

Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme



Recommandation

Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme

Par la Commissaire
aux droits de l'homme
du Conseil de l'Europe

Photo de couverture :
© Shutterstock.com

© Conseil de l'Europe, mai 2019
Imprimé au Conseil de l'Europe

Remerciements:

La Commissaire tient à exprimer sa reconnaissance à Nani Jansen Reventlow, avocate spécialiste des droits de l'homme et directrice du Digital Freedom Fund, et à Jonathan McCully, consultant indépendant et conseiller juridique auprès du Digital Freedom Fund, pour leur précieuse contribution à l'élaboration de la présente recommandation.

Contenus

INTRODUCTION	5
RECOMMANDATIONS	7
1 - Évaluation de l'impact sur les droits de l'homme	7
2 - Consultations publiques	8
3 - Obligation, pour les États membres, de faciliter la mise en œuvre des normes des droits de l'homme dans le secteur privé	9
4 - Information et transparence	10
5 - Contrôle indépendant	10
6 - Non-discrimination et égalité	11
7 - Protection des données et respect de la vie privée	12
8 - Liberté d'expression, liberté de réunion et d'association, et droit au travail	13
9 - Possibilités de recours	14
10 - Promotion de la connaissance et de la compréhension de l'intelligence artificielle	15
CHECKLISTS	17
ANNEXE DÉFINITIONS	26

Introduction

Les effets de l'intelligence artificielle (IA) sur les droits de l'homme constituent l'un des facteurs les plus importants qui définiront la période dans laquelle nous vivons. Les technologies utilisant l'IA sont de plus en plus présentes dans la vie de chacun, par le biais de la domotique ou des médias sociaux, par exemple. Elles sont aussi de plus en plus utilisées par les pouvoirs publics pour évaluer la personnalité ou les compétences des individus, pour allouer des ressources et pour prendre d'autres décisions qui peuvent avoir des conséquences graves et concrètes pour les droits de l'homme. Comme l'a souligné la Commissaire aux droits de l'homme dans un article du Carnet des droits de l'homme, il est donc urgent de trouver le juste équilibre entre les progrès technologiques et la protection des droits de l'homme.

Conformément à son mandat, qui consiste notamment à promouvoir la sensibilisation aux droits de l'homme, le respect effectif et la pleine jouissance de ces droits dans les États membres du Conseil de l'Europe ainsi qu'à fournir des conseils et toute information concernant la protection de ces droits (articles 3 et 8 de la Résolution (99) 50 du Comité des Ministres), la Commissaire aux droits de l'homme publie la présente recommandation sur l'IA et les droits de l'homme, qui met en avant 10 séries de mesures.

Il n'existe pas à l'heure actuelle de définition communément admise de l'« intelligence artificielle ». Toutefois, aux fins de la présente recommandation, l'IA est une expression générique utilisée pour désigner de façon générale un ensemble de sciences, de théories et de techniques dont le but est d'améliorer la capacité des machines à réaliser des tâches requérant des facultés cognitives. Un système d'IA est un système informatique qui formule des recommandations, établit des prévisions ou prend des décisions en fonction d'une série donnée d'objectifs. Pour cela, il (i) recourt à des données informatiques et/ou humaines pour percevoir des environnements réels et/ou virtuels ; (ii) convertit manuellement ou automatiquement ces perceptions en modèles ; et (iii) tire des conclusions à partir de ces modèles, par des moyens humains ou automatisés, sous la forme de recommandations, de prévisions ou de décisions.

Le glossaire en annexe définit d'autres termes pertinents utilisés dans la présente recommandation.

L'IA offre de nouvelles possibilités mais implique aussi des risques ; les droits de l'homme ne devraient pas être fragilisés mais renforcés par l'IA. La présente recommandation sur l'IA et les droits de l'homme donne des orientations sur la manière dont les effets négatifs des systèmes d'IA sur les droits de l'homme peuvent être évités ou atténués, en distinguant 10 grands domaines d'action.

La recommandation se fonde sur les travaux effectués dans ce domaine par le Conseil de l'Europe, en particulier la Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires, les Lignes directrices sur l'intelligence artificielle et la protection des données, la Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques et l'Étude sur les dimensions des droits humains dans les techniques de traitement automatisé des données et les éventuelles implications réglementaires, et sur le rapport dans lequel le Rapporteur spécial des Nations Unies sur la promotion et la protection du droit à la liberté d'opinion et d'expression examine les incidences des technologies d'intelligence artificielle sur les droits de l'homme dans le cyberspace. La recommandation repose sur le système international des droits de l'homme, qui constitue un cadre universel et contraignant, et notamment sur les instruments du Conseil de l'Europe relatifs aux droits de l'homme.

La recommandation s'adresse aux États membres, mais les principes énoncés concernent quiconque a une influence importante – directe ou indirecte – sur le développement, la mise en œuvre ou les effets d'un système d'IA. L'IA mise au point dans le secteur privé devrait être assujettie aux mêmes normes que l'IA mise au point dans le secteur public dès lors qu'il existe une intention de travailler avec des organismes ou services publics.

La liste qui figure en annexe recense les actions à mener dans chaque domaine clé et donne des orientations générales pour la mise en œuvre des recommandations.

Recommandations

1 - Évaluation de l'impact sur les droits de l'homme

Les États membres devraient établir un cadre juridique qui prévoit une procédure à suivre par les autorités publiques pour évaluer l'impact, sur les droits de l'homme, des systèmes d'IA qu'elles ont acquis, développés et/ou déployés. Les études d'impact sur les droits de l'homme (EIDH) devraient être mises en œuvre dans le même esprit que d'autres formes d'évaluation réalisées par les autorités publiques, telles que les évaluations de l'impact de la réglementation et les études d'impact relatives à la protection des données.

Les États membres peuvent définir les types de système d'IA soumis à des EIDH en vertu de la loi, mais de telles définitions doivent être suffisamment générales pour couvrir tous les systèmes d'IA pouvant interférer avec les droits de l'homme à un stade quelconque du cycle de vie du système.

Le cadre juridique des EIDH devrait imposer aux autorités publiques l'obligation de réaliser une auto-évaluation des systèmes d'IA existants et des systèmes envisagés. Ce processus devrait permettre d'évaluer les éventuels effets du système d'IA sur les droits de l'homme en tenant compte de la nature, du contexte, de la portée et de l'objectif du système. Lorsqu'une autorité publique n'a pas encore acquis ou développé un système d'IA dont la création est envisagée, cette évaluation doit être menée avant l'acquisition et/ou le développement du système.

L'EIDH doit également comporter un examen externe rigoureux du système d'IA, mené par un organe de contrôle indépendant ou par un chercheur/auditeur externe doté de l'expertise nécessaire, afin d'aider à détecter, à mesurer et/ou à cartographier les incidences et les risques en matière de droits de l'homme au fil du temps. Les organismes publics devraient envisager d'associer les structures nationales des droits de l'homme (SNDH) à cet examen externe.

Les auto-évaluations et les examens externes ne devraient pas se limiter à une évaluation des modèles ou algorithmes sur lesquels se fonde le système d'IA, mais devraient également inclure une évaluation de la manière dont

les décideurs pourraient collecter ou influencer les données entrantes et interpréter les données sortantes d'un tel système. Les auto-évaluations et les examens externes devraient également viser à déterminer si un système d'IA demeure véritablement sous le contrôle de l'être humain tout au long de son cycle de vie.

Lorsque l'auto-évaluation ou l'examen externe révèle que le système d'IA présente un risque réel d'atteinte aux droits de l'homme, l'EIDH doit décrire les mesures, les garanties et les mécanismes envisagés pour prévenir ou réduire ce risque. Lorsqu'un tel risque a été identifié en relation avec un système d'IA ayant déjà été déployé par une autorité publique, son utilisation devrait être immédiatement suspendue en attendant l'adoption des mesures, garanties et mécanismes susmentionnés. Lorsqu'il n'est pas possible de réduire de manière significative les risques identifiés, le système d'IA ne devrait être déployé ou utilisé par aucune autorité publique. Lorsque l'auto-évaluation ou l'examen externe révèle une violation des droits de l'homme, l'autorité publique doit agir immédiatement pour mettre un terme et remédier à la violation et adopter des mesures pour écarter ou réduire le risque qu'une telle violation se reproduise.

Les EIDH, y compris les résultats des recherches ou les conclusions de l'examen externe, doivent être mises à la disposition du public dans un format facilement accessible et lisible par ordinateur.

Les autorités publiques ne devraient pas acquérir de systèmes d'IA provenant de tiers qui ne sont pas disposés à lever les restrictions à l'échange d'informations (informations confidentielles ou secrets d'affaire, par exemple) lorsque ces restrictions empêchent ou entravent (i) la réalisation des EIDH (y compris la recherche/l'examen externe) et (ii) l'information du public sur les EIDH.

Les autorités publiques devraient être tenues de mener des EIDH à intervalles réguliers, et non pas seulement lorsqu'elles acquièrent et/ou développent des systèmes d'IA. À tout le moins, une EIDH devrait être réalisée à chaque nouvelle phase du cycle de vie du système d'IA et à chaque autre étape importante.

2 - Consultations publiques

L'utilisation des systèmes d'IA par les États devrait être régie par les normes relatives aux marchés publics, appliquées dans le cadre de procédures transparentes, auxquelles tous les acteurs concernés seraient invités à contribuer. Les États membres devraient envisager de mettre à jour leurs lois et leurs politiques sur l'accès à l'information, la transparence et les marchés publics, pour tenir compte des exigences spécifiques de l'IA.

Les États membres devraient permettre l'organisation de consultations publiques à diverses étapes du processus, et au minimum lors de l'acquisition du système d'IA et lors de la réalisation de l'EIDH. Une véritable consultation publique suppose la publication préalable, en temps utile, de toutes les informations relatives au système d'IA qui facilitent une bonne compréhension de son fonctionnement, de son rôle et de ses effets potentiels ou mesurés. Les consultations devraient permettre à l'ensemble des parties prenantes d'apporter leur contribution : acteurs étatiques, représentants du secteur privé, universitaires, monde associatif, médias et représentants de communautés ou de groupes marginalisés et concernés. Les SNDH, qui servent de pont entre la société civile et les autorités étatiques, peuvent favoriser des consultations constructives.

3 - Obligation, pour les États membres, de faciliter la mise en œuvre des normes des droits de l'homme dans le secteur privé

Les États membres devraient effectivement 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. Lors de cette mise en œuvre, les États devraient éviter toute discrimination et tenir dûment compte des risques liés au genre. De plus, chaque État devrait indiquer clairement à l'ensemble des acteurs de l'IA (créateurs, propriétaires, fabricants, gestionnaires, fournisseurs de services, etc.) qui relèvent de sa juridiction ou qui y exercent des activités, qu'ils sont aussi censés appliquer ces principes dans l'ensemble de leurs opérations.

En vue de respecter les obligations positives et procédurales leur incombant au titre de la Convention européenne des droits de l'homme, les États membres devraient appliquer les mesures nécessaires pour protéger les personnes contre les violations des droits de l'homme qui pourraient être commises par des acteurs de l'IA tout au long du cycle de vie des systèmes d'IA. Les États membres devraient en particulier garantir que leur législation crée des conditions propices au respect des droits de l'homme par les acteurs de l'IA et qu'elle n'entrave pas la responsabilité effective de ces derniers ni le recours effectif contre les violations des droits de l'homme dans le domaine de l'IA.

Les États membres devraient appliquer des mesures supplémentaires pour exiger des acteurs de l'IA qu'ils respectent les droits de l'homme, y compris, le cas échéant, en faisant preuve de diligence raisonnable. Ils devraient aussi exiger des acteurs de l'IA qu'ils prennent des mesures effectives

pour prévenir et/ou atténuer les préjudices liés à leurs systèmes d'IA, et les acteurs de l'IA devraient exposer en toute transparence les efforts déployés pour identifier, prévenir et atténuer ces préjudices. Les États membres devraient prévoir des conséquences appropriées dans le cas où des risques identifiés d'effets négatifs sur les droits de l'homme ne seraient pas dûment atténués ou traités.

4 - Information et transparence

L'utilisation d'un système d'IA dans tout processus décisionnel ayant des effets concrets sur les droits des personnes doit être identifiable. Il ne suffit pas que l'utilisation d'un système d'IA soit rendue publique en des termes clairs et accessibles ; encore faut-il que les personnes soient capables de comprendre comment les décisions sont prises et comment ces décisions ont été vérifiées.

Si un système d'IA est utilisé pour interagir avec des individus dans le contexte de services publics, notamment dans les domaines de la justice, de l'aide sociale et de la santé, l'utilisateur doit en être averti. Il doit aussi être informé de la possibilité de faire appel à un professionnel, sur demande et sans délai. Lorsqu'une personne fait l'objet d'une décision d'une autorité publique se fondant entièrement ou de manière significative sur les résultats d'un système d'IA, cette personne devrait en être avisée et recevoir rapidement les informations susmentionnées.

Le contrôle sur un système d'IA dans son ensemble doit aussi être rendu possible par des exigences de transparence. Ce contrôle peut passer par la communication publique d'informations relatives au système en question, à ses processus et à ses effets directs et indirects sur les droits de l'homme, et aux mesures prises pour identifier et atténuer les incidences négatives du système sur les droits de l'homme, ou par un audit indépendant, complet et efficace. Dans tous les cas, les informations rendues publiques devraient permettre une véritable évaluation du système d'IA. Aucun système d'IA ne devrait atteindre un degré de complexité tel qu'il ne puisse plus être surveillé et contrôlé par des êtres humains. Les systèmes qui ne peuvent être soumis à des normes de transparence et de responsabilité appropriées ne devraient pas être utilisés.

5 - Contrôle indépendant

Les États membres devraient établir un cadre législatif qui permette de vérifier de manière indépendante et effective que le développement, le

déploiement et l'utilisation des systèmes d'IA par les autorités publiques et les entités privées respectent les droits de l'homme. Ce cadre législatif pourrait instaurer des mécanismes réunissant plusieurs organes de contrôle, à caractère administratif, judiciaire, quasi judiciaire et/ou parlementaire, coopérant efficacement entre eux. Les États membres devraient envisager d'habiliter les SNDH, le cas échéant, à jouer un rôle dans le contrôle indépendant et effectif du respect des droits de l'homme en matière d'IA.

Les organes de contrôle devraient être indépendants des autorités publiques et des entités privées qui développent, déploient ou utilisent les systèmes d'IA, et être dotés de l'expertise, des compétences et des ressources interdisciplinaires appropriées et adéquates pour exercer leur fonction de contrôle.

Les organes de contrôle indépendants devraient examiner et vérifier de manière proactive le respect des droits de l'homme par les systèmes d'IA, recevoir et traiter les plaintes des personnes concernées, et dresser des bilans périodiques des capacités des systèmes d'IA et des progrès technologiques en général. Ils devraient avoir le pouvoir d'intervenir lorsqu'ils détectent des violations des droits de l'homme ou un risque de violations. Ils devraient également rendre compte au parlement et publier des rapports sur leurs activités de façon régulière.

Les autorités publiques et les acteurs privés devraient être tenus de fournir, sur demande, toutes les informations nécessaires au contrôle effectif des systèmes d'IA et rendre régulièrement compte aux organes de contrôle. Elles devraient mettre en œuvre les recommandations des organes de contrôle en ce qui concerne les effets des systèmes d'IA sur les droits de l'homme. Les processus de contrôle doivent également être transparents et être soumis à un examen approprié du public ; les décisions des organes de contrôle doivent pouvoir faire l'objet d'un recours ou d'un réexamen indépendant.

6 - Non-discrimination et égalité

Dans tous les cas, il faut prévenir et atténuer les risques de discrimination en accordant une attention particulière aux groupes qui présentent un risque accru de voir leurs droits affectés par l'IA de manière disproportionnée. Parmi ces groupes figurent les femmes, les enfants, les personnes âgées, les personnes économiquement défavorisées, les membres de la communauté LGBTI, les personnes handicapées et les groupes « raciaux », ethniques ou religieux. Les États membres doivent s'abstenir d'utiliser des systèmes d'IA qui sont discriminatoires ou qui entraînent des résultats discriminatoires ; ils doivent aussi protéger les personnes relevant de leur juridiction contre

les conséquences de l'utilisation de tels systèmes d'IA par des tiers.

La participation active et la consultation véritable d'une communauté diverse, où sont effectivement représentés les groupes susmentionnés, à toutes les étapes du cycle de vie d'un système d'IA, peuvent beaucoup contribuer à la prévention et à la réduction des effets négatifs sur les droits de l'homme. De plus, une attention particulière doit être accordée à la transparence et à l'accessibilité de l'information sur les données d'apprentissage utilisées pour le développement d'un système d'IA. Les EIDH et d'autres procédures de diligence raisonnable en matière de droits de l'homme devraient être répétées à intervalles réguliers et des moyens de responsabilisation et de recours appropriés et accessibles devraient être disponibles.

Les États membres devraient renforcer les mesures de contrôle lorsqu'ils utilisent des systèmes d'IA dans le contexte de l'application des lois, notamment lorsqu'ils recourent à des méthodes comme la police prédictive ou préventive. Avant de déployer un tel système, il faudrait le soumettre à un audit indépendant pour vérifier qu'il n'entraîne pas d'effets discriminatoires qui se traduiraient, dans les faits, par le profilage de groupes spécifiques. Si de tels effets sont détectés, le système ne doit pas être utilisé.

7 - Protection des données et respect de la vie privée

Le développement, l'apprentissage, la phase d'essai et l'utilisation de systèmes d'IA qui reposent sur le traitement de données à caractère personnel doivent garantir pleinement le droit des personnes au respect de la vie privée et familiale, consacré par l'article 8 de la Convention européenne des droits de l'homme, y compris le « droit à l'autodétermination en matière d'information » en ce qui concerne leurs données.

Le traitement des données dans le contexte des systèmes d'IA doit être proportionné à la finalité légitime poursuivie par ce traitement et devrait refléter, à chaque étape du traitement, un juste équilibre entre, d'une part, les intérêts visés par le développement et le déploiement du système d'IA et, d'autre part, les droits et les libertés en jeu.

Les États membres devraient effectivement mettre en œuvre la Convention modernisée du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (« Convention 108+ ») ainsi que les autres instruments internationaux sur la protection des données et le respect de la vie privée qui sont juridiquement contraignants. Le traitement des données à caractère personnel, à toute étape du cycle de vie d'un système d'IA, doit reposer sur les principes énoncés dans la Convention 108+, en particulier : (i) le traitement des

données à caractère personnel aux étapes pertinentes du cycle de vie du système d'IA doit s'effectuer en vertu d'un fondement légitime prévu par la loi ; (ii) les données à caractère personnel doivent être traitées licitement, loyalement et de manière transparente ; (iii) les données à caractère personnel doivent être collectées pour des finalités explicites, déterminées et légitimes, et ne doivent pas être traitées de manière incompatible avec ces finalités ; (iv) les données à caractère personnel doivent être adéquates, pertinentes et non excessives au regard des finalités pour lesquelles elles sont traitées ; (v) les données à caractère personnel doivent être exactes et, si nécessaire, mises à jour ; (vi) les données à caractère personnel doivent être conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire aux finalités pour lesquelles elles sont traitées.

Les États membres devraient mettre en place un cadre législatif qui prévoit des garanties appropriées lorsque les systèmes d'IA reposent sur le traitement de catégories particulières de données : données génétiques ; données à caractère personnel concernant des infractions, des procédures et des condamnations pénales et des mesures de sûreté connexes ; données biométriques ; données à caractère personnel sur l'origine « raciale » ou ethnique, les opinions politiques, l'appartenance syndicale, les convictions religieuses ou autres convictions, la santé ou la vie sexuelle. De telles garanties doivent également offrir une protection contre le traitement discriminatoire ou biaisé de ces données.

8 - Liberté d'expression, liberté de réunion et d'association, et droit au travail

Dans le cadre de leur obligation de respecter, de protéger et de réaliser les droits de l'homme et les libertés fondamentales de toute personne, les États membres devraient tenir compte de l'ensemble des normes internationales relatives aux droits de l'homme qui peuvent être concernées par l'utilisation de l'IA.

Liberté d'expression : les États membres devraient garder à l'esprit l'obligation de créer un environnement favorisant une information diversifiée et pluraliste, et être conscients des effets négatifs que la modération et la curation de contenus s'appuyant sur l'IA peuvent avoir sur l'exercice du droit à la liberté d'expression, du droit à l'accès à l'information et du droit à la liberté d'opinion. Les États membres sont également encouragés à envisager de prendre des mesures appropriées pour réguler les monopoles dans le domaine des technologies, afin de prévenir les conséquences négatives de la concentration de l'expertise en matière d'IA et du pouvoir

sur la libre circulation de l'information.

Liberté de réunion et d'association : une attention particulière devrait être accordée aux conséquences que l'utilisation de systèmes d'IA dans la modération de contenus peut avoir sur la liberté de réunion et d'association, en particulier dans les contextes où l'exercice de ces libertés est difficile hors ligne. Le recours à la technologie de reconnaissance faciale devrait être strictement encadré par les États membres, y compris au moyen de lois établissant des limites claires et au moyen de garanties de transparence, afin de protéger l'exercice effectif du droit à la liberté de réunion.

Droit au travail : le potentiel de l'IA d'accélérer l'automatisation, et d'avoir ainsi des répercussions négatives sur la disponibilité du travail, devrait être surveillé attentivement. Des évaluations régulières devraient être menées pour connaître le nombre et les types d'emplois créés et détruits en raison des progrès de l'IA. Il faudrait élaborer des plans adéquats pour que les travailleurs clairement touchés par la baisse de la demande de main-d'œuvre humaine puissent suivre une formation et occuper un nouvel emploi. Les États membres devraient également adapter les programmes d'enseignement afin d'assurer l'accès aux emplois requérant des compétences liées aux systèmes d'IA.

9 - Possibilités de recours

Un système d'IA doit toujours rester sous le contrôle de l'être humain, même lorsque l'apprentissage automatique ou des techniques similaires permettent au système de prendre des décisions indépendamment de l'intervention spécifique d'un être humain. Les États membres doivent établir des lignes de responsabilité claires en ce qui concerne les violations des droits de l'homme qui peuvent se produire à différentes phases du cycle de vie d'un système d'IA. La responsabilité des violations des droits de l'homme survenues lors du développement, du déploiement ou de l'utilisation de systèmes d'IA, et l'obligation de rendre des comptes, doivent toujours incomber à une personne physique ou morale, même dans les cas où la mesure portant atteinte aux droits de l'homme n'a pas été directement ordonnée par un dirigeant ou un opérateur humain responsable.

Quiconque se déclare victime d'une violation des droits de l'homme découlant du développement, du déploiement ou de l'utilisation, par une entité publique ou privée, d'un système d'IA devrait pouvoir exercer un recours effectif devant une instance nationale. De plus, les États membres devraient garantir l'accès à un recours effectif aux personnes qui soupçonnent qu'elles pourraient avoir été soumises, de manière non transparente et à leur insu, à une mesure fondée exclusivement ou de

manière significative sur les résultats d'un système d'IA.

Des voies de recours effectives devraient permettre d'obtenir rapidement des réparations et des indemnités adéquates pour tout préjudice causé par le développement, le déploiement ou l'utilisation de systèmes d'IA ; elles peuvent inclure des mesures civiles, administratives ou, le cas échéant, pénales. Les SNDH peuvent incarner une telle source de recours, en rendant leurs propres décisions conformément à leurs mandats respectifs.

Les États membres devraient garantir aux individus le droit de ne pas être soumis à une décision les affectant de manière significative lorsque cette décision est fondée sur un processus décisionnel automatisé sans réelle intervention humaine. À tout le moins, toute personne devrait pouvoir obtenir une intervention humaine dans une telle prise de décision automatisée et faire valoir ses opinions avant que la décision soit appliquée.

Les États membres doivent veiller à ce que quiconque s'estime victime d'une violation des droits de l'homme causée par un système d'IA ait accès aux informations détenues par le défendeur ou par un tiers qui sont pertinentes pour justifier son allégation, y compris, le cas échéant, les données d'apprentissage et de test, les informations sur la façon dont le système d'IA a été utilisé, des informations utiles et compréhensibles sur la manière dont le système d'IA est parvenu à une recommandation, à une décision ou à une prévision, ainsi que des informations expliquant comment les résultats du système d'IA ont été interprétés et quelles suites leur ont été données.

Lorsque les autorités nationales examinent des plaintes relatives à des violations des droits de l'homme causées par le développement, le déploiement ou l'utilisation de systèmes d'IA, elles doivent manifester un scepticisme approprié à l'égard de « l'apparence d'objectivité » que présentent les systèmes d'IA, et veiller à ce que les personnes qui dénoncent les atteintes aux droits de l'homme ne soient pas tenues de satisfaire à des exigences plus élevées en matière de preuve que les personnes responsables de la mesure contestée.

10 - Promotion de la connaissance et de la compréhension de l'intelligence artificielle

La connaissance et la compréhension de l'IA devraient être encouragées dans les institutions gouvernementales, les organes de contrôle indépendants, les structures nationales des droits de l'homme, le système judiciaire et les services répressifs, et auprès du grand public. Les États membres devraient envisager d'établir, au sein du gouvernement, un organe consultatif qui

serait chargé de donner des conseils sur toutes les questions liées à l'IA.

Les personnes participant directement ou indirectement au développement ou à la mise en œuvre d'un système d'IA doivent comprendre comment le système fonctionne et être informées de son impact sur les droits de l'homme. Pour que ces acteurs soient conscients de cet impact, ils doivent aussi être sensibilisés à l'étendue des normes relatives aux droits de l'homme qui pourraient devoir être prises en compte dans le cadre de leur travail.

Les États membres devraient s'employer à familiariser davantage le grand public avec l'IA en déployant d'importants efforts de sensibilisation, de formation et d'éducation, y compris (et en particulier) dans les établissements scolaires. Il faudrait certes expliquer le fonctionnement de l'IA, mais ne pas oublier de traiter aussi la question de ses éventuels effets – positifs et négatifs – sur les droits de l'homme. Des efforts particuliers devraient être faits pour toucher les groupes marginalisés et les personnes qui, de manière générale, sont moins familiarisées avec l'informatique.

Checklists

Évaluation de l'impact sur les droits de l'homme	À faire
	<p>Prendre des mesures pour introduire des dispositions législatives et réglementaires imposant la réalisation d'EIDH en relation avec les systèmes d'IA qui ont été, ou peuvent être, acquis, développés et/ou déployés par les autorités publiques.</p> <p>Mener sans délai des EIDH pour tous les systèmes d'IA qui ont déjà été déployés/sont déjà utilisés par les autorités publiques au moment de l'adoption du cadre juridique pertinent sur les EIDH. Dans les autres cas, une EIDH doit être réalisée avant l'acquisition et/ou le développement de tout système d'IA par une autorité publique.</p> <p>Suivre en permanence les effets du système d'IA sur les droits de l'homme, tout au long de son cycle de vie, et mener régulièrement des EIDH, à chaque nouvelle étape du cycle de vie et lorsque des changements interviennent dans le contexte, la portée, la nature et l'objectif du système.</p>
	À ne pas faire
	<p>Omettre de consulter de manière constructive les parties prenantes concernées, dont les organisations de la société civile et les spécialistes de l'IA et des droits de l'homme, et se priver de leur contribution, lors de la mise en place d'un cadre juridique sur les EIDH.</p> <p>Mener des EIDH de façon non transparente et recourir, ou faciliter le recours, à des lois sur la confidentialité, la vie privée, les secrets d'affaire, les secrets d'État ou la propriété intellectuelle pour entraver la mise en œuvre ou la publication d'EIDH.</p> <p>Acquérir, développer, déployer ou utiliser un système d'IA qui est susceptible d'empiéter sur les droits de l'homme dans les cas où (i) le système d'IA n'a pas été soumis à une EIDH ou (ii) une EIDH a révélé que le système risque réellement de porter atteinte aux droits de l'homme et, pour autant, ni mesures, ni garanties, ni mécanismes n'ont été adoptés pour prévenir ou atténuer les risques identifiés.</p>

Consultations publiques	À faire
	<p>Appliquer à l'utilisation des systèmes d'IA les normes relatives aux marchés publics et un processus transparent.</p> <p>Associer tous les acteurs aux consultations publiques, y compris les communautés ou groupes concernés, au minimum lors de la phase d'acquisition et lors de l'EIDH.</p>
	À ne pas faire
	<p>Organiser des consultations publiques sans prendre de mesures adéquates pour les rendre constructives, c'est-à-dire sans publier au préalable, en temps utile, l'ensemble des informations pertinentes relatives au système d'IA, et sans chercher activement à faire participer tous les acteurs pertinents.</p>

À faire

Examiner les dispositions du droit pénal et civil en vigueur, ainsi que d'autres régimes de responsabilité équivalents, afin d'identifier les lacunes et les obstacles qui empêchent de tenir les acteurs de l'IA pour responsables des violations des droits de l'homme liées à l'IA.

Faire appliquer la législation en vigueur lorsque cela est nécessaire pour respecter l'obligation de l'État de protéger les personnes contre les violations des droits de l'homme commises par des acteurs de l'IA.

Prendre des mesures pour que les acteurs de l'IA « connaissent et manifestent » leur responsabilité en matière de droits de l'homme, y compris par des procédures transparentes de diligence raisonnable qui consistent notamment à identifier les risques que leurs systèmes d'IA présentent pour les droits de l'homme et à adopter des mesures effectives pour prévenir et/ou atténuer les préjudices qui en résultent.

À ne pas faire

Considérer les lois, les politiques et les réglementations applicables au domaine de l'IA de manière isolée, comme si elles n'avaient pas de lien avec les obligations des États membres en matière de droits de l'homme.

Faciliter l'application et le respect des normes des droits de l'homme dans le domaine de l'IA de façon discriminatoire.

Information et transparence	À faire
	Donner toutes les informations qui sont nécessaires aux individus pour comprendre quand et comment les systèmes d'IA sont utilisés, notamment dans le contexte des services publics.
Information et transparence	À ne pas faire
	Utiliser des systèmes d'IA dont le degré de complexité est tel qu'ils ne peuvent plus être surveillés et contrôlés par des êtres humains selon des normes appropriées en matière de transparence et de responsabilité.

Contrôle indépendant	À faire
	<p>Légiférer en vue d'établir un cadre pour le contrôle indépendant et effectif du respect des droits de l'homme par les systèmes d'IA, en s'appuyant sur les organes de contrôle existants, dont les SNDH, dans la mesure du possible.</p> <p>Prendre des mesures pour que tous les organes de contrôle pertinents aient accès à une expertise suffisante, aient reçu une formation appropriée sur les systèmes d'IA et sur leurs conséquences pour les droits de l'homme, et disposent des ressources, financières et autres, dont ils ont besoin pour exercer efficacement leurs fonctions.</p> <p>Veiller à ce que les fonctions des organes de contrôle pertinents leur permettent de mener des enquêtes et d'assurer la surveillance de tous les acteurs, publics ou privés, qui pourraient être responsables de violations des droits de l'homme causées par des systèmes d'IA (y compris celles qui se produisent au cours du développement, de la phase d'essai et de l'utilisation des systèmes).</p>
	À ne pas faire
	<p>Limiter les fonctions et les pouvoirs des organes de contrôle au point qu'ils sont dans l'incapacité d'intervenir efficacement lorsqu'ils identifient des (risques de) violations des droits de l'homme.</p> <p>Compromettre l'indépendance institutionnelle, opérationnelle, financière et personnelle des organes de contrôle chargés de mener des enquêtes et de vérifier si les systèmes d'IA respectent les droits de l'homme.</p> <p>Priver, ou permettre à d'autres de priver, les organes de contrôle des informations nécessaires à leur bon fonctionnement ; les priver notamment de l'accès aux jeux de données (d'apprentissage et de test), aux données d'IA entrantes/sortantes, aux modèles/algorithmes, aux orientations opérationnelles et à la procédure de diligence raisonnable en matière de droits de l'homme.</p>

Non-discrimination et égalité	À faire
	<p>Prévenir et atténuer les risques de discrimination liés à l'utilisation des systèmes d'IA pour les groupes qui présentent un risque accru de voir leurs droits affectés par ces systèmes de manière disproportionnée.</p> <p>Renforcer les mesures de contrôle lors de l'utilisation de systèmes d'IA dans les activités de la police, notamment pour éviter le profilage d'individus appartenant à des groupes spécifiques.</p>
	À ne pas faire
	<p>Utiliser, ou permettre à des tiers d'utiliser, des systèmes d'IA qui discriminent ou qui conduisent à des résultats discriminatoires.</p>

Protection des données et respect de la vie privée	À faire
	<p>Passer en revue et évaluer les lois en vigueur relatives à la protection des données pour déterminer si elles protègent suffisamment le droit au respect de la vie privée et le droit à la protection des données dans le contexte des systèmes d'IA et, en cas de lacunes, engager une réforme législative.</p> <p>Prendre des mesures proactives pour que les entités privées et publiques participant au développement, au déploiement et à l'utilisation des systèmes d'IA respectent les droits des personnes concernées et s'acquittent de leurs obligations au titre des lois applicables relatives à la protection des données.</p>
	À ne pas faire
	<p>Accorder à ceux qui développent, déploient ou utilisent des systèmes d'IA des dérogations ou des immunités disproportionnées en matière de traitement des données.</p> <p>Permettre le développement ou l'utilisation de systèmes d'IA qui reposent sur des jeux de données d'apprentissage ou de test ayant été collectées ou autrement traitées en violation du droit au respect de la vie privée et du droit à la protection des données.</p> <p>Permettre le développement ou l'utilisation de systèmes d'IA qui traitent des données à caractère personnel, qu'il s'agisse de données entrantes ou sortantes, en violation du droit au respect de la vie privée et du droit à la protection des données.</p>

À faire

Prendre en compte l'ensemble des normes internationales relatives aux droits de l'homme qui peuvent être concernées par l'utilisation de l'IA.

Garder à l'esprit les effets que la modération et la curation de contenus s'appuyant sur l'IA peuvent avoir sur l'exercice des droits à la liberté d'expression, à l'accès à l'information et à la liberté d'opinion.

Réglementer strictement l'utilisation de la technologie de reconnaissance faciale pour permettre l'exercice effectif du droit à la liberté de réunion.

Surveiller les effets négatifs potentiels de l'IA sur le droit au travail et prévoir des moyens d'atténuer ces effets, y compris par l'éducation et la formation.

À ne pas faire

Permettre ou faciliter le développement, le déploiement ou l'utilisation de systèmes d'IA qui portent atteinte aux droits de l'homme protégés par les normes internationales pertinentes.

À faire

Mener une évaluation des lois existantes, en matière civile, pénale et administrative, et engager des réformes lorsque ces lois ne prévoient pas de recours effectifs pour les personnes qui s'estiment victimes de violations des droits de l'homme découlant du développement, du déploiement ou de l'utilisation d'un système d'IA.

Veiller à ce que les régimes de responsabilité établissent clairement qui est légalement responsable des violations des droits de l'homme qui peuvent se produire aux différentes phases du cycle de vie d'un système d'IA.

S'assurer que les autorités judiciaires et les autres autorités nationales pertinentes n'accordent pas une importance inappropriée à l'exactitude ou à l'objectivité supposée/perçue d'un système d'IA, et qu'elles garantissent l'égalité des armes entre la victime et le défendeur dans les affaires concernant des violations des droits de l'homme causées par des systèmes d'IA.

À ne pas faire

Permettre le développement, le déploiement ou l'utilisation de systèmes d'IA qui fonctionnent totalement en dehors de tout contrôle humain.

Permettre que des individus soient soumis à des décisions automatisées les affectant de manière significative sans avoir eu la possibilité d'obtenir une réelle intervention humaine et sans avoir vu leurs opinions prises en compte avant l'application de la décision.

Promotion de la connaissance et de la compréhension de l'IA	À faire
	<p>Établir, au sein du gouvernement, un organe consultatif chargé de donner des conseils sur toutes les questions liées à l'IA.</p> <p>Favoriser la connaissance et la compréhension de l'IA et des droits de l'homme pour tous, des personnes participant au développement des systèmes d'IA jusqu'au grand public.</p>
	À ne pas faire
	<p>Dans le cadre de l'éducation à l'IA, se limiter aux aspects technologiques et omettre de traiter la question des effets potentiels de l'IA sur les droits de l'homme.</p>

Annexe

Définitions

Aux fins de la présente recommandation, les termes suivants s'entendent ainsi :

Algorithme : suite finie de règles/instructions formelles, généralement sous la forme d'une logique mathématique, permettant d'obtenir un résultat à partir d'éléments fournis en entrée.

Intelligence artificielle (IA) : expression générique utilisée pour désigner un ensemble de sciences, de théories et de techniques dont le but est d'améliorer la capacité des machines à réaliser des tâches requérant des facultés cognitives.

Système d'IA : système informatique pouvant formuler des recommandations, établir des prévisions ou prendre des décisions en fonction d'une série donnée d'objectifs. À cette fin, il recourt à des données informatiques et/ou humaines pour : (i) percevoir des environnements réels et/ou virtuels ; (ii) convertir manuellement ou automatiquement ces perceptions en modèles ; et (iii) formuler de possibles résultats à partir des interprétations de ces modèles.

Cycle de vie d'un système d'IA : série de phases concernant un système d'IA : (i) la planification et la conception, la collecte et le traitement des données, et l'élaboration de modèles ; (ii) la vérification et la validation ; (iii) le déploiement ; (iv) le fonctionnement et le suivi ; et (v) la fin de vie.

Prise de décision automatisée : processus de prise de décision par des moyens automatisés. Elle consiste généralement à utiliser un raisonnement automatisé pour faciliter ou remplacer un processus décisionnel qui incomberait autrement à des êtres humains. Elle n'implique pas nécessairement l'utilisation de l'IA mais passe généralement par la collecte et le traitement de données.

Apprentissage automatique : domaine de l'IA qui consiste en une série de techniques et d'algorithmes pouvant être utilisés pour « entraîner » une machine à reconnaître automatiquement des régularités dans une série de données. En reconnaissant des régularités dans des données, l'ordinateur peut créer des modèles qui expliquent les données et/ou prévoient des données futures. En résumé, l'apprentissage automatique permet à un ordinateur de réaliser des tâches pour lesquelles il n'a pas été explicitement programmé.

Modèle : représentation exploitable de tout ou partie de l'environnement externe d'un système d'IA qui décrit la structure et/ou la dynamique de l'environnement. Le modèle représente le cœur de tout système d'IA. Un modèle peut se fonder sur des données et/ou des connaissances spécialisées, issues d'êtres humains et/ou d'outils automatisés tels que les algorithmes d'apprentissage automatique.

Données à caractère personnel : informations concernant une personne physique identifiée ou identifiable, directement ou indirectement, par référence à un ou plusieurs éléments propres à cette personne. Les données à caractère personnel sensibles désignent les données à caractère personnel relatives à l'origine « raciale » ou ethnique, aux opinions politiques, aux convictions religieuses ou philosophiques ou à l'appartenance syndicale, ainsi que les données génétiques, les données biométriques, les données concernant la santé et les données concernant la vie sexuelle ou l'orientation sexuelle.

Traitement des données à caractère personnel : toute opération ou tout ensemble d'opérations effectuées avec ou sans l'aide de procédés automatisés et appliquées à des données ou des ensembles de données à caractère personnel, telles que la collecte, l'enregistrement, l'organisation, la structuration, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, la diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, la limitation, l'effacement ou la destruction.

Le Commissaire aux droits de l'homme est une instance 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 œuvre 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

FR

www.coe.int

Le Conseil de l'Europe est la principale organisation de défense des droits de l'homme du continent. Il compte 47 États membres, dont 28 sont également membres de l'Union européenne. Tous les États membres du Conseil de l'Europe ont signé la Convention européenne des droits de l'homme, un traité visant à protéger les droits de l'homme, la démocratie et l'État de droit. La Cour européenne des droits de l'homme contrôle la mise en œuvre de la Convention dans les États membres.



COMMISSIONER
FOR HUMAN RIGHTS

COMMISSAIRE AUX
DROITS DE L'HOMME

