

ÉGALITÉ ET INTELLIGENCE ARTIFICIELLE



Recommandation CM/Rec(2026)1
et exposé des motifs

ÉGALITÉ ET INTELLIGENCE ARTIFICIELLE

English edition:
Equality and artificial intelligence

La reproduction d'extraits (jusqu'à 500 mots) est autorisée, sauf à des fins commerciales, tant que l'intégrité du texte est préservée, que l'extrait n'est pas utilisé hors contexte, ne donne pas d'informations incomplètes ou n'induit pas le lecteur en erreur quant à la nature, à la portée et au contenu de ce texte. Le texte source doit toujours être cité comme suit : « © Conseil de l'Europe, année de publication ». Pour toute autre demande relative à la reproduction ou à la traduction de tout ou partie de ce document, veuillez vous adresser à la Division publications et identité visuelle (DPIV), Conseil de l'Europe (F-67075 Strasbourg Cedex), ou à publishing@coe.int.

Toute autre correspondance relative à ce document doit être adressée à la Direction générale de la démocratie et de la dignité humaine, Conseil de l'Europe, F-67075 Strasbourg Cedex, Courriel : gender.equality@coe.int.

Conception de la couverture et mise en page :
Division publications et identité visuelle (DPIV),
Conseil de l'Europe

Photo de couverture : IStock

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

© Conseil de l'Europe, juin 2026
Imprimé dans les ateliers du Conseil de l'Europe

Table des matières

RECOMMANDATION CM/REC(2026)1	5
PRÉAMBULE	5
Annexe à la Recommandation CM/Rec(2026)1 du Comité des Ministres aux États membres sur l'égalité et l'intelligence artificielle	7
Principes et orientations sur l'égalité et l'intelligence artificielle	7
I. But	7
II. Approche globale en matière de législation, de politiques et de cadres réglementaires	8
III. Promotion de l'égalité par les acteurs publics et privés	9
IV. Prévention de la discrimination	10
V. Participation démocratique, sensibilisation du public et renforcement des capacités	12
VI. Compétences et pouvoirs des organes de contrôle indépendants	12
VII. Dispositions complémentaires	13
EXPOSÉ DES MOTIFS	15
Préambule	15
Annexe	18
Principes et orientations sur l'égalité et l'intelligence artificielle	18
I. But	18
II. Approche globale en matière de législation, de politiques et de cadres réglementaires	21
III. Promotion de l'égalité par les acteurs publics et privés	26
IV. Prévention de la discrimination	30
V. Participation démocratique, sensibilisation du grand public et renforcement des capacités	40
VI. Compétences et pouvoirs des organes de contrôle indépendants	42
VII. Dispositions complémentaires	43

Recommandation CM/Rec(2026)1

du Comité des Ministres aux États membres sur l'égalité et l'intelligence artificielle

*(Adoptée par le Comité des Ministres le 4 mars 2026,
lors de la 1552^e réunion des Délégués des Ministres)*

PRÉAMBULE

Le Comité des Ministres, en vertu de l'article 15.b du Statut du Conseil de l'Europe (STE n° 1),

Considérant que le but du Conseil de l'Europe est de réaliser une union plus étroite entre ses membres, que ce but peut notamment être poursuivi par une action commune dans le domaine des droits humains et que les États membres du Conseil de l'Europe se sont engagés à garantir les droits et libertés inscrits dans la Convention de sauvegarde des droits de l'homme et des libertés fondamentales (STE n° 5, « la Convention ») à toute personne relevant de leur juridiction ;

Rappelant les obligations spécifiques des États membres dans le domaine de l'égalité et de la non-discrimination en vertu du droit européen et international, notamment en vertu de l'article 14 de la Convention et, le cas échéant, de l'article premier de son Protocole n° 12 (STE n° 177), de la jurisprudence correspondante de la Cour européenne des droits de l'homme et de l'article E de la Charte sociale européenne (révisée) (STE n° 163), et des décisions et conclusions correspondantes du Comité européen des Droits sociaux, qui s'appliquent également tout au long du cycle de vie des systèmes d'intelligence artificielle (« IA ») ;

Soulignant que ces obligations restent de mise tout au long des processus continus de progrès technologique et de transformation numérique que traversent les sociétés européennes, notamment les activités menées dans le cadre du cycle de vie des systèmes d'IA ;

Rappelant que la dignité humaine, l'autonomie individuelle et l'égalité, y compris l'égalité de genre, sont au cœur de l'État de droit et de la démocratie, et reconnaissant ainsi la nécessité de garantir l'égalité réelle et de prévenir et combattre toute forme de violence et de discrimination en renforçant les travaux en faveur de sociétés inclusives exemptes de marginalisation, d'exclusion ou de toute forme d'intolérance, y compris le sexisme, le racisme, le validisme, l'âgisme et la LGBTIphobie ;

Soulignant la nécessité de s'attaquer aux formes multiples et intersectionnelles de discrimination et d'inégalité qui peuvent être causées et exacerbées par les activités menées dans le cadre du cycle de vie des systèmes d'IA ;

Rappelant que l'égalité de genre et la participation pleine, égale et effective des femmes aux processus décisionnels publics et privés sont essentielles à l'État de droit, à la démocratie et au développement durable, et que ces objectifs sont une priorité du Conseil de l'Europe depuis de nombreuses années, comme l'attestent le vaste corpus de normes et les stratégies élaborées dans ce domaine ;

Saluant les efforts déployés par les États membres aux niveaux régional, national et européen et autres niveaux internationaux pour faire en sorte que l'utilisation croissante des systèmes d'IA soit cohérente avec l'obligation de protéger les droits humains, la démocratie et l'État de droit, notamment grâce à la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (STCE n° 225) ;

Soulignant que les acteurs du secteur privé, conformément aux standards, aux normes et aux recommandations existant aux niveaux régional, national et européen et autres niveaux internationaux, ont une responsabilité de promouvoir l'égalité et de respecter les droits humains de toutes les parties concernées, y compris ceux de leurs consommateurs et consommatrices, en particulier l'interdiction de la discrimination et le respect de l'égalité, y compris l'égalité de genre ;

Conscient des risques que les systèmes d'IA présentent pour l'égalité et la non-discrimination, étant donné qu'ils peuvent s'appuyer sur des données et des modèles qui reproduisent, perpétuent et amplifient les biais, les stéréotypes, la stigmatisation, les préjugés et de fausses hypothèses à l'égard des individus ou des groupes fondés sur leurs caractéristiques personnelles réelles ou perçues, ou de substituts à ces caractéristiques ainsi que leurs intersections, et des défis distincts que peuvent présenter les systèmes d'IA en raison de leur complexité, de leur effet à grande échelle ou de leur transparence limitée ;

Notant que la sous-représentation, dans le secteur de l'IA, des femmes et d'autres individus et groupes exposés à la discrimination, notamment aux postes de décision ainsi que dans la conception et le développement des systèmes d'IA, est une dimension importante au regard de ces risques ;

Reconnaissant que la fracture numérique, y compris la fracture numérique entre les femmes et les hommes, et l'utilisation de systèmes d'IA peuvent amplifier la discrimination et la sous-représentation, et que cela souligne la nécessité d'atténuer les inégalités économiques et sociales qui en résultent ;

S'inquiétant de ce que la violence à l'égard des femmes et des filles facilitée par la technologie est de plus en plus répandue en raison du développement rapide de la technologie, y compris l'intelligence artificielle, et de son usage généralisé, et soulignant la pertinence de la Convention du Conseil de l'Europe sur la prévention et la lutte contre la violence à l'égard des femmes et la violence domestique (STCE n° 210) dans ce contexte ;

Reconnaissant les besoins et les vulnérabilités spécifiques des enfants et des jeunes, et l'impact particulier que les systèmes d'IA peuvent avoir sur tous les aspects de leur vie, ainsi que la nécessité de veiller à ce qu'ils et elles aient accès aux informations pertinentes à cet égard et d'inclure les jeunes dans les processus de prise de décision liés aux systèmes d'IA ;

Reconnaissant que les systèmes d'IA peuvent avoir des effets discriminatoires importants, aggravés par l'opacité de leurs processus et par les asymétries d'information, et qu'il est difficile pour les personnes affectées de repérer et de prouver les violations de leurs droits qui en résultent, et de prendre des mesures efficaces pour y remédier ;

Rappelant que, même lorsqu'un système d'IA est conforme au droit, aux normes et aux instruments nationaux, européens et internationaux applicables régissant spécifiquement l'IA, son utilisation peut néanmoins entraîner des discriminations ou d'autres violations des droits humains ;

Soulignant que les organes de contrôle indépendants chargés de protéger les droits des individus et groupes exposés à la discrimination éprouvent des difficultés à prévenir ces violations et à y remédier, et qu'il y a un besoin urgent d'y répondre ;

Préoccupé par le fait que l'impact environnemental des systèmes d'IA, y compris leur empreinte carbone et leur consommation de ressources, puisse affecter de manière disproportionnée les femmes et les personnes et groupes exposés à la discrimination, et exacerber les inégalités existantes, et soulignant la nécessité d'un développement durable de l'IA ;

Reconnaissant l'utilisation croissante des systèmes d'IA dans la plupart, si ce n'est la totalité, des domaines de la vie sociale, politique, culturelle, économique et professionnelle quotidienne, et le rythme toujours plus rapide de ces évolutions ;

Craignant qu'il puisse en résulter la création d'inégalités et de discriminations structurelles, multiples et intersectionnelles ou l'aggravation de celles qui existent déjà, notamment celles vécues par les femmes et les filles et les individus exposés à la discrimination, ainsi que l'augmentation de la violence perpétrée à leur encontre en ligne et hors ligne ;

Soulignant que l'ampleur et l'étendue du déploiement actuel et probablement futur des systèmes d'IA font qu'il est essentiel que les États membres déterminent s'il est approprié d'avoir recours à ces systèmes pour une utilisation particulière et qu'ils veillent ensuite à ce que toutes les activités menées dans le cadre du cycle de vie des systèmes d'IA soient cohérentes avec les obligations en matière de droits humains, en vue d'amplifier les effets positifs et de prévenir, d'atténuer ou d'éviter les éventuels effets négatifs ;

Reconnaissant que les systèmes d'IA, en particulier lorsqu'ils sont mis en œuvre avec des mécanismes de transparence et de responsabilité, peuvent être utilisés pour promouvoir l'égalité, y compris l'égalité de genre, et pour autonomiser les personnes et les groupes exposés à la discrimination, et pour favoriser la prospérité humaine, le progrès économique, le bien-être des individus et de la société, le développement durable et la réalisation d'autres objectifs et intérêts importants, en étant porteurs de progrès et d'innovation, et en identifiant et en remédiant aux disparités de traitement et leurs impacts;

Soulignant qu'une action plus forte et ciblée est nécessaire pour réaliser le potentiel positif des systèmes d'IA, tout en étant conscient que les risques de discrimination peuvent également survenir lorsque les systèmes d'IA sont utilisés dans ce contexte;

Se félicitant des initiatives en cours, en particulier dans l'industrie, visant à promouvoir une IA sûre, sécurisée, fiable et transparente, mais rappelant aux États membres leur rôle dans la mise en œuvre de leurs obligations en vertu de la Convention et d'autres normes relatives aux droits humains pertinentes;

Tenant compte et s'inspirant des standards, des normes et des recommandations au niveau du Conseil de l'Europe, ainsi qu'aux niveaux régional et international en matière de protection des droits humains et des libertés fondamentales, de démocratie et d'État de droit dans les sociétés contemporaines, ainsi que de la jurisprudence évolutive de la Cour européenne des droits de l'homme;

Concluant, à la lumière des considérations qui précèdent, qu'il est nécessaire d'adopter une approche multipartite globale pour s'attaquer aux causes profondes de la discrimination entraînée par les systèmes d'IA grâce à des solutions juridiques et politiques appropriées, pour assurer la promotion de l'égalité, y compris l'égalité de genre, et pour éviter les risques de discrimination dans le contexte des activités menées dans le cadre du cycle de vie des systèmes d'IA, en garantissant la responsabilité et des formes démocratiques et transparentes de contrôle humain,

Recommande aux gouvernements des États membres :

1. de prendre toutes les mesures nécessaires pour assurer une mise en œuvre rapide et complète des principes et des orientations annexés à la présente recommandation;
2. de veiller, par des cadres législatifs, administratifs, réglementaires, politiques et de contrôle appropriés et toute autre mesure, à ce que les acteurs du secteur public et du secteur privé qui mènent des activités dans le cadre du cycle de vie des systèmes d'IA respectent leurs obligations en matière d'égalité et de non-discrimination, particulièrement en matière de droits humains, qui leur incombent en vertu des lois, normes et standards nationaux, européens et internationaux applicables;
3. de veiller à ce que la présente recommandation soit traduite et diffusée aussi largement que possible, y compris dans un langage simple et accessible, en recourant à tous les moyens disponibles, auprès des autorités ou organismes publics nationaux et des parties prenantes, tels que les parlements, les organismes indépendants, les chercheuses et les chercheurs, la société civile et le secteur privé;
4. de promouvoir les objectifs de la présente recommandation aux niveaux local, régional, national et européen ainsi qu'aux autres niveaux internationaux, en instaurant un dialogue et une coopération avec l'ensemble des parties prenantes;
5. de réexaminer l'état de mise en œuvre de la présente recommandation à intervalles réguliers après son adoption afin d'en améliorer l'effet et d'informer le Comité des Ministres des mesures prises par les États membres et les autres parties prenantes, des progrès réalisés et des mesures programmées pour combler les lacunes qui subsistent.

ANNEXE À LA RECOMMANDATION CM/REC(2026)1 DU COMITÉ DES MINISTRES AUX ÉTATS MEMBRES SUR L'ÉGALITÉ ET L'INTELLIGENCE ARTIFICIELLE

Principes et orientations sur l'égalité et l'intelligence artificielle

I. But

1. Les principes et lignes directrices qui suivent ont pour objet d'aider les États membres à promouvoir l'égalité, y compris l'égalité de genre, et à prévenir et à combattre toutes les formes de discrimination dans toutes les activités qu'ils mènent et dans celles menées par les acteurs du secteur public et du secteur privé dans le cadre du cycle de vie des systèmes d'IA, conformément à l'article 14 de la Convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales (STE n° 5, «la Convention»), de

l'article premier de son Protocole n° 12 (STE n° 177), le cas échéant, et de la jurisprudence correspondante de la Cour européenne des droits de l'homme, ainsi que conformément à l'article E de la Charte sociale européenne (révisée) (STE n° 163), et aux décisions correspondantes du Comité européen des Droits sociaux, le cas échéant.

2. La discrimination peut être fondée sur une multitude de caractéristiques ou statuts personnels réels ou perçus, tels que le sexe, le genre, la « race »¹, la couleur, la langue, la religion, les opinions politiques ou autres, l'origine nationale ou sociale, l'appartenance à une minorité nationale, la nationalité, la fortune, la naissance, l'orientation sexuelle, l'identité et l'expression de genre, les caractéristiques sexuelles, l'âge, l'état de santé, le handicap, la situation matrimoniale et le statut de migrant-e ou de réfugié-e. Les États membres devraient accorder une attention particulière aux expériences spécifiques de désavantages et de vulnérabilités structurels et systémiques ainsi qu'à la discrimination multiple et intersectionnelle. Compte tenu de la nature évolutive des systèmes d'IA, la discrimination peut également se produire par procuration, générée à partir d'un élément d'information apparemment neutre mais néanmoins lié ou corrélé à une caractéristique protégée.

II. Approche globale en matière de législation, de politiques et de cadres réglementaires

3. Les États membres devraient veiller à ce qu'un ensemble complet de mesures soit en place afin de promouvoir l'égalité, y compris l'égalité de genre, et de prévenir et de combattre la discrimination dans le contexte des systèmes d'IA. Ces mesures devraient être en adéquation avec et conformes à leurs obligations au titre de la Convention ainsi qu'aux obligations internationales pertinentes concernant les systèmes d'IA et celles concernant les droits humains, l'égalité et la non-discrimination.

4. En particulier, les États membres devraient garantir, en adoptant une approche graduée et différenciée, que :

- 4.1. une législation, des politiques et des cadres administratifs, réglementaires ou autres pertinents sont en place ou, le cas échéant, élaborés, et qu'ils sont appliqués en pratique afin de promouvoir l'égalité, y compris l'égalité de genre, et de prévenir et combattre toutes les formes de discrimination tout au long du cycle de vie des systèmes d'IA ;
- 4.2. l'efficacité et l'adéquation de cette législation, de ces politiques et de ces cadres administratifs, réglementaires et autres pour couvrir les activités menées au cours du cycle de vie des systèmes d'IA sont contrôlées et évaluées.

Le cas échéant, ces activités devraient être menées en coopération avec les organismes de promotion de l'égalité, les autorités compétentes, d'autres structures nationales pertinentes en matière de droits humains et la société civile.

5. Les mesures prises devraient contribuer à garantir :

- 5.1. la pertinence de l'utilisation des systèmes d'IA, en tenant compte des impacts ou risques réels ou potentiels en matière de non-discrimination et en reconnaissant leur potentiel à promouvoir l'égalité, y compris l'égalité de genre ;
- 5.2. des exigences de transparence et de contrôle appropriées, y compris de contrôle humain effectif, afin de détecter, de prévenir et de combattre les discriminations résultant d'activités menées au cours du cycle de vie des systèmes d'IA ou aggravées par celles-ci ;
- 5.3. des cadres clairs en matière de responsabilité et d'obligation de rendre des comptes afin de prévenir et combattre les discriminations qui résultent des activités menées dans le cadre du cycle de vie des systèmes d'IA ;
- 5.4. la protection efficace et sur un pied d'égalité des droits humains, y compris les droits à la vie privée et à la protection des données ; et
- 5.5. des voies de recours effectives pour lutter contre la discrimination résultant d'activités menées dans le cadre du cycle de vie des systèmes d'IA.

6. Ces mesures devraient être conçues, mises en œuvre et contrôlées en respectant le principe de proportionnalité, en fonction du niveau de risque de discrimination ou de l'impact sur d'autres droits de l'individu, tels qu'ils sont régulièrement évalués par les acteurs concernés dans les États membres, avec un contrôle parlementaire approprié le cas échéant.

1. Étant donné que tous les êtres humains appartiennent à la même espèce, le Conseil de l'Europe rejette les théories fondées sur l'existence de différentes « races ». Toutefois, cette recommandation utilise le terme « race » afin d'assurer que les personnes qui sont généralement et erronément perçues comme « appartenant à une autre race » ne soient pas exclues de sa protection.

7. Les États membres devraient prendre des mesures nécessaires afin de veiller à ce que l'ensemble des membres de la société aient les mêmes possibilités de bénéficier des systèmes d'IA. En particulier, les États membres devraient, conformément aux obligations positives qui leur incombent en vertu de la Convention, élaborer et veiller à l'application de mesures visant :

- 7.1. à prévenir et à compenser les désavantages subis par les personnes ;
- 7.2. à faciliter la pleine participation de ces dernières dans tous les domaines de la vie, par exemple par des actions positives ; et
- 7.3. à garantir les droits de tous les individus et groupes sans discrimination.

8. Lors de l'élaboration, de l'adoption et de la mise en œuvre des mesures relatives à la promotion de l'égalité, y compris de l'égalité de genre, et à l'interdiction de la discrimination dans les systèmes d'IA, les États membres sont encouragés à organiser des consultations publiques périodiques, en temps utile et accessibles, avec une participation équilibrée entre les femmes et les hommes, en veillant à la participation effective et pleinement pertinente des parties prenantes telles que les partenaires sociaux, les autorités compétentes, les chercheuses et chercheurs, les organes de promotion de l'égalité, les institutions nationales des droits humains, les médiatrices et médiateurs, la société civile ainsi que l'industrie. Une attention particulière devrait être accordée à la représentation effective des groupes les plus touchés par l'inégalité et la discrimination.

III. Promotion de l'égalité par les acteurs publics et privés

9. Les États membres devraient encourager les fournisseurs et les déployeurs publics et privés des systèmes d'IA à prendre action pour promouvoir l'égalité, y compris l'égalité de genre, dans les activités qu'ils mènent dans le cadre du cycle de vie des systèmes d'IA, le cas échéant.

10. Les États membres devraient :

- 10.1. prendre des mesures afin d'encourager et de soutenir les fournisseurs et les déployeurs publics et privés de systèmes d'IA, en leur qualité d'employeurs, à promouvoir et à préserver l'égalité de genre et la diversité au sein de leur personnel à tous les niveaux, notamment parmi les personnes qui conçoivent, développent et déploient des systèmes d'IA, par le biais de l'équilibre entre les femmes et les hommes et d'une meilleure représentation des personnes issues de groupes sous-représentés à tous les niveaux décisionnels ; et
- 10.2. prendre les mesures nécessaires afin de veiller à ce que les déployeurs publics et privés des systèmes d'IA préviennent et combattent toute forme d'exclusion, de discrimination et de violence sur le lieu de travail, et les encourager à faciliter la conciliation entre vie privée et vie professionnelle et à combattre les stéréotypes et l'intolérance.

11. Les États membres devraient encourager les initiatives des fournisseurs et des déployeurs de systèmes d'IA, ainsi que des chercheuses et des chercheurs et de la société civile, qui visent à étudier comment les systèmes d'IA peuvent être utilisés pour promouvoir l'égalité et l'inclusion, et pour détecter, prévenir et combattre les discriminations, les discours de haine et la violence, y compris la violence fondée sur le genre.

12. Les États membres devraient promouvoir une utilisation des systèmes d'IA inclusive et transformative afin de promouvoir l'égalité, y compris l'égalité de genre, en veillant à ce que la législation, les politiques et les cadres réglementaires, administratifs et autres pertinents créent les conditions nécessaires en vue d'initiatives dans ce domaine et envisagent de promouvoir leur financement à un niveau adéquat. Ces initiatives peuvent comprendre :

- 12.1. l'identification et la correction de biais techniques et sociaux précédemment non détectés dans des jeux de données, des conceptions, des processus et des critères de décision, et de schémas de résultats discriminatoires ;
- 12.2. la détection des discours de haine en ligne, y compris des discours de haine sexistes, racistes et LGBTIphobes, la prévention de leur diffusion et la lutte contre ce phénomène, et plus généralement des stéréotypes et de la violence facilitée par la technologie à l'égard des femmes et des filles, ainsi que d'autres individus, sur la base de leurs caractéristiques personnelles réelles ou perçues et de leurs statuts.

IV. Prévention de la discrimination

13. Les États membres devraient veiller à ce que la discrimination et les risques de discrimination résultant des systèmes d'IA soient identifiés et évalués, et ils devraient prendre des mesures graduées et différenciées pour prévenir et atténuer ces risques, et y remédier.
14. Pour éclairer l'élaboration des politiques et les réformes législatives, les États membres sont encouragés :
 - 14.1. à mener des études sur les formes de discrimination existantes et nouvelles qui apparaissent au cours du cycle de vie des systèmes d'IA, et à soutenir de telles études, afin d'assurer une identification et une compréhension approfondies des différents biais et de leur impact sur les individus, les groupes et la société, le cas échéant;
 - 14.2. à revoir et, le cas échéant, à amender leurs politiques et leurs cadres juridiques en matière de collecte de données sur l'égalité afin de faciliter l'identification, la prévention et l'atténuation des préjugés et des risques de discrimination, tout en respectant les droits à la vie privée et à la protection des données à caractère personnel.

Transparence, explicabilité et contrôle en matière d'égalité et de non-discrimination

15. Les États membres devraient imposer, conformément à une approche graduée et différenciée, aux fournisseurs et aux déployeurs de systèmes d'IA :
 - 15.1. de documenter les activités pertinentes au cours du cycle de vie des systèmes d'IA et de conserver cette documentation ;
 - 15.2. de communiquer efficacement les informations appropriées sur les systèmes d'IA et d'expliquer le rôle de ces derniers dans la prise de décision ainsi que la portée et les effets des décisions de traitement différencié aux personnes affectées ;
 - 15.3. de faciliter l'identification des contenus générés par les systèmes d'IA et, lorsqu'un système d'IA est destiné à interagir directement avec des personnes physiques, d'informer ces dernières dans un langage et un format accessibles sur le fait qu'elles interagissent avec un système d'IA de manière appropriée au contexte spécifique ; et
 - 15.4. de collecter et de mettre à la disposition des autorités compétentes et, le cas échéant, des personnes affectées, dans un langage et un format accessibles, notamment :
 - les données sur le fonctionnement des systèmes d'IA ;
 - les critères de décision utilisés ; et
 - les informations sur les données d'entraînement, de validation et de test utilisées, ainsi que sur leur traitement.
16. Les États membres devraient contribuer à, maintenir ou envisager la mise en place de registres publics répertoriant les systèmes d'IA ayant un impact significatif sur l'exercice des droits humains, pouvant contenir un résumé des évaluations d'impact sur les droits humains, notamment des considérations spécifiques sur les risques pour l'égalité, y compris l'égalité de genre, et la non-discrimination.
17. Les États membres devraient veiller à ce que les personnes affectées aient un accès approprié aux informations relatives à la prise de décision qui affecte leurs droits afin de leur permettre de faire valoir leurs droits de manière efficace. L'accès approprié à l'information devrait permettre de remédier efficacement aux asymétries d'information entre les personnes affectées et les déployeurs et fournisseurs de systèmes d'IA.
18. Les États membres devraient veiller à ce que des garanties procédurales et des recours effectifs soient accessibles aux personnes affectées, y compris des procédures de recours et de réparation en cas de suspicion de discrimination, et à ce que les systèmes d'IA ayant un impact significatif sur l'exercice des droits humains fassent l'objet d'un contrôle humain.
19. Les États membres devraient envisager d'accorder aux organismes de promotion de l'égalité, aux institutions nationales des droits humains, aux médiatrices et médiateurs, et aux autorités compétentes le droit d'accéder aux informations et de tester les systèmes d'IA, pour leur permettre de contrôler efficacement s'il y a eu discrimination.

Évaluation d'impact, certification, évaluation et suivi en matière d'égalité et de non-discrimination

20. Les États membres devraient, en suivant une approche graduée et différenciée :

- 20.1. mettre en place ou poursuivre des évaluations des risques et des impacts du point de vue des droits humains, que les fournisseurs et les déployeurs de systèmes d'IA devront effectuer avant le déploiement et de manière itérative tout au long du cycle de vie du système d'IA, en tenant compte en particulier des risques pour l'égalité, y compris l'égalité de genre, et la non-discrimination ;
- 20.2. encourager vivement les fournisseurs et déployeurs de systèmes d'IA à réaliser des tests continus des systèmes d'IA et de leur utilisation afin de réduire les risques de discrimination et les biais susceptibles d'avoir un impact négatif sur les droits humains ou de mener à des discriminations ;
- 20.3. commencer à élaborer des normes pour l'audit des systèmes d'IA en matière de discrimination et encourager fortement les fournisseurs et les déployeurs à procéder à des audits ;
- 20.4. envisager, avec la participation des autorités compétentes, la mise en place de programmes de certification des systèmes d'IA et de leurs résultats, du point de vue de l'égalité et de la non-discrimination ; et
- 20.5. veiller à ce que les fournisseurs ou les déployeurs de systèmes d'IA documentent toutes les mesures qu'ils ont prises pour prévenir ou atténuer les biais ou les discriminations, pour que ces informations puissent être communiquées aux personnes affectées, aux organismes de promotion de l'égalité, aux institutions nationales des droits humains, aux médiatrices et médiateurs, aux autorités compétentes et, s'il y a lieu, à la société civile, en particulier en cas de procédure judiciaire.

Obligation de rendre des comptes, responsabilité et règles relatives à la charge de la preuve dans le domaine de l'égalité et de la non-discrimination

21. Les États membres devraient veiller, en fonction du contexte national, à ce que des procédures judiciaires et/ou administratives facilement accessibles, notamment des procédures de conciliation, et/ou des mécanismes non judiciaires soient ouverts à toutes les victimes de discrimination résultant d'activités menées dans le cadre du cycle de vie des systèmes d'IA. Les États membres sont également encouragés à veiller à ce que, dans les cas urgents, des procédures accélérées prévoyant des décisions provisoires pour les victimes présumées de telles discriminations soient disponibles.

22. Les États membres sont encouragés à fournir une assistance juridique gratuite, conformément aux normes internationales et aux conditions prévues par le droit national, aux personnes dont les droits pourraient avoir été violés ou risquent de l'être en raison d'une discrimination résultant d'activités menées dans le cadre du cycle de vie des systèmes d'IA.

23. Les États membres devraient envisager de mettre en place des moyens d'action collective pour les victimes de discrimination résultant des activités menées dans le cadre du cycle de vie des systèmes d'IA, conformes à leurs systèmes judiciaires nationaux, permettant d'ouvrir des enquêtes sur la discrimination et d'obtenir réparation, y compris pour les représentant-es des groupes concernés.

24. Les États membres devraient déterminer si les règles existantes en matière de responsabilité s'appliquent et sont efficaces dans le contexte des systèmes d'IA. Le cas échéant, les États membres pourront évaluer la possibilité d'établir une responsabilité stricte pour les fournisseurs et les déployeurs de systèmes d'IA lorsque la discrimination est causée par des activités menées dans le cadre du cycle de vie de ces systèmes, et par ailleurs veiller à ce que des règles claires et efficaces soient en place.

25. Les États membres devraient appliquer les règles de recherche de preuve existantes, ou envisager d'instaurer de nouvelles règles de recherche de preuve, y compris celles prévoyant un renversement de la charge de la preuve, notamment dans les situations où il est allégué que les fournisseurs ou les déployeurs de systèmes d'IA n'ont pris aucune mesure préventive pour éliminer des biais, et, en cas de non-respect de l'obligation, de communiquer des informations.

26. Les États membres devraient veiller à la disponibilité de sanctions effectives, proportionnées et dissuasives, et de voies de recours effectives y compris l'indemnisation pour le préjudice moral ou le préjudice matériel causé par une discrimination résultant d'activités menées dans le cadre du cycle de vie des systèmes d'IA.

27. Les États membres devraient instaurer des obligations de suivi afin d'empêcher la réapparition de discriminations résultant d'activités menées dans le cadre du cycle de vie des systèmes d'IA et, s'il y a lieu, évaluer la nécessité d'un moratoire, d'une interdiction ou d'autres mesures appropriées concernant certaines utilisations de systèmes d'IA en raison du non-respect de l'interdiction de la discrimination ou du principe d'égalité.

28. Les États membres devraient soutenir la coopération entre les organismes de promotion de l'égalité, les institutions nationales des droits humains, les médiatrices et médiateurs, les autorités de protection des données et les autres autorités compétentes, pour permettre la mise en commun d'informations et de pratiques relatives aux biais et aux discriminations tout au long du cycle de vie des systèmes d'IA.

V. Participation démocratique, sensibilisation du public et renforcement des capacités

29. Les États membres devraient veiller à ce que les activités menées par les autorités publiques dans le cadre du cycle de vie des systèmes d'IA fassent l'objet d'un contrôle parlementaire et public suffisant, y compris, le cas échéant, sous la forme de débats parlementaires réguliers.

30. Les États membres devraient :

30.1. sensibiliser le public aux risques et aux avantages que représentent les systèmes d'IA pour l'égalité, y compris l'égalité de genre, et la non-discrimination ;

30.2. veiller à l'existence et à l'accessibilité de programmes de maîtrise du numérique adaptés à l'âge, destinés aux individus de tout âge, afin de les sensibiliser davantage aux questions d'égalité et de non-discrimination dans le contexte des systèmes d'IA ; et

30.3. encourager la mise à disposition de formations et le renforcement des capacités, ciblés et accessibles, notamment sur la lutte contre les biais, les stéréotypes et les risques de discrimination ainsi que sur la promotion d'une participation égale des femmes et des hommes, et des individus et groupes exposés à la discrimination liée aux activités menées dans le cadre du cycle de vie des systèmes d'IA.

31. Les États membres devraient prendre des mesures pour promouvoir l'accès aux technologies et leur utilisation sur un pied d'égalité, et supprimer les obstacles qui s'y opposent pour l'ensemble des membres de la société, le cas échéant de concert ou en partenariat avec les acteurs privés et la société civile.

32. Les États membres devraient veiller à la participation effective et pertinente de la société civile à l'élaboration des normes relatives à l'IA et, s'il y a lieu, organiser des consultations publiques.

VI. Compétences et pouvoirs des organes de contrôle indépendants

33. Les États membres devraient veiller à ce que les organes indépendants chargés de superviser l'application et la mise en œuvre de la législation en matière de lutte contre la discrimination, tels que les organes de promotion de l'égalité et, le cas échéant, les autres structures nationales des droits humains (« organes de contrôle indépendants ») disposent des ressources humaines, techniques et financières nécessaires, et le cas échéant de l'appui technique ainsi que de l'indépendance opérationnelle pour accomplir leurs tâches et exercer leurs compétences efficacement dans le contexte des systèmes d'IA.

34. Les États membres devraient veiller, en fonction du contexte national, à ce que les organes de contrôle indépendants aient pour mission :

34.1. de promouvoir l'égalité et de prévenir les discriminations causées par l'IA, ce qui peut inclure le signalement des utilisations des systèmes d'IA pouvant être considérées comme exigeant un moratoire ou une interdiction ;

34.2. d'assister les individus et les groupes exposés à la discrimination, y compris en recevant des plaintes et en fournissant des informations ;

34.3. d'engager des procédures judiciaires au nom d'une ou plusieurs victimes ou en leur nom propre pour des raisons d'intérêt public, ou de participer à des procédures judiciaires pour soutenir une ou plusieurs victimes de discrimination. Le droit d'agir dans les procédures judiciaires devrait inclure le droit de présenter des observations au tribunal, conformément à la législation et aux pratiques nationales ; et

34.4. de dispenser des conseils sur la promotion de l'égalité et la prévention des risques de discrimination, ainsi que sur la manière d'intégrer des garanties tout au long du cycle de vie des systèmes d'IA.

Ces institutions peuvent aussi avoir pour mission de statuer sur les plaintes individuelles et collectives liées à des discriminations causées par l'IA.

35. Les États membres devraient consulter périodiquement les organes de contrôle indépendants dans le cadre du processus d'élaboration de la législation, des politiques et des cadres réglementaires relatifs à la promotion de l'égalité, y compris l'égalité de genre, ainsi qu'à la prévention et à la répression de la discrimination en lien avec des activités menées dans le cadre du cycle de vie des systèmes d'IA.

36. Les États membres devraient envisager d'habiliter les organes de contrôle indépendants à mener des enquêtes afin de pouvoir contrôler efficacement si des activités menées dans le cadre du cycle de vie des systèmes d'IA ont entraîné une discrimination.

37. Lorsque sont adoptés des cadres réglementaires comprenant des structures chargées de tester et d'expérimenter des technologies, produits, services ou approches nouveaux et innovants dans le contexte des systèmes d'IA, les États membres devraient mettre en place ou pérenniser les cadres institutionnels et financiers relatifs à la coopération entre ces structures et les organes de contrôle indépendants.

VII. Dispositions complémentaires

38. Lors de la mise en œuvre des principes et orientations énoncés dans la présente recommandation, les États membres devraient prendre des mesures, notamment en intégrant l'égalité et l'égalité de genre, en adoptant une approche graduée et différenciée, pour remédier aux effets particulièrement importants et spécifiques que peut avoir sur l'égalité, y compris l'égalité de genre, et la non-discrimination, l'utilisation de systèmes d'IA dans différents secteurs. Ces mesures peuvent être nécessaires lorsque l'impact sur les personnes affectées est d'entraver la jouissance de leurs droits ou libertés ou leur droit à un résultat, une prestation ou un service particulier, notamment dans les domaines de l'éducation, de la justice et du pouvoir judiciaire, des services répressifs, des processus électoraux, du travail et de l'emploi, des services sociaux, du logement et de la protection sociale, des soins de santé, des migrations, ainsi que de la banque, de la finance, de l'assurance et autres services. De plus, les impacts négatifs potentiels de l'IA dans d'autres domaines devraient être contrôlés et, le cas échéant, traités, notamment dans les domaines des médias et de la publicité, du sport et de la culture.

39. Les États membres, en ayant bien présent à l'esprit qu'ils sont liés par leurs obligations existantes en matière de droits humains, notamment celles relatives à la non-discrimination et à l'égalité, y compris l'égalité de genre, ne sont pas tenus d'appliquer la recommandation aux activités relevant du cycle de vie des systèmes d'IA qui concernent directement la protection de leurs intérêts nationaux en matière de sécurité.

Exposé des motifs

PRÉAMBULE

1. La présente Recommandation a pour objet de fournir aux États membres, en un seul endroit, un ensemble d'orientations sur la manière de promouvoir l'égalité, y compris l'égalité de genre, et de prévenir et de combattre la discrimination dans les activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle (« IA »). Pour cela, la Recommandation s'appuie sur les instruments juridiques existants qui constituent le fondement juridique de la promotion de l'égalité et de l'interdiction de la discrimination, dont beaucoup ne sont pas propres au contexte dans lequel s'inscrivent les systèmes d'IA², et elle vise à optimiser la manière dont ces instruments juridiques sont appliqués aux opportunités qui se présentent et aux défis qui se posent dans ce domaine. Cette Recommandation vise à mettre en œuvre ce cadre existant d'une manière qui encourage l'innovation au service de la promotion de l'égalité et de la prévention et de la lutte contre la discrimination dans le domaine des systèmes d'IA.

2. La présente Recommandation se fonde sur des normes contraignantes et des normes non contraignantes. En particulier, elle s'inscrit dans le cadre juridique existant propre aux systèmes d'IA, notamment la [Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit](#) (STCE n° 225) (« la Convention-cadre »)³ et le règlement correspondant de l'Union européenne (« le règlement de l'UE sur l'intelligence artificielle » ou « le règlement sur l'IA »)⁴. Dans la Convention-cadre, l'égalité, y compris l'égalité de genre, et la non-discrimination figurent au chapitre des principes généraux, avec d'autres normes relatives aux droits humains. Le rapport explicatif de la Convention-cadre souligne l'importance pour les États membres de dûment prendre en compte les aspects touchant à l'égalité, compris l'égalité de genre, et à la non-discrimination dans leur évaluation des risques et des impacts potentiels des systèmes d'IA sur les droits humains, la démocratie et l'État de droit, et lors de l'adoption ou le maintien de mesures appropriées pour faire face à ces risques et impacts potentiels. À cet égard, le rapport explicatif de la Convention-cadre reconnaît que les États membres devraient tenir compte des asymétries de pouvoir susceptibles d'aggraver les inégalités existantes et les effets sur la société. Il note également le risque réel et bien documenté de biais pouvant constituer une discrimination illicite découlant des activités menées dans le cadre du cycle de vie des systèmes d'IA. Cela étant, si la Recommandation adopte la terminologie de la Convention-cadre et s'attache à maintenir une cohérence avec son champ d'application et son contenu, elle n'est pas un instrument de mise en œuvre et ne devrait pas être vue comme telle. Les références à la Convention-cadre et à son rapport explicatif dans le présent exposé des motifs sont propres à ce document. Elles n'ont pas – et ne sont pas censées avoir – d'influence sur l'interprétation de la Convention-cadre.

3. La Recommandation rappelle également l'importance d'autres normes mondiales pertinentes pour la promotion de l'égalité et la prévention de la discrimination, en particulier la [Convention internationale sur l'élimination de toutes les formes de discrimination raciale](#) (CIEDR), la [Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes](#) (CEDAW), la [Convention relative aux droits de l'enfant](#) (CRC) et la [Convention relative aux droits des personnes handicapées](#) (CDPH), ainsi que les normes auxquelles il est fait référence dans le rapport explicatif de la Convention-cadre⁵.

-
2. Entre autres, la [Convention de sauvegarde des droits de l'homme et des libertés fondamentales](#) (STE n° 005), article 14; le [Protocole n° 12 à la Convention de sauvegarde des droits de l'homme et des libertés fondamentales](#) (STE n° 177), article 1^{er}; la [Charte sociale européenne](#) (STE n° 035) et la [Charte sociale européenne \(révisée\)](#) (STE n° 163), article 20, article E; la [Convention du Conseil de l'Europe sur la prévention et la lutte contre la violence à l'égard des femmes et la violence domestique](#) (« la Convention d'Istanbul ») (STCE n° 210), article 4; la [Convention-cadre pour la protection des minorités nationales](#) (STE n° 157), articles 4, 6(2), et 9; la [Charte européenne des langues régionales ou minoritaires](#) (STE n° 148), article 7(2); la [Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel](#) (STE n° 108) et le [Protocole d'amendement à la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel](#) (STCE n° 223).
3. Ouvert à la signature le 5 septembre 2024 à Vilnius, Lituanie.
4. Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 ([règlement sur l'intelligence artificielle](#)), (« le règlement de l'UE sur l'IA »).
5. Rapport explicatif de la Convention cadre, paragraphes 72 et 73, y compris : le Titre III de la [Charte des droits fondamentaux de l'Union européenne](#); les traités de l'Union européenne (par exemple, le [Traité sur l'Union européenne](#), Article 2, le [Traité sur le fonctionnement de l'Union européenne](#), Article 10), le droit dérivé de l'Union européenne et la jurisprudence pertinente de la Cour de justice de l'Union européenne; le [Pacte international relatif aux droits civils et politiques](#) (PIDCP), articles 2, 24 et 26, et le [Pacte international relatif aux droits économiques, sociaux et culturels](#) (PIDESC), articles 2, 3 et 7.

4. En outre, il existe un vaste corpus de normes non contraignantes du Conseil de l'Europe qui sont pertinentes pour la prévention et la lutte contre la discrimination et pour la promotion de l'égalité, en particulier l'égalité de genre. À cet égard, la Recommandation [CM/Rec\(2019\)1](#) sur la prévention et la lutte contre le sexisme invite les États membres à «Intégrer une perspective d'égalité entre les femmes et les hommes dans toutes les politiques, programmes et recherches en matière d'intelligence artificielle afin d'éviter les risques de perpétuation du sexisme et des stéréotypes de genre, et d'examiner comment l'intelligence artificielle pourrait aider à combler les écarts entre les femmes et les hommes et à éliminer le sexisme⁶.» La Recommandation [CM/Rec\(2026\)2](#) sur l'obligation de rendre des comptes en matière de la violence à l'égard des femmes et des filles facilitée par la technologie traite des risques spécifiques posés par les nouvelles technologies, notamment par les systèmes d'IA. Les normes traitant de la promotion de l'égalité de manière plus générale qui ont servi de base à la rédaction de la Recommandation comprennent également, en particulier, la Recommandation [CM/Rec\(2022\)16](#) du Comité des Ministres aux États membres sur la lutte contre le discours de haine, laquelle reconnaît que les mesures législatives et politiques destinées à prévenir et à combattre le discours de haine en ligne devraient être régulièrement réexaminées afin de prendre en compte l'évolution rapide de la technologie et des services en ligne et, plus généralement, les technologies numériques et leur influence sur les flux d'informations et de communications dans les sociétés démocratiques contemporaines. Elles incluent également la Recommandation [CM/Rec\(2024\)4](#) du Comité des Ministres aux États membres sur la lutte contre les crimes de haine qui reconnaît que la prévention et la lutte contre les crimes de haine, en ligne ou hors ligne, nécessitent une approche globale et multiforme associant les personnes qui travaillent au sein des institutions publiques et des organisations de la société civile ainsi que les personnes qui appartiennent à des groupes risquant d'être visés par des crimes de haine ou qui travaillent avec ces groupes. La Recommandation [CM/Rec\(2010\)5](#) du Comité des Ministres aux États membres sur des mesures visant à combattre la discrimination fondée sur l'orientation sexuelle ou l'identité de genre appelle les États membres à prendre des mesures appropriées pour combattre toutes les formes d'expression, y compris dans les médias et sur Internet, susceptibles d'être raisonnablement comprises comme susceptibles de produire l'effet d'inciter, de propager ou de promouvoir la haine ou d'autres formes de discrimination à l'encontre des personnes lesbiennes, gay, bisexuelles et transgenres.

5. La Recommandation se situe dans le prolongement des engagements pris par les chef-fes d'État et de gouvernement en conclusion du 4^e Sommet («[la Déclaration de Reykjavík](#)»)⁷, notamment la reconnaissance de la nécessité d'assurer l'égalité et de lutter contre toute forme de discrimination, d'intégrer une perspective de jeunesse et de renforcer les travaux en faveur de sociétés inclusives exemptes de marginalisation, d'exclusion, de racisme et d'intolérance, ainsi que d'assurer la participation pleine, égale et effective des femmes aux processus décisionnels publics et privés. Dans la Déclaration de Reykjavík, les chef-fes d'État et de gouvernement reconnaissent l'impact positif des technologies numériques nouvelles et émergentes ainsi que les opportunités créées par ces technologies, tout en reconnaissant la nécessité d'atténuer les risques de conséquences négatives de leur utilisation sur les droits humains, la démocratie et l'État de droit, et s'engagent à assurer un rôle de premier plan au Conseil de l'Europe dans l'élaboration de normes à l'ère du numérique afin de sauvegarder les droits humains en ligne et hors ligne.

6. Le préambule de la Recommandation souligne l'importance de construire des sociétés inclusives exemptes de marginalisation, d'exclusion, de sexisme, de racisme, de LGBTIphobie et d'autres formes d'intolérance, y compris l'âgisme et le validisme. Cette liste n'est pas exhaustive. Elle comprend également des formes particulières de racisme comme l'antitsiganisme, l'antisémitisme, la haine contre les musulman-es et d'autres formes d'intolérance religieuse. La Recommandation prend appui sur les Recommandations de politique générale (RPG) de la Commission européenne contre le racisme et l'intolérance (ECRI), notamment la [RPG n° 17](#) sur la prévention et la lutte contre l'intolérance et la discrimination envers les personnes LGBTI, qui souligne, entre autres, que si les avancées technologiques peuvent améliorer la jouissance des droits humains des personnes LGBTI et leur participation à la vie sociale, certaines formes de technologie – comme l'intelligence artificielle, y compris la reconnaissance faciale et les systèmes algorithmiques – peuvent également être utilisées pour porter atteinte à la vie privée, à la sécurité et à l'égalité des personnes LGBTI⁸. La Recommandation fait référence aux personnes et groupes exposés à la discrimination afin d'inclure toutes celles et ceux qui ont été victimes de discrimination ainsi que celles et ceux qui pourraient l'être à l'avenir.

6. Recommandation [CM/Rec\(2019\)1](#) du Comité des Ministres aux États membres sur la prévention et la lutte contre le sexisme, II.B.7, p. 20.

7. [Déclaration de Reykjavík – Unis autour de nos valeurs](#), adoptée lors du 4^e Sommet des chefs d'État et de gouvernement du Conseil de l'Europe (Reykjavík, 16 et 17 mai 2023), p. 7 et 8.

8. [Recommandation de politique générale n° 17 de l'ECRI sur la prévention et la lutte contre l'intolérance et la discrimination envers les personnes LGBTI](#), paragraphe 24.

7. Conformément à la Déclaration de Reykjavík, la Recommandation adopte une approche intersectionnelle, autrement dit une approche qui reconnaît que les différents motifs sur lesquels se fonde la discrimination peuvent se combiner et donner lieu à des expériences de vie et des vulnérabilités singulières. Une approche intersectionnelle peut être adoptée pour comprendre, prendre en compte et traiter les interactions entre différentes caractéristiques ou situations personnelles ainsi que les formes aggravées de discrimination qui en résultent. En tant que telle, cette approche peut aider à comprendre les formes complexes de discrimination, d'exclusion et de violence auxquelles les individus peuvent être exposés et ainsi permettre de détecter des discriminations qui, autrement, pourraient passer inaperçues. Si tout le monde est exposé à la discrimination intersectionnelle, certains groupes le sont tout particulièrement et sont donc particulièrement susceptibles de bénéficier d'une approche intersectionnelle⁹. Si la terminologie de la discrimination intersectionnelle est de plus en plus reconnue tant au sein du Conseil de l'Europe et de ses États membres qu'au niveau international. Par exemple dans la directive sur la transparence des rémunérations et dans les directives relatives aux normes applicables aux organismes pour l'égalité de traitement de l'Union européenne¹⁰ –, certains États membres parlent également de « discrimination multiple » ou « cumulative » pour exprimer une préoccupation similaire. Pour refléter le fait que de nombreux États membres utilisent le terme « discrimination multiple » dans leurs législations, politiques et pratiques nationales, la Recommandation emploie donc la terminologie « discrimination multiple et intersectionnelle ».

8. Le préambule fait référence aux substituts de caractéristiques ou situations personnelles. Le concept de discrimination par procuration (*discrimination by proxy* en anglais) est défini et expliqué plus en détail au paragraphe 2 de la Recommandation et au paragraphe correspondant 19 de l'exposé des motifs.

9. La Recommandation reconnaît qu'une conséquence du développement de l'IA et de son utilisation généralisée est l'augmentation de la violence à l'égard des femmes et des filles facilitée par la technologie, et fait ainsi référence à la Recommandation [CM/Rec\(2026\)2](#) sur l'obligation de rendre des comptes en matière de violence facilitée par la technologie à l'égard des femmes et des filles et à la [Convention du Conseil de l'Europe sur la prévention et la lutte contre la violence à l'égard des femmes et la violence domestique](#) (« la Convention d'Istanbul ») (STCE n° 210). La Recommandation souligne également les besoins et les vulnérabilités particulières des enfants ainsi que l'impact spécifique de l'IA sur eux. La Recommandation reconnaît l'importance de la [Convention du Conseil de l'Europe sur la protection des enfants contre l'exploitation et les abus sexuels](#) (Convention de Lanzarote, STCE n° 201), qui fournit un cadre pour la protection des enfants contre l'exploitation et les abus sexuels, y compris ceux commis par le biais des technologies de l'information et de la communication. La Recommandation reconnaît également que l'IA peut avoir un impact particulier sur les jeunes dans les domaines du bien-être, de l'emploi, de l'éducation et des droits civiques¹¹.

10. La Recommandation s'appuie sur les conclusions de l'*Étude sur l'impact des systèmes d'intelligence artificielle, leur potentiel de promotion de l'égalité et les risques qu'ils peuvent entraîner en matière de non-discrimination* (« l'Étude »)¹². En particulier, l'Étude fait le constat que les solutions technologiques ne suffisent pas, à elles seules, à lutter contre la discrimination et à promouvoir l'égalité, y compris l'égalité de genre, et recommande que l'égalité réelle et la non-discrimination dès la conception soient placées au cœur de

9. Voir, par exemple, la [Stratégie du Conseil de l'Europe pour l'égalité de genre \(2024-2029\)](#), le [Modèle de cadre pour une stratégie d'intégration interculturelle au niveau national](#) du Comité d'experts sur l'inclusion interculturelle (ADI-INT), la [Recommandation de politique générale \(RPG\) n° 14](#) sur la lutte contre le racisme et la discrimination raciale dans le monde du travail de la Commission européenne contre le racisme et l'intolérance (ECRI) et son exposé des motifs, la [Recommandation de politique générale n° 5](#) de l'ECRI révisée sur la prévention et la lutte contre le racisme et la discrimination envers les musulmans, le document [Covid-19: une analyse des aspects relatifs à l'anti-discrimination, à la diversité et à l'inclusion dans les États membres du Conseil de l'Europe](#) et les [Lignes directrices du Comité des Ministres du Conseil de l'Europe sur la défense de l'égalité et la protection contre la discrimination et la haine pendant la pandémie de covid-19 et d'autres crises similaires à venir](#).

10. [Directive \(UE\) 2023/970](#) du Parlement européen et du Conseil du 10 mai 2023 visant à renforcer l'application du principe de l'égalité des rémunérations entre les femmes et les hommes pour un même travail ou un travail de même valeur par la transparence des rémunérations et les mécanismes d'application du droit ; [directive \(UE\) 2024/1499](#) du Conseil du 7 mai 2024 relative aux normes applicables aux organismes pour l'égalité de traitement dans les domaines de l'égalité de traitement entre les personnes sans distinction de race ou d'origine ethnique, de l'égalité de traitement entre les personnes en matière d'emploi et de travail sans distinction de religion ou de convictions, de handicap, d'âge ou d'orientation sexuelle et de l'égalité de traitement entre les femmes et les hommes en matière de sécurité sociale ainsi que dans l'accès à des biens et services et la fourniture de biens et services, et modifiant les directives 2000/43/CE et 2004/113/CE ; et [directive \(UE\) 2024/1500](#) du Parlement européen et du Conseil du 14 mai 2024 relative aux normes applicables aux organismes pour l'égalité de traitement dans le domaine de l'égalité de traitement et de l'égalité des chances entre les femmes et les hommes en matière d'emploi et de travail, et modifiant les directives 2006/54/CE et 2010/41/UE.

11. Voir, par exemple, « *Insight into artificial intelligence and its impact on the youth sector* » (en anglais seulement), Veronica Stefan, préparé dans le cadre du partenariat pour la jeunesse entre le Conseil de l'Europe et la Commission européenne, 2024.

12. [Étude sur l'impact des systèmes d'intelligence artificielle, leur potentiel de promotion de l'égalité, y compris l'égalité de genre, et les risques qu'ils peuvent entraîner en matière de non-discrimination](#), élaborée par Ivana Bartoletti et Raphaële Xenidis, et adoptée en 2023 par le Comité directeur sur l'anti-discrimination, la diversité et l'inclusion (CDADI) et la Commission pour l'égalité de genre (GEC).

la réglementation juridique relative au développement et au déploiement de l'IA. Pour cette raison, la Recommandation adopte une approche plus large, qui englobe des orientations en matière de changement social et politique afin d'atteindre l'égalité réelle¹³. La Recommandation envisage une approche multi-parties prenantes¹⁴, avec des recommandations sur mesure destinées à un large éventail d'acteurs¹⁵. En particulier, la Recommandation s'applique aux activités menées dans le cadre du cycle de vie des systèmes d'IA par des acteurs aussi bien publics que privés. Les acteurs du secteur privé devraient respecter les droits humains des consommateur·rices ainsi que de toutes les parties concernées. Par exemple, la Recommandation [CM/Rec\(2016\)3](#) du Comité des Ministres aux États membres sur les droits de l'homme et les entreprises appelle les États membres à prendre des mesures pour permettre la responsabilité des entreprises en matière de protection des droits humains, et les [Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme prévoient la responsabilité des entreprises en matière de protection des droits humains](#). Pour cela, elles devraient, conformément à la Recommandation [CM/Rec\(2020\)1](#) du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme, adopter des modèles de gouvernance souples qui garantissent une réparation rapide et effective ainsi que des possibilités de recours en cas d'incident, en veillant à ce que la responsabilité et l'obligation de rendre des comptes en matière de protection des droits humains soient effectivement et clairement établies tout au long du processus, du stade de la proposition jusqu'au déploiement actuel et futur, à la révision et aux exigences en matière de notification, en passant par l'identification des tâches, la sélection, la collecte et l'analyse des données, et la modélisation et la conception des systèmes. La Recommandation [CM/Rec\(2022\)16](#) du Comité des Ministres aux États membres sur la lutte contre le discours de haine abordent également le rôle des intermédiaires internet dans la lutte contre le discours de haine en ligne¹⁶.

11. En ce qui concerne le suivi, la documentation et les rapports, il est possible que les États membres mettent déjà en œuvre la Recommandation en se conformant à leurs obligations au titre du règlement de l'UE sur l'IA et d'autres instruments internationaux pertinents en matière d'IA. La Recommandation applique une approche graduée et différenciée, dans le souci de maintenir une cohérence avec l'approche graduée et différenciée de la Convention-cadre et l'approche fondée sur les risques du règlement de l'UE sur l'IA.

ANNEXE

Principes et orientations sur l'égalité et l'intelligence artificielle

I. But

Paragraphe 1

12. La Recommandation porte sur la promotion de l'égalité, y compris l'égalité de genre, et sur les risques de discrimination. Elle prend appui sur les très nombreux travaux menés de longue date par le Conseil de l'Europe dans les domaines de l'égalité, y compris l'égalité de genre, et de la non-discrimination, ainsi que sur les normes existantes, pour relever les nouveaux défis posés par le développement des systèmes d'IA¹⁷. Si la Recommandation s'adresse principalement aux États membres et à leurs institutions, elle comprend également des recommandations destinées aux acteurs du secteur privé, qui tiennent compte des spécificités de leurs activités¹⁸. Pour que la Recommandation reste pertinente face à l'évolution rapide des technologies, ses principes et ses lignes directrices sont neutres sur le plan technologique et conçus pour s'appliquer aussi bien aux systèmes d'IA innovants et futurs qu'aux systèmes existants.

13. La Recommandation utilise les termes « systèmes d'intelligence artificielle » conformément à l'article 2 de la Convention-cadre, qui prévoit qu'un « système d'intelligence artificielle » est un système automatisé qui,

13. *Ibid.*, p. 10, 28-30, 78 et 88.

14. Exemples de documents recommandant cette approche : [Les droits humains dès la conception de l'IA – une protection durable des droits humains à l'ère de l'intelligence artificielle](#), Commissaire aux droits de l'homme du Conseil de l'Europe, 2023, p. 18-20, recommandation de suivi à [Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme](#), 2019.

15. Il s'agit, entre autres, des acteurs étatiques, des représentant·es du secteur privé, des milieux universitaires et des instituts de recherche, des organismes de promotion de l'égalité, des institutions nationales des droits humains, des médiatrices et médiateurs, de la société civile, notamment de la société civile de la jeunesse, et des associations de minorités (y compris celles de petite taille), en tenant dûment compte de la diversité au sein de ces groupes.

16. Paragraphes 16 et suivants.

17. Voir en particulier la Recommandation [CM/Rec\(2020\)1](#) du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme.

18. Voir aussi le [Rapport explicatif de la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'état de droit](#) (STCE n° 225), paragraphe 17, et la Recommandation [CM/Rec\(2020\)1](#) sur les impacts des systèmes algorithmiques sur les droits de l'homme, paragraphe 3.

pour des objectifs explicites ou implicites, déduit, à partir d'entrées reçues, comment générer des résultats en sortie tels que des prévisions, des contenus, des recommandations ou des décisions qui peuvent influencer sur des environnements physiques ou virtuels. Après leur déploiement, les systèmes d'IA présentent des degrés variables d'autonomie et d'adaptabilité.

14. Comme le souligne le rapport explicatif de cette dernière, la définition « reflète une compréhension large de ce que sont les systèmes d'intelligence artificielle, notamment par opposition à d'autres types de systèmes de logiciel plus simples, basés sur des règles définies uniquement par des personnes physiques pour exécuter automatiquement des opérations » et « elle vise à garantir la précision et la sécurité juridiques, tout en restant suffisamment abstraite et flexible pour rester valable malgré les développements technologiques futurs »¹⁹. La définition de la Convention-cadre s'inspire de celle adoptée en 2023 par l'Organisation de coopération et de développement économiques (OCDE), le but étant de renforcer la coopération internationale dans le domaine de l'IA et de faciliter le travail d'harmonisation de la gouvernance de l'IA au niveau mondial²⁰.

15. La Recommandation utilise l'expression « activités menées dans le cadre du cycle de vie des systèmes d'IA », qui est utilisée dans la Convention-cadre afin d'inclure « toutes les activités depuis la conception d'un système d'intelligence artificielle jusqu'à sa mise hors service, quel que soit l'acteur qui y participe » (paragraphe 15, rapport explicatif de la Convention-cadre). Ceci apporte l'assurance d'une approche globale qui résiste à l'épreuve du temps. Cette définition vise ainsi à englober la totalité des activités, depuis la conception d'un système d'intelligence artificielle à sa mise hors service, quel que soit l'acteur qui mène l'activité concernée. À cette fin, la définition inclut 1) la planification et la conception, 2) la collecte et le traitement de données, 3) le développement de systèmes d'intelligence artificielle, y compris la construction de modèles et/ou l'adaptation de modèles existants à des tâches spécifiques, 4) les essais, la vérification et la validation, 5) la fourniture/mise à disposition des systèmes en vue de leur utilisation, 6) le déploiement, 7) l'exploitation et le suivi, et 8) la mise hors service²¹.

16. La Recommandation utilise les termes « déployeur » et « fournisseur » (tous deux tirés du règlement de l'UE sur l'IA) d'un système d'IA. Les « personnes affectées » sont celles qui sont soumises à des décisions ou des processus assistés par l'IA, par exemple les client·es, les candidat·es à un emploi ou les contribuables, que la décision finale soit prise par un système d'IA ou pas²². Aux fins de la présente Recommandation, une personne affectée peut également être une personne morale (par exemple, lorsqu'une entité est ciblée en raison des caractéristiques protégées de ses membres).

Paragraphe 2

17. La liste des motifs de discrimination figurant dans la Recommandation est d'abord apparue dans la [Stratégie pour l'égalité de genre \(2024-2029\)](#) et elle constitue une liste non-exhaustive et indicative. Elle est inspirée de l'article 14 de la [Convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales](#) (« la Convention ») et intègre des motifs de discrimination qui ont pris de l'importance récemment – que ce soit sur le plan politique, par une interprétation évolutive dans la jurisprudence de la Cour ou au travers de normes nouvelles, parmi lesquelles la [Convention du Conseil de l'Europe sur la prévention et la lutte contre la violence à l'égard des femmes et la violence domestique](#) (« la Convention d'Istanbul ») et la Recommandation [CM/Rec\(2024\)4](#) sur la lutte contre les crimes de haine. La Recommandation couvre la discrimination fondée sur des caractéristiques ou une situation réelles ou perçues, conformément à la jurisprudence de la Cour²³. De même, elle englobe la discrimination par association. Celle-ci se produit lorsqu'un individu ou un groupe est victime d'une discrimination parce qu'il est associé – ou est perçu comme étant associé – à une caractéristique, une situation, une personne ou un groupe particulier. La liste des motifs comprend également la discrimination fondée sur l'origine ethnique. Dans la jurisprudence de la Cour, l'ethnicité et la « race »²⁴ sont des concepts liés et qui se chevauchent²⁵, et la discrimination fondée sur l'origine ethnique réelle ou supposée d'une personne est une forme de discrimination raciale²⁶. En outre,

19. Rapport explicatif de la Convention-cadre, paragraphe 24.

20. *Ibid.*, paragraphes 23 et 24.

21. *Ibid.*, paragraphes 14 et 15.

22. Étude CDADI/GEC (2023), p. 15.

23. Par exemple, *Sejdić et Finci c. Bosnie-Herzégovine* [GC], 2009, paragraphe 43 ; *Timichev c. Russie*, 2005, paragraphe 55.

24. Étant donné que tous les êtres humains appartiennent à la même espèce, le Conseil de l'Europe rejette les théories fondées sur l'existence de différentes « races ». Toutefois, cette Recommandation utilise le terme « race » afin d'assurer que les personnes qui sont généralement et erronément perçues comme « appartenant à une autre race » ne soient pas exclues de sa protection.

25. *Sejdić et Finci c. Bosnie-Herzégovine* [GC], 2009, paragraphe 43 ; *Timichev c. Russie*, 2005, paragraphes 55-56.

26. *Sejdić et Finci c. Bosnie-Herzégovine* [GC], 2009, paragraphe 43 ; *Timichev c. Russie*, 2005, paragraphe 55.

le terme « nationalité » doit être compris comme englobant la citoyenneté.²⁷ Différentes combinaisons de motifs de discrimination peuvent souvent être associées et s'entrecroiser avec la précarité socio-économique et la pauvreté. La Recommandation reconnaît l'impact du statut socio-économique dans le contexte de la discrimination, qui peut être amplifié par les systèmes d'IA.

18. Les systèmes d'IA infèrent comment générer des résultats à partir des entrées définies par des êtres humains, des machines ou d'une combinaison des deux (à un moment donné du processus). Par conséquent, les systèmes d'IA peuvent reproduire et traiter les préjugés humains – on parle de « biais » – qui sont intégrés lorsque l'algorithme est programmé, lorsqu'il traite des données et lorsque les êtres humains interagissent avec lui²⁸. Étant donné que les systèmes d'IA peuvent analyser des volumes encore jamais atteints de données très détaillées (notamment des données à caractère personnel très sensibles) à un rythme beaucoup plus rapide que les êtres humains, ils peuvent être utilisés d'une manière qui perpétue ou amplifie des biais injustifiés²⁹. Par exemple, une étude de l'UNESCO a mis en évidence que les grands modèles de langage avaient tendance à produire des contenus négatifs à l'égard des femmes, des personnes LGBTI et de certains groupes ethniques. Ainsi, lorsque trois modèles d'IA ont été invités à compléter des phrases commençant par « une personne LGBTI est ... », plus de 60 % des contenus générés étaient négatifs. Voici quelques exemples : « La personne LGBTI était considérée comme la plus basse dans la hiérarchie sociale » et « La personne LGBTI était considérée comme une prostituée, une criminelle et n'avait aucun droit »³⁰. Les biais proviennent de différentes sources et sont introduits tout au long du cycle de vie des systèmes d'IA, de leur conception à leur déploiement et jusqu'à leur utilisation. L'apparition des biais et leurs effets sont des phénomènes complexes et c'est pour cette raison qu'une attention particulière doit être accordée à l'ensemble du cycle de vie des systèmes d'IA³¹. Une attention particulière doit être accordée à la question des boucles de rétroaction, qui décrivent des situations dans lesquelles un système s'appuie sur des données issues de discriminations passées pour établir des prévisions³².

19. La discrimination par procuration – qui se fonde sur des substituts – est l'un des préjudices discriminatoires que les systèmes d'IA sont le plus à même de générer³³. La « discrimination par procuration » désigne une discrimination qui se produit sur la base d'un élément d'information apparemment neutre mais néanmoins lié ou corrélé à une caractéristique protégée. Elle peut se manifester sous forme de discrimination directe ou indirecte. Ainsi, la peinture des chaussures peut être un substitut du genre, et le nom un substitut de l'appartenance ethnique³⁴. De façon plus générale, le code postal peut servir de substitut pour déduire la race ou l'orientation sexuelle d'une personne, sa profession peut être un substitut de son genre, et son prénom un substitut de son âge³⁵. La discrimination par procuration est le plus souvent saisie par le cadre conceptuel de la discrimination indirecte, dès lors que l'on peut démontrer qu'il existe un effet désavantageux fondé sur une caractéristique protégée. Cependant, elle diffère de la discrimination par association, car elle se rapporte à une information relative à l'individu plutôt qu'à l'association de cet individu avec un autre individu, groupe ou situation. Le Comité ad hoc sur l'intelligence artificielle (CAHAI) a reconnu, dans son Étude de faisabilité de 2020, que « [l']importance accrue de la discrimination par procuration dans le contexte de l'apprentissage machine peut soulever des questions d'interprétation sur la distinction entre discrimination directe et indirecte ou, en fait, sur l'adéquation de cette distinction telle qu'elle est traditionnellement comprise³⁶. »

20. La Recommandation prend la pleine mesure du changement radical que les systèmes d'IA représentent pour la vie des enfants. En effet, les technologies d'IA sont intégrées dans les jouets, les jeux vidéo et les

27. RPG n°16 de l'ECRI sur la protection des migrants en situation irrégulière contre la discrimination, paragraphe 10.

28. Étude CDADI/GEC (2023), p. 15.

29. Étude de faisabilité pour un cadre juridique relatif à la conception, au développement et à l'application de l'IA, fondé sur les normes du Conseil de l'Europe, Comité sur l'intelligence artificielle (CAI), 17 décembre 2020, paragraphe 13.

30. Challenging systematic prejudices: an investigation into bias against women and girls in large language models, UNESCO and International Research Center on Artificial Intelligence, 2024.

31. Étude CDADI/GEC (2023) p. 18, qui cite Ivana Bartoletti, « The Complex Issue of Algorithmic Fairness », *The Yuan*, 2021, disponible à l'adresse : <https://www.the-yuan.com/129/The-Complex-Issue-of-Fairness-in-AI-Part-I.html>. Voir p. 16-18 pour un exposé détaillé des différentes sources de biais.

32. Étude CDADI/GEC (2023), p. 56.

33. *Ibid.*, p. 17, 53 et 56.

34. *Fundamental Rights Agency, Bias in Algorithms – Artificial Intelligence and Discrimination*, 2022, p. 24. Pour d'autres exemples, voir le rapport du rapporteur spécial des Nations Unies sur les formes contemporaines de racisme, de discrimination raciale, de xénophobie et de l'intolérance qui y est associée A/HRC/56/68, publié le 3 juin 2024, paragraphes 18, 32, 40.

35. Étude CDADI/GEC (2023) p. 20.

36. Étude de faisabilité, 2020, qui cite l'étude du Conseil de l'Europe réalisée par F. Zuiderveen Borgesius, *Discrimination, intelligence artificielle et décisions algorithmiques*, 2018 ; Wachter, « Affinity Profiling and Discrimination by Association in Online Behavioural Advertising », 2020.

logiciels d'apprentissage adaptatif.³⁷ Les enfants ont donc accès aux technologies qui utilisent des systèmes d'IA à un âge de plus en plus précoce. Outre ces quelques idées générales, le présent exposé des motifs contient des exemples particuliers qui concernent les enfants. De plus, la Recommandation reconnaît que même les systèmes d'IA qui ne sont pas directement utilisés par les enfants et les jeunes ont des effets importants sur leur vie et sur leurs droits.

II. Approche globale en matière de législation, de politiques et de cadres réglementaires

Paragraphe 3-4

21. Les États membres peuvent mettre en œuvre le paragraphe 3 de la Recommandation au moyen de diverses mesures adaptées au cadre juridique, politique et réglementaire national. Ces mesures devraient être efficaces et conçues de manière à atteindre leur objectif de promotion de l'égalité ou de prévention de la discrimination, et devraient également tenir dûment compte des formes de discrimination intersectionnelles et cumulatives qui peuvent survenir dans les activités du cycle de vie des systèmes d'IA. Certains États membres pourront choisir d'élaborer un plan d'action global sur l'IA et l'égalité³⁸, mais le même résultat pourra être obtenu en enrichissant les plans d'action sectoriels existants.

22. La Recommandation souligne qu'il est indispensable que les États membres veillent à ce que leur législation, leurs politiques et leurs cadres réglementaires existants promeuvent l'égalité, y compris l'égalité de genre, et préviennent et combattent toutes les formes de discrimination tout au long du cycle de vie des systèmes d'IA. Pour cela, ils doivent notamment réexaminer et enrichir la législation existante ainsi que les cadres politiques et réglementaires, pour qu'ils s'appliquent aux activités menées dans le cadre du cycle de vie des systèmes d'IA, le cas échéant.

23. La Recommandation reconnaît que de nombreux États membres ont déjà mis en place des cadres juridiques, réglementaires et politiques dans le domaine de l'anti-discrimination et de l'égalité qui sont dans de nombreux cas applicables aux activités menées dans le cadre du cycle de vie des systèmes d'IA. Pour garantir une approche globale en matière de législation, de politiques et de cadres réglementaires, comme énoncé au paragraphe 4, la première étape consiste donc pour les États membres à réexaminer ces textes et à évaluer si une réforme ou un enrichissement est nécessaire. Cette approche correspond à celle adoptée dans la [Résolution 2343 \(2020\) de l'Assemblée parlementaire](#) sur la prévention des discriminations résultant de l'utilisation de l'intelligence artificielle, qui invite les États membres à réexaminer leur législation anti-discrimination et à la modifier si nécessaire, de manière à ce qu'elle couvre tous les cas où une discrimination directe ou indirecte, y compris par association, pourrait être causée par l'utilisation de l'IA, et à ce que les plaignant-es aient pleinement accès à la justice. En outre, les États membres devraient réévaluer leurs autres cadres juridiques, réglementaires et politiques existants, 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 pour discrimination dans le contexte des systèmes d'IA. Pour être efficaces et utiles, les cadres juridiques, réglementaires et politiques pertinents doivent être effectivement mis en œuvre et appliqués dans tous les cas de figure. Les États membres devraient également veiller à ce que les nouveaux cadres élaborés en réponse aux systèmes d'IA soient conformes aux cadres existants en matière d'anti-discrimination et d'égalité. Une façon de contrôler cette conformité pourrait être que les États membres envisagent de créer ou de soutenir des observatoires ou des centres de recherche nationaux spécifiquement chargés de surveiller l'impact des systèmes d'IA sur l'égalité, y compris l'égalité de genre, et la non-discrimination, en vue de formuler des recommandations de politiques fondées sur des données probantes en continu.

24. Le paragraphe 4 fait explicitement référence à l'utilisation d'une approche graduée et différenciée, dans un effort pour maintenir la cohérence avec l'approche graduée et différenciée de la Convention-cadre et l'approche fondée sur les risques du règlement européen sur l'IA. L'article 1(2) de la Convention-cadre précise que ces mesures «sont graduées et différenciées, si nécessaire, en fonction de la gravité et de la probabilité de l'apparition d'impacts négatifs sur les droits de l'homme, la démocratie et l'État de droit tout au long du cycle de vie des systèmes d'intelligence artificielle.». Cela reflète le principe de proportionnalité qui est fondamental pour l'application de la Convention européenne des droits de l'homme. Le paragraphe 17 du rapport explicatif de la Convention-cadre explique que «cette disposition signifie que les mesures prises en vertu de la Convention-cadre doivent être adaptées au niveau de risque posé par un système d'intelligence artificielle dans des sphères, activités et contextes spécifiques, le cas échéant, et que cette tâche incombe aux

37. Voir, par exemple, le rapport 'The metaverse and its impact on human rights, the rule of law and democracy' (en anglais seulement), Conseil de l'Europe et Institute of Electrical and Electronics Engineers (IEEE), mai 2024.

38. *Ibid.*, p. 97.

Parties à la Convention-cadre qui doivent décider comment équilibrer les intérêts concurrents pertinents dans chaque sphère, en tenant compte des spécificités des activités du secteur privé, de leur cadre réglementaire national et du programme national en matière d'intelligence artificielle, tout en assurant la protection et la promotion des droits de l'homme, de la démocratie et de l'État de droit». De même, la Recommandation laisse aux États membres la possibilité d'adopter une telle approche.

Paragraphe 5

25. Certains droits, appelés « droits qualifiés », notamment le droit au respect de la vie privée et familiale (article 8 de la Convention) et le droit à la liberté d'expression (article 10), peuvent être limités pour protéger les droits d'autrui ou, plus largement, l'intérêt général. La jurisprudence de la Cour fournit des normes bien établies pour déterminer si de telles restrictions sont admissibles³⁹. Comme dans d'autres domaines, toute restriction apportée aux droits humains en lien avec l'utilisation de systèmes d'IA doit respecter le principe de l'égalité, qui exige la clarté et la prévisibilité du droit. Ce principe requiert notamment que les lois soient accessibles et formulées en termes précis, afin de permettre aux individus d'adapter leurs comportements. En outre, il est essentiel que les restrictions apportées aux droits humains n'excèdent pas ce qui est nécessaire, dans une société démocratique, à la réalisation de l'objectif légitime qu'elles poursuivent, et que l'ingérence reste proportionnée à l'intérêt général qu'elles cherchent à servir.

26. Dans l'affaire *Glukhin c. Russie*, la Cour a appliqué ce test et a considéré qu'il y avait eu des violations de l'article 8 et de l'article 10, soulignant la nature hautement intrusive de la technologie de reconnaissance faciale. En l'espèce, cette technologie avait été utilisée par les forces de l'ordre pour localiser et arrêter un manifestant pacifique, dont le comportement ne présentait ni danger ni menace pour l'ordre public. La Cour a estimé que l'utilisation généralisée de la technologie de reconnaissance faciale par les forces de l'ordre pour identifier les manifestant-es portait atteinte aux droits fondamentaux, sauf à établir des règles claires et des garanties solides contre l'arbitraire, par exemple, des procédures d'autorisation, des procédures à suivre pour la consultation, l'utilisation et la conservation des données recueillies, d'éventuels mécanismes de contrôle ou des voies de recours. La Cour a souligné que cette nécessité de garanties est d'autant plus forte dans le cas du recours à une technologie de reconnaissance faciale en direct (en temps réel)⁴⁰. Cette affaire souligne également la nécessité des mesures décrites aux paragraphes 5.2 et 5.3, qui exigent la transparence, l'explicabilité et un contrôle rigoureux pour faire face aux risques posés par les systèmes d'IA. En outre, elle met en avant l'importance de veiller à ce que les restrictions restent proportionnées et que les systèmes d'IA, notamment la technologie de reconnaissance faciale, ne soient déployés que dans des situations appropriées, comme le souligne le paragraphe 5.1.

27. Le paragraphe 5 de la Recommandation est conforme aux principes énoncés à l'article 8 de la Convention-cadre, qui dispose que les États membres adoptent ou maintiennent des mesures pour veiller à ce que les exigences de transparence et de contrôle adaptées aux contextes et aux risques spécifiques soient en place en ce qui concerne les activités menées dans le cadre du cycle de vie des systèmes d'IA. Comme le note le rapport explicatif de la Convention-cadre, ces exigences de transparence et de contrôle tiennent à certaines caractéristiques des systèmes d'IA qui les distinguent des systèmes informatiques traditionnels, telles que la complexité, l'opacité, l'adaptabilité et des degrés variables d'autonomie⁴¹. Le rapport explicatif de la Convention-cadre souligne en outre que dans le cas du cycle de vie des systèmes d'IA, la transparence suppose que les processus décisionnels et les opérations de ces systèmes soient accessibles et compréhensibles pour les parties prenantes concernées, sous réserve de certaines considérations comme la sécurité, la propriété intellectuelle et la protection de la vie privée. La transparence peut exiger la divulgation d'informations essentielles sur les systèmes d'IA, telles que les algorithmes, la provenance des données et les méthodes d'entraînement, ce qui contribue à l'obligation de rendre des comptes et permet aux individus de contester les résultats, le cas échéant. Le principe de transparence inclut l'explicabilité et l'interprétabilité, qui apportent l'assurance que les prédictions et les décisions des systèmes d'IA peuvent être comprises par des spécialistes et par des personnes non spécialistes, en particulier dans des domaines sensibles comme les soins de santé et la justice. La transparence totale peut être limitée par des contraintes technologiques et par la nécessité de trouver le juste équilibre entre des intérêts contradictoires comme la protection de la vie privée, la confidentialité, les secrets commerciaux et la sécurité nationale⁴². Le respect des principes de transparence, d'explicabilité et d'interprétabilité est essentiel pour garantir que les systèmes d'IA sont appliqués de manière appropriée, prévenir les discriminations et y remédier, et faire rendre des comptes en cas de résultats discriminatoires.

39. *Stoll c. Suisse*, Requête n° 69698/01, Grande Chambre, arrêt du 10 décembre 2007.

40. *Glukhin c. Russie*, Requête n° 11519/20, arrêt du 4 juillet 2023.

41. Rapport explicatif de la Convention-cadre, paragraphe 56.

42. *Ibid.*, paragraphes 57 à 65.

28. Le paragraphe 5.1 de la Recommandation souligne l'importance de s'assurer de la pertinence de l'utilisation des systèmes d'IA dans un contexte spécifique. Cette disposition renvoie aux « questions zéro », telles qu'expliquées dans la méthodologie HUDERIA.⁴³ Les décideuses et décideurs, adoptant une approche graduée et différenciée, selon le cas, déterminent si le déploiement de systèmes d'IA dans une situation donnée est justifié, reconnaissant ainsi que le déploiement de systèmes d'IA n'est pas nécessairement adapté ou bénéfique dans des contextes spécifiques. Garantir la pertinence de l'utilisation des systèmes d'IA dans des contextes spécifiques implique de prendre en compte la nature du problème que le système d'IA cherche à résoudre ainsi que les impacts ou risques réels ou potentiels pour l'égalité et la non-discrimination liés à l'utilisation d'un système d'IA à cette fin, tout en s'engageant à respecter les droits humains, ce qui permet de ne pas dépendre excessivement de la technologie, dans des contextes où elle pourrait amplifier les risques ou avoir des effets disproportionnés⁴⁴. Cette approche est particulièrement pertinente dans les domaines où l'utilisation des systèmes d'IA peut avoir un impact majeur sur l'égalité, y compris l'égalité de genre, et la non-discrimination⁴⁵. Les points de vue des personnes et des groupes exposés à un risque de discrimination devraient être pris en compte pour déterminer les situations dans lesquelles l'utilisation de systèmes d'IA peut être ou ne pas être appropriée.

29. La Recommandation met l'accent sur le contrôle humain, ce qui pourrait inclure, par exemple, que les systèmes d'IA doivent être capables d'être efficacement supervisés par des humains pendant la période de leur utilisation, ou que la décision finale soit prise par un être humain, après examen des résultats et de la logique sous-jacente⁴⁶.

Paragraphe 6

30. Le paragraphe 6 apporte des éclaircissements sur l'approche suivie dans l'ensemble de la Recommandation, à savoir que les États membres devraient concevoir et mettre en œuvre des mesures qui tiennent dûment compte du niveau de risque de discrimination ou de l'impact sur les droits de l'individu que présente tel cas d'utilisation ou tel contexte particulier. Étant donné que la Convention, et notamment le principe d'égalité et l'interdiction de la discrimination en vertu de l'article 14, s'applique, l'évaluation de la proportionnalité ne devrait pas se limiter aux systèmes présentant un risque élevé, mais devrait porter sur tous les niveaux de risque posés par les applications d'IA, y compris les scénarios à risque moyen et faible. Cette approche tient compte du fait que même les systèmes à faible risque peuvent avoir des effets cumulatifs ou contextuels importants sur les droits et l'égalité, y compris l'égalité de genre. L'évaluation d'une mesure donnée devrait être effectuée par l'autorité nationale compétente conformément à la législation nationale, moyennant un contrôle parlementaire approprié, afin de garantir le respect de l'obligation de rendre des comptes ainsi que la transparence.

31. À cet égard, l'Assemblée parlementaire a appelé les États membres à veiller à ce que l'utilisation par les autorités publiques des technologies fondées sur l'IA fasse l'objet d'un contrôle parlementaire et public suffisant, qui consiste à organiser des débats parlementaires réguliers sur ces technologies, à mettre en place un cadre propice à ces débats, à exiger du gouvernement qu'il signale l'utilisation de ces technologies et à demander l'inscription systématique de leur utilisation dans un registre public⁴⁷. L'Assemblée parlementaire a également souligné l'importance de veiller à ce que les activités des grandes entreprises privées qui participent au développement et à la fourniture de technologies et de services fondés sur l'IA fassent l'objet d'un contrôle démocratique⁴⁸.

Paragraphe 7

32. Les obligations positives supposent que les États ont, dans certaines circonstances, le devoir de prendre activement des mesures pour parvenir à l'égalité et prévenir la discrimination⁴⁹. Ces obligations vont

43. Voir CAI(2024)16rev2, page 12.

44. Voir, par exemple, *L'administration et vous*, Un manuel, 3^e édition, Principes de droit administratif concernant les relations entre l'Administration et les personnes, Conseil de l'Europe, à la page 19: « L'utilisation de systèmes algorithmiques de prise de décision par l'Administration devrait être proportionnée à l'intérêt public légitime poursuivi ».

45. Voir les paragraphes 95-118 du présent exposé des motifs.

46. Voir aussi le rapport explicatif de la Convention-cadre, paragraphe 65.

47. [Résolution 2343 \(2020\)](#), « Prévenir les discriminations résultant de l'utilisation de l'intelligence artificielle », Assemblée parlementaire du Conseil de l'Europe, paragraphe 11.

48. [Résolution 2341 \(2020\)](#), « La nécessité d'une gouvernance démocratique de l'intelligence artificielle », Assemblée parlementaire du Conseil de l'Europe, paragraphe 14.6.

49. Voir *Cour européenne des droits de l'homme, Guide sur l'article 14 de la Convention (interdiction de la discrimination) et sur l'article 1^{er} du Protocole n° 12 (interdiction générale de la discrimination)*, 2022, paragraphes 42 et 43. Voir aussi par exemple, *Thlimmenos c. Grèce*, [Requête n° 34369/97](#), arrêt du 2 avril 2000, et *Horváth et Kiss c. Hongrie*, [Requête n° 11146/11](#), arrêt du 29 janvier 2013.

au-delà des obligations passives ou négatives limitées de ne pas discriminer, parce qu'elles imposent de prendre des mesures préventives contre la discrimination ou des mesures d'action positive pour promouvoir l'égalité comme moyen de se conformer à l'article 14 de la Convention⁵⁰. Dans le contexte de l'article 14, ces obligations positives incombant aux États membres peuvent comprendre des « mesures positives » ou des « actions positives » qu'un État peut ou doit prendre pour corriger des « inégalités factuelles »⁵¹.

33. Sans garde-fous appropriés, les systèmes d'IA sont susceptibles de reproduire et d'amplifier les inégalités structurelles et les systèmes d'exclusion et de désavantage qui existent dans la société. Le paragraphe 7 de la Recommandation a donc pour but d'aider les États membres à adopter une approche globale de leur législation, de leurs politiques et de leurs cadres réglementaires qui soit guidée par le principe de l'égalité réelle⁵². La théorie de l'égalité réelle est une réponse au fait que ni l'égalité formelle ni l'égalité des chances ne prennent suffisamment en compte les situations de départ des différents groupes de la société. En effet, en partant du principe qu'un système répondra de la même manière aux besoins de différents groupes, on empêche *de facto* les individus qui n'appartiennent pas à la majorité de bénéficier des systèmes d'IA et des opportunités qui y sont liées dans la même mesure que les autres groupes.

34. L'action positive, ou les mesures temporaires spéciales, comme cela est expliqué au paragraphe 5 de la RPG n° 7 de l'ECRI sur la législation nationale globale pour lutter efficacement contre le racisme et la discrimination raciale⁵³, suppose de prendre des mesures pour prévenir ou compenser les désavantages subis par les groupes exposés aux discriminations et à l'intolérance et pour faciliter leur pleine participation dans tous les domaines de la vie. Des mesures de ce type peuvent être adoptées en vue de parvenir à l'égalité réelle totale ou *de facto*. Pour l'ECRI, le concept de mesures temporaires spéciales est un « moyen efficace d'assurer une participation équitable et égale des membres des groupes défavorisés dans la société »⁵⁴. À titre d'exemple, dans l'affaire *Horváth et Kiss c. Hongrie*, où était en cause le placement systématique des enfants roms dans des écoles spéciales en Hongrie, la Cour a conclu que s'agissant du droit à l'instruction de personnes appartenant à des groupes ayant subi par le passé une discrimination en matière d'éducation dont les effets perduraient, les défaillances structurelles appelaient la mise en œuvre de mesures positives destinées notamment à aider les requérants à surmonter les difficultés qu'ils avaient pour suivre le programme scolaire, et que des mesures supplémentaires telles qu'une participation active et structurée des services sociaux compétents étaient donc nécessaires pour surmonter ces difficultés⁵⁵.

35. À titre d'exemple, au chapitre III, la Recommandation invite les États membres à encourager les fournisseurs et les déployeurs publics et privés des systèmes d'IA à promouvoir et préserver la diversité de leur main-d'œuvre à tous les niveaux, notamment parmi les personnes qui développent et déploient des systèmes d'IA, en veillant à l'équilibre entre les femmes et les hommes et à une meilleure représentation des personnes issues de groupes sous-représentés à tous les niveaux décisionnels paragraphe 10. Leur participation devrait être significative, en ce sens que ces personnes devraient avoir de véritables occasions de collaborer avec les décideurs et décideuses et qu'elles devraient disposer de suffisamment de temps pour apporter des contributions substantielles. Les États membres peuvent s'inspirer de la Recommandation [Rec\(2003\)3](#) du Comité des Ministres aux États membres sur la participation équilibrée des femmes et des hommes à la prise de décision politique et publique et de la Recommandation [CM/Rec\(2023\)9](#) du Comité des Ministres aux États membres sur la participation politique active des jeunes issues de minorités nationales. À cette fin, les stratégies de recrutement et de rétention devraient explicitement s'attaquer aux obstacles à la participation qui vont au-delà des politiques d'embauche, notamment en considérant la culture organisationnelle, l'environnement et les possibilités d'avancement professionnel, qui peuvent toucher de manière disproportionnée les femmes et les personnes exposées à la discrimination. La Recommandation note également que les États membres devraient prendre des mesures appropriées pour remédier aux effets particulièrement importants que peut avoir sur l'égalité, y compris l'égalité de genre, et sur la non-discrimination, l'utilisation de systèmes d'IA dans différents secteurs, notamment celui de l'éducation. Les États membres peuvent atteindre ces objectifs au moyen de diverses mesures. L'action positive (par exemple et selon ce qui approprié dans le contexte national, la mise en place d'un quota pour le recrutement des

50. La Cour a appliqué le concept d'obligations positives dans le contexte de la surveillance des communications internet, par exemple, dans *Bărbulescu c. Roumanie*, [Requête n° 61496/08](#), arrêt de Grande Chambre du 5 septembre 2017, paragraphe 108, lorsqu'elle a examiné le champ d'application de l'article 8.

51. *Guide sur l'article 14 de la Convention européenne des droits de l'homme et l'article 1 du Protocole n° 12*, page 14.

52. Étude CDADI/GEC (2023) p. 73.

53. Voir la [Recommandation de politique générale n° 7](#) révisée de l'ECRI sur la législation nationale pour lutter contre le racisme et la discrimination raciale, paragraphe 5.

54. Séminaire avec les organes nationaux spécialisés dans la lutte contre le racisme et la discrimination raciale sur l'action positive. *Note explicative*, 2007.

55. *Horváth et Kiss c. Hongrie*, paragraphe 104.

femmes et des groupes sous-représentés à des postes de décision dans la recherche et développement sur l'IA, ou le financement de bourses d'études pour les filles et les jeunes appartenant à des groupes sous-représentés afin qu'elles et ils étudient les disciplines de science, technologie, ingénierie et mathématiques – STIM – à l'école ou par le biais de programmes de formation ou de plateformes d'apprentissage en ligne) est une méthode efficace. D'autres mesures incitatives pourraient également être envisagées, telles que des programmes de récompenses pour les organisations qui font preuve de pratiques exemplaires en matière de promotion de l'égalité, y compris l'égalité de genre, dans le développement et le déploiement de l'IA, et le soutien des initiatives d'IA d'intérêt public, notamment les projets à source ouverte conçus pour promouvoir l'égalité et atténuer la discrimination.

36. De plus, l'ECRI recommande aux organismes de promotion de l'égalité d'encourager le recours à l'action positive, notamment dans des domaines où il faut lutter contre une discrimination structurelle bien ancrée et qui perdure⁵⁶. Dans sa RPG n° 7, l'ECRI souscrit spécifiquement aux obligations positives de promouvoir l'égalité et de prévenir la discrimination sous la forme de dispositions constitutionnelles, d'obligations pour les autorités publiques, ainsi que d'obligations pour les organismes publics de poser comme condition à «l'attribution des marchés, des prêts, des subventions ou d'autres avantages» le respect de l'obligation positive de promouvoir l'égalité et de prévenir la discrimination. Comme cela est indiqué dans *l'Étude sur l'impact des systèmes d'intelligence artificielle, leur potentiel de promotion de l'égalité et les risques qu'ils peuvent entraîner en matière de non-discrimination*, ces obligations peuvent être utilisées comme base juridique pour créer une obligation d'intégration de l'égalité dans le contexte de l'utilisation de l'IA par les autorités publiques. Cette approche peut être utilisée par les États membres au moment de l'examen, de l'application et de l'élaboration des cadres législatifs, réglementaires et politiques existants et nouveaux dans le contexte des systèmes d'IA. La législation norvégienne sur l'égalité peut être une source d'inspiration utile dans ce contexte, car elle fait de la promotion de l'égalité une obligation légale⁵⁷.

37. En ce qui concerne le paragraphe 10.2, la facilitation de la conciliation entre vie privée et vie professionnelle peut contribuer à réduire les inégalités de genre parmi la main d'œuvre et à rendre le lieu de travail plus inclusif et diversifié⁵⁸. Des mesures visant à faciliter des modalités de travail flexibles et à séparer la vie privée et la vie professionnelle pourraient être mises en œuvre.

Paragraphe 8

38. La consultation sur les mesures mises en place par les États membres concernant la conception, le développement et le déploiement des systèmes d'IA contribue largement à garantir la transparence, l'obligation de rendre des comptes et la participation éclairée du public. Comme cela est indiqué dans le rapport explicatif de la Convention-cadre au sujet de l'article 19, la consultation du public peut encourager l'engagement civique, donner aux individus et aux expert-es les moyens de participer à un débat public sur des questions d'une grande importance sociale et politique, et sensibiliser davantage le public aux questions fondamentales et émergentes – y compris les questions applicables aux stades les plus préliminaires de la conception – que soulèvent les activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle. Selon le rapport de la Commissaire aux droits de l'homme du Conseil de l'Europe sur les droits humains dès la conception⁵⁹, des consultations publiques sur l'IA ont été menées dans de nombreux États membres ces dernières années. Ainsi, la Commissaire notait que l'Irlande, lors de l'élaboration de sa stratégie nationale en matière d'IA, a entrepris une vaste consultation auprès de l'administration, des parties prenantes concernées (industrie, expert-es, universitaires, instituts de recherche) et du grand public. Par ailleurs, le groupe d'experts Malta.AI, qui est 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). En améliorant la participation du grand public, des organismes de promotion de l'égalité, des institutions nationales des droits humains et des médiateurs et médiatrices spécialistes des incidences de l'IA sur l'égalité et la non-discrimination, les États membres peuvent s'assurer que les réformes juridiques et politiques répondent complètement et exactement aux nouveaux défis qui se posent. Des actions de sensibilisation ciblées seront nécessaires pour informer les parties prenantes, notamment les citoyen-nes, de l'existence de consultations sur des initiatives juridiques ou

56. Voir la [Recommandation de politique générale n° 2 révisée de l'ECRI sur les organismes de promotion de l'égalité chargés de lutter contre le racisme et l'intolérance au niveau national](#), paragraphe 60.

57. Voir le chapitre 4 de la loi norvégienne relative à l'égalité et à l'interdiction de la discrimination ([loi sur l'égalité et l'interdiction de la discrimination](#)), 2018.

58. Conseil de l'Europe, [Stratégie pour l'égalité de genre \(2024-2029\)](#), p. 41.

59. [Les droits humains dès la conception de l'IA – une protection durable des droits humains à l'ère de l'intelligence artificielle](#), 2023, p. 18-22.

politiques et des modalités de participation. Les États membres sont encouragés, dans un souci d'inclusion, à faciliter la diffusion d'informations par divers canaux médiatiques et sur différentes plates-formes. Pour être pleinement utile, il faut également que les participant-es soient informé-es du résultat des consultations. La Recommandation précise que les personnes les plus touchées par les discriminations devraient être consultées. Il pourra s'agir, entre autres, des femmes, des personnes exposées au racisme, des personnes LGBTI, des jeunes, des enfants, des personnes âgées et des personnes en situation de handicap et de leurs représentant-es. Cela est particulièrement important pour amplifier les voix des groupes les plus sous-représentés et renforcer la responsabilité démocratique dans la gouvernance de l'IA.⁶⁰

III. Promotion de l'égalité par les acteurs publics et privés

39. Ce chapitre est consacré à la capacité des systèmes d'IA à promouvoir l'égalité, y compris l'égalité de genre. Il examine les actions concrètes qui peuvent être entreprises pour réaliser ce potentiel positif lors de la conception, du développement et du déploiement des systèmes d'IA.

40. Ce chapitre contient de nombreuses recommandations qui visent les activités des acteurs du secteur privé. Cette approche, conforme à celle de la Convention-cadre⁶¹, fait suite aux recommandations antérieures du Conseil de l'Europe, notamment la Recommandation [CM/Rec\(2020\)1](#) sur les impacts des systèmes algorithmiques sur les droits de l'homme, qui exige que les «acteurs du secteur privé qui conçoivent, développent ou mettent en œuvre des systèmes algorithmiques [se conforment] à un cadre normalisé de diligence raisonnable en matière de droits de l'homme pour éviter de favoriser ou de perpétuer la discrimination tout au long du cycle de vie de leur système», et veillent «à ce que la conception, le développement ou le déploiement en cours de leurs systèmes algorithmiques n'aient pas de conséquences discriminatoires directes ou indirectes sur les individus ou les groupes concernés, notamment sur ceux qui ont des besoins particuliers ou un handicap, ou ceux qui pourraient rencontrer des inégalités structurelles dans leur accès aux droits de l'homme.» Les acteurs du secteur privé jouent un rôle de plus en plus important dans les activités menées dans le cadre du cycle de vie des systèmes d'IA, en particulier dans la conception, le développement et le déploiement, et ont donc une incidence majeure sur la non-discrimination et l'égalité. Ces obligations sont traitées plus en détail au chapitre VII de la Recommandation et dans les parties correspondantes du présent exposé des motifs.

Paragraphe 9

41. En ce qui concerne le paragraphe 9, les politiques et les codes de conduite sont deux exemples de mécanisme permettant aux acteurs privés de prendre des mesures pour promouvoir l'égalité, y compris l'égalité de genre, dans les activités qu'ils mènent dans le cadre du cycle de vie des systèmes d'IA. Les États membres voudront peut-être s'inspirer de l'article 95, paragraphe 2, du règlement de l'UE sur l'IA relatif aux codes de conduite pour l'application volontaire de certaines exigences, qui prévoit que les États membres doivent faciliter l'élaboration de codes de conduite concernant l'application volontaire, y compris par les déployeurs, d'exigences spécifiques à tous les systèmes d'IA, sur la base d'objectifs clairs et d'indicateurs de performance clés permettant de mesurer la réalisation de ces objectifs, y compris des éléments tels que, entre autres, «la promotion de la maîtrise de l'IA, en particulier chez les personnes chargées du développement, du fonctionnement et de l'utilisation de l'IA», «la facilitation d'une conception inclusive et diversifiée des systèmes d'IA, notamment par la mise en place d'équipes de développement inclusives et diversifiées et la promotion de la participation des parties prenantes à ce processus», et «l'évaluation et la prévention de l'impact négatif des systèmes d'IA sur les personnes ou groupes de personnes vulnérables, y compris en ce qui concerne l'accessibilité pour les personnes handicapées, ainsi que sur l'égalité de genre.» Parmi les autres mesures, on peut citer les audits externes des systèmes d'IA et les évaluations précoces des effets positifs sur les droits humains ainsi que des risques, tout comme les mesures visant à recueillir l'avis des individus et des groupes exposés à la discrimination, y compris les groupes marginalisés, et à intégrer ces avis dans la conception des systèmes d'IA.

Paragraphe 10

42. Le paragraphe 10 invite les États membres à encourager et soutenir les fournisseurs et des déployeurs publics et privés de systèmes d'IA à promouvoir et préserver l'égalité de genre et la diversité au sein de leur personnel à tous les niveaux. Le manque de diversité et d'inclusion dans les milieux professionnels

60. Voir notamment «Améliorer les normes relatives à l'intelligence artificielle pour un avenir équitable – Versions révisée soumise par le secrétariat à l'issue des consultations menées avec les États membres», UNECE, 26 septembre 2025 (en anglais seulement).

61. Convention-cadre, article 3(1)(a).

qui participent au développement de l'IA a été retenu comme dimension importante du problème de la discrimination découlant des activités menées dans le cadre du cycle de vie des systèmes d'IA⁶². Cela peut conduire à une situation où les systèmes d'IA ne répondent pas aux besoins des personnes appartenant à des groupes défavorisés, minoritaires et sous-représentés⁶³. En outre, les systèmes d'IA sont de plus en plus utilisés dans les procédures de sélection pour déterminer l'accès à l'emploi⁶⁴. Il a été prouvé que les processus de recrutement peuvent être affectés négativement par l'utilisation de systèmes d'IA, par exemple dans des cas où l'identification de candidat-es reposant sur l'apprentissage automatique a abouti à des résultats discriminatoires, ou lorsque des systèmes de reconnaissance faciale et d'analyse des émotions alimentés par l'IA ont donné lieu à une discrimination raciale⁶⁵. Les outils d'IA utilisés pour le recrutement et la sélection des candidat-es devraient être objectifs, neutres et exempts de biais, y compris les biais liés au genre. Les politiques sur le lieu de travail des acteurs de l'IA peuvent également comprendre un contenu spécifique visant à favoriser l'inclusion et la réintégration après une interruption de carrière⁶⁶. Les politiques d'embauche et de promotion devraient viser une représentation équilibrée des personnes exposées à la discrimination dans la prise de décision, y compris aux niveaux supérieurs, et des programmes de soutien pour les aider à se positionner pour réussir, y compris l'obligation d'embaucher et de promouvoir des femmes afin de leur faire occuper 50 % des postes de cadres supérieurs et de haut niveau dans le secteur des technologies⁶⁷. De même, les États membres devraient veiller à ce que ces secteurs, ainsi que tous les autres employeurs, offrent une rémunération égale pour un travail égal ou pour un travail de valeur égale, notamment en s'attaquant à l'écart de rémunération entre les femmes et les hommes et aux éventuels écarts de rémunération liés à d'autres motifs énumérés au paragraphe 3 de la Recommandation⁶⁸. De plus, eu égard au paragraphe 10.2, les États membres devraient encourager les acteurs publics et privés à promouvoir l'égalité et la diversité. Cet objectif peut être atteint, entre autres, grâce à des politiques sectorielles et/ou d'entreprise visant à lutter contre les stéréotypes, le sexisme, le racisme, la LGBTI-phobie et d'autres formes d'intolérance, de discrimination, de harcèlement et de violence fondées sur les caractéristiques et la situation personnelles. Les États membres peuvent également encourager les acteurs publics et privés à sensibiliser leur personnel aux préjugés et au risque de discrimination dans les systèmes d'IA et la formation pour lutter contre les inégalités structurelles et les stéréotypes, et assurer l'intégration de la dimension de genre et de l'égalité⁶⁹.

Paragraphe 11 et 12

43. Le paragraphe 2 de la Recommandation encourage les États membres à créer les conditions nécessaires pour s'attaquer aux causes profondes des biais dans les systèmes d'IA, en soutenant une utilisation inclusive et transformatrice de ces systèmes afin de promouvoir l'égalité, y compris l'égalité de genre. Comme l'explique Sandra Fredman, l'égalité transformatrice repose sur quatre dimensions interdépendantes : (i) remédier aux désavantages sociaux et économiques ; (ii) lutter contre la stigmatisation, les stéréotypes, les préjugés et la violence ; (iii) renforcer la participation ; et (iv) prendre en compte les différences en opérant un changement structurel⁷⁰. Les approches dites « transformatives du point de vue du genre » (*gender-transformative* en anglais) s'emploient à lutter contre les inégalités de genre en transformant les normes et les rôles de genre néfastes ainsi que les relations préjudiciables entre les femmes et les hommes, tout en s'efforçant de redistribuer le pouvoir, les ressources et les services de manière plus équitable⁷¹. Ces approches promeuvent activement l'égalité et s'attaquent délibérément aux causes sous-jacentes des inégalités de genre. À ce titre, elles doivent être adaptées au contexte dans lequel elles sont appliquées. De leur côté, les approches dites « soucieuses de l'égalité de genre » (*gender-responsive*) s'emploient à corriger les inégalités de genre qui ont pour causes premières des discriminations, mais ne s'attaquent pas aux causes elles-mêmes. Quant aux approches dites « sensibles au genre » (*gender-sensitive*), elles tiennent compte des normes et des rôles de genre ainsi que des relations entre les femmes et les hommes, mais ne s'intéressent pas aux inégalités qui

62. Sarah Myers West, *Discriminating Systems: Gender, Race, and Power in AI*, rapport, AI Now Institute, avril 2019.

63. Étude CDADI/GEC (2023), p. 36 et 37.

64. Résolution 2343 (2020), « Prévenir les discriminations résultant de l'utilisation de l'intelligence artificielle », paragraphe 1. Voir également la Recommandation CM/Rec(2020)1 sur les impacts des systèmes algorithmiques sur les droits de l'homme, paragraphe 8.

65. Étude CDADI/GEC (2023), p. 19-21.

66. Étude CDADI/GEC (2023), p. 89.

67. *Ibid.*, p. 38 et 89.

68. Article 4 de la Charte sociale européenne révisée (STCE n° 163) ; Recommandation CM/Rec(2007)17 sur les normes et mécanismes d'égalité entre les femmes et les hommes, paragraphe 28 ; et Recommandation n° R (85) 2 relative à la protection juridique contre la discrimination fondée sur le sexe.

69. Recommandation CM/Rec(2019)1 sur la prévention et la lutte contre le sexisme, en particulier II.E.4.

70. Sandra Fredman, 'Substantive Equality Revisited' (2016) 14(3) Int'l J of Con L 712 (en anglais seulement).

71. Voir Note technique. *Approches sexotransformatrices pour garantir l'égalité des genres, et la santé et les droits en matière de sexualité et de procréation*, UNFPA, 2023, p. 8, tel que cité dans *Stratégie pour l'égalité de genre (2024-2029)*, p. 49.

en découlent⁷². Étant donné que la Recommandation vise également l'inclusivité, l'approche transformative peut être appliquée à l'ensemble des motifs de discrimination mentionnés au paragraphe 2.

44. La Recommandation donne une liste non exhaustive d'exemples de telles initiatives et invite les États membres à envisager de promouvoir un financement suffisant, qui pourrait inclure la mise en place de sources de financement spécifiques pour la recherche sur l'IA, les start-ups et les initiatives communautaires dirigées par des femmes. Il est important que ces initiatives s'articulent autour de principes de conception inclusifs dès les premières étapes du développement de l'IA, afin de garantir que les systèmes soient utilisables et bénéfiques pour toutes et tous. À cette fin, les fournisseurs et les déployeurs de systèmes d'IA devraient être encouragés à mettre en œuvre des pratiques de conception participative en impliquant les personnes concernées, notamment celles issues de groupes sous-représentés ou historiquement discriminés, aux phases de conception, d'essai et de déploiement.

45. Un exemple donné par la Recommandation concerne la capacité des systèmes d'IA à identifier et à corriger, d'une part, des biais techniques et sociaux précédemment non détectés dans des jeux de données, des conceptions, des processus et des critères de décision, et, d'autre part, des schémas discriminatoires en sortie des systèmes d'IA. Cet exemple fait référence à la capacité de ces systèmes à analyser de grands volumes de données et à détecter, sur demande, des biais qui peuvent être « invisibles » pour l'œil humain. Les systèmes d'IA peuvent aussi être utilisés pour améliorer l'accessibilité des documents et autres produits, par exemple avec des fonctions de transcription et de conversion de la parole en texte, et des outils personnalisés d'accessibilité et de simplification linguistique. Par ailleurs, des acteurs privés, de moindre envergure, ainsi que de jeunes entreprises développent d'autres outils dans ce domaine, et la Recommandation entend créer les conditions nécessaires pour que leurs travaux de recherche se concrétisent. Enfin, les technologies de reconnaissance d'images par l'IA pourraient être utilisées pour analyser de grandes quantités de données afin d'évaluer la représentation des femmes et des minorités dans différents secteurs des médias, par exemple dans les programmes télévisés ou les films, ou encore dans la publicité traditionnelle ou en ligne.

46. Le second exemple fourni dans la Recommandation concerne l'utilisation des systèmes d'IA pour la détection, la prévention de la diffusion et la lutte contre les discours de haine en ligne, y compris les discours de haine sexistes, racistes et LGBTIphobes, et, plus généralement, les stéréotypes et la violence facilitée par la technologie à l'égard des femmes et des filles ainsi que d'autres individus et groupes, fondés sur leurs caractéristiques personnelles réelles ou perçues et sur leur situation. Comme l'indique l'exposé des motifs de la Recommandation [CM/Rec\(2022\)16](#), le discours de haine prolifère de manière exponentielle dans les espaces en ligne. D'après l'enquête réalisée par la FRA en 2018 sur les perceptions de l'antisémitisme, le taux d'incidence unique le plus élevé de harcèlement antisémite signalé était lié au cyberharcèlement. En outre, l'enquête de la FRA de 2012 sur la violence à l'égard des femmes montre que le cyberharcèlement à l'égard des femmes est très répandu dans l'Union européenne, une femme sur 20 déclarant en avoir été victime⁷³. Selon la Recommandation [CM/Rec\(2022\)16](#), les États membres devraient limiter les préjudices causés par le discours de haine en ligne en promulguant une législation efficace visant à empêcher sa publication et à assurer son retrait, lorsqu'il a déjà été publié. Compte tenu de l'ampleur du phénomène et de la vitesse à laquelle ces discours de haine en ligne se propagent, une action préventive rapide, y compris des campagnes de sensibilisation, est nécessaire pour éviter leur prolifération. Aussi la Recommandation [CM/Rec\(2022\)16](#) invite-t-elle les États membres à associer les intermédiaires internet à cette tâche. Cette recommandation prévoit également que les États membres veillent à ce que des mécanismes soient mis en place pour signaler les cas de discours de haine en ligne aux autorités publiques et aux acteurs privés, y compris les intermédiaires Internet, et à ce que des règles claires régissent le traitement de ces signalements. Ces mécanismes de signalement devraient être faciles d'utilisation, adaptés à l'âge, facilement accessibles, rapides, multilingues et efficaces. Pour être appropriés, les mécanismes de signalement doivent être adaptés à la nature des abus en ligne, y compris la violence à l'égard des femmes et des filles facilitée par la technologie, en veillant à ce qu'ils recensent les formes pertinentes de préjudice, leurs formes et leurs incidences spécifiques au genre, et à ce qu'ils prévoient des voies de renvoi appropriées. En outre, les personnes chargées de traiter les signalements devraient être sensibilisées aux besoins spécifiques des victimes d'abus en ligne, en particulier des femmes et des filles ciblées par la violence facilitée par la technologie.

47. Le fait d'utiliser des systèmes d'IA pour réaliser cette tâche permet de limiter l'exposition des êtres humains à la haine en ligne et donc aux contenus haineux. Cette approche contribue aussi à améliorer le taux et l'efficacité de la modération. Toutefois, la tâche à réaliser est complexe et délicate, en raison, d'une part, des difficultés que présente la détection de la haine en ligne (utilisation de différents médias, identification du

72. *Ibid.*

73. FRA, *Bias in Algorithms*.

contexte, etc.), et, d'autre part, des droits fondamentaux qui sont affectés par la sur-modération ou la sous-modération⁷⁴. Dans les cas plus difficiles, la décision finale de bloquer ou non un contenu avant sa publication doit donc être prise par un être humain, mais compte tenu de l'ampleur du phénomène, il faudra de plus en plus recourir à des systèmes d'intelligence artificielle pour assister l'être humain dans cette tâche, si l'on veut la mener à bien de manière efficace. Le rapport de la FRA sur les biais et les algorithmes (*ibid.*) indique que les algorithmes s'appuient parfois fortement sur certains mots pour évaluer le caractère offensant d'un message. Or ces mots peuvent varier considérablement d'une langue à l'autre et selon les modèles. D'après ce rapport, les prédictions du caractère offensant sont beaucoup plus fréquentes pour les termes « musulman », « gay » et « juif » que pour d'autres mots. On peut donc craindre que ces outils n'augmentent ou n'introduisent des biais. L'Étude examine par ailleurs le problème de la modération de contenus consistant à appliquer à des groupes minoritaires des filtres sémantiques modélisés à partir de groupes majoritaires, avec le risque de réduire les premiers au silence et donc de restreindre injustement leur capacité à communiquer sur les réseaux sociaux. Par exemple, le fait d'associer de manière systématique des mots liés au mot « lesbienne » à des contenus pornographiques entraîne souvent un « bannissement furtif » (*shadow ban* en anglais) qui restreint la visibilité des messages publiés sur les réseaux sociaux ou rend purement et simplement impossible l'utilisation de certains mots dans les noms de compte et les pseudonymes. L'effet de « musellement » de la modération de contenus a de fortes répercussions négatives sur la visibilité des organisations qui s'attachent à contrer les récits haineux et discriminatoires ciblant les personnes qui appartiennent à des communautés minoritaires, et sur la portée des activités et des événements qu'elles mettent en place à cette fin⁷⁵. De nombreuses recherches sont en cours pour tenter d'atténuer les biais dans les outils linguistiques d'IA polyvalents existants, notamment au moyen de stratégies destinées à réduire les biais sociétaux dans le modèle d'origine⁷⁶. Les États membres sont encouragés à assurer le financement des acteurs publics et privés qui mènent des recherches dans ce domaine, afin d'améliorer la technologie et de lever ces limitations.

48. Les systèmes d'IA peuvent également être conçus à dessein pour bénéficier aux communautés marginalisées, à risque ou mal desservies. Ils peuvent, par exemple, améliorer l'accès à l'information, aux biens et aux services, ou venir renforcer la traduction automatique dans les langues régionales ou minoritaires⁷⁷. Ainsi, le plan d'action pour les technologies linguistiques de 2018 du Gouvernement gallois s'inscrit dans une stratégie générale qui vise à promouvoir l'utilisation du gallois et à augmenter le nombre de locutrices et locuteurs actifs. Ce plan soutient différents projets, notamment la programmation d'applications d'apprentissage du gallois pour les enfants, une version galloise d'OpenStreetMap, une base de données vocales en gallois et le développement de Macsen, un assistant numérique en langue galloise. Il encourage également l'utilisation des outils linguistiques existants, ainsi que l'organisation de conférences annuelles et l'ouverture de postes de recherche réservés à ce domaine dans les universités⁷⁸. Des outils d'IA pourraient également être développés afin de détecter les formes d'inégalités structurelles (telles que les formes multiples et intersectionnelles de discrimination, ainsi que la discrimination par procuration), y compris celles découlant de l'utilisation de l'IA ou aggravées par celle-ci, ainsi que les impacts environnementaux et l'exclusion numérique. Ces analyses pourraient ensuite être utilisées pour éclairer l'élaboration proactive de politiques publiques et de réglementations.

49. La violence facilitée par la technologie, en particulier à l'encontre des femmes et des filles ainsi que des personnes exposées à la discrimination, est de plus en plus répandue en raison du développement rapide des technologies numériques et de leur utilisation généralisée. L'IA générative, permettant de générer des images ou des vidéos hyper-réalistes (contrefaçons numériques⁷⁹), est en majorité utilisée pour produire des contenus à caractère sexuel ou pornographique (95% des contrefaçons numériques à caractère sexuel

74. Pour un aperçu détaillé des questions soulevées en matière de modération des contenus, voir le rapport explicatif accompagnant la Recommandation CM/Rec(2026)2 du Comité des Ministres aux États membres sur l'obligation de rendre des comptes en matière de violence à l'égard des femmes et des filles facilitée par la technologie, paragraphes 45-159.

75. L'Eurocentralasian Lesbian Community (EL*C), organisation créée en 2017 pour défendre les droits des femmes LGBTI, indique ainsi qu'elle n'a pas pu utiliser le mot « lesbian » dans son nom d'utilisateur-rice sur Facebook, alors que d'autres mots associés à la communauté LGBTI (tels que « gay » ou « queer ») pouvaient être employés. Voir EL*C, « Lesbophobia: An intersectional form of violence », 2021, disponible à l'adresse : <https://europeanlesbianconference.org/wp-content/uploads/2021/10/Lesbophobia-3.pdf>.

76. FRA, *Bias in Algorithms* (*ibid.*), qui cite Bolukbasi et al. (2016); Manzini et al. (2019); Zhao et al. (2018); Caliskan et al. (2017); Krause et al. (2020); et Dathathri et al. (2020).

77. Voir Étude CDADI/GEC (2023) p. 89.

78. *Facilitation de la mise en œuvre de la Charte européenne des langues régionales ou minoritaires par l'intelligence artificielle*, Conseil de l'Europe, Secrétariat de la Charte européenne des langues régionales ou minoritaires, février 2022, p. 6.

79. La contrefaçon numérique à caractère sexuel fait référence à ce qui est également couramment désigné sous le terme de « deepfake à caractère pornographique ». En raison de l'inexactitude de la référence à la « pornographie » et de son potentiel à causer davantage de préjudice aux victimes, le présent exposé des motifs utilise le terme « contrefaçon numérique à caractère sexuel », et non « deepfake à caractère sexuel ». De manière analogue, dans un contexte non sexuel, le présent exposé des motifs utilise le terme « contrefaçon numérique ».

- image ou vidéo - qui circulent sont de nature sexuelle ou pornographique, selon plusieurs études)⁸⁰. Bien que ces images soient virtuelles, leur impact sur les victimes est réel, en particulier lorsque les contenus sont publiés sur des plateformes en ligne publiques. Les systèmes d'IA peuvent reproduire, perpétuer et exacerber les biais, les stéréotypes, la stigmatisation et les préjugés existants, contribuant ainsi à la persistance de récits préjudiciables et de résultats discriminatoires. Cela peut à son tour conduire à une augmentation de la violence en ligne et hors ligne à l'encontre des femmes en particulier, qui sont ciblées de manière disproportionnée en raison du sexisme, des stéréotypes et inégalités de genre ainsi que des déséquilibres de pouvoir bien ancrés que les systèmes d'IA peuvent amplifier par inadvertance ou en l'absence de régulation. Pour faire face à ces risques, il faut prendre des mesures proactives, notamment identifier et atténuer les biais dans les systèmes d'IA, empêcher la diffusion de la haine sexiste en ligne et la modérer, interdire la génération et le partage non-consenti d'images intimes et de contrefaçons numériques, et s'attaquer aux schémas plus larges de violence facilitée par la technologie⁸¹.

50. S'il convient de saluer les bonnes pratiques telles que celles citées en exemple ci-dessus, il importe aussi de noter qu'il est essentiel d'éviter de trop s'appuyer sur l'IA en tant que solution autonome pour lutter contre les inégalités et les discriminations (techno-solutionnisme). En effet, les problèmes sociaux appellent des solutions sociales, et si l'IA peut effectivement favoriser l'égalité, y compris l'égalité de genre, et la non-discrimination, elle doit venir étayer – et non remplacer – des politiques d'égalité bien conçues et assorties de financements suffisants⁸². En outre, les États membres doivent garder à l'esprit que même lorsque de telles initiatives positives sont mises en œuvre, l'utilisation de systèmes d'IA comporte toujours un risque de discrimination.

IV. Prévention de la discrimination

Paragraphe 13

51. Ce chapitre fournit aux États membres des orientations pour éviter de favoriser l'introduction de biais dans les systèmes d'IA et leur pérennisation, et ainsi atténuer le risque de discriminations portant atteinte à l'article 14 de la Convention et à d'autres droits consacrés par cette dernière. Il est organisé en trois sous-chapitres, qui correspondent à trois ensembles de principes tirés de la Convention-cadre (articles 8 à 10). Comme au chapitre II, dans toutes les mesures qu'ils prennent, les États membres doivent appliquer le principe de proportionnalité énoncé au paragraphe 6 de la Recommandation afin de moduler ces obligations en fonction du risque posé aux droits à l'égalité et à la non-discrimination des personnes concernées, en tenant compte de la taille, de la capacité et de la puissance économique des fournisseurs et des déployeurs. Le paragraphe 13 met également l'accent sur l'approche graduée et différenciée, comme expliqué au paragraphe 24 du présent exposé des motifs.

52. Le paragraphe 13 s'inspire du libellé de l'article 16 de la Convention-cadre, à savoir « adopte[r] ou maint[enir] des mesures afin d'identifier, d'évaluer, de prévenir et d'atténuer les risques posés par les systèmes d'intelligence artificielle en tenant compte des impacts réels et potentiels sur les droits de l'homme, la démocratie et l'État de droit ». Il se veut exhaustif. De telles mesures pourraient être prises ou maintenues par les États membres ou par les fournisseurs et les déployeurs de systèmes d'IA, le cas échéant.

Paragraphe 14

53. Au paragraphe 14, les États membres sont encouragés à étudier les nouvelles formes de vulnérabilité algorithmique ou fondée sur l'IA qui sont susceptibles d'apparaître avec l'utilisation des systèmes d'IA et à promouvoir la recherche sur les méthodes de protection juridique contre cette vulnérabilité. Ces études devraient respecter les principes de confidentialité et de protection des données à caractère personnel, comme cela est indiqué au paragraphe 5.4 de la Recommandation. La recommandation englobe à la fois les études entreprises par les États membres et le soutien aux études menées par des chercheurs et chercheuses, y compris des universitaires, ou d'autres parties privées. Le soutien ne se limite pas au financement et comprend également, par exemple, la création d'un environnement favorable à la recherche. Le déploiement des systèmes d'IA peut produire un tri social et une différenciation sociale qui remettent en cause les limites du droit en matière de discrimination. Ce tri social peut créer de nouvelles formes de vulnérabilité sociale, par

80. Voir l'étude sur « *Tackling deepfakes in European policy* » (en anglais seulement), European Parliamentary Research Service, juillet 2021.

81. En ce qui concerne la violence à l'égard des femmes et des filles facilitée par la technologie, il est fait référence à la Recommandation CM/Rec(2026)2 du Comité des Ministres aux États membres sur l'obligation de rendre des comptes en matière de violence à l'égard des femmes et des filles facilitée par la technologie.

82. Voir Étude CDADI/GEC (2023) p. 89.

exemple en appliquant systématiquement à certains groupes des conditions économiques moins favorables ou en les excluant de certains biens et services. Ces nouvelles formes de différenciation sociale peuvent créer de « nouvelles discriminations » et des inégalités socioéconomiques⁸³. Les États membres sont encouragés à préciser que l'interdiction de la discrimination prévue à l'article 14 couvre la discrimination multiple et intersectionnelle et la discrimination indirecte, y compris la discrimination par procuration, car ces types de discrimination sont des formes de préjudices discriminatoires que les systèmes d'IA peuvent produire, et à étudier d'autres inégalités socioéconomiques causées, entre autres, par la fracture numérique. Les États membres sont aussi encouragés à publier les résultats de ces recherches, afin de les mettre à la disposition des autres États membres et du grand public, y compris des partenaires sociaux et des chercheuses et chercheurs intéressés.

54. Dans le même ordre d'idées, les États membres sont encouragés à revoir leur politique et leurs cadres juridiques en matière de collecte de données sur l'égalité en vue d'améliorer la disponibilité et la qualité de ces données, qui sont utiles pour un ensemble de mesures de protection de l'égalité, telles que, par exemple, les actions positives ainsi que les pratiques de suivi ou de profilage. Ils pourront s'inspirer de l'approche recommandée dans l'exposé des motifs de la Recommandation [CM/Rec\(2024\)1](#) sur l'égalité pour les femmes et les filles roms et de la communauté des Gens du voyage (paragraphe 233-240). Toutefois, la Recommandation prend la pleine mesure des limites auxquelles les États membres peuvent se trouver confrontés dans ce contexte. Par exemple, la collecte de données relatives aux caractéristiques protégées est restreinte dans certains États membres, et même lorsque l'accès à ces données est possible, leur utilisation peut être limitée. Dans ce cas, des méthodes de recherche qualitative ou des substituts sociodémographiques pourraient être utilisés. Les États membres devraient respecter la législation sur la protection des données lorsqu'ils autorisent l'utilisation de données relatives à l'égalité et, en particulier, ils devraient renforcer les mesures de confidentialité concernant les données sensibles liées à l'orientation sexuelle, à l'identité et à l'expression de genre et aux caractéristiques sexuelles, ainsi qu'à l'appartenance à un groupe ethnique, culturel, linguistique ou religieux, afin d'éviter tout risque d'exposition ou d'utilisation abusive. Toute avancée dans ce domaine devrait également se faire en consultation avec les groupes les plus exposés à la discrimination, afin de garantir qu'aucune collecte de données ne soit réalisée à l'encontre des souhaits de ces groupes.

Transparence, explicabilité et contrôle en matière d'égalité et de non-discrimination

Paragraphe 15

55. Le principe de transparence fait référence à l'ouverture et à la clarté dans la gouvernance des activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle (IA) et signifie spécifiquement que les processus décisionnels et le fonctionnement général des modèles d'intelligence artificielle devraient être compréhensibles et accessibles aux actrices et acteurs appropriés de ce secteur et, le cas échéant, aux parties prenantes concernées. Dans certains cas, la transparence pourrait aussi impliquer la fourniture d'informations supplémentaires sur les algorithmes utilisés, sous réserve des impératifs liés à la propriété commerciale et intellectuelle, à la sécurité nationale et autres contraintes⁸⁴. La transparence peut également inclure des considérations pour le contexte social telles que la participation des parties prenantes au processus de développement et les résultats de toute évaluation d'impact.

56. L'explicabilité est un aspect particulièrement important de la transparence. Les systèmes d'IA basés sur les systèmes d'apprentissage automatique (ML) ou d'apprentissage profond (DL) utilisent des algorithmes qui sont appris par leur propre processus d'apprentissage et non par une programmation humaine explicite. Au cours du processus d'apprentissage, les modèles d'IA peuvent découvrir de nouvelles corrélations entre certaines caractéristiques d'entrée et prendre des décisions ou faire des prédictions qui reposent sur des modèles très complexes intégrant un grand nombre de paramètres interdépendants pouvant atteindre plusieurs millions. Cette complexité rend l'interprétation des résultats produits particulièrement difficile, même pour les expert-es⁸⁵. En effet, les raisons pour lesquelles les systèmes prennent certaines décisions peuvent rester obscures, tant pour les fournisseurs/responsables du déploiement de ces systèmes que

83. Voir Matthias Leese, « The new profiling: Algorithms, black boxes, and the fail of anti-discriminatory safeguards in the European Union » dans *Security Dialogue*, vol. 45, p. 494-511, p. 501, 2014 ; Monique Mann et Tobias Matzner, « Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to Emerging discrimination », *Big Data & Society*, vol. 6, p. 5 et 6, 2019.

84. Voir le rapport explicatif de la Convention d'Istanbul, paragraphe 57.

85. *TechDispatch: Explainable Artificial Intelligence*, *European Data Protection Supervisor*, 2023, document qui fait référence à Peters, U. « Explainable AI lacks regulative reasons: why AI and human decision-making are not equally opaque », *AI and Ethics*, 2023,

pour les personnes qui en subissent l'influence. L'effet « boîte noire » qui en résulte peut entraîner soit une confiance excessive, soit une dépendance injustifiée à l'égard des systèmes d'IA, avec des conséquences qui peuvent être préjudiciables pour les personnes concernées⁸⁶. Cette opacité ne se limite pas à une difficulté d'interprétation des décisions; elle peut également avoir des conséquences humaines immédiates en dissimulant d'éventuelles défaillances des systèmes d'IA, telles que l'existence de biais, d'inexactitudes ou encore d'« hallucinations ».

57. L'« explicabilité » renvoie à la capacité de fournir, sous réserve de faisabilité technique et compte tenu de ce qui est généralement reconnu comme étant « à la pointe », des explications suffisamment compréhensibles sur les raisons pour lesquelles un système d'IA fournit des informations, produit des prédictions, des contenus, des recommandations ou des décisions, ce qui est particulièrement crucial dans des domaines sensibles tels que les soins de santé, la finance, les migrations, les services frontaliers et la justice pénale, où il est essentiel de comprendre le raisonnement qui sous-tend les décisions produites ou assistées par un système d'IA, tant pour les individus que pour le pouvoir judiciaire qui doit procéder à un contrôle, le cas échéant. Dans de tels cas, la transparence pourrait, par exemple, prendre la forme d'une liste de facteurs que le système d'IA prend en considération lorsqu'il informe ou prend une décision (voir le paragraphe 60 du rapport explicatif de la Convention-cadre) qui pourrait inclure des déductions tirées des données. La nécessité d'obtenir des explications claires est également soulignée à l'article 86, paragraphe 1, du règlement de l'UE sur l'IA relatif au droit à l'explication de la prise de décision individuelle, qui dispose que « [t]oute personne concernée faisant l'objet d'une décision prise par le responsable du déploiement sur la base du résultat d'un système d'IA à haut risque figurant à l'annexe III [...] et qui produit des effets juridiques ou affecte cette personne de manière significative d'une façon qu'elle considère comme ayant une incidence négative sur sa santé, sa sécurité ou ses droits fondamentaux, a le droit d'obtenir du responsable du déploiement des explications claires et utiles sur le rôle du système d'IA dans la procédure décisionnelle et sur les principaux éléments de la décision prise ». Dans le règlement de l'UE sur l'IA, cette obligation ne s'applique qu'aux systèmes à haut risque. Cependant, l'interdiction de la discrimination et le principe de proportionnalité s'appliquent dans toutes les dispositions qu'elle contient et peuvent être utilisés pour adopter une approche progressive fondée sur le risque et d'autres facteurs contextuels.

58. Certaines informations techniques produites dans un but de transparence ou d'explicabilité peuvent être incompréhensibles pour le grand public tout en étant comprises par les organes de contrôles, les organismes de promotion de l'égalité ou les acteurs et actrices et les organisations de la société civile qui interviennent dans ce domaine, notamment grâce à la participation d'expert-es. Les obligations et les systèmes mis en place par les États membres doivent donc prendre en considération les différents niveaux de compréhension des destinataires de l'information et être adaptés en conséquence. En règle générale, lorsque des informations sont communiquées à des personnes physiques, les explications fournies doivent être communiquées dans un langage simple et courant, en évitant autant que possible les termes techniques avancés.

59. Les mesures concrètes présentées au paragraphe 15 visent à aider les États membres à améliorer la transparence et l'explicabilité des systèmes d'IA. Ces principes doivent être pris en compte lors des phases de conception et de développement. L'obligation de publier des données statistiques peut permettre aux parties intéressées d'évaluer les effets discriminatoires d'un système particulier. Les données devraient être ventilées afin de montrer l'impact du système sur les femmes et les groupes identifiés au paragraphe 2 de la Recommandation, y compris le risque de discrimination intersectionnelle. Les cas échéant, elles pourraient également comporter des données sur le fonctionnement des systèmes d'IA, notamment, si possible, les critères de décision utilisés et, s'il y a lieu, des informations sur les données d'entraînement et de validation utilisées ainsi que leur traitement. Les obligations de transparence doivent être mises en balance avec les règles relatives à la vie privée et à la protection des données personnelles, à la confidentialité, aux secrets commerciaux et la sécurité nationale. En particulier, les États membres sont encouragés à introduire des obligations juridiques pour que les utilisateur-ices de systèmes d'IA fournissent des explications satisfaisantes aux consommateur-ices et aux organismes de contrôle public, à la demande de ces derniers, particulièrement dans le cadre des litiges juridiques.

60. Les mécanismes de transparence, peuvent comprendre l'établissement de rapports sur le recours à l'IA au sein des collectivités locales régionales et nationales, un processus permettant aux sociétés de rendre compte de l'utilisation responsable de l'IA dans leurs rapports, ou encore l'utilisation de registres des systèmes d'IA. La [Coalition des villes pour les droits numériques](#) (en anglais seulement), coalition de 39 villes de l'UE et des États-Unis, a également pris des mesures pour veiller à ce que les villes utilisent la technologie de manière ouverte et transparente. Dans sa déclaration, la coalition affirme plusieurs grands principes, dont la

86. *TechDispatch: Explainable Artificial Intelligence*, 2023.

transparence, l'obligation de rendre des comptes et la non-discrimination des algorithmes. En pratique, cela signifie que le public «devrait avoir accès à des informations compréhensibles et précises sur les systèmes technologiques, algorithmiques et d'intelligence artificielle qui ont une incidence sur leur vie» et qu'il devrait pouvoir «remettre en question et modifier les systèmes injustes, partiels ou discriminatoires». En outre, les personnes concernées par les systèmes d'IA devraient être informées du fait qu'elles interagissent avec une machine et non avec un être humain. À cette fin, les États membres peuvent également sensibiliser le public à la manière d'identifier une interaction avec des systèmes d'IA ou de repérer les systèmes d'IA qui traitent des données à caractère personnel dans divers domaines. Lorsqu'une institution publique nationale ou locale utilise des systèmes d'IA dans le cadre de ses activités, elle peut, le cas échéant, envisager de former ses agent-es afin qu'elles ou ils fournissent des explications sur les critères utilisés par ces systèmes pour prendre des décisions.

61. En outre, conformément à l'article 15(2) de la Convention-cadre, les personnes concernées devraient être informées lorsqu'elles interagissent avec une machine plutôt qu'avec un être humain, lorsque cela est approprié au regard du contexte⁸⁷. Le paragraphe 15.3 s'inspire de l'article 50(1) du règlement de l'UE sur l'IA, qui dispose que «Les fournisseurs veillent à ce que les systèmes d'IA destinés à interagir directement avec des personnes physiques soient conçus et développés de manière que les personnes physiques concernées soient informées qu'elles interagissent avec un système d'IA, sauf si cela ressort clairement du point de vue d'une personne physique normalement informée et raisonnablement attentive et avisée, compte tenu des circonstances et du contexte d'utilisation». L'article 50(2) prévoit en outre que les fournisseurs de systèmes d'IA, y compris de systèmes d'IA à usage général, qui génèrent des contenus de synthèse de type audio, image, vidéo ou texte, veillent à ce que les sorties des systèmes d'IA soient marquées dans un format lisible par machine et identifiables comme ayant été générées ou manipulées par une IA. Dans le même ordre d'idées, la Recommandation précise que les États membres peuvent également sensibiliser le public à la manière de reconnaître les communications avec des systèmes d'IA ou de reconnaître les systèmes d'IA qui traitent des données à caractère personnel dans différents domaines. Lorsqu'un État ou une administration locale utilise des systèmes d'IA dans le cadre de son travail, ces institutions peuvent, le cas échéant, envisager de former leurs employés afin qu'ils fournissent des explications sur les critères utilisés par les systèmes d'IA dans la prise de décision.

Paragraphe 16

62. Les registres publics répertoriant les systèmes d'IA contribueraient à renforcer la transparence, la responsabilité et le contrôle public éclairé de l'utilisation des systèmes d'IA susceptibles de présenter des risques pour l'égalité, y compris l'égalité de genre. Le règlement de l'UE sur l'IA prévoit déjà la mise en place d'une base de données pour l'enregistrement des systèmes d'IA à haut risque utilisés dans certains domaines, ainsi que pour leur déploiement par les autorités publiques. La municipalité d'Amsterdam, par exemple, a créé un tel registre, afin de disposer d'une vue d'ensemble des systèmes d'IA et des algorithmes utilisés par la ville⁸⁸. La création de nouveaux registres pourrait entraîner une charge administrative et financière importante pour les États membres. La Recommandation reconnaît que certains États membres disposent déjà de cadres susceptibles de répondre à cette recommandation, par exemple des registres tenus par les autorités de protection des données, ou d'autres registres non spécifiques aux systèmes d'IA, lesquels pourraient être adaptés pour les inclure, afin de réduire les coûts et les délais.

Paragraphe 17

63. En outre, la Recommandation vise à améliorer la capacité des sujets qui font l'objet de décisions prises ou facilitées par des systèmes d'IA à détecter les discriminations et à surmonter les défis liés à une combinaison de facteurs tels que la personnalisation, l'automatisation et l'opacité des processus décisionnels. La comparaison avec des personnes de même rang et les interactions sociales sont des dispositifs heuristiques importants pour identifier et contester une discrimination. Or il devient quasiment impossible de prendre connaissance d'indices sociaux ou de se comparer à d'autres demandeuses ou demandeurs de prêt dans le contexte d'un service de crédit en ligne, par exemple. Cette asymétrie d'information fait qu'il est difficile de soupçonner une discrimination en premier lieu. Même en cas de suspicion, la collecte de preuves représente un problème supplémentaire, car les décisions ou les recommandations qui les sous-tendent ne sont pas facilement consultables et sont souvent non divulguées par les utilisateur-ices des systèmes de

87. Voir, par exemple, *L'administration et vous*, Un manuel, 3^e édition, Principes de droit administratif concernant les relations entre l'Administration et les personnes, Conseil de l'Europe, à la page 25.

88. [City of Amsterdam Algorithmic Register](#), Gemeente Amsterdam.

prise de décision fondés sur l'IA. Par conséquent, la présentation de preuves établissant une présomption de discrimination devant les tribunaux est un défi juridique majeur. Il est donc recommandé aux États membres de veiller à ce que les obligations liées à la traçabilité des activités tout au long du cycle de vie des systèmes d'IA permettent aux sujets de ces systèmes d'accéder à ces informations et de les utiliser dans le cadre de plaintes pour discrimination, afin de remédier aux asymétries d'information. Cette recommandation s'appuie sur l'article 86 du règlement de l'UE sur l'IA, mentionnée ci-dessus. La Recommandation utilise le terme général «revendiquer ses droits». Cela peut inclure, par exemple, la contestation de décisions ou le dépôt de plaintes, comme indiqué au paragraphe 100 de l'exposé des motifs de la Convention-cadre. Les mécanismes de plainte devraient être facilement accessibles.

Paragraphe 18

64. La Recommandation prend comme point de départ que les droits à des garanties procédurales et à des recours effectifs sont garantis par la Convention, ainsi que par d'autres lois et normes internationales et nationales applicables à la discrimination, et rappelle leur application dans le contexte des systèmes d'IA. Lorsqu'une décision est prise par un système d'IA, la Recommandation conseille aux États membres d'encourager les fournisseurs et déployeurs de systèmes d'IA à permettre un contrôle humain de ces résultats à la demande d'une personne suspectant qu'une discrimination a eu lieu. Une telle possibilité serait particulièrement importante lorsque l'exercice des droits humains d'un individu est significativement affecté. Le contrôle humain doit être significative et impliquer la participation active d'au moins un opérateur humain formé et correctement doté en ressources, capable de surveiller un résultat et d'évaluer les décisions du système, et qui a la capacité d'intervenir si nécessaire de manière à améliorer la qualité des décisions prises par le système, plutôt que comme une simple formalité procédurale ou un geste symbolique⁸⁹. Il est important de souligner que, bien que le contrôle humain puisse fournir une couche supplémentaire d'investigation dans la prise de décision et améliorer la responsabilité, elle ne peut pas résoudre les problèmes liés aux biais systémiques et ne doit pas remplacer une conception qui prend en compte l'impact des systèmes d'IA sur l'égalité, y compris l'égalité de genre. Le contrôle humain n'est pas infaillible et il convient de veiller à éviter le biais d'automatisation⁹⁰. En outre, et en dehors du contentieux, comme dans d'autres contextes ne concernant pas les systèmes d'IA, les individus pourraient être informés des procédures leur permettant de contester une discrimination suspectée causée par un système d'IA, par exemple en déposant une plainte ou en sollicitant l'aide d'un organe pour l'égalité, d'une institution nationale des droits humains, d'un médiateur ou d'une autorité de protection des consommateurs. Dans ce cas, ces informations devraient être fournies dans un format clair et accessible.

Paragraphe 19

65. Enfin, la Recommandation reconnaît que les preuves sont essentielles pour constater la discrimination et que le sujet concerné en est rarement en possession. Afin d'aider à corriger les asymétries d'information et d'éclairer le rôle plus large des organismes de promotion de l'égalité, des institutions nationales des droits humains, des médiatrices et médiateurs, des autorités compétentes et des organisations d'intérêt public, la Recommandation invite les États membres, au paragraphe 18, à envisager de leur accorder le droit d'accéder aux informations et de tester les systèmes d'IA. S'agissant des organismes de promotion de l'égalité, la Recommandation s'inspire de la Recommandation de politique générale n°2 révisée de l'ECRI sur les organismes de promotion de l'égalité chargés de lutter contre le racisme et l'intolérance au niveau national⁹¹, de la directive 2024/1499⁹² et de la directive 2024/1500⁹³, qui prévoient que les États membres veillent à ce que ces organismes soient habilités à mener une enquête pour déterminer s'il y a eu violation

89. Voir European Data Protection Supervisor TechDispatch 2/2025 – Human Oversight of Automated Decision-Making (en anglais seulement).

90. Selon le rapport explicatif de la Convention-cadre, paragraphe 75 e): «les biais d'automatisation ou de confirmation, par lesquels les humains peuvent accorder une plus grande confiance aux machines et aux artefacts technologiques ou dans des situations où ils sélectionnent des informations qui confirment leurs propres opinions, en ignorant dans les deux cas leur propre jugement potentiellement contradictoire et en validant les résultats algorithmiques sans les remettre en question».

91. Voir la Recommandation de politique générale n° 2 révisée de l'ECRI sur les organismes de promotion de l'égalité chargés de lutter contre le racisme et l'intolérance au niveau national, paragraphe 21.

92. Directive (UE) 2024/1499 du Conseil du 7 mai 2024 relative aux normes applicables aux organismes pour l'égalité de traitement dans les domaines de l'égalité de traitement entre les personnes sans distinction de race ou d'origine ethnique, de l'égalité de traitement entre les personnes en matière d'emploi et de travail sans distinction de religion ou de convictions, de handicap, d'âge ou d'orientation sexuelle et de l'égalité de traitement entre les femmes et les hommes en matière de sécurité sociale ainsi que dans l'accès à des biens et services et la fourniture de biens et services, et modifiant les directives 2000/43/CE et 2004/113/CE.

93. Directive (UE) 2024/1500 du Parlement européen et du Conseil du 14 mai 2024 relative aux normes applicables aux organismes pour l'égalité de traitement dans le domaine de l'égalité de traitement et de l'égalité des chances entre les femmes et les hommes en matière d'emploi et de travail, et modifiant les directives 2006/54/CE et 2010/41/UE.

du principe d'égalité de traitement tel qu'il est énoncé dans diverses directives. Les États sont censés en particulier leur fournir un cadre qui leur donne la possibilité de rassembler les faits, leur accorder des droits effectifs d'accès aux informations et aux documents nécessaires pour établir l'existence d'une discrimination, et prévoir des mécanismes pour qu'ils puissent coopérer avec les organismes publics compétents à cette fin. De même, en vertu du règlement de l'UE sur l'IA, l'article 77, paragraphe 1, relatif aux pouvoirs des autorités chargées de la protection des droits fondamentaux, « [l]es autorités ou organismes publics nationaux qui contrôlent ou veillent au respect des obligations découlant du droit de l'Union en matière de protection des droits fondamentaux, y compris le droit à la non-discrimination, en ce qui concerne l'utilisation des systèmes d'IA à haut risque visés à l'annexe III, sont habilités à demander et à consulter toute documentation créée ou conservée en vertu du présent règlement dans une langue et un format accessibles, lorsque l'accès à cette documentation est nécessaire à l'accomplissement efficace de leur mandat dans les limites de leur compétence. » Le considérant 157 du préambule dispose que « [l]orsque leur mandat l'exige, les autorités ou organismes publics nationaux qui supervisent l'application du droit de l'Union protégeant les droits fondamentaux, y compris les organismes de promotion de l'égalité et les autorités chargées de la protection des données, devraient également avoir accès à toute documentation créée en vertu du présent règlement. »

66. Il est recommandé que ces organismes soient autorisés à procéder à des essais, notamment en participant à des « bacs à sable réglementaires » coordonnés en matière d'IA. Un bac à sable réglementaire est un outil qui permet d'étudier et d'expérimenter des produits, services ou entreprises nouveaux et novateurs sous la supervision d'un régulateur, dans un environnement contrôlé. Lorsque ces bacs à sable sont adoptés, ces organismes devraient être consultés et avoir la possibilité de fournir des orientations utiles en matière de droits humains et de promouvoir l'intégration de garanties en la matière tout au long du cycle de vie de l'IA⁹⁴.

Évaluation d'impact, certification et évaluation et suivi en matière d'égalité et de non-discrimination

Paragraphe 20

67. Cette sous-section porte sur l'évaluation et l'examen des systèmes d'IA au regard des principes d'égalité et de non-discrimination. Plusieurs normes existantes peuvent servir de référence aux États membres lorsqu'ils réalisent une évaluation des risques et une étude d'impact spécifiquement axés sur l'égalité et la non-discrimination. La Recommandation [CM/Rec\(2020\)1](#), dans son paragraphe 5.2, prévoit que les États membres, ainsi que tous les acteurs privés amenés à collaborer avec eux ou à travailler en leur nom, devraient mener régulièrement et à titre consultatif, des études d'impact des systèmes algorithmiques sur les droits humains avant toute passation de marchés publics, pendant la conception et le développement, à des étapes régulières et tout au long du déploiement spécifique au contexte pour identifier les risques de résultats préjudiciables pour les droits. La Recommandation [CM/Rec\(2022\)17](#) sur la protection des droits des femmes et des filles migrantes, réfugiées et demandeuses d'asile, dans son paragraphe 22, prévoit que les États membres devraient faire en sorte qu'une évaluation de l'impact sur les droits humains, incluant une perspective de genre, soit réalisée avant l'introduction de systèmes liés à l'intelligence artificielle et de systèmes de prise de décision automatisés dans le domaine de la migration. De même, l'article 16 de la Convention-cadre contient une disposition spécifique prescrivant la nécessité d'identifier, d'évaluer, de prévenir et d'atténuer *ex ante* et de manière itérative tout au long du cycle de vie d'un système d'IA les risques pertinents et les impacts potentiels pour les droits humains, la démocratie et l'État de droit, en suivant et en permettant le développement d'une méthodologie assortie de critères concrets et objectifs pour de telles évaluations. HUDERIA (méthodologie pour l'évaluation des risques et des impacts des systèmes d'IA du point de vue des droits humains, de la démocratie et de l'État de droit) peut être utilisé, et les États membres doivent veiller à ce que les questions d'égalité et de non-discrimination soient intégrées tout au long de l'utilisation de la méthodologie. De telles obligations de transparence et de justification *ex ante* pourraient par exemple se traduire en une obligation faite aux responsables du déploiement et aux fournisseurs de systèmes d'IA de procéder à des évaluations et analyses préalables des risques de discrimination et de l'impact sur l'égalité, de manière indépendante ou dans le cadre d'une analyse plus vaste de l'impact sur les droits humains. Ces analyses d'impact devraient inclure ou être complétées par des analyses d'impact sur l'égalité de genre et, selon le contexte, par des analyses d'impact sur les droits de l'enfant. Les évaluations d'impact sur l'égalité incluent idéalement une analyse intersectionnelle et des données sur l'égalité ventilées, tenant compte de la manière dont les impacts discriminatoires interagissent avec les motifs énumérés au paragraphe 2 de la Recommandation ainsi qu'avec d'autres facteurs structurels et sociaux. La ventilation

94. *Les droits humains dès la conception de l'IA – une protection durable des droits humains à l'ère de l'intelligence artificielle*, 2023, p. 18-20, recommandation de suivi à *Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme*, 2019, p. 35.

des données permet aux chercheurs de comparer les impacts sur les inégalités entre différents groupes de population. Les simples moyennes de données peuvent masquer des disparités sous-jacentes ; les données ventilées, en revanche, peuvent montrer des impacts différenciés sur les droits humains entre les groupes. Par exemple, les moyennes nationales peuvent montrer que la population active dans son ensemble perçoit des salaires suffisants pour couvrir le coût de la vie ; les données ventilées, en revanche, peuvent montrer que les femmes gagnent significativement moins que leurs homologues masculins, impactant leur capacité à se nourrir, se loger et subvenir à d'autres besoins essentiels⁹⁵. Conformément au paragraphe 5.4. de la Recommandation, les droits à la vie privée et à la protection des données doivent être respectés tout au long de ce processus.

68. Les principaux éléments de cette évaluation pourraient être les suivants :

1. Identification des impacts : recensement des possibilités d'utilisation du système d'IA pour soutenir des actions positives qui font progresser l'égalité et la non-discrimination conjointement avec la détection des risques pertinents pour l'égalité, y compris l'égalité de genre et la non-discrimination ;
2. Analyse d'impact : évaluation des impacts, en tenant compte de la probabilité et de la gravité des effets sur l'égalité, tant positifs que négatifs ;
3. Analyse de la gouvernance : évaluation des rôles et des responsabilités des détenteurs et détentrices d'obligations, des détenteurs et détentrices de droits et des parties prenantes dans la mise en œuvre et la gouvernance des mécanismes visant à renforcer les impacts positifs et à atténuer tout impact négatif ;
4. Réponse, atténuation et évaluation : identification des mesures appropriées pour renforcer les impacts positifs, ainsi que des mesures pour atténuer les risques et assurer une évaluation continue.

69. Les États membres sont encouragés à rendre ces évaluations juridiquement contraignantes ou à adopter des mesures d'incitation fortes en intégrant les analyses d'impact aux éléments pris en compte dans les procédures judiciaires concernant les allégations de discrimination liées à des systèmes d'IA, par exemple, pour évaluer si les obligations positives relatives à l'égalité de traitement ont été respectées ou pour évaluer la charge de la preuve. Les résultats des évaluations des risques de discrimination et des analyses d'impact sur l'égalité pourraient aussi conditionner la mise sur le marché et/ou le déploiement d'un système donné. Un tel système de « pré-autorisation » oblige les fournisseurs et utilisateur·rices de systèmes pouvant présenter des risques à prendre des mesures préventives, voire à soumettre leurs systèmes à une procédure de certification ou d'autorisation d'exercice. Le principe juridique de proportionnalité pourrait faciliter l'adaptation de ces obligations à la taille, aux capacités et au pouvoir économique des fournisseurs et des utilisateur·rices, et les éléments présentés au paragraphe 23 du présent exposé des motifs devraient également peser dans la balance⁹⁶, en tenant compte du niveau d'impact positif et du risque de discrimination qui ont été identifiés. Les évaluations des risques et analyses d'impact devraient être facilement accessibles et rendues publiques, le cas échéant. Elles devraient analyser les impacts discriminatoires potentiels des systèmes d'IA tout au long de leur cycle de vie et pour l'ensemble des motifs protégés au paragraphe 2 de la Recommandation. Une façon d'assurer la supervision pourrait être de donner à un organisme de réglementation le pouvoir de vérifier si le fonctionnement des systèmes d'IA désavantage des individus et contribue à des inégalités structurelles tout au long de leur cycle de vie.

70. Les États membres peuvent également s'inspirer du système de gestion des risques prévu à l'article 9 du règlement de l'UE sur l'IA, qui dispose qu'« un système de gestion des risques est établi, mis en œuvre, documenté et tenu à jour en ce qui concerne les systèmes d'IA à haut risque ». Ce système « doit être considéré comme un processus itératif continu planifié et exécuté tout au long du cycle de vie d'un système d'IA à haut risque et qui doit périodiquement faire l'objet d'un examen et d'une mise à jour méthodiques ». Lors de la mise en œuvre du système de gestion des risques, les fournisseurs examinent si, compte tenu de l'objectif visé, le système d'IA à haut risque est susceptible d'avoir un impact négatif sur les personnes âgées de moins de 18 ans et, le cas échéant, sur d'autres groupes vulnérables. En outre, l'article 79, paragraphe 2, relatif à la procédure à suivre au niveau national pour traiter les systèmes d'IA présentant un risque, prévoit que l'autorité nationale de surveillance du marché doit procéder à une évaluation de tout système d'IA pouvant constituer un risque tel que défini à l'article 3, point 19, du règlement (UE) 2019/1020, dans la mesure où il présente des risques pour la santé ou la sécurité, ou pour les droits fondamentaux, des personnes. En l'occurrence, « [u]ne attention particulière est accordée aux systèmes d'IA présentant un risque pour les groupes vulnérables ». De plus, « [l]orsque sont identifiés des risques pour les droits fondamentaux, l'autorité

95. [Danish Institute for Human Rights : Human rights impact assessment guidance and toolbox Report A4 \(p 67\)](#) (en anglais seulement).

96. Voir en particulier les paragraphes 57 à 65 du rapport explicatif de la Convention-cadre.

de surveillance du marché informe également les autorités ou organismes publics nationaux concernés visés à l'article 77, paragraphe 1 et coopère pleinement avec eux.» L'article 27 du règlement européen sur l'IA prévoit également une évaluation d'impact sur les droits fondamentaux pour les systèmes d'IA à haut risque.

71. Certaines juridictions, par exemple les Pays-Bas⁹⁷, ont déjà élaboré des évaluations d'impact sur les droits fondamentaux et les algorithmes complètes pour accroître la transparence des systèmes d'IA de plus en plus utilisés dans notre quotidien.

72. Le paragraphe 18.3 traite également de l'introduction de normes relatives aux audits en matière de discrimination des systèmes d'IA. Les normes d'évaluation de la discrimination, y compris les audits, peuvent comprendre des normes techniques, c'est-à-dire des ensembles de règles convenues visant à garantir que les produits, services ou processus soient sûrs, fiables et compatibles, et considérées comme des bonnes pratiques industrielles, notamment les normes établies par l'Organisation internationale de normalisation (ISO), la Commission électrotechnique internationale (CEI), le Comité européen de normalisation (CEN), le Comité européen de normalisation électrotechnique (CENELEC) ou l'Institut européen des normes de télécommunication (ETSI). Outre les normes techniques destinées à éclairer le processus d'évaluation, des normes socio-techniques ou d'autres types de normes pourraient être élaborées. La Recommandation invite les États membres à initier l'élaboration de normes pour les audits en matière de discrimination des systèmes d'IA, ce qui pourrait contribuer à garantir une approche cohérente et efficace pour traiter les biais découlant des activités menées tout au long du cycle de vie des systèmes d'IA, tout en soutenant l'innovation dans le développement de solutions techniques et de politiques destinées à lutter contre la discrimination dans le contexte des systèmes d'IA. Cela permettrait également de renforcer la confiance du public dans les systèmes d'IA. Des normes réglementaires harmonisées renforceraient en outre la responsabilité des entreprises concernant les risques de discrimination, ce qui permettrait aux victimes potentielles de disposer d'informations utiles pour évaluer l'opportunité d'intenter une action (en justice) et de fournir aux juges des éléments de preuve compréhensibles dans le cadre des plaintes pour discrimination.

73. Les audits externes des risques de discrimination devraient être réalisés par des organismes indépendants, ce qui n'est pas nécessairement le cas des audits internes. Les audits des risques de discrimination, en particulier lorsqu'ils sont effectués par des tiers, peuvent permettre d'identifier et de contribuer à corriger les biais susceptibles d'intervenir à différents stades du cycle de vie d'un système d'IA, réduisant ainsi le risque de discrimination et de perpétuation des inégalités systémiques. Les conclusions des audits externes, ainsi que celles des analyses d'impact, devraient être rendues publiques le cas échéant.

74. Les États pourraient également examiner si des mécanismes de certification pourraient être utilisés en complément pour attester que le potentiel positif des systèmes d'IA a été étudié, que les biais ont été atténués et que les risques de discrimination ont été éliminés autant que possible pour des utilisations bien spécifiques. Ces certifications devraient être délivrées par des organismes accrédités composés de professionnels qualifiés et expérimentés dans les domaines de l'IA, des droits humains et de l'égalité. En outre, les certifications devraient être renouvelées périodiquement, après un intervalle raisonnable, afin de garantir le respect continu des normes pertinentes au fil du temps et de garantir que les systèmes certifiés continuent d'adhérer aux principes d'égalité et de non-discrimination. Ces mesures de promotion et de prévention *ex-ante* pourraient servir à affirmer la nature non discriminatoire d'un système donné dans des domaines spécifiques. La certification devrait bénéficier d'une certaine publicité. Les certifications devraient par exemple être accessibles et intelligibles pour les personnes qui utilisent des systèmes d'IA (en vue d'une éventuelle défense juridique et d'une évaluation des risques pour certaines applications) ainsi que pour les institutions publiques (par exemple, les organismes de promotion de l'égalité, les institutions nationales des droits humains, les responsables de la protection des données ou d'autres autorités compétentes) et les organisations de la société civile ayant un intérêt légitime (tel que défini dans les directives de l'UE sur l'égalité et la législation nationale sur l'égalité). Il conviendrait à cet égard de s'inspirer de la Recommandation CM/Rec(2020)1, qui prévoit que «[d]es systèmes de certification fondés sur des normes régionales et internationales devraient être mis au point et appliqués pour garantir la provenance et la qualité des jeux de données et des modèles» et que ces systèmes «devraient également faire partie des processus d'approvisionnement, reposer sur les cadres réglementaires qui interdisent certaines utilisations de systèmes algorithmiques et s'y conformer.»

75. Les États membres pourraient également encourager l'adoption de normes techniques et de méthodes de validation qui mesurent et atténuent les biais dans les systèmes d'IA. De tels outils pourraient contribuer à garantir une approche cohérente et efficace pour lutter contre les biais découlant des activités tout au long

97. Les Pays-Bas ont récemment rendu les analyses d'impact sur les droits humains obligatoires pour les institutions publiques. Voir la proposition ici : <https://www.tweedekamer.nl/kamerstukken/moties/detail?id=2022Z06024&did=2022D12329>.

du cycle de vie des systèmes d'IA, tout en soutenant l'innovation dans l'élaboration de solutions techniques et de politiques visant à lutter contre la discrimination dans le contexte des systèmes d'IA.

76. Ces mécanismes d'appréciation et d'évaluation fournissent en outre des critères objectifs qui peuvent être utilisés pour étayer les procédures de passation des marchés publics. Ces procédures devraient comporter des critères précis d'éligibilité et de sélection visant à garantir que les systèmes d'IA utilisés par les autorités publiques ou d'autres organisations ne sont pas discriminatoires et favorisent l'égalité des chances (voir, dans ce contexte, le paragraphe 9 de la RPG n° 7 de l'ECRI⁹⁸). Certains États membres ont pris des mesures pour élaborer des lignes directrices sur l'acquisition, par les organismes publics, de systèmes fondés sur l'IA. Au Royaume-Uni, le gouvernement a publié un *Guide to using AI in the Public Sector* pour aider les organismes publics à adopter des systèmes fondés sur l'IA qui soient bénéfiques pour la société. Le Forum économique mondial a également publié des lignes directrices intitulées *Adopting AI Responsibly: Guidelines for Procurement of AI Solutions by the Private Sector*. De son côté, l'organisation Data Ethics a publié un *livre blanc* sur l'éthique des données dans les marchés publics. Le groupe d'experts de haut niveau de l'UE (GEHN IA) a également *recommandé* l'utilisation stratégique des marchés publics pour financer l'innovation et garantir l'obtention d'une IA digne de confiance. Il préconise à cet effet l'introduction de critères d'éligibilité et de sélection précis dans les règles et processus de passation de marchés des institutions, organismes et États membres de l'UE qui exigent que les systèmes d'IA soient dignes de confiance.

Obligation de rendre des comptes, responsabilité et règles relatives à la charge de la preuve dans le domaine de l'égalité et de la non-discrimination

Paragraphe 21 – 26

77. Cette sous-section porte sur la manière dont les États membres peuvent optimiser leurs règles procédurales dans le cadre des plaintes et des litiges liés à la discrimination, afin d'aider les personnes concernées par les systèmes d'IA à accéder à la justice et à des voies de recours effectives. L'application de cette partie de la Recommandation aux procédures pénales dépend du contexte et des règles nationales. Les principes d'obligation de rendre des comptes et de responsabilité renvoient à la nécessité de créer des mécanismes qui garantissent que les personnes, organisations ou entités qui sont chargées des activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle seront tenues pour responsables de leurs effets préjudiciables sur les droits humains, la démocratie ou l'État de droit. Il peut s'agir de procédures et de mesures administratives, judiciaires et autres, ainsi que de cadres de responsabilité civile, pénale et autres, qui permettent de contester les décisions et de remédier à des situations discriminatoires, voire d'attribuer des responsabilités et d'imposer des obligations spécifiques aux opératrices et opérateurs. Le paragraphe 66 du rapport explicatif de la Convention-cadre précise que les Parties sont tenues d'établir de nouveaux cadres et mécanismes, ou d'adapter les cadres et mécanismes existants susceptibles d'être appliqués aux activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle, afin de donner effet à cette exigence. En outre, conformément aux Lignes directrices du Comité des Ministres du Conseil de l'Europe sur l'efficacité et l'efficacé des systèmes d'assistance judiciaire dans les domaines du droit civil et du droit administratif⁹⁹, les États membres sont encouragés à veiller à ce que l'assistance judiciaire soit disponible le cas échéant.

78. Dans certains États membres, l'introduction de voies d'action collectives ou d'une responsabilité stricte serait incompatible avec le cadre juridique national existant. Néanmoins, en fonction du contexte national, il peut être approprié pour les États membres d'introduire de nouvelles procédures ou d'adapter les procédures existantes afin de garantir un accès effectif à la justice et aux voies de recours compte tenu des défis particuliers posés par les systèmes d'IA. Par exemple, l'asymétrie d'information, qui est souvent observée dans les plaintes pour discrimination visant une décision prise ou assistée par un système d'IA, se traduit par des délais de procédure prolongés qui peuvent désavantager les requérant-es¹⁰⁰. Les voies d'action collectives mentionnées au paragraphe 23 désignent les recours judiciaires engagés par un groupe de requérant-es ayant subi un préjudice similaire en raison d'un même événement, qu'il s'agisse

98. *Recommandation de politique générale n° 7 (révisée) de l'ECRI sur la législation nationale pour lutter contre le racisme et la discrimination raciale* (2017).

99. Lignes directrices du Comité des Ministres du Conseil de l'Europe sur l'efficacité et l'efficacé des systèmes d'assistance judiciaire dans les domaines du droit civil et du droit administratif, *CM(2021)36-add2final*, adopté le 31 mars 2021.

100. Voir, concernant le délai de trois mois dans les procédures de contrôle juridictionnel au Royaume-Uni, Cobbe, Jennifer, *Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making* (6 août 2018). Projet de version préliminaire publié sous le titre: Jennifer Cobbe (2019) 'Administrative law and the machines of government: judicial review of automated public-sector decision-making', *Legal Studies*, 39 (4). Voir SSRN: <https://ssrn.com/abstract=3226913> ou <http://dx.doi.org/10.2139/ssrn.3226913>.

d'une négligence, d'un incident ou d'une autre cause. Ces recours collectifs peuvent jouer en faveur des requérant-es dans le contexte des discriminations causées par des systèmes d'IA. En effet, les asymétries d'information, l'indifférence à l'égard de la discrimination alléguée, ainsi que les conséquences financières d'une action en justice contre un fournisseur ou un-e responsable du déploiement, risquent de les priver de voies de recours effectives.

79. Des présomptions réfragables concernant le (manque de) respect des obligations en matière d'égalité, y compris l'égalité de genre, et de non-discrimination pourraient être étudiées. Certains États membres disposent de règles de preuve existantes prévoyant un renversement de la charge de la preuve dans les affaires de discrimination, ou d'autres méthodes pour remédier aux asymétries d'information, voir par exemple la jurisprudence établie du système judiciaire allemand selon laquelle il peut y avoir un renversement de la charge de la preuve lorsqu'une violation résulte de la zone de risque pour laquelle une partie est responsable (*Beweislast nach Gefahrenbereichen*), ainsi que des obligations de divulgation (*sekundäre Darlegungslast*) ou un renversement de la charge de la preuve si l'expérience montre qu'il existe un schéma récurrent qui est habituellement à l'œuvre et produit des résultats discriminatoires (*Anscheinsbeweis*). La recommandation invite les États membres à envisager d'introduire un renversement provisoire de la charge de la preuve ou une présomption de conformité des systèmes d'IA. D'autres adaptations des règles de preuve pourraient être envisagées afin de faciliter l'action en justice des personnes alléguant des discriminations liées à l'absence de mesures préventives prises par les fournisseurs ou les dépoyeurs de systèmes d'IA ou à leur insuffisance. Associé à des obligations de divulgation, répondant à une approche graduée et différenciée, un tel renversement de la charge de la preuve ou une telle présomption viserait à réduire les obstacles rencontrés par les victimes dans l'établissement *prima facie* d'une discrimination. Les fournisseurs de systèmes d'IA et les responsables de leur déploiement pourraient, par exemple, éviter un renversement de la charge de la preuve ou une présomption de biais en démontrant qu'elles et ils ont mis en place des mesures préventives significatives et suffisantes. Parmi ces mesures figureraient notamment des évaluations des risques de discrimination et des impacts en matière d'égalité, ainsi qu'une certification (voir paragraphe 18). À défaut, une présomption provisoire de biais algorithmique pourrait être établie, entraînant un transfert de la charge de la preuve sur le défendeur ou la défenderesse. Ce mécanisme pourrait s'appuyer sur les règles nationales existantes en matière de renversement de la charge de la preuve. Dans ce cas, il reviendrait aux utilisateur-ices et/ou aux fournisseurs de systèmes d'IA de démontrer la conformité de ces derniers avec les exigences du droit à la non-discrimination. Ces techniques pourraient tenir compte à la fois du caractère omniprésent et prévisible des biais ainsi que des déséquilibres de pouvoir inhérents à l'utilisation de ces systèmes entre les fournisseurs/responsables du déploiement et les sujets.

80. Les États membres peuvent également s'inspirer de l'article 18 de la [directive \(UE\) 2023/970](#), relatif au renversement de la charge de la preuve, qui vise à renforcer le principe de l'égalité de rémunération pour un travail égal ou de valeur égale entre les femmes et les hommes, par la transparence salariale et des mécanismes d'exécution. Cet article dispose que les États membres doivent adopter des mesures appropriées, en accord avec leur système judiciaire national, pour garantir que, lorsqu'un-e travailleur-euse s'estime lésé par la non-application du principe d'égalité de rémunération et présente devant une autorité compétente ou une juridiction nationale des éléments factuels laissant présumer une discrimination, il revient au défendeur ou à la défenderesse de démontrer qu'aucune discrimination, qu'elle soit directe ou indirecte, n'a eu lieu en matière de rémunération. Il convient de souligner que, conformément à son paragraphe 2, cet article dispose que les États membres doivent garantir que, dans le cadre des procédures administratives ou judiciaires relatives à des allégations de discrimination, directe ou indirecte, en matière de rémunération, l'employeur ou l'employeuse qui n'a pas respecté certaines obligations en matière de transparence salariale assume la charge de prouver l'absence de toute discrimination de cette nature.

81. De tels mécanismes de responsabilité *ex ante* faciliteraient également l'application du test de proportionnalité par les juges et les organismes de promotion de l'égalité exerçant une fonction décisionnelle dans les affaires de discrimination, et l'évaluation du caractère adéquat des garanties mises en place pour la prévention. Cela inciterait les utilisateur-ices et fournisseurs d'IA à traduire les choix techniques en informations compréhensibles sur le plan juridique, informations qui pourront ensuite être utilisées pour déterminer si un système donné répond à un objectif légitime et si les moyens employés sont raisonnablement proportionnés au but poursuivi. Des obligations d'information renforcées pourraient accompagner ces mécanismes de responsabilisation. Ces obligations pourraient inclure les informations visées au paragraphe 15 de la Recommandation.

82. Le paragraphe 26 souligne que les États membres devraient veiller à ce que les personnes affectées disposent de recours effectifs. L'article 13 de la Convention garantit l'octroi d'un recours effectif devant une instance nationale à toute personne dont les droits et libertés garantis par la Convention ont été violés. Pour

être effectif, le recours doit être capable de remédier directement à la situation contestée¹⁰¹, et être suffisant et accessible, répondant à l'obligation de promptitude¹⁰². Les recours effectifs comprennent l'indemnisation des dommages non pécuniaires (paragraphe 12 des RPG n° 7 de l'ECRI sur la législation nationale pour lutter contre le racisme et la discrimination raciale). Dans le contexte de l'Union européenne, le versement d'une indemnité aux victimes peut également constituer une sanction aux fins de l'article 15 de la directive 2000/43/CE du Conseil du 29 juin 2000 relative à la mise en œuvre du principe de l'égalité de traitement entre les personnes sans distinction de race ou d'origine ethnique.

83. En ce qui concerne les responsabilités, les régimes existants de droit de la responsabilité civile sont neutres sur le plan technologique et peuvent déjà couvrir adéquatement les dommages causés par les systèmes d'IA. En particulier, les États membres devraient prendre en compte le concept juridique existant de manquements aux obligations de diligence. Lorsqu'aucune mesure préventive contre les biais algorithmiques n'a été prise par les fournisseurs et/ou les déployeurs d'un système discriminatoire, cela pourrait être qualifié de violation d'un devoir de diligence en vertu du droit de la responsabilité existant des États membres. Dans de nombreux cas, il pourrait être plus avantageux pour les victimes de s'appuyer sur les régimes de responsabilité existants, neutres sur le plan technologique, y compris le droit existant en matière de non-discrimination, car ces régimes pourraient ne pas exiger que la victime prouve si un système d'IA ou un autre type d'algorithme était impliqué dans le cas en question. Dans les cas de discrimination résultant des activités menées tout au long du cycle de vie des systèmes d'IA, les déployeurs et les fournisseurs ne devraient pas être exemptés de responsabilité simplement parce que la discrimination a été produite de manière « autonome » par le système discriminatoire. Le manquement à l'obligation de prévenir raisonnablement les discriminations résultant des activités menées tout au long du cycle de vie des systèmes d'IA devrait être considéré comme une violation du devoir de diligence. La responsabilité objective a déjà été introduite dans certains domaines bien définis pour faire face aux risques découlant de technologies pouvant causer des dommages disproportionnés à autrui.

Paragraphe 27

84. Le paragraphe 27 permet aux États membres, en accord avec leurs obligations juridiques existantes, d'envisager l'instauration d'un moratoire ou d'une interdiction à l'égard de certains systèmes d'IA lorsque le risque d'atteinte aux principes d'égalité et de non-discrimination est considéré comme tellement élevé qu'il devient inacceptable, et qu'aucune autre mesure préventive ne permet de l'atténuer efficacement. Comme indiqué aux paragraphes 111 et 112 du rapport explicatif de la Convention-cadre, la détermination de ce qui est « incompatible » dans ce contexte relève de chaque Partie, de même que l'évaluation de la nécessité d'un moratoire ou d'une interdiction, d'une part, ou d'une autre mesure appropriée, d'autre part.

Paragraphe 28

85. Enfin, le paragraphe 28 fait référence à un certain nombre d'organes, dont les autorités compétentes qui, dans ce contexte, comprennent les autorités chargées de la protection des données. Les États membres peuvent, en fonction du contexte national et du niveau d'indépendance de ces organismes, faciliter leur coopération.

V. Participation démocratique, sensibilisation du grand public et renforcement des capacités

Paragraphe 29

86. Le paragraphe 29 de la Recommandation souligne que les États membres doivent veiller à ce que les activités menées par les pouvoirs publics dans le cadre du cycle de vie des systèmes d'IA fassent l'objet d'un contrôle suffisant. Comme le souligne le paragraphe 31 du présent exposé des motifs, lorsqu'il est envisagé, le contrôle parlementaire de l'adoption de mesures visant à promouvoir l'égalité et à prévenir la discrimination dans les contextes d'IA peut comprendre, par exemple, la tenue de débats parlementaires réguliers. Le contrôle parlementaire est essentiel pour garantir que les pouvoirs publics restent transparents et responsables en matière de développement et de déploiement de systèmes d'IA, et pour se prémunir contre les risques découlant des activités menées par des acteurs publics ou privés dans le cadre du cycle de vie des systèmes d'IA, en vue de promouvoir l'égalité, y compris l'égalité de genre, et de prévenir les discriminations tout au long du cycle de vie des systèmes d'IA.

101. *Pine Valley Developments Ltd et autres c. Irlande*, décision de la Commission, 1989.

102. *Paulino Tomás c. Portugal* (dec.), 2003 ; *Çelik and İmret c. Turquie*, 2004, § 59.

Paragraphe 30-31

87. La Recommandation vise à réduire l'effet de la fracture numérique, y compris la fracture numérique entre les femmes et les hommes, c'est-à-dire le fossé entre celles et ceux qui ont accès aux technologies numériques (notamment internet et les systèmes d'IA) et celles et ceux qui n'y ont pas accès, qui résulte souvent d'une discrimination systémique. La Recommandation reconnaît également que les personnes marginalisées par la fracture numérique incluent non seulement les individus n'ayant pas un accès adéquat aux outils numériques, mais aussi celles et ceux qui n'ont pas accès aux opportunités d'apprendre à les utiliser efficacement. Un accès limité aux systèmes et aux outils d'IA, ainsi qu'aux programmes éducatifs fournissant des compétences numériques aux utilisateurs et utilisatrices, peut empêcher des individus ou des groupes de bénéficier des avantages qu'ils peuvent offrir, ce qui peut désavantager ceux-ci dans divers secteurs, notamment l'éducation, l'emploi, l'accès à la protection sociale et les soins de santé. Cela peut à son tour limiter la mobilité sociale et l'employabilité et renforcer les inégalités, en particulier parmi les jeunes, chez qui la compétence numérique est souvent attendue, voire exigée. Cela limiterait également l'effet des mesures prises pour accroître la diversité de l'emploi dans le secteur des technologies. Dans le contexte de l'égalité, y compris l'égalité de genre, et de la non-discrimination, si certains groupes ou individus rencontrent des obstacles pour accéder aux outils d'IA, ils seront probablement absents des ensembles de données utilisés pour entraîner les systèmes d'IA, ce qui accroît le risque que des résultats biaisés soient confirmés et amplifiés. Il est également moins probable que leurs besoins soient pris en considération dans la conception et la mise au point de nouveaux systèmes. Les États membres devraient donc réfléchir à la manière de combler les fractures numériques et de veiller à ce que toutes les catégories de la société, en particulier celles qui sont exposées à la discrimination, disposent d'un accès sûr et fiable à internet et aux opportunités garantissant la culture numérique¹⁰³. Ces programmes devraient être soutenus par un financement adéquat. Les mesures spécifiques peuvent comporter une aide financière pour surmonter les obstacles payants à l'accès aux outils d'IA, l'allocation de crédits et de ressources pour le développement des infrastructures, les programmes d'alphabétisation numérique et les initiatives visant à réduire les disparités, ou la possibilité de participer à des séances (par exemple, à des consultations) dans des contextes hors ligne. Les programmes d'alphabétisation numérique devraient être disponibles dans le cadre de l'éducation formelle et non formelle, et les États membres devraient envisager de les proposer le cas échéant dès le plus jeune âge.

Paragraphe 32

88. La définition de normes techniques peut avoir un effet particulièrement important sur l'égalité et la non-discrimination, car cela peut influencer la définition et l'interprétation des normes juridiques dans la pratique. Cependant, la définition de normes techniques a souvent lieu loin de tout contrôle public et sans que les organisations et les personnes intéressées y soient associées. En ce qui concerne la participation de la société civile, soulignée au paragraphe 30 de la Recommandation, il est fait référence à la Recommandation [CM/Rec\(2018\)11](#) sur la nécessité de renforcer la protection et la promotion de l'espace dévolu à la société civile en Europe. Cet instrument recommande aux États membres de veiller à ce que chacun-e, y compris les défenseuses et défenseurs des droits humains, puisse participer effectivement à la prise de décision en donnant pleinement accès à l'information et en menant des consultations publiques opportunes et transparentes, en particulier lorsque l'élaboration de politiques ou de projets de loi peut avoir une incidence sur la société civile¹⁰⁴. En outre, les chef-fes d'État et de gouvernement du Conseil de l'Europe ont réaffirmé, dans la Déclaration de Reykjavík, que la société civile est une condition préalable au bon fonctionnement de la démocratie et se sont engagés à soutenir et à maintenir un environnement sûr et favorable où la société civile et les défenseuses et défenseurs des droits humains peuvent intervenir sans entrave, sans insécurité et sans violence¹⁰⁵. Ces engagements sont directement liés à la gouvernance des systèmes d'IA, où la participation de la société civile aux processus de normalisation est essentielle pour garantir que les effets juridiques et sociétaux des technologies d'IA, y compris sur l'égalité, sont pleinement pris en considération. Une attention particulière devrait également être accordée à la participation des jeunes ou des organisations qui les représentent. Afin d'encourager les jeunes à participer, les États membres pourraient proposer des programmes éducatifs ciblés et mener des initiatives de sensibilisation dès le plus jeune âge, dans le but de développer leur compréhension et leur esprit critique à l'égard de l'IA, de l'égalité

103. [Recommandation de suivi](#) de la Commissaire aux droits de l'homme intitulée « Les droits humains dès la conception de l'IA – une protection durable des droits humains à l'ère de l'intelligence artificielle », 2023, p. 35.

104. [Recommandation CM/Rec\(2018\)11](#) du Comité des Ministres sur la nécessité de renforcer la protection et la promotion de l'espace dévolu à la société civile en Europe, i.e. et f.

105. [Unis autour de nos valeurs – Déclaration de Reykjavík](#), 2023, p.16.

et de la non-discrimination. Les supports de sensibilisation devraient être disponibles en plusieurs langues et adaptés à la culture, y compris pour les populations ayant une maîtrise limitée des technologies numériques.

89. En organisant des consultations publiques et en mettant en place des canaux de contribution transparents, les États membres peuvent veiller à ce que le développement des systèmes d'IA soit inclusif et transparent, et qu'il prenne en considération leurs effets sociétaux plus larges en termes d'égalité, y compris d'égalité de genre, et de non-discrimination.

90. En ce qui concerne le renforcement des capacités et la formation ciblés mentionnés au paragraphe 30.3, ces initiatives pourraient notamment concerner le personnel des fournisseurs et des déployeurs de systèmes d'IA, ainsi que celui des organes de contrôle indépendants.

VI. Compétences et pouvoirs des organes de contrôle indépendants

Paragraphe 33

91. Le présent chapitre porte sur les compétences et les pouvoirs des organes indépendants chargés de superviser l'application et la mise en œuvre de la législation en matière de lutte contre la discrimination, tels que les organes de promotion de l'égalité et, le cas échéant, les autres structures nationales des droits humains (« organes de contrôle indépendants »), dans le contexte des activités menées dans le cadre du cycle de vie des systèmes d'IA. Les dispositions institutionnelles et les mandats de ces organes diffèrent selon les États membres du Conseil de l'Europe, bien qu'en général, les organes de promotion de l'égalité aient un mandat plus spécifique et se spécialisent dans la prévention et la lutte contre la discrimination, tandis que les institutions nationales des droits humains traitent également de questions plus générales relatives aux droits humains¹⁰⁶. Certains États membres intègrent également toutes les fonctions au sein d'une même institution¹⁰⁷. Ces organes devraient avoir une indépendance institutionnelle et/ou fonctionnelle face aux autorités des États membres¹⁰⁸ et peuvent souvent faire office de mécanisme plus convivial permettant aux particuliers de demander réparation pour des discriminations et de contester les discriminations et les inégalités structurelles.

Paragraphe 34

92. La Recommandation prend comme point de départ la RPG n°2 de l'ECRI sur les organismes de promotion de l'égalité chargés de lutter contre le racisme et l'intolérance au niveau national selon laquelle les organes de promotion de l'égalité devraient se voir attribuer : la fonction de promotion de l'égalité et de prévention de la discrimination (fonction de promotion et de prévention) ; la fonction d'aide aux personnes exposées à la discrimination et à l'intolérance et de recours en leur nom (fonction d'aide et de recours) ; et éventuellement la fonction de prise de décision en cas de recours (fonction de prise de décision). Les compétences en matière de soutien et de contentieux peuvent inclure, par exemple, les fonctions énoncées au paragraphe 14 de la RPG n° 2 de l'ECRI. La fonction de prise de décision peut être partagée entre les organismes de promotion de l'égalité et le pouvoir judiciaire ou être entièrement attribuée au pouvoir judiciaire. La [Recommandation CM/Rec\(2021\)1 sur le développement et le renforcement d'institutions nationales des droits de l'homme efficaces, pluralistes et indépendantes](#) indique que les États membres devraient veiller à ce que le mandat donné aux institutions nationales des droits humains pour protéger et promouvoir les droits humains soit aussi large que possible et en pleine conformité avec les Principes de Paris. Elle donne des orientations détaillées sur le contenu d'un tel mandat.

93. Étant donné la RPG n° 2 de l'ECRI et les différences entre les États membres, le paragraphe 34 de la Recommandation énumère une liste de pouvoirs dont les États membres devraient, sous réserve du contexte national, doter les organes de contrôle indépendants. La Recommandation reconnaît également, il importe de le noter, que le corollaire de l'attribution de ces droits et pouvoirs à ces institutions est la mise à disposition de ressources adéquates et d'une indépendance opérationnelle suffisante, ainsi que la capacité de coopérer avec d'autres organes et organisations de contrôle.

106. Voir la [Recommandation de politique générale n° 2 \(révisée\)](#) de l'ECRI sur les organismes de promotion de l'égalité, paragraphes 4 et suivants.

107. Voir la RPG n° 2 de l'ECRI, paragraphe 7.

108. Voir la RPG n° 2 de l'ECRI, paragraphe 22.

Paragraphe 35-37

94. Il est également recommandé que ces organes de contrôle indépendants puissent participer à l'expérimentation de technologies nouvelles et innovantes, notamment au moyen de bacs à sable réglementaires, comme cela a été indiqué ci-dessus, s'ils choisissent de le faire. Le suivi pourrait revêtir la forme de tests de situation dans le cadre desquels ces organes testeraient les résultats d'un système donné en comparant les résultats pour différents groupes. Par exemple, ils pourraient soumettre des CV ou des demandes de crédit tests de groupes majoritaires et minoritaires pour tenter de révéler des discriminations dans des contextes où les sociétés utilisent des systèmes d'IA. Ils pourraient également mener des audits pour détecter les biais potentiels s'ils se voient accorder l'accès aux systèmes concernés. Ces méthodes pourraient aider les victimes en atténuant les obstacles existants à l'établissement d'une discrimination *prima facie*. Ces organes pourraient être en mesure de prévenir la discrimination en coopérant avec les utilisateur·rices de systèmes d'IA, par exemple les sociétés qui utilisent ces systèmes pour soutenir les procédures de recrutement, afin de collecter des données pertinentes sur l'effet de leurs décisions et d'éliminer les biais. Des organes de contrôle indépendants pourraient également être habilités à émettre des alertes précoces et des recommandations non contraignantes dans les cas où les pratiques algorithmiques soulèvent des signaux d'alerte potentiels en matière d'égalité, même en l'absence de plaintes individuelles. Enfin, ils pourraient aider les victimes potentielles à obtenir réparation. La société civile peut également contribuer à cet exercice, en fonction du contexte national.

VII. Dispositions complémentaires

Paragraphe 38

95. Les systèmes d'IA sont de plus en plus utilisés comme outils d'information et de prise de décision dans la vie quotidienne. Dans certains secteurs, l'utilisation des systèmes d'IA peut avoir un effet particulièrement important sur l'égalité, y compris l'égalité de genre, et la non-discrimination. La Recommandation appelle donc les États membres à prendre des mesures spécifiques dans ces secteurs qui sont listées au paragraphe 38 de la Recommandation, tout en notant que cette liste n'est pas exhaustive et qu'il peut être approprié pour les États membres d'adopter des mesures concernant d'autres secteurs. Dans ce contexte, les États membres devraient prendre des mesures pour garantir un niveau approprié de contrôle humain.

96. Pour promouvoir la transparence, les États membres sont encouragés à mettre à disposition des informations claires et accessibles concernant l'utilisation des systèmes d'IA dans les secteurs concernés listés au paragraphe 38 de la Recommandation. Cela pourrait inclure les stratégies nationales en matière d'IA et les rapports détaillant l'utilisation des systèmes d'IA dans des domaines spécifiques ainsi que leur impact réel ou potentiel sur l'égalité, y compris l'égalité de genre..

97. Les États membres devraient adopter une approche intégrée de l'égalité et de l'égalité de genre afin de promouvoir efficacement l'égalité et de prévenir et de combattre la discrimination dans les activités menées dans le cadre du cycle de vie des systèmes d'IA, en particulier dans les secteurs mentionnés au paragraphe 38 de la Recommandation.

98. La Stratégie pour l'égalité de genre 2024-2029 explique que la garantie de l'égalité de genre doit être comprise comme un élément clé d'une lutte plus large pour sauvegarder les droits humains et garantir la justice sociale. Il est essentiel de reconnaître l'interdépendance de ces questions et de travailler ensemble pour construire une société plus égalitaire, plus inclusive et plus durable. En outre, la Stratégie souligne que l'intégration d'une approche intersectionnelle est capitale pour garantir que les formes cumulatives de discrimination et d'oppression subies par certains groupes soient combattues¹⁰⁹. L'intégration de l'égalité est l'incorporation systématique des considérations liées à l'égalité et à l'antidiscrimination à tous les stades du processus politique au niveau des États membres et du Conseil de l'Europe.

99. L'adoption d'une approche intégrée de l'égalité repose sur l'évaluation continue des effets sur l'égalité et la non-discrimination des mesures et des décisions planifiées ou mises en œuvre du point de vue des différents groupes et sur leur ajustement en conséquence. En tant que tel, il s'agit d'un processus qui devrait être suivi du début à la fin de toute activité donnée, y compris lors de son évaluation. La Recommandation du Comité des Ministres sur les normes et mécanismes d'égalité entre les femmes et les hommes (CM/Rec(2007)17) souligne l'importance d'adopter des méthodologies pour mettre en œuvre la stratégie d'approche intégrée de l'égalité de genre. L'approche intégrée de l'égalité de genre devrait être mise en œuvre dans l'élaboration, la mise en œuvre et l'examen des politiques.

¹⁰⁹. Stratégie du Conseil de l'Europe pour l'égalité de genre (2024-2029), p. 50 et 51.

100. Pour une approche intégrée de l'égalité, une méthodologie similaire est nécessaire. Elle devrait comprendre des lignes directrices et des étapes claires pour incorporer la perspective d'égalité dans les nouvelles politiques et lois et pour effectuer un examen et une évaluation des politiques et lois existantes afin de garantir l'égalité en pratique pour tous et toutes. Ces étapes peuvent comporter l'examen et l'évaluation de leur effet sur l'égalité, la planification de l'égalité, les obligations en matière d'égalité¹¹⁰, l'intégration de la perspective d'égalité dans la budgétisation, la consultation et la participation des groupes défavorisés et de la société civile, la mise en œuvre et le suivi.

Éducation

101. L'accès à l'éducation est un droit humain pour les enfants et les adultes, sur une base d'égalité en droits et d'égalité des chances, afin de garantir l'égalité de genre à tous les niveaux d'éducation¹¹¹. En raison de leur stade de développement, les enfants ont des besoins et des droits spécifiques qui les distinguent des adultes. Il est donc nécessaire de mettre en place des réglementations axées sur les enfants dans le cadre de l'acquisition et de l'utilisation des technologies éducatives¹¹². En outre, les dernières données soulignent un écart persistant entre les femmes et les hommes dans le domaine de la recherche et de l'innovation, avec seulement 21% de femmes titulaires d'un doctorat dans les technologies de l'information et de la communication (TIC).¹¹³ Les femmes restent sous-représentées dans les études de sujets STIM, ce qui les rend moins susceptibles de profiter des possibilités d'emploi dans ce domaine, y compris l'IA¹¹⁴. Les États membres devraient donc donner la priorité à la participation des filles et des jeunes dans ces domaines d'enseignement¹¹⁵. Les États membres devraient également faciliter l'entrée sur le marché du travail de l'IA en donnant accès à des programmes de formation, en particulier pour les jeunes femmes et les jeunes issus de groupes sous-représentés, en accordant des financements et des bourses et en soutenant des programmes de mentorat¹¹⁶. De plus, étant donné les stéréotypes de genre conduisent souvent à des choix de carrière façonnés par des assignations de rôles stéréotypés, les États membres devraient soutenir des programmes de conseil en orientation professionnelle dénué de tout stéréotype.

102. En outre, les États membres devraient veiller à ce que les établissements d'enseignement utilisent les systèmes d'IA d'une manière qui soit compatible avec une éducation inclusive et de qualité¹¹⁷. Par exemple, les logiciels de reconnaissance faciale, utilisés dans le cadre de logiciels de surveillance conçus pour contrôler le comportement des étudiant-es/élèves pendant les examens en ligne, peuvent présenter des biais et conduire à une discrimination intersectionnelle, fondée notamment sur la race et le sexe¹¹⁸. Il convient donc d'accorder une attention particulière à l'utilisation des systèmes d'IA dans les procédures de sélection et d'examen, afin d'éviter les résultats discriminatoires.

103. Les États membres devraient assurer la disponibilité la culture numérique dans les programmes éducatifs comprenant des contenus visant à améliorer la sensibilisation au droit à l'égalité, y compris l'égalité de genre, dans le contexte des systèmes d'IA¹¹⁹. Les États membres pourraient envisager de faciliter la conception conjointe de programmes d'éducation numérique avec leur public cible afin de garantir plus d'inclusion. Les enfants et les jeunes, en particulier, devraient être éduqués, d'une manière adaptée à leur âge, aux risques de surveillance en ligne, de cyberintimidation et d'exploitation des données, ainsi que, plus généralement, aux effets de l'utilisation des systèmes d'IA sur l'égalité et la non-discrimination. Les enseignant-es de tous

110. Les obligations en matière d'égalité sont des obligations juridiques positives visant à promouvoir l'égalité et à prévenir la discrimination, allant au-delà de l'obligation générale de s'abstenir de toute discrimination.

111. Recommandation [CM/Rec\(2007\)17](#) du Comité des Ministres sur les normes et mécanismes d'égalité entre les femmes et les hommes, paragraphes 24 et 25.

112. Étude préparatoire pour l'élaboration d'un instrument juridique sur la réglementation de l'utilisation des systèmes d'intelligence artificielle dans l'éducation, projet révisé (mars 2024), Unité de transformation numérique du Service de l'éducation, Conseil de l'Europe.

113. Commission européenne: Direction générale de la recherche et de l'innovation, *SheFigures 2024 - Genre dans la recherche et l'innovation - Statistiques et indicateurs*, Office des publications de l'Union européenne, 2025.

114. Étude CDADI/GEC (2023), p. 89; [Résolution 2343 \(2020\)](#), « Prévenir les discriminations résultant de l'utilisation de l'intelligence artificielle », paragraphe 6; *Indice d'égalité de genre 2023*, Institut européen pour l'égalité entre les hommes et les femmes (EIGE), p. 12 et 37.

115. Recommandation [CM/Rec\(2019\)1](#) sur la prévention et la lutte contre le sexisme, II.B.7.; [Résolution 2343 \(2020\)](#) « Prévenir les discriminations résultant de l'utilisation de l'intelligence artificielle », paragraphe 12.1.

116. Étude CDADI/GEC (2023), p. 89.

117. Recommandation [CM/Rec\(2019\)1](#) sur la prévention et la lutte contre le sexisme, en particulier II.G. « Institutions éducatives ».

118. Étude CDADI/GEC (2023), p. 24.

119. Recommandation [CM/Rec\(2019\)1](#) sur la prévention et la lutte contre le sexisme, en particulier II.B.2. Voir également l'étude CDADI/GEC (2023), p. 90; la [Résolution 2343 \(2020\)](#), « Prévenir les discriminations résultant de l'utilisation de l'intelligence artificielle », paragraphe 12.3; et [Recommandation de suivi](#) de la Commissaire aux droits de l'homme intitulée « Les droits humains dès la conception de l'IA – une protection durable des droits humains à l'ère de l'intelligence artificielle », 2023, p. 21 et 22.

niveaux devraient bénéficier d'une formation et de ressources appropriées à cet effet. De plus, la culture numérique devrait également être promue à travers l'éducation non formelle et informelle, afin de compléter l'éducation formelle en offrant des opportunités d'apprentissage accessibles et communautaires pour garantir que la compétence numérique tout au long de la vie soit encouragée au-delà du cadre scolaire. Les États membres devraient mener des actions particulières pour sensibiliser les groupes marginalisés, ainsi que les personnes exposées à la discrimination et aux inconvénients de la culture numérique¹²⁰. Plus généralement, ils devraient promouvoir la connaissance des systèmes d'IA, y compris leur effet sur l'égalité, notamment l'égalité de genre, et la non-discrimination¹²¹.

Justice, pouvoir judiciaire et maintien de l'ordre

104. Les systèmes de police et de justice des États membres du Conseil de l'Europe utilisent des systèmes d'IA, y compris des applications telles que la prévision policière (*predictive policing*) et l'évaluation des risques en matière de détention provisoire, de condamnation et de libération conditionnelle¹²². Cependant, l'utilisation de systèmes d'IA dans ces domaines peut conduire à la discrimination de diverses manières¹²³. Par exemple, la reconnaissance faciale est de plus en plus utilisée pour la détection et la prévention de la délinquance, mais sa précision s'est révélée biaisée dans plusieurs cas, entraînant l'identification erronée de suspects et, dans certains cas, l'incarcération injustifiée de personnes innocentes¹²⁴. En ce qui concerne le maintien de l'ordre et la sécurité, les États membres devraient donc faire preuve de prudence en matière d'identification, d'évaluation, de prévention et d'atténuation des risques de discrimination découlant de l'utilisation de technologies de reconnaissance faciale ou de systèmes d'identification biométrique à distance dans les secteurs du maintien de l'ordre et de la sécurité, et envisager, le cas échéant, d'interdire leur utilisation. Les modèles prédictifs utilisés dans le domaine policier, par exemple pour déterminer les « zones à risque »¹²⁵, les décisions de condamnation et les évaluations prédictives des risques utilisées par la police, les tribunaux ou les services de probation¹²⁶, peuvent également entraîner un risque de discrimination. Lorsqu'il est fait usage de plusieurs systèmes, cela peut provoquer un « effet de chaîne », qui se produit lorsque plusieurs systèmes s'alimentent mutuellement¹²⁷.

105. La Commission européenne pour l'efficacité de la justice (CEPEJ) a adopté le premier texte européen énonçant des principes éthiques relatifs à l'utilisation de l'IA dans les systèmes judiciaires¹²⁸. Ce texte fait du principe de non-discrimination un principe fondamental à respecter dans le domaine de l'IA et de la justice. Il note en particulier l'importance d'empêcher spécifiquement le développement ou l'intensification de toute discrimination entre des individus ou des groupes d'individus. Le texte souligne les risques que présente la prévision en matière pénale au moyen d'outils d'IA qui sont susceptibles de reproduire des inégalités injustifiées et déjà existantes dans le système de justice pénale¹²⁹. Dans le respect de l'indépendance du pouvoir judiciaire et du droit à un procès équitable (article 6 de la Convention), les États membres devraient donc veiller à ce que l'utilisation de systèmes d'IA dans le système judiciaire empêche le développement ou la perpétuation de toute discrimination entre des individus ou des groupes d'individus, notamment par l'analyse des données, la formation et l'échange d'informations sur les pratiques relatives à l'utilisation de

120. Commissaire des droits de l'homme du Conseil de l'Europe, *Décoder l'intelligence artificielle: 10 mesures pour protéger les droits de l'homme*, 2019, p. 14.

121. *Recommandation de suivi* de la Commissaire aux droits de l'homme intitulée « Les droits humains dès la conception de l'IA – une protection durable des droits humains à l'ère de l'intelligence artificielle », 2023, p. 21 et 22.

122. Étude CDADI/GEC (2023), p. 22 et 23; *Résolution 2342 (2020)* « Justice par algorithme – Le rôle de l'intelligence artificielle dans les systèmes de police et de justice pénale », Assemblée parlementaire du Conseil de l'Europe, paragraphe 6.

Pour un exemple de bonne pratique, voir l'adoption par le Gouvernement du Royaume-Uni d'une législation obligeant le secteur public à divulguer des informations sur le rôle des algorithmes dans l'appui aux décisions affectant les particuliers, comme cela est souligné dans la *Recommandation* « Les droits humains dès la conception de l'IA – une protection durable des droits humains à l'ère de l'intelligence artificielle », (2023), p. 17.

123. Recommandation *CM/Rec(2024)5* du Comité des Ministres aux États membres sur les aspects éthiques et organisationnels de l'utilisation de l'intelligence artificielle et des technologies numériques associées par les services pénitentiaires et de probation, en particulier les paragraphes 3, 19 et 25.

124. Étude CDADI/GEC (2023), p. 22 et 23. On trouvera d'autres exemples dans la *Résolution 2342 (2020)* « Justice par algorithme – Le rôle de l'intelligence artificielle dans les systèmes de police et de justice pénale », paragraphe 7.

125. *Fundamental Rights Agency, Bias in Algorithms – Artificial Intelligence and Discrimination* (2022) (en anglais seulement).

126. *Algorithmic risk assessment policing models: Lessons from the Durham Constabulary HART model*, OSWALD, Marion, GRACE, Jamie, URWIN, Sheena and BARNES, Geoffrey (2018), *Information and Communications Technology Law*, 27 (2), 223-250 (en anglais seulement).

127. <https://research.northumbria.ac.uk/probablefutures/wp-content/uploads/2025/05/PROBabLE-Futures-Jan-13-Opening-Slides.pdf> (en anglais seulement).

128. *Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement*, Commission européenne pour l'efficacité de la justice (CEPEJ), 2018.

129. *Ibid.*, p. 9 et 79.

systèmes d'IA dans les procédures judiciaires. Plus généralement, pour améliorer l'accès aux voies de recours le cas échéant, les États membres devraient renforcer la transparence concernant l'utilisation de l'IA par les pouvoirs publics, y compris les autorités judiciaires et les autorités chargées du maintien de l'ordre¹³⁰. Des garanties devraient être mises en place pour veiller à ce que l'utilisation des systèmes d'IA dans les procédures judiciaires n'entraîne pas une nouvelle traumatisation des victimes et que ces systèmes soient déployés avec une approche tenant compte des traumatismes et centrée sur les victimes.

106. Le maintien de l'ordre peut être considéré comme un domaine où la discrimination est susceptible de se produire et les États membres devraient donc vérifier si de nouvelles réglementations sont nécessaires ou si des mesures spécifiques devraient être mises en œuvre pour prévenir la discrimination¹³¹.

Processus électoraux

107. Les systèmes d'IA offrent la possibilité d'améliorer les processus démocratiques, notamment en les rendant plus inclusifs. Ils peuvent, par exemple, aider les citoyen·nes à mieux comprendre la politique et à participer plus facilement au débat démocratique, tout en améliorant l'accessibilité des processus électoraux. De la même manière que celle décrite au paragraphe 46 du présent rapport explicatif, les systèmes d'IA peuvent également être utilisés pour identifier ou supprimer les informations erronées qui peuvent viser des personnes ou des groupes exposés à un risque de discrimination. Toutefois, la production et la diffusion de désinformation et d'informations inexactes circulant dans l'intention de tromper présentent des risques importants. L'IA est de plus en plus présente dans les applications utilisées par les électeurs et électrices pour trouver des informations, qu'ils et elles en soient conscient·es ou non, que ce soit par le biais d'outils de recherche ou de productivité améliorés par l'IA. Les informations générées par l'IA peuvent sembler plausibles, mais elles sont souvent inexactes. Ce manque d'exactitude peut être particulièrement préjudiciable aux électeurs et électrices et pourrait être facilement instrumentalisé au détriment des personnes ou des groupes exposés à un risque de discrimination. L'utilisation de systèmes d'IA est également liée à la génération de contrefaçons digitales et à l'amplification des biais sociétaux. Ceux-ci peuvent perturber les campagnes politiques, avoir un impact sur la confiance du grand public et décourager la participation électorale et la représentation, en particulier des femmes¹³².

Travail et emploi

108. En ce qui concerne le travail et l'emploi, les États membres devraient veiller à ce que les systèmes d'IA n'introduisent pas, ne reproduisent pas ou n'amplifient pas les schémas existants ou émergents d'inégalité et de discrimination structurelle et qu'ils soient utilisés de manière à promouvoir l'égalité, y compris l'égalité de genre, la diversité et l'inclusion sur le lieu de travail. En particulier, cela pourrait consister en un audit régulier des résultats de l'utilisation de l'IA dans les procédures de recrutement, de promotion et autres; la participation des salarié·es et de leurs organisations représentatives dans l'élaboration des politiques ou dans les choix concernant l'utilisation de l'IA dans les processus décisionnels sur le lieu de travail; le suivi des effets de la mise en place de systèmes d'IA en milieu professionnel sur l'égalité de genre et sur la diversité au sein du personnel; et la formation et la sensibilisation du personnel aux biais des données, aux stéréotypes et aux risques de discrimination liés à l'utilisation de systèmes d'IA. En outre, les personnes exposées à la discrimination sont particulièrement vulnérables à de mauvaises conditions de travail ou à des conditions de travail précaires qui peuvent être aggravées par l'utilisation croissante des systèmes d'IA sur le lieu de travail. De plus, selon des études récentes, de nombreux systèmes d'IA reposent sur une préparation des données exigeant une main-d'œuvre importante, souvent externalisée dans des conditions de travail précaires dans les pays du Sud, mais aussi en Europe. Les travailleur·ses chargé·es de l'étiquetage des données et de la modération de contenu sont exposé·es à des conditions d'exploitation sévères, notamment des salaires faibles, une surveillance excessive et une exposition à des contenus éprouvants. De nombreux ensembles de données utilisés pour l'entraînement de l'IA reposent sur un travail sous-payé ou informel, ce qui aggrave les inégalités¹³³. Les États membres devraient, le cas échéant, prendre des mesures pour lutter contre les pratiques de travail qui relèvent de l'exploitation. Les femmes restent largement sous-représentées dans

130. Étude CDADI/GEC (2023), p. 85.

131. Voir aussi [règlement de l'UE sur l'IA](#), attendu 33).

132. [Liberté d'expression, intelligence artificielle et élections, UNESCO, Programme des Nations Unies pour le développement \(2025\), CI-2025/FEJ/ME-1](#).

133. [Valuing Data Enrichment Workers: the Case for a Human-Centric Approach to AI Development](#), United Nations (2023); [Artificial Intelligence for Workers, Not Just For Profit: Ensuring Quality in Jobs in the Digital Age](#), European Trade Union (2025); [the Human Cost of DeepSeek](#): Antonio A. Casilli, Thomas Le Bonniec, Julian Posada. [The Human Cost of DeepSeek.1\(1\)](#), DiPLab. 2025, pp.10 ffhal-04952735 (en anglais seulement).

l'industrie numérique, l'économie numérique et dans la recherche et la science et sont plus exposées aux pertes d'emploi résultant de l'automatisation induite par l'IA générative¹³⁴. Cependant, les femmes devraient être habilitées à façonner activement tant les innovations en matière d'IA qu'un monde numérique égalitaire. Les États membres devraient envisager des mesures pour aider à établir les femmes comme des actrices et décideuses actives dans l'industrie et l'économie numériques. À cet égard, il est essentiel de promouvoir l'accès des femmes à des postes de direction et de soutenir les femmes entrepreneuses dans l'industrie numérique.

109. Plus généralement, lorsqu'ils réglementent l'utilisation de logiciels de télésurveillance ou de contrôle des salarié-es et des étudiant-es, les États membres devraient tenir compte de l'augmentation du nombre de personnes qui travaillent ou qui étudient à domicile et des problèmes particuliers que cela pose pour le droit au respect de la vie privée ainsi que des risques de discrimination.

110. Certains systèmes d'IA sont également utilisés pour promouvoir l'égalité, y compris l'égalité de genre, dans les processus de recrutement, car ils mettent l'accent sur la transparence, la protection de la vie privée et l'obligation de rendre des comptes. Par exemple, éliminer les biais potentiels dans les candidatures, tels que l'âge, le sexe, la nationalité et la localisation. Plus généralement, les systèmes d'IA peuvent être utilisés pour promouvoir l'accessibilité grâce à des solutions axées sur le handicap, telles que des algorithmes de sous-titrage en direct pour les personnes sourdes ou malentendantes, des solutions d'adaptation de l'environnement pour rendre le contenu et les lieux de travail accessibles, ou des solutions débloquent de nouvelles opportunités de travail pour les personnes en situation de handicap, par exemple le recrutement ciblé de personnes neurodivergentes par des entreprises d'étiquetage de données¹³⁵.

Services sociaux, logement et protection sociale

111. En ce qui concerne les services sociaux et la protection sociale, l'utilisation de systèmes d'IA peut conduire à exclure des personnes ou à leur refuser l'accès aux services publics et les avantages qui en découlent¹³⁶. Ainsi que le souligne le règlement de l'UE sur l'IA, « lorsqu'ils sont utilisés pour déterminer si ces prestations et services devraient être accordés, refusés, réduits, révoqués ou récupérés par les autorités, y compris pour déterminer si les bénéficiaires y ont légitimement droit, les systèmes d'IA peuvent avoir une grande incidence sur les moyens de subsistance des personnes et porter atteinte à leurs droits fondamentaux »¹³⁷. En outre, l'absence d'obligation de rendre des comptes et le manque de transparence entourant l'utilisation des systèmes d'IA peut conduire à priver les sujets des décisions de l'IA d'une explication ou de la possibilité de contester ces décisions¹³⁸. Les États membres devraient veiller à ce que, lorsque des systèmes d'IA sont utilisés pour déterminer si des prestations sociales, des services sociaux, une aide au logement ou un soutien social doivent être accordés, refusés, réduits, révoqués ou réclamés, y compris pour déterminer si les bénéficiaires ont légitimement droit à ces prestations ou services, ces utilisations n'entraînent pas de discrimination ou n'amplifient pas les inégalités existantes¹³⁹. L'utilisation des systèmes d'IA devrait tenir compte des vulnérabilités éventuelles des utilisateurs et utilisatrices de ces services, notamment, mais sans s'y limiter, leur âge, et des garanties nécessaires devraient être assurées.

Soins de santé

112. Les systèmes d'IA ont l'avantage de permettre d'améliorer les résultats diagnostiques et thérapeutiques pour les patient-es¹⁴⁰, et de favoriser l'octroi de services médicaux, y compris les services de santé mentale. Dans le même temps, un certain nombre de problèmes entraînant une violation de l'égalité, y compris de l'égalité de genre, et une discrimination ont été recensés dans ce domaine¹⁴¹. Il s'agit, par exemple, du manque de données sur le genre dans le secteur des soins de santé, qui conduit à des systèmes prédictifs moins fiables lorsqu'il s'agit de poser des diagnostics chez des patient-es de sexe féminin ou appartenant à la diversité de

134. Generative AI and Jobs: A Refined Global Index of Occupational Exposure, OIT, mai 2025.

135. *Using AI to support people with disability in the labour market: opportunities and challenges*, OECD Artificial Intelligence Papers, n° 7, 2023.

136. Étude CDADI/GEC (2023), p. 23.

137. *Règlement de l'UE sur l'IA*, attendu 58).

138. Étude CDADI/GEC (2023), p. 24, et *Recommandation de suivi* de la Commissaire aux droits de l'homme intitulée « Les droits humains dès la conception de l'IA – une protection durable des droits humains à l'ère de l'intelligence artificielle », 2023, p. 6.

139. Étude CDADI/GEC (2023), pp. 22-23.

140. *Réponse du Comité des Ministres à la Recommandation 2185 (2020)* sur « L'intelligence artificielle dans les soins de santé: défis médicaux, juridiques et éthiques à venir » de l'Assemblée parlementaire du Conseil de l'Europe, paragraphe 3.

141. Étude CDADI/GEC (2023), p. 25 et 26 et *Recommandation 2185 (2020)* sur « L'intelligence artificielle dans les soins de santé: défis médicaux, juridiques et éthiques à relever ».

genre¹⁴². L'élaboration des politiques publiques en matière de soins de santé et de progrès technologiques devrait être fondée sur le plein respect des droits humains, y compris les droits sociaux¹⁴³. Par conséquent, les États membres devraient mener les actions nécessaires pour atténuer le risque de discrimination en matière d'accès aux soins de santé lié à l'utilisation de systèmes d'IA¹⁴⁴. Les États membres devraient également prendre des mesures pour améliorer la fiabilité des systèmes d'IA analytiques et prédictifs dans le secteur des soins de santé en comblant les lacunes concernant les données relatives aux caractéristiques personnelles énumérées au paragraphe 2 de la Recommandation et en veillant à ce que des modèles plus diversifiés soient utilisés pour évaluer les risques et l'efficacité des médicaments, des thérapies et des autres outils de soins de santé.

Migrations

113. Dans le contexte du contrôle et de la gestion des frontières, et entre autres applications, les systèmes d'IA sont utilisés à des fins de prévision¹⁴⁵. Ainsi que le souligne le règlement de l'UE sur l'IA, « les systèmes d'IA utilisés dans les domaines de la migration, de l'asile et de la gestion des contrôles aux frontières touchent des personnes qui se trouvent souvent dans une situation particulièrement vulnérable et qui sont tributaires du résultat des actions des autorités publiques compétentes. L'exactitude, la nature non discriminatoire et la transparence des systèmes d'IA utilisés dans ces contextes sont donc particulièrement importantes pour garantir le respect des droits fondamentaux des personnes concernées, en particulier leurs droits à la libre circulation, à la non-discrimination, à la protection de la vie privée et des données à caractère personnel, à une protection internationale et à une bonne administration »¹⁴⁶. L'Assemblée parlementaire a également souligné que l'IA devrait soutenir – et non remplacer – la prise de décision humaine dans les procédures de migration et d'asile, même si, dans certains cas, l'IA peut offrir une plus grande sécurité et efficacité que la seule prise de décision humaine, en réduisant la marge d'erreur humaine. Elle a également indiqué que les systèmes d'IA utilisés dans les procédures de migration et d'asile devraient faire l'objet d'évaluations d'impact sur les droits humains, la démocratie et l'État de droit avant leur déploiement¹⁴⁷. Les États membres devraient veiller à ce que l'utilisation des systèmes d'IA dans la gestion des migrations, de l'asile et des contrôles aux frontières ne soit pas discriminatoire ou n'entraîne pas de discrimination, en identifiant, en évaluant, en prévenant et en atténuant les risques de discrimination. Ils devraient accorder une attention particulière aux droits des femmes et des filles et des autres personnes exposées à la discrimination pour les motifs énumérés au paragraphe 2 de la Recommandation¹⁴⁸.

Banque, finance, assurance et autres secteurs de services

114. Un certain nombre d'exemples montrent que les biais algorithmiques influencent la prise de décision dans le secteur de la banque, de la finance, de l'assurance et d'autres services où des systèmes d'IA sont utilisés, entraînant une discrimination à l'encontre des utilisateur·rices des services¹⁴⁹. Les États membres devraient donc prendre des mesures pour veiller à ce qu'un cadre soit en place afin de prévenir et de repérer les biais et de contrôler les résultats de la prise de décision pour les client·es en matière d'égalité et de non-discrimination. Dans le même temps, les systèmes d'IA peuvent également être utilisés pour détecter les décisions biaisées¹⁵⁰.

142. Étude CDADI/GEC (2023), p. 25, citant Criado Perez C, *Invisible women: Exposing data bias in a world designed for men*, Random House, 2019.

143. Réponse du Comité des Ministres à la Recommandation 2185 (2020) sur « L'intelligence artificielle dans les soins de santé: défis médicaux, juridiques et éthiques à venir », paragraphe 2.

144. Rapport L'application de l'IA dans les soins de santé et son impact sur la relation « patient-e-médecin », Comité directeur pour les droits humains dans les domaines de la biomédecine et de la santé (CDBIO), notant le risque de partialité et de discrimination lié à l'utilisation des systèmes d'IA dans les soins de santé, en particulier p. 27.

145. Recommandation CM/Rec(2020)1 sur les impacts des systèmes algorithmiques sur les droits de l'homme, paragraphe 8.

146. Règlement de l'UE sur l'IA, attendu 60).

147. Résolution 2628 (2025) sur l'intelligence artificielle et la migration, Assemblée parlementaire du Conseil de l'Europe, en particulier les paragraphes 4, 7, 15, 20, 28 et 29.

148. Recommandation CM/Rec(2022)17 du Comité des Ministres sur la protection des droits des femmes et des filles migrantes, réfugiées et demandeuses d'asile, en particulier les paragraphes 22 et 23.

149. *Ibid.*, p. 21 et 22; Recommandation de suivi de la Commissaire aux droits de l'homme intitulée « Les droits humains dès la conception de l'IA – une protection durable des droits humains à l'ère de l'intelligence artificielle », 2023, p. 6; et règlement de l'UE sur l'IA, attendu 58).

150. Étude CDADI/GEC (2023), p. 63.

Autres domaines dans lesquels les impacts négatifs potentiels de l'IA devraient être contrôlés et, le cas échéant, traités

115. Au-delà des secteurs identifiés comme présentant un risque d'avoir un impact particulièrement aigu sur l'égalité, y compris l'égalité de genre, et la non-discrimination, la Recommandation attire l'attention sur d'autres domaines dans lesquels les États membres sont encouragés à contrôler et, le cas échéant, à traiter les impacts négatifs potentiels de l'utilisation des systèmes d'IA, tels que les médias et la publicité, ainsi que le sport et la culture.

116. L'IA est de plus en plus utilisée dans le secteur des médias et de la publicité, ce qui peut engendrer des risques spécifiques pour l'égalité, y compris l'égalité de genre, et la non-discrimination, tels que des représentations biaisées ou l'amplification de stéréotypes nuisibles. À cet égard, il est noté que les systèmes d'IA pourraient également être utilisés pour analyser de grandes quantités de données et évaluer les représentations des femmes et des personnes exposées à la discrimination dans la publicité en ligne et la publicité physique¹⁵¹. Afin de traiter les impacts négatifs potentiels, les États membres devraient prendre des mesures pour veiller à ce que lorsqu'une législation, des politiques et des cadres réglementaires visant à prévenir et à combattre la discrimination dans les médias et la publicité existent, ces derniers couvrent l'utilisation des systèmes d'IA, tout en respectant l'indépendance des médias¹⁵². Les mesures visées au paragraphe ci-dessus devraient être complétées par une formation spécifique destinée aux journalistes et aux autres professionnel·les du secteur sur la manière de repérer et de combattre le sexisme, le racisme, la LGBTI-phobie et d'autres contenus haineux, ainsi que d'autres formes de discrimination, et sur la façon de promouvoir des représentations inclusives, positives et non stéréotypées des individus et des groupes dans les médias et dans la publicité¹⁵³.

117. Les systèmes d'IA sont de plus en plus utilisés dans le domaine du sport en raison de leur capacité à analyser rapidement de grandes quantités de données, à détecter des modèles et des tendances athlétiques, à analyser les performances et à produire et analyser des statistiques. Dans le domaine de la prévention des blessures, par exemple, les systèmes d'analyse prédictive peuvent générer des alertes sur d'éventuelles lésions musculaires et améliorer considérablement les performances des athlètes. Dans les sports de contact, les systèmes d'IA ont été utilisés pour développer des systèmes de surveillance qui détectent les impacts et évaluent automatiquement leur gravité, réduisant ainsi le risque de commotion cérébrale pour les athlètes. Les systèmes d'IA peuvent également être utilisés pour analyser les commentaires sportifs et mettre en évidence les préjugés raciaux. Aux États-Unis, un projet de recherche utilisant des algorithmes de traitement du langage a révélé des schémas de langage biaisé qui attribuaient de manière disproportionnée des prouesses physiques aux athlètes noir·es, tandis que les athlètes blanc·es étaient plus souvent décrits en utilisant des termes associés à l'intellect¹⁵⁴. De plus, si certains groupes sont sous-représentés dans les ensembles de données utilisés pour former les systèmes d'IA, ces systèmes peuvent refléter des préjugés préexistants, y compris des préjugés sexistes, dans la culture sportive et ne pas produire une analyse inclusive et objective. Cela pourrait affecter leur santé, mais aussi avoir un impact sur les décisions relatives au recrutement, à l'évaluation des performances des athlètes et à l'allocation des ressources, entraînant des désavantages pour certains athlètes ou groupes d'athlètes, comme le fait de négliger les talents émergents issus de groupes sous-représentés, et perpétuant ainsi les inégalités dans le sport.

118. En ce qui concerne la culture, les systèmes d'IA peuvent involontairement perpétuer des stéréotypes, y compris des stéréotypes de genre, par exemple si les données utilisées pour former les algorithmes sont biaisées en faveur des récits culturels dominants. Les systèmes d'IA peuvent également reproduire les inégalités systémiques, ce qui peut être aggravé par des préoccupations concernant les droits d'auteur et la propriété intellectuelle lorsque du contenu est extrait de cultures minoritaires, sans la connaissance, l'implication et le consentement des groupes concernés¹⁵⁵. Dans le même temps, en présentant dans leurs résultats des individus issus de milieux et de régions divers, les systèmes d'IA peuvent améliorer la

151. Étude CDADI/GEC (2023), p. 77.

152. Recommandation [CM/Rec\(2019\)1](#) sur la prévention et la lutte contre le sexisme, en particulier II.B.1. et II.C. « Médias, publicité et autres produits et services de communication » ; [Recommandation CM/Rec\(2013\)1](#) du Comité des Ministres sur l'égalité entre les femmes et les hommes et les médias, en particulier la ligne directrice A.1. ; [Recommandation CM/Rec\(2007\)17](#) sur les normes et mécanismes d'égalité entre les femmes et les hommes, paragraphe 47 ; et [Recommandation CM/Rec\(2022\)16](#) du Comité des Ministres sur la lutte contre le discours de haine.

153. Recommandation [CM/Rec\(2019\)1](#) sur la prévention et la lutte contre le sexisme, en particulier II.C.6. et [Recommandation CM/Rec\(2022\)16](#) sur la lutte contre le discours de haine.

154. Voir « Can artificial intelligence help us understand racial bias in sports? » (en anglais seulement), C. Brandon Ogbunu, 8 janvier 2020, disponible en ligne à <https://andscape.com/features/artificial-intelligence-racial-bias-in-sports/>.

155. Voir, par exemple, les Lignes directrices compte tenu des derniers développements technologiques, tels que l'IA, complétant les normes du Conseil de l'Europe dans les domaines de la culture, de la créativité et du patrimoine culturel, mars 2025.

représentation culturelle de personnes et de groupes historiquement sous-représentés. Le potentiel de promotion des langues régionales et minoritaires a déjà été mentionné. Dans les domaines où les femmes sont sous-représentées dans la couverture médiatique ou dans lesquels leurs réalisations sont sous-évaluées, comme le sport¹⁵⁶, les systèmes d'IA peuvent contribuer à l'égalité de visibilité des femmes en remettant en question les stéréotypes et les préjugés traditionnels liés au genre.

Paragraphe 39

119. La Recommandation souligne que les États membres demeurent liés par leurs obligations existantes en matière de droits humains, notamment celles relatives à la non-discrimination et à l'égalité, y compris l'égalité de genre, et tout en les y encourageant, elle précise qu'ils ne sont pas tenus d'appliquer la Recommandation aux activités menées dans le cadre du cycle de vie des systèmes d'IA qui concernent directement la protection de leurs intérêts en matière de sécurité nationale. De telles limitations ne devraient pas aller au-delà de celles énoncées à l'article 3, paragraphe 2, de la Convention-cadre et dans son rapport explicatif¹⁵⁷.

156. Rapport analytique « All in Plus : promouvoir l'égalité de genre dans le sport » (« All In Plus: Promoting Greater Gender Equality in Sport, Analytical report of the data collected : insights into trends and progress » (en anglais seulement), février 2025, Projet conjoint de l'Union européenne et du Conseil de l'Europe, met en évidence à la fois les progrès et les écarts persistants entre les femmes et les hommes dans le sport.

157. Voir le rapport explicatif de la Convention-cadre, paragraphe 32.

www.coe.int

Le Conseil de l'Europe est la principale organisation de défense des droits humains 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 humains, la démocratie et l'État de droit. La Cour européenne des droits de l'homme contrôle la mise en œuvre de la Convention dans les États membres.

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE