

Les droits humains dès la conception de l'IA

une protection durable des droits humains à l'ère de l'intelligence artificielle



Jane Doe	
Age 32	Height 5'10"
Gender	Female
Occupation	Teacher in public health
Interests	Environmental activism, Hiking
Location	London
ID	38713597845
MATCH 97%	
database generalPublic2B	

Recommandation de suivi

à "Décoder l'IA"
(2019)

Les droits humains dès la conception de l'IA une protection durable des droits humains à l'ère de l'intelligence artificielle

Recommandation de suivi
à "Décoder l'IA" (2019)
par la Commissaire aux droits de l'homme
du Conseil de l'Europe

Toute demande concernant
la reproduction ou la traduction
de tout ou partie de ce document
doit être adressée à la
Direction de la Communication
(publishing@coe.int).

Tout autre correspondance
concernant ce document doit
être adressée au Bureau du/de la
Commissaire aux droits de l'homme.

Les publications sont disponibles
sur le site web :
www.commissioner.coe.int

Photo de couverture :
photo éditée ©Shutterstock
© Conseil de l'Europe,
Mai 2023

Remerciements:

La Commissaire souhaite exprimer
sa gratitude à Nani Jansen
Reventlow et Jonathan McCully
pour leurs recherches et leur
aide dans la préparation de cette
Recommandation de suivi.

Contenu

INTRODUCTION	5
CHAPITRE 1 - ÉTUDES D'IMPACT DE L'IA SUR LES DROITS HUMAINS	9
CHAPITRE 2 - LES NORMES EN MATIÈRE DE DROITS HUMAINS DANS LE SECTEUR PRIVÉ	13
CHAPITRE 3 - INFORMATION ET TRANSPARENCE	17
CHAPITRE 4 - CONSULTATIONS PUBLIQUES	21
CHAPITRE 5 - PROMOTION DE LA CONNAISSANCE ET DE LA COMPRÉHENSION DE L'IA	23
CHAPITRE 6 - CONTRÔLE INDÉPENDANT	25
CHAPITRE 7 - VOIES DE RECOURS EFFECTIVES	29
OBSERVATIONS FINALES ET RECOMMANDATIONS	33

Introduction

Le développement technologique s'accélère et l'intelligence artificielle (IA) joue un rôle de plus en plus important dans tous les aspects de notre vie, y compris dans l'administration publique. Il est indispensable que les États membres du Conseil de l'Europe ne se laissent pas distancer face à cette dépendance toujours plus grande à l'égard des processus automatisés et de l'apprentissage automatique, et veillent, dans ce contexte de mutation rapide, à préserver les droits humains de tous les membres de la société.

Ce besoin pressant a incité la Commissaire aux droits de l'homme du Conseil de l'Europe, Dunja Mijatović, (ci-après, la Commissaire) à publier en 2019 la Recommandation « [Décoder l'IA : 10 mesures pour protéger les droits de l'homme](#) » (ci-après, la Recommandation de 2019), qui fournit aux États membres des orientations sur les grands principes à suivre pour prévenir ou atténuer les effets négatifs des systèmes d'IA sur les droits humains. La Recommandation de 2019 met en avant les risques bien spécifiques que fait peser l'IA sur le droit à la non-discrimination et à l'égalité et le droit à la protection des données et à la vie privée, et sur la liberté d'expression, la liberté de réunion et le droit au travail. De plus, elle met en évidence sept domaines clés, auxquels il convient de porter une attention particulière : la nécessité de réaliser des études d'impact sur les droits humains (EIDH) avant l'acquisition, le développement et/ou le déploiement d'un système d'IA ; le respect des normes en matière de droits humains dans le secteur privé ; l'information et la transparence ; la nécessité de consultations publiques pleinement pertinentes ; la promotion de l'éducation à l'IA ; le besoin d'un contrôle indépendant ; et l'accès à des voies de recours effectives.

Les États membres ont pris des mesures dans certains des domaines clés identifiés dans la Recommandation de 2019, mais sans cohérence globale. Ainsi, sur de nombreux fronts, la réglementation des systèmes d'IA centrée sur les droits humains est-elle encore absente, et le plus souvent, les autorités publiques s'investissent trop tard – et avancent trop lentement – pour que leur contribution soit véritablement utile. Les normes et garanties en matière de droits humains, bien que technologiquement neutres et applicables à tous les contextes, y compris ceux mettant en œuvre des systèmes d'IA, sont rarement appliquées et les contrôles restent sporadiques.

La présente Recommandation de suivi examine les défis auxquels sont confrontés les États membres dans la mise en œuvre de la Recommandation de 2019, notamment en ce qui concerne la pertinence de l'évaluation des risques et de l'étude d'impact sur les droits humains, la mise en place de garanties de transparence plus solides et la nécessité d'un contrôle indépendant. Dans quelle mesure les États membres sont-ils parvenus à utiliser l'IA comme un moyen de renforcer les droits humains plutôt que de leur nuire, et quelles sont, plus généralement, les tendances qui influent sur leurs pratiques? Quel rôle jouent les structures nationales des droits humains (SNDH), notamment les institutions nationales des droits humains (INDH), les organismes de promotion de l'égalité et les institutions des ombudsmans, et comment renforcer leur rôle pour qu'une réponse efficace soit apportée aux multiples atteintes aux droits fondamentaux causées par l'IA?

Cette Recommandation considère les effets négatifs que les systèmes d'IA peuvent avoir sur la capacité des personnes à exercer leurs droits fondamentaux comme une possible [atteinte aux droits humains](#). Pour bien comprendre les éventuels préjudices causés par les nouvelles technologies comme l'intelligence artificielle, il est essentiel de prendre en compte le contexte global dans lequel elles sont déployées, et notamment les inégalités existantes et les déséquilibres de pouvoir. Les technologies d'apprentissage automatique exploitent les schémas et les hypothèses qui prédominent dans les données qu'elles utilisent. En conséquence, elles renforcent et exacerbent les biais et préjugés systémiques qui préexistent, par exemple à l'égard des femmes, des jeunes, des personnes handicapées ou des personnes issues d'une minorité. Les applications d'IA étant utilisées dans un grand nombre de secteurs, notamment pour la distribution des prestations sociales, les décisions concernant la solvabilité de futurs clients, les processus de recrutement et de maintien du personnel, les procédures pénales, l'immigration, le contrôle aux frontières, le maintien de l'ordre, la publicité ciblée et les fils d'actualité, les effets négatifs de l'IA ne se traduisent pas seulement par des violations individuelles et éventuellement collectives des droits humains : ils nuisent aussi à la justice sociale, modifient la relation et la confiance entre l'État et les citoyens, et retentissent sur l'intégrité et même le résultat des élections. Enfin, la multiplication des systèmes d'IA soulève d'importantes questions quant aux conditions de travail de personnes employées – souvent de façon informelle – sur plateforme numérique et quant aux quantités exorbitantes d'énergie et d'eau nécessaires à leur fonctionnement. Ce n'est qu'en adoptant une approche holistique face aux menaces que l'IA fait peser sur les droits humains et en tenant compte du caractère pluridimensionnel, intersectionnel et dynamique des préjudices causés, que les États membres pourront dûment remplir leurs obligations internationales en matière de droits humains.

Depuis la Recommandation de la Commissaire, en 2019, les SNDH et leurs réseaux ont pris plusieurs initiatives louables en vue de renforcer leur capacité à traiter les questions que soulèvent la conception, le développement et le déploiement de l'IA sur le plan des droits humains. Ainsi, en 2020, le réseau des organismes de promotion de l'égalité EQUINET a publié un [rapport](#) détaillé traitant de l'impact de l'IA sur l'égalité et du rôle qu'ont à jouer les organismes de promotion de l'égalité à cet égard. En mars 2020, l'Institut international de l'Ombudsman a organisé un [atelier](#) pour examiner les défis, les rôles et les outils des institutions des ombudsmans qui traitent de l'IA et des droits humains. En outre, pour renforcer la capacité de ses membres à investir ce nouveau domaine, le Réseau européen des institutions nationales des droits de l'homme (REINDH) a axé son [Académie des INDH 2022](#) sur l'IA et les droits humains, et a créé, en octobre 2022, un groupe de travail des INDH sur l'IA, qui servira de plate-forme d'échange et d'apprentissage entre pairs. Parallèlement, le réseau EQUINET a aussi organisé plusieurs formations. Le réseau REINDH et ses membres ont également entrepris des démarches pour introduire des considérations touchant aux droits humains dans les processus législatifs concernés au niveau régional, et ont présenté des contributions pour le Livre blanc de la [Commission européenne](#) sur l'intelligence artificielle et pour le projet de [Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit](#). Les deux réseaux, REINDH et EQUINET, qui ont le statut d'observateur auprès du [Comité sur l'intelligence artificielle du Conseil de l'Europe](#) (CAI), assistent régulièrement aux réunions et apportent leur [contribution](#) au projet de convention-cadre. La Commissaire se félicite vivement de ces avancées et encourage à poursuivre les efforts en vue d'une participation accrue des SNDH à la défense et à la promotion des droits humains dans le domaine de la gouvernance et de la réglementation de l'IA.

Un [échange de vues](#) entre la Commissaire et trente dirigeants et hauts représentants de SNDH des États membres du Conseil de l'Europe a été organisé en mars 2023 dans le cadre de la préparation de la présente Recommandation de suivi. Certains des points soulevés par les participants lors de cet échange sont cités dans les chapitres I à VII, qui donnent un aperçu de certaines des mesures prises pour mettre en œuvre la Recommandation de 2019. Les conclusions recensent les lacunes actuelles et dégagent des tendances plus générales, amenant à la formulation de recommandations sur la manière de progresser vers une meilleure application des normes en matière de droits humains, en lien avec la conception, le développement et le déploiement de l'IA.

Chapitre 1

Études d'impact de l'IA sur les droits humains

Les systèmes d'IA peuvent avoir des effets négatifs importants sur les droits humains et constituer une menace pour [la vie humaine](#), [l'environnement](#), [la démocratie](#) et [l'État de droit](#). Il est donc souhaitable que les États membres réalisent des EIDH avant la conception, le développement ou le déploiement de tout système d'IA. Pour être efficace, il importe de bien comprendre que les atteintes aux droits humains ne résultent pas uniquement de l'utilisation de ces systèmes, mais qu'elles peuvent apparaître à tout moment de leur [cycle de vie](#), y compris lors des phases d'acquisition des données et de modélisation. Il faut aussi prêter attention au caractère dynamique des environnements réels dans lesquels ces systèmes sont amenés à fonctionner, et notamment à interagir avec d'autres systèmes présents simultanément.

Depuis 2019, les évaluations des risques et les études d'impact occupent une place importante dans les débats de politique générale sur la gouvernance et la réglementation de l'IA en Europe. Pourtant, bon nombre des initiatives mises en œuvre n'ont pas abouti à la réalisation systématique d'EIDH pour tous les systèmes d'IA qui sont achetés, développés et/ou déployés par les autorités publiques. De fait, l'approche par secteur qui a été privilégiée restreint l'obligation de mener des EIDH en la limitant à des branches bien spécifiques de l'industrie ou à certains droits uniquement. Cette approche, qui peut être à l'origine de lacunes réglementaires et conduire à une mise en œuvre parcellaire des EIDH, doit être évitée. Si les projets pilotes, notamment dans le domaine de la [santé](#), sont les bienvenus, ils devraient être étendus à d'autres secteurs utilisant l'IA, notamment la [police](#), la [migration](#) et la [protection sociale](#), et tous devraient évaluer les effets éventuels des systèmes d'IA sur l'ensemble des droits humains.

En effet, le plus souvent, les États membres n'ont pas pris en considération la totalité des droits humains menacés ou ont examiné de vagues cadres éthiques plutôt que de se référer à des normes spécifiques et établies en matière de droits fondamentaux. En outre, la plupart des cadres

d'évaluation sont conçus comme des exercices facultatifs du secteur privé relevant de son devoir général de vigilance, et non comme des obligations contraignantes des États membres destinées à prévenir les atteintes aux droits humains causées par l'IA, y compris lorsqu'elles sont le fait de tiers.

Le champ d'application des EIDH doit être suffisamment large pour couvrir la totalité des droits humains. Les risques qui font l'objet de ces études ne doivent donc pas être définis de façon trop restrictive, une telle approche étant incompatible avec les principes d'indivisibilité et d'interdépendance des droits. Les algorithmes de modération de contenu, par exemple, n'ont pas pour seul effet d'entraver la liberté d'expression. De fait, des erreurs lors du retrait de certains contenus peuvent venir alimenter des algorithmes utilisés pour sélectionner les candidats à un emploi, risquant ainsi de bloquer des accès de manière temporaire ou permanente et donc de porter atteinte au droit de s'associer et de se désengager. Le CAI, soulignant la nécessité d'une approche globale des évaluations des risques et des études d'impact de l'IA, conçoit en ce moment une [méthodologie](#) contenant les paramètres essentiels d'un processus de gestion des risques et des effets des systèmes d'IA du point de vue des droits humains, de la démocratie et de l'État de droit. Les répercussions sur les groupes jouissant d'une protection particulière, notamment les enfants, qui représentent un tiers des utilisateurs des services numériques dans le monde, doivent également être examinées et atténuées de manière efficace.

Les SNDH, parce qu'elles ont une connaissance spécifique de la situation des droits humains dans les États membres et des groupes à risque, peuvent fournir des conseils utiles et importants sur les types de risques susceptibles de se présenter – qu'il s'agisse de violations individuelles des droits humains ou, plus largement, de menaces qui pèsent sur certaines catégories de personnes – et sur les mesures d'atténuation qui pourraient être prises. Dans ce contexte, il est important que les SNDH soient elles-mêmes bien au fait des types d'atteintes aux droits humains que peut causer l'utilisation de l'IA dans leur contexte national spécifique. À titre d'exemple, en 2019, l'Agence fédérale allemande de lutte contre la discrimination a financé une étude intitulée «[Les risques de discrimination liés à l'utilisation des algorithmes](#)» dans le but de porter ces risques à l'attention des autorités. En 2022, l'Autorité suédoise pour l'égalité de genre a réalisé une [étude](#) dans laquelle elle examine comment les organismes publics utilisent l'IA et dans quelle mesure ils prennent en compte les risques de discrimination. D'après les conclusions de cette étude, malgré une certaine prise de conscience du grand nombre d'individus touchés par les systèmes d'IA et des risques de discrimination qui existent, les autorités publiques ne prennent pas en compte, de façon systématique, la dimension de la discrimination dans les analyses de risque ou les processus de mise en œuvre, mais elles privilégient

l'aspect éthique et les questions d'intégrité.

Les initiatives prises pour évaluer les risques et analyser les effets des systèmes d'IA font souvent référence à l'éthique et aux principes éthiques plutôt qu'aux normes reconnues en matière de droits humains. Cela a pour effet de décentrer les protections de ces droits et de diluer la responsabilité juridique en matière de conception, de développement et de déploiement des systèmes d'IA. Le droit international en matière de droits humains a été décrit comme la [cristallisation de principes éthiques en normes](#), et il est le résultat de décennies de débats minutieux et d'examen approfondis de diverses considérations relatives aux droits humains, mais aussi d'autres intérêts, notamment l'éthique. Même s'il ne garantit pas la protection de toutes les valeurs éthiques d'intérêt sociétal qui sont concernées par les systèmes d'IA, le droit international en matière de droits humains impose aux États membres du Conseil de l'Europe une liste claire d'obligations juridiques auxquelles ils ne peuvent échapper, y compris en matière de conception, de développement et de déploiement de l'IA.

“ *Bon nombre de préjudices algorithmiques sont cachés. Ainsi, en essayant de réduire le risque de discrimination fondée sur le genre, on pourrait aboutir à une élévation du risque de discrimination fondée sur le handicap ou l'appartenance ethnique. Ces caractéristiques inhérentes aux algorithmes ne sont pas encore suffisamment comprises. Aussi, pour poser les bonnes questions et comprendre les compromis qui sont faits au sein des systèmes, nous avons besoin d'être formés davantage.*

Les normes établies en matière de droits humains offrent des méthodes et des processus solides pour évaluer les droits individuels par rapport aux droits et intérêts concurrents. Pour cela, elles comportent des critères de légalité, de nécessité et de proportionnalité, lesquels sont largement connus et admis et peuvent donc être appliqués par l'ensemble des organisations et acteurs intervenant dans le développement, la construction et l'exploitation des systèmes d'IA. Les normes internationales en matière de droits humains devraient donc être le point de référence dans les études d'impact et les évaluations des risques réalisées par les autorités publiques en lien avec les systèmes d'IA. Ces normes devraient leur servir de guide lorsqu'elles examinent les risques que ces systèmes font peser sur les droits humains et évaluent si ces risques, une fois suffisamment atténués, sont toujours nécessaires et proportionnés pour atteindre un objectif ou un

intérêt légitime. Alors seulement, ils pourront être considérés comme des risques justifiables en droit.

En 2020, l'Institut danois des droits humains a élaboré [un guide et une boîte à outils sur la manière de réaliser des EIDH sur les activités numériques](#). Ces aides sont utilisées par un nombre croissant d'entreprises et par certains acteurs du secteur public pour évaluer les effets de leurs activités numériques sur les détenteurs de droits, notamment les travailleurs, les populations locales et les consommateurs.

En 2022, les Pays-Bas sont [devenus le premier État membre](#) à demander aux organismes publics de réaliser une [EIDH obligatoire](#) avant d'utiliser des algorithmes à des fins d'évaluation ou de prise de décisions concernant des personnes. La publication des résultats des EIDH a également été rendue obligatoire, ce qui est conforme à la Recommandation de 2019 de la Commissaire. Ce [cadre juridique](#) fournit des indicateurs permettant de déterminer si un algorithme porte atteinte à une liberté ou un droit fondamental lors de la phase de conception, de développement ou de déploiement, et s'il existe une base juridique adéquate en vertu des normes de protection des droits humains. Il présente également diverses mesures de prévention et d'atténuation pouvant être adoptées en réponse aux risques identifiés, ce qui, comme indiqué plus haut, favorise l'adoption des EIDH par rapport à d'autres cadres d'évaluation plus limités.

Pour être efficaces et pleinement pertinents, les cadres juridiques qui prévoient la réalisation d'EIDH obligatoires et accessibles au public pour les systèmes d'IA doivent être strictement mis en œuvre, quelle que soit la taille ou la part de marché du développeur d'IA concerné. Les États membres doivent donc doter les autorités publiques de ressources suffisantes, de l'expertise nécessaire et d'un accès adéquat à l'information. De leur côté, les autorités publiques ne doivent pas acquérir ni utiliser de systèmes d'IA provenant de tiers si elles n'ont pas la possibilité d'effectuer une EIDH effective, notamment si le développeur refuse de communiquer des informations importantes sur le système d'IA qu'il commercialise. De plus, les États membres doivent veiller à ce que, dans leurs cadres juridiques, tous les obstacles à la réalisation effective d'EIDH soient surmontés.

Chapitre 2

Les normes en matière de droits humains dans le secteur privé

Les technologies d'IA développées par le secteur privé progressent à une telle vitesse que les États membres se doivent d'agir au plus vite, en tant que principaux garants des obligations en matière de droits humains, pour que leur législation ne devienne pas obsolète face à la menace que représentent ces avancées technologiques pour les droits fondamentaux. Pourtant, depuis la Recommandation de 2019, les États membres tardent à adopter des cadres juridiques permettant de traiter, prévenir et réparer les violations des droits humains commises par les acteurs de l'IA dans le secteur privé. Ce retard pourrait s'expliquer en partie par le fait que les initiatives déployées au niveau du Conseil de l'Europe et de l'Union européenne pour élaborer des cadres juridiques appropriés en matière de conception, développement et déploiement de l'IA sont toujours en cours, et que les États membres attendent les résultats pour agir. Cela dit, les négociations qui ont cours au niveau régional ne diminuent en rien les obligations permanentes qui leur incombent de protéger les personnes et les systèmes contre les violations des droits humains commises par des tiers, et notamment par les acteurs de l'IA.

Comme le soulignent les [Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme, l'évaluation de l'impact des activités commerciales sur les droits humains](#), notamment dans la sphère numérique, est une composante essentielle du devoir de vigilance des entreprises. Mais il reste encore beaucoup à faire, dans de nombreux États membres, pour que ces Principes directeurs et la [Recommandation complémentaire CM/Rec\(2016\)3 du Comité des ministres aux États membres sur les droits humains et les entreprises](#) soient effectivement mis en œuvre. Comme le reconnaît le [Groupe de travail des Nations Unies](#) sur la question des droits de l'homme et des sociétés transnationales et autres entreprises, aucune réponse ni réglementation n'est apportée au niveau national face aux

violations des droits humains en lien avec les entreprises qui sont constatées dans le monde. De plus, les informations fournies par les États membres sur la mise en œuvre de ces Principes directeurs et recommandations restent **souvent limitées**. Ces questions doivent être traitées de toute urgence.

Pour prévenir, traiter et réparer les violations des droits humains commises par les acteurs de l'IA dans le secteur privé, les États membres continuent de s'appuyer presque exclusivement sur des cadres législatifs de protection des données. Si ces législations sont un moyen important d'obliger les entreprises à rendre compte des violations du droit à la vie privée et de les sanctionner par des amendes, elles ne sauraient suffire à encadrer les violations des autres droits fondamentaux. Elles ne sont pas non plus adaptées aux situations dans lesquelles des systèmes d'IA représentent des risques pour les droits humains alors qu'ils ne contribuent pas au traitement de données à caractère personnel et ne traitent aucune information identifiante ou identifiable. Les États membres doivent combler ces lacunes de toute urgence lors de la mise en œuvre des normes relatives aux entreprises et aux droits humains dans le contexte de l'IA.

21 des 46 États membres du Conseil de l'Europe ont élaboré des **plans d'action nationaux** sur les entreprises et les droits humains ou PAN. Toutefois, peu de PAN abordent expressément les risques que posent l'IA et les technologies analogues pour les droits humains dans le contexte des entreprises. Certains d'entre eux, dont celui de la Norvège, mentionnent les effets de certaines technologies sur les droits fondamentaux, notamment les **technologies militaires** et **de surveillance** qui sont soumises à des régimes de licences d'exportation renforcés afin d'éviter les abus. D'autres, dont les PAN lituanien, polonais et italien, font référence à la promotion, entre autres, des technologies **renouvelables, respectueuses de l'environnement et écologiquement rationnelles**. De son côté, le PAN de la **Suisse** mentionne la collaboration avec des instances internationales en vue de l'élaboration de « lignes directrices faisant autorité pour l'application des Principes directeurs de l'ONU à des questions clés liées au développement, à l'utilisation et à la gouvernance des technologies numériques ». Enfin, le PAN du **Luxembourg** traite expressément des risques pour les droits humains pouvant naître du développement et de l'utilisation des systèmes d'IA, mais il semble prendre pour cadre juridique la législation relative à la protection des données, même s'il fait aussi référence à d'autres types de conséquences pour les droits humains.

“ Dans tous les débats sur l’IA, la voix des entreprises est exagérément amplifiée. Les stratégies d’IA ayant été conçues sans tenir suffisamment compte des droits humains, elles n’imposent pas d’obligations contraignantes en la matière, mais sont généralement axées sur l’industrie et font éventuellement référence à de vagues notions d’éthique et d’auto-régulation. Or la réglementation ne tuera pas l’industrie ; au contraire, elle stimulera l’innovation.

Comme l’indique la [Déclaration d’Édimbourg](#) de l’Alliance mondiale des institutions nationales des droits de l’homme (GAHNRI), qui précède les Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l’homme, les INDH contribuent de manière essentielle à ce que les acteurs économiques respectent les droits humains, notamment grâce à l’éducation, à la surveillance des atteintes aux droits fondamentaux, au traitement des plaintes et à la médiation. De manière plus générale, les SNDH peuvent apporter aux États membres des conseils sur les lacunes réglementaires qui subsistent dans leur droit national et qui empêchent la création d’un cadre réglementaire holistique régissant les atteintes aux droits humains et à l’égalité causées par les acteurs du secteur privé.

Depuis 2019, les acteurs de l’IA eux-mêmes [prennent les devants et appellent](#) les organes législatifs du monde entier à réglementer leur secteur. Malheureusement, ces appels minimisent le rôle des droits humains dans la réglementation visée et semblent être, au moins en partie, motivés par le désir des acteurs du secteur d’orienter les efforts de réglementation en cours. Il est important que les États membres sensibilisent, mobilisent et consultent les acteurs de l’IA au moment de mettre en œuvre les normes relatives aux entreprises et aux droits humains et de prendre des mesures législatives. Ce faisant, ils doivent veiller à ce que le secteur n’influence pas déraisonnablement les travaux de réglementation en cours et que la protection des droits humains reste au cœur de tout cadre juridique issu d’un tel processus. Une fois les cadres réglementaires en place, il convient d’imposer au secteur privé un devoir de transparence quant à leur application.

Chapitre 3

Information et transparence

Pour sauvegarder efficacement les droits humains, la première chose est de savoir qu'il existe un risque que ces droits soient menacés. Il faut donc que l'utilisation de systèmes d'IA dans tout processus décisionnel ayant une incidence significative sur la vie des personnes soit rendue publique en des termes clairs et accessibles. Cela dit, la transparence véritable ne se limite pas à la communication d'informations sur l'existence ou l'utilisation d'un système d'IA. Il faut aussi expliquer pourquoi la décision de mettre en place un tel système a été prise, quels sont les avantages de l'automatisation, quel est le nombre d'erreurs attendues (faux positifs et faux négatifs) et quels sont les risques possibles pour les droits humains, afin que chacun soit en mesure de comprendre les mécanismes qui régissent le système d'IA, les arbitrages qu'il opère et les processus qu'il met en œuvre pour prendre des décisions et s'assurer de leur exactitude. De même, les processus de contrôle des systèmes d'IA doivent être transparents et accompagnés d'informations accessibles au public.

Certains États membres manifestent la volonté de rendre l'utilisation de l'IA par les autorités publiques plus transparente, s'agissant notamment des procédures d'achat et de la préservation d'un droit de regard du public sur ces dernières.

Ainsi, en janvier 2021, la Chambre des représentants des Pays-Bas a adopté une motion demandant au gouvernement de créer un [registre des algorithmes d'intelligence artificielle](#) utilisés par les organismes publics. Ce registre doit, d'ici à 2023, « décrire quels algorithmes sont utilisés par les administrations publiques, dans quel but, et sur quels jeux de données ils s'appuient, afin que chacun puisse contrôler s'ils sont discriminatoires ou pas. » En novembre 2021, le gouvernement britannique a adopté la [norme de transparence algorithmique](#), qui met le secteur public dans l'obligation de fournir davantage d'informations sur le rôle des algorithmes dans l'appui aux décisions qui concernent des personnes, en particulier dans les services de répression.

[En Grèce, une nouvelle législation](#) impose de multiples obligations de

transparence aux organismes publics, notamment l'obligation de mettre en place un registre des systèmes de prise de décision algorithmique, de respecter le principe de transparence dans l'utilisation des systèmes d'IA et de fournir des informations au public sur l'existence et la méthodologie de ces systèmes. Le respect de ces obligations sera contrôlé par l'Autorité nationale pour la transparence. De leur côté, Malte et le Danemark ont élaboré des programmes de certification en matière d'IA. Le [cadre national maltais de certification des technologies d'IA](#), premier instrument de ce genre au monde lancé en 2019, permet de reconnaître les systèmes d'IA qui ont été développés d'une manière « conforme à l'éthique, transparente et socialement responsable ». Le [D-Seal](#) mis au point par le gouvernement danois permet aux consommateurs de savoir quelles entreprises traitent les données et l'IA de manière « fiable, éthique et sécurisée ».

Les autorités locales aussi ont développé des outils créatifs pour communiquer, de manière plus transparente, la date et le lieu de déploiement des systèmes d'IA. Ainsi, les villes d'[Helsinki](#) en Finlande et d'[Amsterdam](#) aux Pays-Bas ont lancé conjointement leur registre public des technologies d'IA en septembre 2020. D'autres villes néerlandaises ont suivi, notamment Utrecht, La Haye et Rotterdam. Ces initiatives positives constituent des exemples de bonnes pratiques pour tous les États membres.

“ *Les algorithmes sont créés dans des environnements totalement opaques et sans la participation du public. La protection de la propriété intellectuelle prime donc sur le droit à l'information, ce qui pose un grave problème de transparence. L'accès à l'information pour les SNDH doit passer par des obligations de coopération.* ”

Mais il est aussi essentiel que les SNDH soient habilitées à contrôler et à promouvoir la transparence, la responsabilité et la sensibilisation du public, à toutes les étapes des processus, et que les cadres juridiques correspondants exigent des acteurs du secteur privé qu'ils coopèrent avec les organes judiciaires et non judiciaires de manière transparente, y compris avec les SNDH. L'institution de l'ombudsman néerlandais, par exemple, a publié un [rapport](#) qui fournit aux autorités publiques des indications sur la bonne utilisation des données et des algorithmes par les administrations. Ce rapport contient notamment des recommandations concernant la clarté, l'accessibilité et une gestion des problèmes axés sur la recherche de solutions. Dans son [rapport annuel 2022](#), l'ombudsman croate recommande au ministère de l'Économie et du Développement durable de créer un registre des systèmes d'IA utilisés dans le secteur public.

Les droits procéduraux devraient permettre d'améliorer la transparence effective, notamment le droit d'accéder, sans avoir à justifier sa demande, aux informations détenues par les autorités sur les systèmes d'IA utilisés dans la sphère publique. Les éléments d'information sur les systèmes qui revêtent une importance particulière devraient être rendus publics de manière proactive. En outre, les SNDH devraient être habilitées à demander aux organismes publics et privés toute information nécessaire à l'accomplissement de leur mandat, y compris des documents relatifs à des EIDH.

Chapitre 4

Consultations publiques

Les consultations publiques sur la conception, le développement et le déploiement de systèmes d'IA dans l'administration contribuent largement à la transparence, à l'obligation de rendre des comptes et à la participation éclairée du public. La Recommandation de 2019 invite les États membres à mener des consultations publiques au minimum lors de l'acquisition d'un système d'IA et à l'étape de l'EIDH. Lorsqu'elles sont menées correctement, les consultations publiques offrent à toutes les parties prenantes, notamment les acteurs étatiques, les représentants du secteur privé, le monde universitaire et la société civile, l'occasion d'apporter leur contribution.

Des consultations publiques sur l'IA ont été menées dans de nombreux États membres ces dernières années. Ainsi, lors de l'élaboration de sa stratégie nationale en matière d'IA, [l'Irlande](#) a entrepris une vaste consultation auprès de l'administration, des parties prenantes concernées (industrie, experts, universitaires, instituts de recherche) et du grand public. Si aucun État membre ne semble avoir mené de consultations publiques portant spécifiquement sur l'IA et les droits humains ou sur l'acquisition ou le déploiement d'un système d'IA en particulier, on relève plusieurs exemples positifs de prise en compte des droits humains dans certains processus de consultation de portée plus générale. Par exemple, le groupe d'experts Malta.AI chargé d'élaborer la stratégie de Malte en matière d'IA a lancé une consultation publique lors de la rédaction du [cadre maltais pour une IA éthique](#), ensemble de lignes directrices sur l'IA éthique et digne de confiance. L'un des objectifs de ce cadre est que toutes les lois et réglementations en vigueur, les droits humains et les valeurs démocratiques soient pleinement respectés.

“ Chaque mois, de nouveaux logiciels – souvent achetés à des pays tiers – sont mis en place dans les transports, les soins de santé, la sécurité et d’autres secteurs, sans que cela suscite de débats sur les droits humains, car le public ne mesure pas les risques liés à ces changements. Dans de nombreux pays, il n’existe aucune protection juridique.

Le Gouvernement espagnol a mené deux séries de consultations publiques au cours de l’élaboration de la [Charte des droits numériques](#), laquelle contient un article consacré aux « droits relatifs à l’IA ». Cette consultation a permis de recueillir l’avis des citoyens, directement ou par l’intermédiaire d’organisations représentatives, sur les éventuels problèmes, les solutions et les objectifs. Les citoyens ont pu formuler des commentaires sur le projet de charte et sur les règles qui doivent s’appliquer à l’IA.

Les systèmes d’IA étant alimentés par des données susceptibles de contenir des partis pris du passé, ils peuvent renforcer les inégalités et les divisions au sein de la société et accentuer la discrimination ou la marginalisation de certains groupes de population. Les États membres devraient donc accorder une attention particulière aux mécanismes de participation multi-parties prenantes et à la consultation proactive et en temps utile des personnes ou des groupes qui sont les plus concernés par ces systèmes d’IA, notamment les enfants. Les SNDH peuvent servir de pont entre la société civile et les autorités publiques et contribuer à ce que les consultations soient fructueuses, par exemple en identifiant les groupes qui sont touchés de près et en facilitant les contacts avec eux. Pour que les consultations publiques soient facilement accessibles, il convient de simplifier les modalités de contribution non écrite et de permettre la participation dans différentes langues.

Enfin, pour que les consultations publiques soient pleinement pertinentes, les États membres devraient s’employer plus activement à actualiser leurs règles en matière d’accès à l’information et de transparence de l’administration, notamment en ce qui concerne la passation des marchés publics. Le Royaume-Uni a été le premier État membre du Conseil de l’Europe à publier des « [Lignes directrices pour l’acquisition de technologies d’IA](#) ». Ces lignes directrices comprennent une obligation de mener des consultations publiques.

Chapitre 5

Promotion de la connaissance et de la compréhension de l'IA

Pour créer des systèmes de contrôle efficaces et mobiliser le plus grand nombre autour de la question de l'IA et des droits humains, il est indispensable que le public ait été sensibilisé aux systèmes d'IA et qu'il en comprenne les rouages. La Recommandation de 2019 souligne que les États membres devraient encourager la connaissance de l'IA dans les institutions gouvernementales, les organes de contrôle indépendants, les structures nationales des droits humains, le système judiciaire et les services de répression. La connaissance et la compréhension de l'IA par le grand public, notamment de l'impact des systèmes d'IA sur les droits humains, est un autre domaine dans lequel les États membres devraient investir massivement, en lançant des actions de sensibilisation, de formation et d'éducation générales et ciblées, notamment dans les écoles, et en s'engageant auprès des groupes marginalisés. Les SNDH, qui ont une mission de promotion, peuvent contribuer à faciliter la collaboration sur ce plan, non seulement entre les autorités nationales et la société civile, mais aussi entre le niveau national et l'échelon régional.

“ *Les modèles économiques sont très problématiques. Le public ne voit pas qu'ils sont le résultat du développement de l'IA. Pour comprendre comment les jeunes générations – qui sont profondément touchées à tous les niveaux – utilisent les systèmes d'IA, nous devons mieux écouter ce qu'ils disent. Nous devons tous coopérer pour sensibiliser le public.* ”

Dans les États membres, la connaissance et la compréhension de l'IA ont été améliorées grâce à des formations nationales gratuites, des programmes d'étude mis en place dans des établissements d'enseignement ou des formations de service public. Aux Pays-Bas, par exemple, une formation nationale sur « [l'IA et l'éthique](#) » a été lancée en octobre 2022 pour donner

suite à un précédent cours sur les rudiments de l'IA. Cette formation est gratuite et accessible à tous ceux qui disposent d'un accès en ligne. En Irlande, le programme de formation pour les fonctionnaires, lancé en 2021, permet à des agents publics d'obtenir un «[certificat de connaissances élémentaires](#)» en intelligence artificielle. Ce programme comprend un module sur les pratiques éthiques en matière d'IA. Toutefois, jusqu'à présent, ces cours ont principalement porté sur des principes et concepts généraux, au lieu de présenter les cadres juridiques et les normes en matière de droits fondamentaux. À plus forte raison, l'accent n'a pas été mis expressément sur les droits humains. À noter également une initiative bienvenue : la [formation](#) sur l'IA et les discriminations élaborée par le Conseil de l'Europe pour les SNDH. Déjà dispensée au Royaume-Uni et en France, elle est programmée pour l'automne 2023 en Belgique.

En règle générale, les campagnes d'éducation à l'IA qui traitent de la question des droits humains ne font référence qu'à certains de ces droits, par exemple la non-discrimination, le droit à la vie privée ou la liberté d'expression. Il n'existe pas d'approche holistique qui tienne compte des effets – positifs ou négatifs – que l'IA peut avoir sur l'éventail complet des droits fondamentaux. Pourtant, ce besoin se fait urgemment ressentir, car les systèmes d'IA sont de plus en plus utilisés dans tous les aspects de notre vie, et les autorités publiques, comme le grand public, ont besoin de mieux comprendre comment ces systèmes fonctionnent pour gérer leurs effets au niveau individuel comme à l'échelle de la collectivité. Les SNDH, qui ont une connaissance précise de la situation des droits humains et de l'égalité dans les États membres, peuvent jouer un rôle important en fournissant des conseils sur la conception et le lancement de campagnes d'éducation et de sensibilisation générales, mais aussi de campagnes ciblant certaines catégories de personnes, notamment, et sans s'y limiter, les acteurs de l'IA comme les codeurs et les ingénieurs, ou les groupes de population les plus touchés par les éventuels préjudices causés par cette nouvelle technologie.

Chapitre 6

Contrôle indépendant

Pour veiller au bon respect des droits humains, il est essentiel de mettre en place un contrôle des processus d'IA qui soit indépendant et efficace. Mais la question du cadre juridique et la mise en place de structures de contrôle adaptées suscitent toujours de vastes débats en Europe, et pendant ce temps, l'IA continue de progresser à une vitesse exponentielle. Il est donc plus nécessaire que jamais d'examiner et de suivre de manière proactive les effets de cette évolution technologique sur les droits humains.

La législation sur l'intelligence artificielle proposée par l'Union européenne (AI Act) prévoit la création d'autorités de contrôle nationales chargées de sa mise en œuvre. Ce règlement devrait laisser une certaine marge de manœuvre aux États, qui pourraient décider de confier ces nouveaux pouvoirs à une instance créée à cet effet ou à une autorité de contrôle existante. Aussi, une fois cette législation entrée en vigueur, différentes approches pourraient coexister dans l'Union quant au contrôle de sa mise en œuvre. Quelle que soit la décision prise par les États membres à cet égard, il est essentiel que les SNDH soient étroitement consultés et associées aux décisions et que la coopération existante entre les SNDH et les autres parties prenantes, y compris les autorités de réglementation, soit institutionnalisée dans tous les processus de contrôle nationaux et régionaux, afin de permettre une approche véritablement pluridisciplinaire qui tienne pleinement compte des droits humains.

On ne sait pas encore si la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit contiendra une disposition imposant aux États parties de mettre en place des autorités de contrôle nationales chargées de surveiller sa bonne application. Comme le souligne la Recommandation de 2019, il est essentiel de prévoir un mécanisme permettant de contrôler, de manière indépendante et effective, que les systèmes d'IA sont conformes aux droits humains tout au long de leur cycle de vie, afin que les États membres du Conseil de l'Europe s'acquittent de leurs obligations. Les autorités de contrôle devraient jouir d'une indépendance formelle et fonctionnelle totale et se voir dotées de larges attributions, notamment en ce qui concerne

les pouvoirs d'investigation, le traitement des plaintes, l'établissement de rapports et les actions de sensibilisation. Elles devraient être habilitées à suspendre le déploiement ou l'utilisation d'un processus d'IA en cas de violation avérée des droits humains.

“ *Nous avons besoin d'instaurer une supervision et un contrôle humain beaucoup plus holistiques, s'agissant notamment de la manière dont les EIDH sont réalisées, mais aussi des risques qui pèsent sur l'État de droit et la démocratie.* ”

Certains États ont pris très tôt des mesures pour mettre en place des autorités chargées de contrôler les technologies d'IA. Ainsi, l'Espagne est l'un des premiers États en Europe à avoir amorcé la création d'une instance séparée, l'[Agence espagnole de contrôle de l'intelligence artificielle](#) (AESIA), qui devrait être opérationnelle fin 2023. Cette agence sera chargée de surveiller la création et l'utilisation des systèmes d'IA, en particulier ceux qui risquent de porter atteinte aux droits fondamentaux. Elle prendra en outre des mesures pour réduire les risques auxquels ces droits pourraient être exposés. Une coordination étroite avec l'ombudsman et les autres institutions concernées sera un élément important de ce dispositif.

D'autres États membres ont aussi pris des mesures pour instaurer des mécanismes de contrôle de l'IA, mais ces mécanismes ne se réfèrent pas aux normes consacrées en matière de droits humains, s'appuyant souvent sur des cadres éthiques pour contrôler et évaluer la conformité dans des secteurs particuliers. Ainsi, en 2019, le Premier ministre français a mis en place un [comité pilote d'éthique du numérique](#), qui est chargé d'examiner comment l'IA est utilisée dans les agents conversationnels, dans les véhicules autonomes et pour les diagnostics médicaux. De son côté, le Gouvernement finlandais a créé un [comité d'éthique de l'IA](#) pour améliorer la compréhension des principes éthiques et veiller à ce que le développement de l'IA en Finlande soit « axé sur l'humain et fondé sur la confiance ». Comme indiqué précédemment, les cadres éthiques, s'ils jouent un rôle important dans la gestion et la gouvernance de l'IA au sens large, ne devraient pas être pris comme référence au détriment des normes en matière de droits humains, droits que les États membres sont légalement tenus de sauvegarder.

On observe en Europe une tendance à considérer que les autorités chargées de la protection des données peuvent être de bons mécanismes de contrôle en matière d'intelligence artificielle. S'il est vrai que ces organismes sont dotés de larges attributions et habilités à infliger des amendes en cas de non-

respect des garanties en matière de protection des données, ils disposent rarement des connaissances approfondies, de l'expertise technique et du mandat nécessaires pour contrôler, mener des investigations et traiter efficacement les plaintes qui concernent des atteintes causées par des systèmes d'IA aux droits humains de manière plus générale.

Les autorités de contrôle de l'IA doivent être dotées de ressources suffisantes et d'une expertise et de compétences interdisciplinaires appropriées pour faire face à la diversité des systèmes d'IA et à la complexité des formes d'atteinte aux droits humains causées par ces systèmes. En outre, ces autorités devraient être expressément chargées d'analyser et de contrôler ce que font les autorités publiques et les acteurs du secteur privé, afin que les risques pour les droits humains soient mis en évidence dès que possible et que les atteintes à ces droits soient évitées. Cette approche permettrait de limiter les recours et les demandes d'indemnisation. La société civile et les SNDH devraient être étroitement associées à la mise en place et au fonctionnement de ces mécanismes de contrôle, afin de permettre une transparence totale et une pleine mise en œuvre des responsabilités. Les SNDH peuvent fournir aux États membres des éléments d'orientation afin que les mécanismes mis en place pour surveiller les systèmes d'IA – que ce soit par la création d'organismes distincts, l'intégration dans des organismes existants ou la création de mécanismes multi-institutionnels bien coordonnés – disposent d'un mandat et de pouvoirs suffisants pour faire face aux diverses formes d'atteintes aux droits humains qui peuvent être causées par les systèmes d'IA. Quel que soit le modèle de contrôle choisi, il est essentiel que ces mécanismes soient dotés de ressources suffisantes pour fonctionner de manière efficace.

Chapitre 7

Voies de recours effectives

Les États membres devraient, avant toute chose, chercher à prévenir les atteintes aux droits humains en agissant de manière préventive, notamment en mettant en place des cadres juridiques bien conçus et des mécanismes de contrôle efficaces, plutôt qu'en adoptant des approches de type « tester puis remédier ». Cela étant, il faut aussi prévoir des voies de recours effectives pour les personnes dont les droits fondamentaux ont été bafoués lors de la conception, du développement ou du déploiement d'un système d'IA.

Comme on l'a dit plus haut, depuis 2019, la législation a évolué lentement par rapport aux avancées technologiques. Quelques affaires ayant trait aux incidences des systèmes d'IA sur les droits humains ont néanmoins été portées devant des organes judiciaires et non judiciaires. Si ces affaires mettent en évidence le besoin de disposer de voies de recours et de réparation en cas d'atteinte aux droits humains liée aux technologies d'IA, elles montrent aussi que la législation actuelle est inadaptée. Le petit nombre d'affaires portées devant les tribunaux – dont la plupart s'appuient sur des réglementations en matière de protection des données – semble indiquer que les atteintes aux droits humains causées par l'utilisation de systèmes d'IA sont insuffisamment signalées, d'où la nécessité d'un meilleur accès à l'information, aux connaissances et à l'expertise pour détecter efficacement ces préjudices et y répondre.

En 2020, le tribunal de district de La Haye a annulé une loi autorisant le gouvernement à utiliser un modèle de risque pour identifier les personnes soupçonnées d'avoir commis divers types de fraude, notamment des fraudes aux prestations sociales. Le tribunal a estimé que cette loi constituait une violation du droit au respect de la vie privée, relevant par ailleurs ses effets discriminatoires, car certains groupes de bénéficiaires faisaient l'objet d'un profilage automatique qui les associait à un risque de fraude plus élevé. Les personnes concernées recevaient par erreur des demandes de remboursements importants, ce qui ruinait leur famille et les conduisait souvent à la dépression et au désespoir.

La même année, la [Cour d'appel d'Angleterre et du Pays de Galles](#) a jugé que la base juridique de l'utilisation d'une technologie de reconnaissance faciale automatique par la police était insuffisante, car elle laissait une trop grande marge de manœuvre aux policiers. Elle a en outre estimé que la police n'avait pas procédé à des évaluations adéquates de l'impact de cette technologie sur la protection des données et l'égalité. Toujours en 2020, [un tribunal de Bologne](#), en Italie, a jugé que le système algorithmique utilisé par une entreprise de livraison alimentaire pour fixer les conditions d'accès au travail des livreurs à deux-roues s'analysait en une discrimination indirecte. En 2021, la [Cour administrative suprême de Pologne](#) a jugé que l'algorithme du système d'attribution aléatoire et automatique des affaires aux juges pouvait être rendu public, en vertu de la législation sur la liberté de l'information. La même année, la [Cour suprême italienne](#) a demandé que dans le contexte des systèmes d'évaluation de la réputation fondés sur l'IA, le consentement éclairé des personnes concernées soit renforcé. En 2022, le tribunal de district d'Amsterdam et la [magistrates' court de La City de Londres](#) ont ordonné que des chauffeurs victimes d'un « licenciement par algorithme » soient réintégrés dans la société de VTC qui les employait et qu'une indemnité leur soit versée.

Des situations analogues peuvent être observées dans des contextes non judiciaires. En 2017, dans l'une des premières affaires importantes, l'ombudsman finlandais pour la non-discrimination a décidé de traduire en justice, devant le tribunal national pour la non-discrimination et l'égalité, une banque qui avait eu recours à la prise de décision automatisée pour octroyer des prêts. Le tribunal [a conclu](#) que la pratique était discriminatoire pour plusieurs raisons et a imposé une amende substantielle à la partie reconnue coupable. En mai 2022, [l'Autorité britannique de protection des données a infligé](#) une amende de 7,5 millions de livres sterling à une société spécialisée dans la reconnaissance faciale qui avait utilisé des images de personnes au Royaume-Uni, et lui a enjoint de supprimer de ses systèmes toutes les données appartenant à des résidents britanniques. [L'Autorité néerlandaise de protection des données a infligé](#) son amende la plus élevée (3,7 millions d'euros) à l'administration fiscale néerlandaise pour avoir traité illégalement des données à caractère personnel pendant six années, dans le cadre de son « dispositif de détection des fraudes » fondé sur des algorithmes. Cette amende s'ajoute à celle de 2,75 millions d'euros infligée au fisc néerlandais en 2021, pendant le scandale des allocations familiales, pour utilisation d'un algorithme discriminatoire.

Le droit à un recours effectif comprend le droit à une réparation et un redressement rapides et suffisants. Pourtant, dans les cas précités, plusieurs années se sont écoulées avant qu'une décision de première instance soit rendue. L'ONG [noyb](#), spécialisée dans le droit en matière de données, [a](#)

signalé que les plaintes relatives à la protection des données qu'elle a déposées à l'encontre d'un certain nombre de grandes plates-formes en ligne utilisant l'IA dans leurs modèles économiques ont été examinées par les autorités nationales de protection des données pendant plus de trois ans sans qu'aucune décision définitive n'ait été rendue. Il faut donc s'employer plus activement à accélérer les mécanismes de réparation et à améliorer leur efficacité.

“ Nous n'avons pas de cadre général offrant des voies de recours effectives, mais seulement des réponses disparates. Dans les cas où une discrimination algorithmique est alléguée, la charge de la preuve ne devrait pas incomber au plaignant, mais à l'utilisateur de l'algorithme.

Pour bien comprendre comment les systèmes d'IA peuvent porter atteinte aux droits humains, il est indispensable que les organes judiciaires, les organes non judiciaires ainsi que les personnes qui portent plainte puissent accéder aux informations nécessaires à l'évaluation correcte des griefs relatifs à l'IA. Or, on l'a vu plus haut, il a été démontré que les acteurs de l'IA protègent et retiennent excessivement ces informations, ce qui contrarie les mécanismes d'accès aux voies de recours. Ainsi, dans l'affaire susmentionnée mettant en cause l'algorithme de l'entreprise de livraison alimentaire, le tribunal de Bologne a dénoncé l'impossibilité d'obtenir des informations complémentaires sur le fonctionnement du modèle statistique utilisé. Il a fait valoir que l'entreprise n'avait pas publié les règles de fonctionnement de son algorithme ni les critères permettant de calculer les statistiques de chaque livreur, ce qui l'empêchait d'examiner l'affaire de manière plus approfondie.

Le droit à un recours effectif est une composante essentielle et incontournable du droit relatif aux droits humains. Les États membres doivent donc veiller à ce que des voies de recours effectives soient disponibles et accessibles à toute personne se déclarant victime d'une violation des droits humains causée par un système d'IA, et accorder une attention spéciale aux groupes particulièrement vulnérables, notamment aux enfants. Comme le montrent les cas présentés dans ce chapitre, les recours peuvent varier en nature. Cela dit, ils devraient tous permettre de remédier directement, en théorie et dans les faits, à la situation contestée résultant de l'utilisation d'un système d'IA. Étant donné que ces systèmes empiètent en général sur les droits d'un grand nombre de personnes, les États membres devraient envisager d'adopter des cadres juridiques pour régir les recours collectifs.

Observations finales et recommandations

La technologie fait partie du monde dans lequel nous vivons et elle est présente dans tous les aspects de notre vie, ou presque. Ce phénomène va encore s'amplifier à l'avenir, pour autant qu'on puisse le prévoir, l'intelligence artificielle (IA) devenant de plus en plus complexe et sophistiquée. La conception, le développement et le déploiement des systèmes d'IA peuvent avoir des répercussions importantes sur nos droits humains et nos cadres de vie. Les États membres doivent donc être très attentifs à la préservation effective de nos droits tout au long du cycle de vie de ces systèmes. Les SNDH, compte tenu de leur rôle majeur dans la surveillance et l'application des droits consacrés par la Convention européenne des droits de l'homme et d'autres instruments relatifs aux droits humains, contribuent de façon essentielle à ce que cette responsabilité des États soit pleinement assumée.

En tant que sociétés, nous nous employons sans relâche à protéger et à sauvegarder les droits et les intérêts des êtres humains dans des domaines que beaucoup d'entre nous jugent pourtant complexes. Ainsi, nous ne sommes pas tous ingénieurs ou concepteurs de voitures, mais nous comprenons et nous respectons les exigences de fabrication et le code de la route, afin de protéger la vie humaine et les droits individuels. De même, nous ne sommes pas tous économistes, mais nous parvenons à nous mettre d'accord sur la réglementation de nos marchés financiers. Le même principe s'applique à l'IA : nul besoin que chacun d'entre nous comprenne le fonctionnement détaillé de cette technologie pour que soient mis en place des systèmes et des processus qui protègent efficacement les droits humains, pendant son développement et lors de son utilisation. Ce n'est pas parce qu'une décision administrative peut être formulée ou prise au moyen d'un processus algorithmique opaque que l'État est libéré de son obligation légale de veiller à ce que toutes les décisions soient fondées sur des critères légaux et qu'elles puissent être réexaminées.

L'IA serait si technique et impénétrable qu'elle échapperait au contrôle humain et à toute réglementation efficace dans le respect des droits fondamentaux, du moins sans qu'il y ait un risque pour la croissance économique et la prospérité. Cette idée continuellement reprise domine le

débat et est souvent entretenue par le secteur privé lui-même, qui privilégie souvent la maximisation des profits au détriment des préoccupations du public. Malheureusement, du fait de cette conception erronée de la réalité, les hauts responsables de l'action publique hésitent, cela est frappant, à s'investir pleinement dans la lutte contre les risques d'atteintes aux droits humains causées par le développement et le déploiement croissants de l'IA, ce qui freine l'application effective des normes juridiques existantes et la création de mécanismes adaptés pour atténuer ces menaces. Comme l'illustre l'utilisation continue de [logiciels espions très intrusifs](#), les États membres avancent encore trop lentement dans la mise en place d'une réponse efficace face aux risques que présente l'IA pour les droits fondamentaux, en particulier lorsqu'elle est développée et déployée par de puissants acteurs du secteur privé.

Dans la Recommandation de 2019, la Commissaire donne des orientations pratiques sur la manière dont les États membres pourraient protéger et promouvoir les droits humains, dans le contexte de la conception, du développement et du déploiement des systèmes d'IA. Quatre années se sont écoulées depuis. L'Europe a été confrontée à de multiples crises des droits humains, qui sont interdépendantes et qui posent aux gouvernements comme aux populations des défis importants. Pendant ces crises, la dépendance à l'égard des technologies numériques et de l'automatisation renforcée par les données n'a fait qu'augmenter et le secteur des technologies n'a cessé de croître.

Les États membres mènent actuellement une réflexion sur la réglementation de l'IA au niveau des politiques publiques et prennent des mesures en vue de l'adoption de cadres législatifs ou de la mise en place de mécanismes de contrôle des systèmes d'IA. Des organes judiciaires et non judiciaires examinent dans quelle mesure les cadres juridiques existants s'appliquent aux situations mettant en jeu des systèmes d'IA et étudient les plaintes dont ils ont été saisis. Dans le contexte des négociations en cours au niveau régional en vue de l'adoption de nouveaux cadres juridiques relatifs aux systèmes d'IA et à leur impact sur les droits humains, il est essentiel d'examiner dans quelle mesure les normes et garanties existantes en matière de droits fondamentaux, notamment celles contenues dans la Recommandation de 2019, sont actuellement mises en œuvre. La Commissaire, sur la base de ses constats et en concertation avec les SNDH, a relevé trois tendances interdépendantes qui constituent des obstacles à la pleine mise en œuvre des normes internationales en matière de droits humains dans le contexte de l'IA en Europe. Pour y répondre, elle adresse aux États membres du Conseil de l'Europe les recommandations suivantes, qui visent à combler les lacunes encore présentes et à renforcer

l'application des obligations en matière de droits humains dans le contexte de la conception, du développement et du déploiement des systèmes d'IA.

Absence d'approches globales fondées sur les droits humains

D'une manière générale, pour mettre en œuvre la Recommandation de 2019 et d'autres normes internationales en matière de droits humains dans le contexte de l'utilisation de l'IA, les États membres ont adopté des approches sectorielles, au lieu de veiller à l'application cohérente et générale de ces textes dans tous les secteurs concernés et par l'ensemble des acteurs. De plus, les cadres juridiques, lorsqu'ils existent, ont rarement été appliqués de manière efficace et rapide, du fait de la dépendance des infrastructures à l'égard des grandes plates-formes, qui a pour effet de ralentir l'application de la réglementation, et en raison du contrôle qui reste parcellaire. Les États membres ont tendance, lorsqu'ils évaluent l'impact des systèmes d'IA, à privilégier certains droits (droits relatifs à la protection des données ou principes de non-discrimination par exemple), ou à se référer, de manière évasive, à des cadres éthiques qui n'intègrent pas l'ensemble des normes et obligations en matière de droits humains. Ainsi a-t-il été demandé aux autorités chargées de la protection des données d'assurer un contrôle indépendant des systèmes d'IA et de leur impact sur les droits humains, ce qui a eu pour effet d'orienter les décideurs vers des problématiques de protection des données et d'empêcher une compréhension plus large des risques et des conséquences pour les droits fondamentaux. Du fait de ces approches sectorielles de la réglementation, les acteurs judiciaires et non judiciaires chargés d'examiner les plaintes relatives aux atteintes aux droits humains causées par les systèmes d'IA et d'y remédier ont, eux aussi, appliqué des cadres juridiques de type sectoriel.

- » Les États membres devraient réévaluer leurs cadres juridiques, notamment ceux qui concernent les voies de recours et le contrôle, pour qu'ils tiennent pleinement compte de la portée et des particularités des plaintes relatives aux violations des droits humains par des systèmes d'IA. Les nouveaux cadres juridiques relatifs à l'IA devraient être en phase avec les garanties existantes en matière de droits humains, telles que le renversement de la charge de la preuve en cas de plainte pour discrimination. Pour être efficaces et utiles, tous les cadres juridiques doivent être rigoureusement mis en œuvre et appliqués dans tous les cas de figure.
- » Les cadres juridiques devraient être applicables à tous les systèmes d'IA susceptibles de porter atteinte aux droits humains, quel que soit le secteur dans lequel il est prévu d'exploiter ces systèmes, et tenir pleinement compte du fait que les droits humains doivent être protégés à toutes les étapes du cycle de vie de ces systèmes et en

permanence lorsqu'ils interagissent avec d'autres systèmes.

- » Les États membres devraient, en étroite liaison avec les SNDH, les détenteurs de droits et la société civile, élaborer une procédure spécifique et indépendante, à l'usage des autorités publiques, pour réaliser une EIDH en lien avec des systèmes d'IA. Ces EIDH devraient être obligatoires et réalisées a priori et a posteriori, et leurs processus et résultats devraient être accessibles au public. Les autorités publiques ne devraient pas acquérir ou utiliser des systèmes d'IA provenant de tiers lorsqu'elles sont dans l'incapacité de réaliser une EIDH effective en raison d'un manque d'informations sur ces systèmes. Tous les obstacles à la réalisation d'EIDH efficaces doivent être levés de manière proactive.
- » Les États membres devraient veiller à ce que la conception, le développement et le déploiement de l'IA soient supervisés par une ou plusieurs autorités totalement indépendantes, dotées de ressources suffisantes et mandatées pour contrôler, mener des investigations et traiter les plaintes relatives à des mesures prises par les autorités publiques et les acteurs du secteur privé. Dans leur fonction de contrôle, les autorités de surveillance devraient être tenues de coopérer étroitement avec d'autres institutions indépendantes, notamment les SNDH.
- » Les organes judiciaires et non judiciaires, notamment les organismes de réglementation et les SNDH, devraient disposer de ressources suffisantes et être en mesure d'intervenir efficacement et rapidement lorsque la question du respect d'un droit humain, quel qu'il soit, est soulevée au sujet d'un système d'IA, et sous un angle holistique et intersectionnel des droits humains.

Manque de transparence et de partage de l'information

Malgré les initiatives notables qui ont été prises pour améliorer la transparence autour des systèmes d'IA dans certains États membres, il est encore difficile d'avoir accès, partout en Europe, à des informations claires et à jour sur l'utilisation de ces systèmes et leur impact potentiel sur les droits humains. Les mesures mises en place pour protéger la propriété intellectuelle et les secrets de fabrication font toujours obstacle à l'application des droits à l'information, ceux notamment du pouvoir judiciaire, des SNDH et des autorités de régulation; or ces organismes ont besoin d'un accès total et effectif à toutes les informations – jeux de données, codes sources, etc. – qui sont nécessaires à leur mission de contrôle indépendant et à la détection et à l'investigation d'éventuelles atteintes aux droits humains. Les consultations publiques – lorsqu'il y en a eu – ont peu souvent abordé la question générale des répercussions des systèmes

d'IA sur les droits fondamentaux, et rares ont été les démarches proactives visant à prendre en compte les points de vue des groupes vulnérables tels que les communautés marginalisées et les enfants. Les États membres n'ont pas encore dispensé de programmes d'éducation solides et structurés qui mettent expressément l'accent sur les répercussions des systèmes d'IA sur les droits humains et soient réellement accessibles à toutes les couches de la société.

- » Les États membres devraient encourager les initiatives qui, comme la création de registres publics, contribuent à promouvoir la transparence sur l'utilisation des systèmes d'IA dans la vie quotidienne et sur les possibles effets préjudiciables de ces systèmes sur les droits humains, notamment la transparence quant à leurs mécanismes de base et aux objectifs qui les sous-tendent. Les autorités de surveillance nouvellement créées ou nommées devraient régulièrement rendre compte de leurs activités au public, dans des formats facilement accessibles et compréhensibles.
- » Les organes judiciaires et non judiciaires, y compris les autorités de régulation et les SNDH, devraient être habilités à exiger des acteurs de l'IA qu'ils communiquent des informations détaillées sur les systèmes d'IA à l'étude et à prendre des mesures à l'encontre des organismes publics ou privés qui ne respectent pas la réglementation.
- » Lors de l'élaboration, de l'adoption et de la mise en œuvre des cadres juridiques relatifs aux répercussions des systèmes d'IA sur les droits humains, les États membres devraient organiser des consultations publiques, régulièrement et en temps utile, non seulement avec les experts, le secteur industriel, les chercheurs, les SNDH, la société civile et le monde universitaire, mais aussi, et cela est important, avec le grand public et les représentants des groupes de population les plus touchés par ces systèmes, notamment les enfants et leurs représentants.
- » Pour informer ces parties prenantes de l'existence de consultations publiques et de la possibilité qui leur est offerte d'y participer, les États membres devraient mener des actions de sensibilisation ciblées en recourant à des méthodes de participation multi-parties prenantes, et faciliter l'apport de contributions sous divers formats.
- » Les États membres devraient prendre immédiatement des mesures pour concevoir, financer et mettre en œuvre des campagnes globales de sensibilisation à l'IA et de connaissance et compréhension de l'IA, afin que les personnes concernées par l'utilisation et le contrôle des systèmes d'IA et celles qui sont touchées par ces systèmes aient une bonne compréhension des systèmes d'IA et de leurs multiples effets sur les droits humains. Les États membres devraient étroitement associer les SNDH à ces processus et leur fournir les ressources nécessaires pour renforcer effectivement leurs capacités dans le domaine de l'IA.

Initiatives visant à utiliser l'IA pour renforcer les droits humains

Étant donné que le marché de l'IA, avec tout le potentiel qu'il offre et les risques qu'il présente, est très largement dominé par des acteurs du secteur privé et que le développement de l'IA est pour l'essentiel tiré par des considérations commerciales, les pouvoirs publics ont, dans l'ensemble, adopté une approche réactive plutôt que proactive. Le secteur privé a donc pu orienter l'évolution de l'IA en s'appuyant sur des modèles économiques qui considèrent les individus comme des consommateurs et des cibles potentielles de développement. En retardant la mise en place d'une réglementation favorisant l'émergence de solutions différentes et innovantes, les États membres risquent de passer à côté de l'opportunité qu'offre l'IA de mettre en œuvre et de renforcer les protections des droits humains et les principes fondamentaux de la démocratie et de l'État de droit, la promotion de l'égalité et la bonne gouvernance. Malgré un certain nombre d'initiatives encourageantes, notamment en faveur de la [promotion de l'égalité et de la protection des données dès la conception](#), il reste encore des possibilités inexploitées de concevoir, développer et déployer des technologies d'IA en poursuivant des objectifs fondés sur des valeurs, par exemple pour détecter et démanteler les préjugés et les stéréotypes, encourager la participation publique, donner plus de place à la parole de ceux que l'on entend peu et lutter contre les inégalités en donnant la priorité aux plus démunis. Enfin, pour que le développement de l'IA puisse profiter à tous, il faut surmonter les fractures numériques et donner à tous les citoyens, dans toutes les couches de la société, un accès sûr et fiable à l'internet.

- » Les États membres, les autorités publiques et les autres acteurs de l'IA devraient expressément se référer au droit et aux obligations juridiques en matière de droits humains lorsqu'ils élaborent, adoptent ou mettent en œuvre des mesures réglementaires concernant les systèmes d'IA, et inciter le secteur privé à orienter les innovations vers des systèmes conformes aux droits humains. Lorsque des [bacs à sable réglementaires](#) sont adoptés, les SNDH devraient être consultées et avoir la possibilité de fournir aux développeurs d'IA des orientations utiles en matière de droits humains et de promouvoir l'intégration de garanties en matière de droits humains tout au long du cycle de vie de l'IA.
- » Les systèmes d'IA doivent être soumis à une analyse complète et minutieuse du respect des droits humains, à un contrôle ainsi qu'à des cadres de référence en matière de responsabilité du fait des produits, et ce grâce à une coopération institutionnalisée entre les différents organismes non judiciaires qui participent à leur supervision, notamment les SNDH en leur qualité de principaux gardiens des droits humains.

- » Les organes judiciaires et non judiciaires, notamment les autorités de régulation et les SNDH, devraient être habilités à empêcher toute nouvelle atteinte aux droits humains par des systèmes d'IA, notamment grâce à l'imposition de moratoires sur la poursuite de l'utilisation de tels systèmes, la suppression d'ensembles de données ou la destruction de systèmes qui sont le résultat d'un traitement de données illégal.
- » Les États membres devraient, de manière proactive, entamer un dialogue avec les développeurs d'IA, la société civile et des experts indépendants pour réfléchir aux nouvelles possibilités qu'offrent les technologies d'IA de renforcer les protections des droits humains et de promouvoir l'égalité effective. Les SNDH devraient être étroitement associés à ces processus. Un financement ciblé devrait être alloué à des travaux de recherche sur le potentiel des systèmes d'IA à promouvoir les droits humains et l'égalité.
- » Les États membres devraient prendre d'urgence des mesures pour mettre en œuvre les [Principes directeurs de l'ONU relatifs aux entreprises et aux droits de l'homme](#) et la [Recommandation CM/Rec\(2016\)3 du Comité des Ministres aux États membres sur les droits de l'homme et les entreprises](#).

Le Commissaire aux droits de l'homme est une institution non judiciaire, indépendante et impartiale, créée en 1999 par le Conseil de l'Europe. Sa mission est de promouvoir la sensibilisation aux droits de l'homme et leur respect dans les États membres. Ses activités s'articulent autour de trois grands axes étroitement liés :

- des visites dans les pays et un dialogue avec les autorités nationales et la société civile,
- un travail thématique et de conseil sur la mise en oeuvre systématique des droits de l'homme, et
- des activités de sensibilisation.

La Commissaire actuelle, Dunja Mijatović, a pris ses fonctions en avril 2018. Elle a succédé à ce poste à Nils Muižnieks (2006-2012), Thomas Hammarberg (2006-2012) et Álvaro Gil-Robles (1999-2006).



www.commissioner.coe.int

FRA

www.coe.int

Le Conseil de l'Europe est la principale organisation de défense des droits de l'homme du continent. Il comprend 46 États membres, dont l'ensemble des membres de l'Union européenne. Tous les États membres du Conseil de l'Europe ont signé la Convention européenne des droits de l'homme, un traité visant à protéger les droits de l'homme, la démocratie et l'État de droit. La Cour européenne des droits de l'homme contrôle la mise en oeuvre de la Convention dans les États membres.



COMMISSIONER
FOR HUMAN RIGHTS

COMMISSAIRE AUX
DROITS DE L'HOMME

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE