadre pour la protection des minorités nationales.	- Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques Voir aussi le Recueil Chapitre A (Séparation des pouvoirs / bonne gouvernance) Chapitre B (Diversité et pluralisme des médias ; protection de la liberté d'expression sur Internet) Chapitre C (participation de la société civile) Chapitre E (participation des citoyens)
Bonne gouvernance	
- Déclaration universelle des droits de l'homme - Pacte international relatif aux droits civils et politiques - Convention internationale sur l'élimination de toutes les formes de discrimination raciale - Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes	-12 principes de bonne gouvernance démocratique - Recommandation du Comité des Ministres aux États membres sur le contrôle des actes des collectivités locales CM/Rec(2019)3 Voir aussi le Recueil

- Convention de sauvegarde des droits de l'homme et des libertés fondamentales et ses protocoles - Charte des droits fondamentaux de l'Union européenne - Convention 108+ - Charte européenne de l'autonomie locale et ses protocoles - Convention du Conseil de l'Europe sur l'accès aux documents publics	Chapitre A (bonne gouvernance) Chapitre E (Politiques d'intégration - normes et mécanismes) Voir également https://www.coe.int/en/web/good-governance/conventions-recommendations
Égalité femmes-hommes, y compris les questions relatives à la violence à l'égard des femmes	
- Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH) - Convention du Conseil de l'Europe sur la prévention et la lutte contre la violence à l'égard des femmes et la violence domestique (article 17§1 sur la participation du secteur des technologies de l'information et de la communication à la prévention et à la lutte contre la violence à l'égard des femmes, article 34 sur le harcèlement en ligne) - Charte sociale européenne - Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes - Déclaration universelle des droits de l'homme - Pacte international relatif aux droits civils et politiques - Pacte international relatif aux droits économiques, sociaux et culturels - Convention internationale sur l'élimination de toutes les formes de discrimination raciale - Charte des droits fondamentaux de l'Union européenne	Recommandation (2019)1 du Comité des Ministres Prévention et lutte contre le sexisme Recommandation (2013)1 du Comité des Ministres sur l'égalité entre les femmes et les hommes et les médias Recommandation de politique générale n° 15 de l'ECRI sur le discours de haine
Culture, créativité et patrimoine	
Déclaration universelle des droits de l'homme Pacte international relatif aux droits économiques, sociaux et culturels Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH)	
Convention-cadre pour la protection des minorités nationales	Nombreuses recommandations et résolutions du Comité des Ministres, de l'Assemblée parlementaire et du Congrès des pouvoirs locaux et régionaux du Conseil de l'Europe sur les questions d'identité culturelle et de diversité
Charte européenne des langues régionales ou minoritaires[158]	Nombreuses recommandations et résolutions du Comité des Ministres, de l'Assemblée parlementaire et du Congrès des pouvoirs

[158] Étude à venir (printemps 2020) sur l'impact de l'IA sur les domaines couverts par la Charte européenne des langues régionales ou minoritaires

	locaux et régionaux du Conseil de l'Europe sur l'identité culturelle, la diversité culturelle et le dialogue culturel, ainsi que les questions relatives aux minorités
Conventions du CdE dans le domaine du patrimoine culturel (Convention de Nicosie, non entrée en vigueur, Convention de Faro, Convention de La Valette, Convention de Grenade)	Nombreuses recommandations du Comité des Ministres et de l'Assemblée parlementaire du Conseil de l'Europe sur le patrimoine culturel
Convention de l'Unesco sur la protection et la promotion de la diversité des expressions culturelles	Déclaration du CdE sur la diversité culturelle Recommandation du Comité des Ministres du CdE sur la Convention de l'Unesco
Convention du Conseil de l'Europe sur la coproduction cinématographique (révisée) Directive UE « Services de médias audiovisuels » / Directive (UE) 2018/1808 Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE	Résolution (88)15 modifiée instituant un Fonds européen de soutien à la coproduction et à la diffusion des œuvres de création cinématographiques et audiovisuelles Recommandation CM/Rec(2017)9 du Comité des Ministres aux États membres sur l'égalité entre les femmes et les hommes dans le secteur audiovisuel
Déclaration universelle des droits de l'homme Pacte international relatif aux droits économiques, sociaux et culturels Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH)	
GRECO	
Convention civile sur la corruption et Convention pénale sur la corruption ; évaluations du GRECO	Recommandations CM sur le modèle de code de conduite pour les agents publics, les activités de lobbying, la protection des lanceurs d'alerte, la transparence du financement des partis politiques, etc.
Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH)	Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH)
CSE	
Charte sociale européenne (Charte de 1961, protocole de 1988 et Charte révisée de 1996)	
Droits contenus dans la Charte sociale européenne	Quelques exemples : - Nouvelle stratégie et plan d'action du Conseil de l'Europe pour la cohésion sociale (approuvés par le CM le 7 juillet 2010) - CM Rec(2000)3 proposant un droit individuel, universel et justiciable à la satisfaction des besoins matériels élémentaires - etc.
Par ailleurs, de nombreux sujets relevant d'autres domaines de travail du CdE comportent des aspects qui ont trait aux droits sociaux (et droits contenus dans la Charte).	Quelques exemples : - CM Rec(93)1 relative à l'accès effectif au droit et à la justice des personnes en situation de grande pauvreté - volet « droits sociaux » des règles pénitentiaires (santé, conditions de vie, emploi, éducation, droits familiaux, etc.) - etc.

Annexe 2 – Domaines concernés

Domaines concernés (applications)
Biomédecine • Surveillance, prévention, diagnostic et interventions dans le domaine des soins de santé, reposant sur l'IA • Surveillance, diagnostic, suivi, financement de traitements (assurance) basés sur la prédiction (par exemple applications de reconnaissance faciale des utilisateurs et services en ligne en dehors du domaine des soins de santé)
Anti-discrimination • Prise de décisions automatisée dans différents domaines des secteurs public et privé (exemple : candidatures à des postes, prestations sociales, accès aux biens et services, tels que prêts bancaires, assurances) • Police prédictive (avec un risque élevé de profilage racial) • Justice prédictive • Reconnaissance faciale • Technologies de prédiction des comportements, par exemple reconnaissance émotionnelle et détection de mensonges par intelligence artificielle • Assistants personnels (par ex. Siri) • Modération de contenus • Protection des données
Cybercriminalité et preuves électroniques • Cybercriminalité et cyberattaques automatisées, par exemple : - Attaques par déni de service distribué (attaques DDOS) - Attaques contre les infrastructures d'information critiques - Attaques « man-in-the-middle » - Hameçonnage et techniques similaires d'ingénierie sociale - Repérage des vulnérabilités - etc. • Enquêtes sur la cybercriminalité et informatique judiciaire : - Collecte et analyse de preuves électroniques (pour toutes les infractions) - Attribution - Rétro-ingénierie • Cybersécurité et prévention de la cybercriminalité : - Détection de logiciels malveillants, d'intrusions, etc. - Correction automatisée des vulnérabilités
Secteur de la justice • Traitement des décisions et des données judiciaires : - pour étayer les décisions judiciaires ou la recherche judiciaire - règlement des litiges en ligne - conseils juridiques aux parties à un litige • Police prédictive
Congrès des pouvoirs locaux et régionaux • Prestation de services publics au niveau local • Instruments de promotion de la participation citoyenne • Grande diversité d'applications numériques et électroniques dans les villes et les collectivités locales. • Application des technologies de l'information et de la communication (TIC) pour améliorer la qualité de vie et les environnements de travail en ville. • Gouvernance intelligente des villes • Intégration des TIC dans les systèmes du secteur public

- Mise en œuvre locale de pratiques qui rapprochent la population des TIC pour favoriser l'innovation et améliorer les connaissances qu'elles offrent.

Liberté d'expression

- Communication individuelle (par modération et restriction automatisées des contenus ; systèmes de tri, classification, optimisation et recommandation par algorithmes)
- Production et diffusion de médias (robot-journalisme, journalisme de données, TALN, micro-ciblage de la base de lecteurs, fils d'actualité automatisés selon le profil du lecteur)
- Communication sociale et politique, fragmentation/polarisation du discours public, redlining politique (micro-ciblage de la base électorale, recours à des bots pour influencer l'opinion, prolifération de sites de médias locaux automatisés)

Élections

Période préélectorale :

- Planification du calendrier électoral
- Formation des acteurs électoraux
- Délimitation des circonscriptions électorales
- Inscription des électeurs et des candidats
- Accréditation des observateurs (internationaux et nationaux)
- Mise à jour de la liste d'électeurs
- Actualisation du cadre juridique
- Financement des partis politiques
- Propagande électorale par l'administration et par les partis politiques/candidats

Période électorale :

- Financement des campagnes électorales
- Accès aux médias
- Vote
- Décompte des bulletins
- Proclamation des résultats

Période post-électorale :

- Publication des résultats électoraux
- Règlement du contentieux électoral

Démocratie (hors questions relatives aux élections et au cycle électoral)

- Séparation des pouvoirs
- Participation de la société civile
- Participation des citoyens
- Vie privée
- Citoyenneté
- Protection des minorités
- Pluralisme et diversité
- Légitimité

Bonne gouvernance

- Collectivités locales
- Administration régionale
- Prestation de services
- Dotations budgétaires
- Sécurité sociale et mécanismes de prestations sociales
- Police et justice
- Villes intelligentes
- Appels d'offres publics et attribution des marchés
- Capacités institutionnelles

Égalité femmes-hommes, y compris les questions relatives à la violence à l'égard des femmes

Question générale des préjugés sexistes dont héritent les algorithmes des systèmes de données lors de la phase d'entraînement (valable pour de nombreux domaines), pouvant aggraver les inégalités entre femmes et hommes et les inégalités sociales.

Questions générales liées à l'IA en tant que secteur d'emploi

- Le manque de participation/la sous-représentation des femmes accentue les éventuels préjugés sexistes et exclut les femmes d'un secteur influent
- Exploitation des « travailleurs du clic » en Europe et dans le monde (bas salaires, absence de protection sociale, pas de droits des travailleurs, exposition durable à des contenus préjudiciables dans le cas des modérateurs de contenus, etc.)

Problèmes spécifiques

- Discrimination dans la sélection des candidatures
- Décisions automatisées pour des services publics et privés
- Reconnaissance faciale et vocale (fonctionnant moins bien pour les femmes, en particulier certains groupes)
- Surveillance / harcèlement facilités par les outils d'IA, par exemple dans le contexte de la violence domestique
- Les décisions automatisées aggravent encore le risque de discrimination multiple fondée sur le sexe, la race et l'origine sociale en combinant des données secondaires comme le niveau d'études, l'adresse et le niveau de revenu.
- Justice prédictive (par ex. violences faites aux femmes)
- Santé prédictive reposant sur des données empreintes de préjugés sexistes (ex. certaines maladies qualifiées de « féminines » ou « masculines »)
- Préjugés hérités dans la modération de contenus par les machines (tolérance élevée au sexisme, au discours de haine sexiste et aux violences contre les femmes)
- Assistants virtuels/robots genrés perpétuant les stéréotypes fondés sur le sexe
- Marketing genré perpétuant les stéréotypes fondés sur le sexe
- Application de prix différenciés selon le sexe

Impacts positifs

- Utilisation de dispositifs de géolocalisation pour assurer le respect des ordonnances de protection dans le contexte des violences faites aux femmes
- Recours à l'IA par la police et la justice pour évaluer les risques dans les affaires de violence domestique
- Recours à l'IA pour repérer et signaler les préjugés sexistes et bloquer ou supprimer la diffusion de propos haineux (sexistes) sur les plateformes
- Développement d'applications d'information et d'aide aux femmes victimes de violence
- Utilisation d'outils d'IA pour analyser les contenus et repérer les préjugés sexistes / analyser les représentations (par ex. dans les films ou autres médias)

Culture, créativité et patrimoine

- Accès et participation à la vie publique/culturelle ;
- Liberté d'expression (y compris liberté d'expression artistique)
- Accès à de l'information impartiale ?
- Décisions automatisées, ciblage, profilage
- Apprentissage de langues menacées pour les préserver/protéger
- Assistance automatisée dans l'administration, la santé, etc. pour les locuteurs de groupes/langues minoritaires
- Géolocalisation, police prédictive, analyse criminelle (destruction, pillage, trafic de biens culturels ; ciblage ; sensibilisation au patrimoine menacé pour contribuer à sa protection)
- Création automatisée de contenus, ciblage, profilage (création culturelle, échanges culturels, consommation culturelle)
- Conception et production de contenus audiovisuels :

- Analyse prédictive d'audience
- Analyse automatique de scénarii
- Écriture de scénarii automatique ou assistée
- Images de synthèse (effets spéciaux, animation...)
- Repérage, programmation et budgétisation automatisés (impact à évaluer)

- Diffusion de contenus
  - Algorithmes de recommandation
  - Publicité ciblée
  - Contrôle automatique des contenus (respect des réglementations) / censure (étude de référence « Entering the new paradigm of artificial intelligence and series » commandée par la DG2 et Eurimages)
- Accès et participation à la vie publique/culturelle
- Liberté d'expression (y compris liberté d'expression artistique)
- Accès à de l'information impartiale ?

GRECO

- Lutte contre la corruption
- Responsabilité pénale liée à l'utilisation de véhicules automatisés
- Article 8 : droit au respect de la vie privée et familiale

CSE

Tous les domaines des droits sociaux, sécurité sociale, cohésion sociale, etc. incluant, sans s'y limiter :

- de nombreux aspects de l'emploi (y compris, mais sans s'y restreindre, le contrôle et le suivi, la sélection des candidatures et le travail dans l'économie de plateformes, etc.)
- différents aspects de la santé (droit de jouir du meilleur état de santé possible)
- idem pour l'éducation
- la protection, l'intégration et la participation sociales
- la lutte contre la discrimination
- le logement et la protection contre l'exclusion sociale

par exemple :

- justice (à la fois administration de la justice, justice pénale et milieu carcéral)
- traite des êtres humains (travail forcé et exploitation, etc.)
- migrations et réfugiés
- égalité femmes-hommes et violence à l'égard des femmes
- enfants et jeunes, éducation
- bioéthique
- non-discrimination, Roms et Gens du voyage, orientation sexuelle et identité de genre
- politique en matière de drogues
- participation et culture
- sport

Annexe 3 – Principes

Principes directeurs et valeurs juridiques	Principes manquants
Biomédecine	
Primauté de l'être humain Respect de la vie privée et confidentialité Consentement éclairé Autonomie Non-discrimination Non-malfaisance / bienfaisance Obligation de rendre compte Transparence et accès équitable Débat public	Principe de précaution Contrôle/surveillance humains Explicabilité Responsabilité quant aux décisions reposant sur l'IA Égalité/équité hommes-femmes
Anti-discrimination	
Non-discrimination et égalité Diversité et inclusion Intersectionnalité Droit à un recours effectif Droit à un procès équitable Droit au respect de la vie privée Présomption d'innocence et charge de la preuve Transparence Impartialité Équité Contrôle/surveillance humains Accès aux compétences numériques	Explicabilité des systèmes IA Approche inclusive dans la conception, le développement et le déploiement des systèmes IA
Cybercriminalité et preuves électroniques	
Conduites à ériger en infraction pénale Données à sécuriser dans certaines enquêtes pénales pour être utilisées comme preuves. Pouvoir effectif de sécuriser des preuves électroniques limité par les conditions et garanties de l'État de droit.	Problème des preuves dans le nuage et de la compétence territoriale en matière pénale (question abordée dans le deuxième protocole additionnel à la Convention de Budapest).
Secteur de la justice	
Non-discrimination Qualité et sécurité des données Transparence Impartialité Équité Liberté de choix/indépendance des juges (processus décisionnel) Contrôle/surveillance humains Garanties du droit d'accès à un juge Garanties du droit à un procès équitable	Principe de précaution pour les applications ne respectant pas les exigences fondamentales de transparence
Congrès	
Transparence Contrôle humain (surveillance) Impartialité Droit au respect de la vie privée Sécurité des données Cybersécurité Non-discrimination Villes inclusives	Démocratie et participation – *deep fakes* (hypertrucage), micro-ciblage et propagande dans le cadre des processus électoraux

Pérennité financière Contrôle de la sécurité Organisation rationnelle des services Éducation au numérique	
Démocratie et participation	
Droit à des élections libres Liberté d'expression Droit d'accès des personnes à Internet Droit au respect de la vie privée Protection des données Égalité des chances pour les partis et les candidats Exigence de neutralité des autorités de l'État vis-à-vis des campagnes électorales, de la couverture médiatique et du financement public des partis politiques et des campagnes Exigence d'un accès minimum à des médias audiovisuels privés pour la campagne et les annonces électorales, pour tous les participants aux élections Transparence du financement des campagnes Prévention de l'influence injustifiée sur les décisions politiques par des dons financiers Couverture responsable, correcte et équitable des campagnes électorales par les médias ; droit de réponse, modalités de diffusion des sondages d'opinion, exigences de transparence quant aux contenus publicitaires payants ; pluralisme des médias Neutralité du réseau Protection des individus à l'égard de la collecte et du traitement de données à caractère personnel sur les autoroutes de l'information Non-discrimination Qualité et sécurité des données Transparence Impartialité Équité Liberté de choix/indépendance des juges (processus décisionnel) Contrôle/surveillance humains Garanties du droit d'accès à un juge Garanties du droit à un procès équitable	Équilibre entre des droits parfois concurrents comme : - le droit à des élections libres / la liberté d'expression - le droit d'accès à l'information y compris sur Internet / le droit au respect de la vie privée et à la protection des données Normes qui seraient applicables et appropriées pour les annonces/campagnes numériques, par exemple en ce qui concerne - l'égalité des chances pour les partis et les candidats - les campagnes électorales et le contrôle/la transparence de leur financement - la couverture médiatique équitable, le pluralisme des médias - la responsabilité des intermédiaires d'Internet en matière de transparence et d'accès aux données pour améliorer la transparence des dépenses, notamment en ce qui concerne les publicités politiques - la neutralité du réseau - la protection des données
Liberté d'expression	
Autonomie individuelle Égalité Sécurité démocratique Transparence et obligation de rendre compte Indépendance des médias Diversité et pluralisme	Principe de précaution pour les applications ne respectant pas les exigences fondamentales de transparence
Élections	
• élections libres et équitables • liberté de choix/d'opinion/d'expression • suffrage universel • suffrage égal	• principe de l'utilisation de systèmes IA dans les processus électoraux (en particulier systèmes de vote électronique, etc.)

• suffrage libre • suffrage secret • suffrage direct • fréquence des élections • transparence du processus électoral • processus électoral inclusif • participation et représentation équilibrées des femmes et des hommes dans la prise de décisions publiques	• possibilités d'améliorer le caractère inclusif des processus électoraux grâce à l'IA (comme outil pour les organes de gestion des élections et les commissions électorales ou comme assistant pour les électeurs).
Démocratie (hors questions relatives aux élections et au cycle électoral)	
Transparence Impartialité Équité Liberté de choix Liberté d'expression Liberté de réunion et d'association Accès à l'information Contrôle/surveillance humains Diversité Égalité Non-discrimination Qualité et sécurité des données Protection des données Indépendance	- rôle des intermédiaires - éducation aux technologies et au numérique - question de la propriété des médias - contrôle démocratique - données ouvertes et gouvernement ouvert - évaluation du risque
Bonne gouvernance	
- non-discrimination - qualité et sécurité des données - impartialité - équité - participation, représentation, déroulement équitable des élections - capacité d'adaptation - organisation rationnelle et efficacité - ouverture et transparence - état de droit - conduite éthique - compétence et capacité - innovation et ouverture au changement - durabilité (2) - gestion financière saine - droits de l'homme, diversité culturelle et cohésion sociale - obligation de rendre compte - mécanismes de réparation - accès à un recours - indépendance	- contrôle démocratique - accès à des mécanismes de recours et de réparation dans le cas d'une prise de décisions automatisée, reposant sur des algorithmes, par les agents de la fonction publique - rôle des intermédiaires - éducation et compétences en matière de technologies - question de la propriété réelle des données - données ouvertes et gouvernement ouvert - responsabilité civile et pénale - évaluations du risque et gestion du risque
Égalité femmes-hommes, y compris les questions relatives à la violence à l'égard des femmes	
Égalité et non-discrimination Intégrité / élimination de la violence (à l'égard des femmes) Égalité d'accès à la justice Garanties du droit à un procès équitable et à une réparation	Données non biaisées Caractère inclusif (pour les femmes) de l'IA en tant que secteur d'activité Respect des droits du travail et des droits sociaux dans l'IA en tant que secteur d'emploi Qualité et sécurité des données Transparence et explicabilité Obligation de rendre compte Impartialité Équité Contrôle/surveillance humains

	Éducation au numérique et résorption de la fracture numérique existante (entre femmes et hommes), essentielle pour le droit à réparation : si les citoyens et consommateurs ne comprennent pas l'IA, ils ne seront pas en mesure de faire valoir leurs droits Principe de précaution pour les applications ne respectant pas les exigences fondamentales de transparence Des principes éthiques comme la non-malfaisance ne sont pas respectés car certains logiciels espions sont mis au point et commercialisés dans le seul but de « savoir ce que fait son conjoint ».
Culture, créativité et patrimoine	
Non-discrimination Accès, liberté d'association, droit de participer à la vie culturelle et de créer et d'apprendre (pacte) Liberté d'expression Accès à de l'information impartiale	Principe de précaution pour les applications ne respectant pas les exigences fondamentales de transparence Nécessité de trouver des paradigmes culturels et des techniques pour traiter la question de l'autonomisation (n'existent que pour l'automatisation) « Éviter une centralisation encore plus forte du savoir et du pouvoir entre les mains de ceux qui les ont déjà et continuent de tenir à l'écart ceux qui en sont dépourvus » (M. Whitaker) Nécessité d'insister sur les règles et droits en matière d'accès aux biens communs et de participation à la vie publique (pratiques centrées sur les citoyens)
Non-discrimination – Impartialité (Protection des minorités nationales)	
Non-discrimination – Impartialité Protection des minorités et de leurs expressions culturelles (langues / diversité linguistique, patrimoine culturel)[159]	Propriété et éventuels biais des informations entrées dans les applications d'apprentissage par IA
Promotion/protection de l'identité, de la diversité et de la coopération européennes Accès et participation au patrimoine culturel ; protection du patrimoine culturel	Protection de la créativité humaine (et sa nature spécifique)
Contrôle/surveillance humains du processus créatif, transparence Protection et promotion de la diversité culturelle Créer les conditions permettant aux cultures de s'épanouir et interagir librement Reconnaître la nature spécifique des activités, biens et services culturels en tant que porteurs d'identité, de valeurs et sens	Gestion des adresses IP et du droit d'auteur Protection de la créativité humaine (et sa nature spécifique)
Diversité culturelle Coopération culturelle en Europe et au-delà Disponibilité des œuvres Non-discrimination Protection des données	Visibilité des œuvres Transparence de la prise de décision (de concevoir et produire/censurer/recommander une œuvre)

[159] Une machine permettant de recréer des langues par IA pourrait combler les lacunes et contribuer à créer une archive globale des langues vivantes, un « Louvre des langues ».

Liberté d'expression et de création Surveillance/contrôle humains, transparence	Propriété des adresses IP, questions de droits d'auteur et de droits moraux
Non-discrimination Accès, liberté d'association, droit de participer à la vie culturelle et de créer et d'apprendre (Pacte) Liberté d'expression Accès à de l'information impartiale	Principe de précaution pour les applications ne respectant pas les exigences fondamentales de transparence Nécessité de trouver des paradigmes culturels et des techniques pour traiter la question de l'autonomisation (n'existent que pour l'automatisation) « Éviter une centralisation encore plus forte du savoir et du pouvoir entre les mains de ceux qui les ont déjà et continuent de tenir à l'écart ceux qui en sont dépourvus » (M. Whitaker) Nécessité d'insister sur les règles et droits en matière d'accès aux biens communs et de participation à la vie publique (pratiques centrées sur les citoyens)
GRECO	
Principes directeurs pour la lutte contre la corruption	Rien de spécifique sur : les applications d'IA pour prévenir la corruption ; la nécessité de s'assurer que les algorithmes ne sont pas corrompus
	Travail en cours du CDPC sur la responsabilité pénale liée à l'utilisation de véhicules automatisés
Article 8 : droit au respect de la vie privée et familiale	Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH)
CSE	
Plusieurs grands principes peuvent être tirés de la Charte et des activités de suivi menées dans le cadre de la Charte sur la transparence et la participation à la prise de décisions.	Une prise de décision automatisée, assistée par ordinateur ou reposant sur l'IA nécessiterait : - un contrôle humain obligatoire pour éviter ou limiter les erreurs dans la gestion, l'attribution ou le retrait de prestations, aides et autres avantages, pouvant accentuer les inégalités et les privations de droits. - des dispositifs efficaces pour protéger les personnes vulnérables contre le dénuement, l'extrême pauvreté ou la perte de logement et contre les blessures graves ou préjudices irréparables découlant de la mise en œuvre de décisions assistées par ordinateur ou reposant sur l'IA dans le domaine des services sociaux. - une approche proactive pour que les personnes concernées par des décisions assistées par ordinateur ou reposant sur l'IA dans le domaine des services sociaux, et en particulier les personnes en situation d'extrême pauvreté ou de vulnérabilité, puissent effectivement faire valoir leurs droits et obtenir réparation.

Annexe 4 – Protection des données

Instruments contraignants et non contraignants dans le domaine de la protection des données

Lignes directrices sur les incidences de l'intelligence artificielle sur la protection des données[160] Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (Convention 108), 2019.	**Contrôle humain** I.6 Les applications de l'IA devraient permettre aux personnes concernées d'exercer un contrôle significatif sur le traitement des données et leurs effets connexes tant au niveau individuel que sur la société. **Conception centrée sur les valeurs** II.1. Les développeurs, fabricants et prestataires de service en IA devraient adopter une approche de conception des produits et services centrée sur les valeurs, conformément à la Convention 108+, notamment son article 10.2, et aux autres instruments pertinents du Conseil de l'Europe. **Approche de précaution** II.2 Les développeurs, fabricants et prestataires de service en IA devraient évaluer les éventuelles conséquences négatives des applications d'IA sur les droits de l'homme et libertés fondamentales des personnes concernées et au regard de ces conséquences, adopter une approche de précaution basée sur des mesures de prévention et de réduction des risques appropriées. **Approche fondée sur les droits de l'homme dès la conception et détection des biais** II.3 Les développeurs, fabricants et prestataires de service en IA devraient, à tous les stades du traitement des données, y compris lors de la collecte, adopter une approche des droits de l'homme dès la conception (by-design) et éviter tout biais potentiel, y compris les biais non intentionnels ou cachés, ainsi que les risques de discrimination ou d'autres effets négatifs sur les droits de l'homme et libertés fondamentales des personnes concernées. **Qualité et minimisation des données** II.4 Les développeurs en IA devraient évaluer de manière critique la qualité, la nature, l'origine et la quantité de données à caractère personnel utilisées, en réduisant les données inutiles, redondantes ou marginales lors des phases de conception et d'apprentissage, puis en vérifiant l'exactitude du modèle lorsqu'il est alimenté par de nouvelles données. Le recours à des données synthétiques pourrait être considéré comme une solution possible pour minimiser la quantité de données personnelles traitées par des applications de l'IA. **Risque de décontextualisation** II.5 Les risques d'impact négatif sur les personnes et la société inhérents aux données décontextualisées et aux modèles algorithmiques décontextualisés devraient être dûment pris en compte lors du développement et de l'utilisation d'applications de l'IA. **Comités d'experts indépendants** II.6 Les développeurs, fabricants et prestataires de service en IA sont encouragés à recourir à des comités d'experts indépendants issus de différents domaines ainsi qu'à des institutions universitaires indépendantes qui peuvent contribuer à concevoir des applications de

[160] Voir également T-PD(2019)01, Lignes directrices sur l'intelligence artificielle et la protection des données ; T-PD(2017)1, Lignes directrices sur la protection des individus à l'égard du traitement des données à caractère personnel à l'ère des mégadonnées.

	l'IA fondées sur les droits de l'homme et orientées de façon éthique et sociale, ainsi qu'à détecter des biais potentiels. Le rôle de ces comités peut être particulièrement important dans les domaines où la transparence et la mobilisation des parties prenantes peuvent être plus difficiles en raison d'intérêts et de droits concurrents, par exemple dans les domaines de la justice prédictive, de la prévention du crime et de la détection des infractions. III.7 Des mécanismes appropriés visant à garantir l'indépendance des comités d'experts mentionnés à la section II.6 devraient être mis en place. **Participation et contrôle démocratique du développement de l'IA** II.7 Des démarches participatives d'évaluation des risques, reposant sur l'engagement actif des personnes et groupes potentiellement affectés par les applications de l'IA, devraient être encouragées. III. 8. Les personnes, les groupes et les autres parties prenantes devraient être informés et impliqués de façon active dans le débat relatif au rôle que l'IA devrait jouer dans le modelage des dynamiques sociales, et dans les processus décisionnels les affectant. **Contrôle humain** II.8 Tous les produits et services de l'IA devraient être conçus de manière à garantir le droit des personnes à ne pas être soumises à des décisions qui les affectent de manière significative prises uniquement sur le fondement d'un traitement automatisé de données, sans que leur point de vue soit pris en compte. **Liberté de choix** II.9 Afin d'accroître la confiance des utilisateurs, les développeurs, fabricants et prestataires de services en IA sont encouragés à concevoir leurs produits et services de manière à préserver la liberté de choix de l'utilisateur concernant l'usage de l'IA en proposant des alternatives réalistes aux applications de l'IA. **Vigilance algorithmique** II.10 Les développeurs, fabricants et prestataires de service en IA devraient adopter des formes de vigilance algorithmique qui fassent la promotion de la responsabilité de toutes les parties prenantes et ce tout au long du cycle de vie des applications afin d'assurer leur conformité avec les principes et la législation relatifs à la protection des données personnelles et aux droits de l'homme. **Transparence et extensibilité** II.11 Les personnes concernées devraient être informées si elles interagissent avec des applications de l'IA et ont le droit de connaître du raisonnement qui sous-tend les opérations de traitement des données qui les concernent. Ceci devrait inclure les conséquences de ce raisonnement. **Droit d'opposition** II.12 Le droit d'opposition devrait être garanti par rapport au traitement basé sur des technologies qui influencent les opinions et le développement personnel des individus. **Responsabilité et vigilance** III.2 Sous réserve des secrets protégés par la loi, les procédures de passation des marchés publics devraient imposer aux développeurs, fabricants et prestataires de service en IA des devoirs spécifiques de transparence et d'évaluation préalable de l'impact des traitements de

	données personnelles sur les droits de l'homme et libertés fondamentales des systèmes d'IA, ainsi que de vigilance à l'égard des effets défavorables potentiels et des conséquences des applications d'IA (ci-après dénommée « vigilance algorithmique »). **Liberté des décideurs humains** III.4. Une dépendance excessive des solutions fournies par les applications de l'IA, de même que la crainte de contester des décisions suggérées par des applications de l'IA risquent d'altérer l'autonomie de l'intervention humaine dans la prise de décision. Le rôle de l'intervention humaine dans le processus décisionnel et la liberté des décideurs humains de ne pas suivre les résultats de recommandations fondées sur l'utilisation de l'IA devraient en conséquence être préservés. **Évaluation préalable** III.5. Les développeurs, fabricants et prestataires de service en IA devraient consulter les autorités de contrôle dès lors que les applications de l'IA peuvent avoir un impact significatif sur les droits de l'homme et les libertés fondamentales des personnes concernées. **Coopération** III.6. La coopération devrait être encouragée entre autorités de contrôle de la protection des données et d'autres instances ayant des compétences liées à l'IA, telles que les autorités de régulation en matière de protection des consommateurs ; concurrence ; anti-discrimination ; médias et autorités sectorielles. **Éducation, éducation au numérique et formation professionnelle** III.9. Les décideurs devraient affecter des ressources à l'éducation au numérique afin de renforcer la sensibilisation des personnes concernées et la compréhension des applications de l'IA et de leurs effets. Ils devraient également encourager la formation professionnelle des développeurs en IA pour qu'ils soient sensibilisés aux effets potentiels de l'IA sur les personnes et la société. Ils devraient soutenir la recherche en matière d'IA orientée vers les droits de l'homme.
Recommandation Rec(2019)2 du Comité des Ministres du Conseil de l'Europe aux États membres en matière de protection des données relatives à la santé.	Le traitement des données relatives à la santé devrait toujours servir la personne concernée ou conduire à améliorer la qualité et l'efficacité des soins de santé, ainsi que les systèmes de santé lorsque cela est possible, tout en respectant les droits fondamentaux de la personne. **Interopérabilité** 1. […] Elle souligne à cette fin l'importance du développement de systèmes d'information sécurisés **interopérables.** **Normes professionnelles** 4.4 Les responsables du traitement et leurs sous-traitants qui ne sont pas des professionnels de santé ne devraient traiter des données relatives à la santé que dans le respect de règles de confidentialité et des mesures de sécurité garantissant un niveau de protection équivalant à celui qui incombe aux **professionnels** de santé. **Retrait du consentement** 5.b Les données relatives à la santé peuvent être traitées dès lors que la personne concernée a donné son consentement, sauf dans les cas où le droit prévoit qu'une interdiction de traiter les données de santé ne peut être levée par le seul consentement de la personne concernée. Lorsque le consentement de la personne concernée au traitement de données relatives à la santé est requis, conformément

	au droit, celui-ci devrait être libre, spécifique, éclairé et explicite. La personne concernée doit être informée de son droit de retirer son consentement à tout moment et du fait qu'un tel retrait ne compromet pas la licéité du traitement fondé sur le consentement effectué avant ce retrait. Il doit être **aussi simple de retirer son consentement que de le donner.** **Droit de ne pas savoir** 7.6 La personne concernée a le droit de connaître toute information relative à ses données génétiques, sous réserve des dispositions des principes 11.8 et 12.7. Toutefois, pour des raisons qui lui appartiennent, la personne concernée peut souhaiter ne pas connaître certains éléments relatifs à sa santé et toute personne devrait être informée, préalablement à la réalisation d'une analyse, de la possibilité dont elle dispose de ne pas être informée de résultats, y compris de découvertes inattendues. Le **souhait de ne pas savoir** peut, dans des circonstances exceptionnelles, **faire l'objet de restrictions prévues par la loi, notamment dans l'intérêt de la personne concernée ou au regard de l'obligation de soigner qui incombe aux médecins**. **Transparence** 11.3. L'information doit, le cas échéant, afin de garantir la loyauté et la transparence du traitement, également porter sur : […] - l'**existence de décisions automatisées,** y compris le profilage qui n'est acceptable que si la loi le permet et sous réserve de garanties appropriées. **Interopérabilité** 14.1. L'interopérabilité peut permettre de répondre à des impératifs relevant du domaine de la santé et peut apporter des moyens techniques qui facilitent la mise à jour, qui évitent la duplication de données identiques dans de multiples bases de données et qui contribuent à la portabilité. 14.2. Il est cependant nécessaire que l'interopérabilité soit mise en œuvre conformément aux principes contenus dans cette recommandation, notamment les principes de licéité, de nécessité et de proportionnalité, et que des mesures de sauvegarde de la protection des données à caractère personnel soient prises lorsque des systèmes interopérables sont utilisés. 14.3. Des référentiels fondés sur des normes internationales et offrant un cadre technique qui facilite l'interopérabilité devraient assurer qu'un haut niveau de sécurité est garanti tout en offrant une telle interopérabilité. Leur mise en œuvre peut être suivie au moyen de schémas de certification. **Intégrité de la recherche scientifique** 15.10. Lorsqu'une personne décide de se retirer d'une recherche scientifique, ses données relatives à la santé traitées dans le cadre de cette recherche devraient être détruites ou anonymisées de manière à **ne pas compromettre la validité scientifique de la recherche** et la personne concernée devrait en être informée.
Recommandation CM/Rec(2016)8 du Comité des Ministres aux États membres sur le traitement de données à caractère personnel relatives à la santé à des fins	8. Le traitement à des fins d'assurance de données à caractère personnel relatives à la santé obtenues dans le cadre d'une recherche impliquant l'assuré(e) ne devrait pas être permis.

d'assurance, y compris les données résultant de tests génétiques	
Recommandation Rec(2010)13 du Comité des Ministres du Conseil de l'Europe aux États membres sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage	**Risque de ré-identification** 8.5. Des mesures appropriées devraient être mises en place pour éviter que des résultats statistiques anonymes et agrégés utilisés dans le cadre du profilage ne puissent déboucher sur une ré-identification des personnes concernées[161]
UNESCO 2019. Étude préliminaire concernant un éventuel instrument normatif sur l'éthique de l'intelligence artificielle.	[Approche fondée sur les principes] • Diversité, inclusion et pluralisme (y compris promotion d'une approche multilingue) • Autonomie • Explicabilité • Transparence • Sensibilisation et éducation • Responsabilité • Obligation de rendre compte • Démocratie (l'IA devrait être élaborée, mise en œuvre et utilisée conformément aux principes démocratiques) • Bonne gouvernance (« Les gouvernements devraient présenter des rapports réguliers sur leur utilisation de l'IA dans la police, le renseignement et la sécurité ») • Durabilité • Contrôle humain • Liberté d'expression (y compris accès universel à l'information ; journalisme de qualité ; médias libres, indépendants et pluralistes ; éviter la diffusion de fausses informations)
OCDE. 2019. Recommandation du Conseil sur l'intelligence artificielle	[Approche fondée sur les principes] • Valeurs centrées sur l'humain et équité • Transparence et explicabilité (**informer** les parties prenantes de leurs interactions avec les systèmes IA ; permettre aux personnes concernées d'appréhender le résultat de l'IA ; permettre à ceux qui subissent les effets néfastes d'un système d'IA de **contester** son résultat sur la base d'**informations** claires et **facilement compréhensibles** sur les facteurs et la **logique** ayant servi à la formulation de prévisions, recommandations ou décisions) • Robustesse, sûreté et sécurité (ne font pas peser un **risque de sécurité démesuré ; traçabilité**, notamment pour ce qui est des ensembles de données, des processus et des décisions prises au cours du cycle de vie des systèmes d'IA ; **approche de**

[161] Voir également le rapport explicatif de la Convention 108+ : par. 19 et 20 (« Les données ne peuvent être considérées comme anonymes que lorsque la ré-identification de la personne concernée est impossible ou nécessiterait des délais, efforts ou ressources déraisonnables au vu des technologies disponibles au moment du traitement et de l'évolution de celles-ci. [...] Lorsque des données sont rendues anonymes, des moyens appropriés doivent être mis en place pour empêcher toute ré-identification des personnes concernées ; en particulier, tous les moyens techniques doivent être mis en œuvre pour garantir que la personne n'est pas ou plus identifiable. Étant donné la rapidité des évolutions techniques, ces moyens techniques devraient être réévalués régu

	gestion des risques à chaque phase du cycle de vie des systèmes d'IA) • Obligation de rendre compte
40e Conférence internationale des commissaires à la protection des données et de la vie privée (ICDPPC). 2018. Déclaration sur l'éthique et la protection des données dans le secteur de l'intelligence artificielle [ICDPPC]	[Approche fondée sur les principes] • Attention et vigilance permanentes (« mettre en place pour tous les acteurs concernés des processus de gouvernance dont on peut apporter la preuve, par exemple en s'appuyant sur des tiers de confiance ou en créant des **comités d'éthique** indépendants ») • Transparence et intelligibilité (**IA explicable**, **transparence** des algorithmes et **vérifiabilité** des systèmes, **connaissance** des interactions avec les systèmes IA ; informations adéquates sur les objectifs et les effets des systèmes IA, **contrôle humain global)** • Évaluation du risque et application des principes de protection de la vie privée par défaut (*privacy by default*) et de protection intégrée de la vie privée (*privacy by design*) (« évaluer et décrire les **impacts** attendus **sur les personnes et la société** au début d'un projet d'intelligence artificielle et au cours des développements pertinents **durant tout son cycle de vie** ») • **Participation publique** • **Réduction et atténuation des préjugés et des discriminations illicites**