

GOUVERNANCE DE L'IA GÉNÉRATIVE : COMBLER LES LACUNES POUR PROTÉGER LES DROITS HUMAINS

ÉLABORÉ PAR LE COMMISSAIRE AUX DROITS DE L'HOMME
DU CONSEIL DE L'EUROPE

TABLE DES MATIÈRES

Contexte	3
Avertissement	3
DISCUSSION ET CONCLUSIONS	3
Thème 1 : Risques pour les droits humains liés au déploiement de l'intelligence artificielle généralive	3
Thème 2 : Lacunes dans les cadres de gouvernance européens et propositions de solutions	5
Annexe I	10
Annexe II	12

Contexte

Les 6 et 7 novembre 2025, le Commissaire aux droits de l'homme du Conseil de l'Europe a réuni un groupe d'expert·es en intelligence artificielle (IA) issu·es d'organisations de la société civile afin d'examiner les risques de l'intelligence artificielle générative (GenAI) et les lacunes relatives à sa gouvernance. Chacun·e a apporté une expertise spécifique dans des domaines tels que les droits de l'enfant, le racisme et la liberté d'expression, afin d'enrichir la discussion de points de vue différents. Le Commissaire a présidé la réunion.

Cette consultation visait à identifier les risques que présente le déploiement de l'IA générative pour les droits humains et les principales lacunes des cadres de gouvernance qui régissent actuellement le développement de la technologie GenAI en Europe. Les participant·es ont également examiné les moyens de renforcer la gouvernance de l'IA dans le respect du droit et des normes en matière de droits humains. La consultation a porté sur les deux grands thèmes suivants :

Thème 1 : Risques pour les droits humains liés au déploiement de l'intelligence artificielle générative

Thème 2 : Lacunes dans les cadres de gouvernance européens et propositions de solutions

L'ordre du jour figure à l'annexe I et la liste des participant·es à l'annexe II du rapport.

Avertissement

Ce rapport présente un aperçu non exhaustif des principaux points et conclusions de la consultation sous la forme d'un résumé du président. Les points de vue et opinions qui y sont exprimés reflètent la discussion entre les participant·es et ne correspondent pas nécessairement à une position ou une politique du Commissaire aux droits de l'homme du Conseil de l'Europe.

DISCUSSION ET CONCLUSIONS

Thème 1 : Risques pour les droits humains liés au déploiement de l'intelligence artificielle générative

1. Les défis et les risques que présente l'IA — y compris l'intelligence artificielle générative (GenAI) — pour les droits humains se répartissent en trois catégories : les risques liés à un secteur spécifique, par exemple ceux relatifs à la confidentialité des données, aux biais et à la discrimination ; les risques liés à la qualité du produit, comme la fiabilité des résultats produits par les grands modèles de langage (LLM), et l'instrumentalisation des systèmes d'IA — par exemple, l'utilisation délibérée de l'IA pour propager la désinformation à des fins politiques à travers des contenus numériques artificiellement créés ou manipulés (médias synthétiques), tels que les « deepfakes » et autres outils génératifs.
2. L'utilisation de l'IA dans le secteur militaire suscite de plus en plus d'inquiétudes. Le recours croissant aux systèmes d'intelligence artificielle dans les conflits en cours est présenté comme une avancée positive qui réduit les risques au combat. Cependant, l'utilisation de systèmes d'IA automatisés pour identifier des cibles dans les conflits transfère le risque d'erreur à la technologie. Cela pose deux problèmes majeurs. Premièrement, un problème de fiabilité : l'absence de supervision ou d'intervention humaine (suffisante) peut entraîner des erreurs fatales. Cela mène au second problème, à savoir un transfert de responsabilité : lorsqu'une machine agit de manière autonome, la

responsabilité de ses erreurs n'est pas claire et il devient plus difficile d'identifier les responsables. Pour des raisons de sécurité et de confidentialité, les secteurs de l'armée et de la sécurité échappent souvent aux réglementations en vigueur et au contrôle public, ce qui se traduit par une surveillance réduite. L'automatisation des systèmes d'IA utilisés au combat soulève donc d'importantes questions d'ordre éthique et pratique. Les risques directs pour les droits humains, y compris le droit à la vie, exigent une surveillance et une responsabilité accrues dans ce secteur.

3. La GenAI, en particulier, joue un rôle majeur dans la propagation de la désinformation dans les conflits internationaux. La société civile s'est dite très préoccupée par la dégradation de la qualité de l'information causée par la propagation de la désinformation et les contenus de mauvaise qualité générés par l'IA (comme les « AI slops »), qui servent d'outils dans les guerres hybrides. La guerre d'agression menée par la Russie contre l'Ukraine en est une illustration parfaite : une multitude de sites web se faisant passer pour des médias ont vu le jour, dans le but de dominer les flux d'information. Ce phénomène engendre une surabondance d'informations qui rend plus difficile la distinction entre la réalité et la fiction.
4. Cette pollution de l'écosystème médiatique se manifeste également dans la vie quotidienne. La surabondance d'informations générées par l'IA dégrade l'environnement médiatique, altère les habitudes de consommation et les façons d'interagir avec les contenus, et érode souvent la confiance dans l'information. Par exemple, l'ingérence électorale au moyen de campagnes de désinformation est devenue courante ces dernières années. L'IA — y compris les médias synthétiques, tels que les « deepfakes » — est utilisée pour amplifier la désinformation à grande échelle et manipuler l'opinion publique, dans le but d'influer sur les résultats des élections. Les très grandes plateformes en ligne (VLOP) contribuent à ce phénomène et les efforts actuellement déployés pour réglementer les contenus générés par l'IA accusent un retard par rapport aux besoins en temps réel.
5. Les contenus générés par l'IA peuvent s'avérer particulièrement préjudiciables pour les groupes marginalisés, notamment les femmes, les personnes migrantes et les personnes LGBTI, et ainsi contribuer à perpétuer les discriminations et les inégalités dans la vie réelle. Les systèmes algorithmiques aussi : par exemple, l'utilisation de l'IA pour la prise de décision automatisée dans les services publics a des répercussions significatives sur les personnes marginalisées dans des secteurs tels que l'aide sociale, la police et les migrations. Les enfants sont aussi touchés de près par les systèmes algorithmiques et les agents conversationnels. Qu'il s'agisse de troubles de l'attention ou de troubles cognitifs (ce que l'on appelle le « brain rot » ou « pourriture du cerveau »), de dépendance ou de problèmes de santé mentale, nous constatons de plus en plus les conséquences négatives de l'utilisation de l'IA chez les enfants. Il convient de redoubler d'efforts pour protéger les enfants contre ces risques, tout en garantissant une participation significative et adaptée à leur âge aux décisions et aux politiques qui les concernent.
6. L'IA utilise de vastes quantités de données, y compris des données à caractère personnel. Le stockage et l'accès à ces données soulèvent de nombreuses préoccupations en matière de confidentialité. La société civile met en avant deux risques principaux. Premièrement, les données à caractère personnel stockées peuvent être inexactes : lorsque les systèmes d'IA, tels que les LLM, génèrent des données inexactes sur des individus, les personnes concernées peuvent ne pas en avoir connaissance et ne pas être en mesure d'en demander la rectification. Deuxièmement, il existe un risque de perte de contrôle sur les données à caractère personnel : les utilisateurs et utilisatrices ne bénéficient souvent pas de

garanties contractuelles solides, car de nombreuses entreprises ont la possibilité de mettre à jour leurs politiques à tout moment et de conserver le contrôle sur les données. Toute personne peut exercer son droit d'opposition à l'utilisation de ses données pour les systèmes d'IA, mais si les données ont déjà été stockées et utilisées — en particulier pour l'entraînement de l'IA — il n'est pas facile de les supprimer. Le recours à des modèles de consentement explicite pour la collecte et le traitement des données peut permettre d'éviter ce risque. En outre, il convient d'être prudent lors de l'intégration de systèmes d'IA existants dans différents environnements, tels que les assistants d'IA intégrés dans des logiciels ou des plateformes de recherche, qui peuvent donner accès à des données privées ou confidentielles sans l'autorisation explicite de l'utilisateur ou l'utilisatrice. Cela peut également se produire lors de mises à jour logicielles, sans que l'utilisateur ou l'utilisatrice ne le sache ou n'ait la possibilité de s'y opposer.

7. Les incidences environnementales et sociales de l'IA à l'échelle mondiale sont un sujet de préoccupation croissante au sein de la société civile. D'une part, les chaînes d'approvisionnement de l'IA reposent sur l'extraction de minerais (parfois dans des conditions abusives, notamment le recours au travail des enfants), et les dommages environnementaux et sanitaires qui en résultent touchent de manière disproportionnée les communautés marginalisées pour des raisons raciales et socio-économiques dans les régions minières, en particulier dans les pays du Sud. D'autre part, les centres de données consomment de grandes quantités d'eau et d'énergie. De plus, leur taux d'utilisation augmente rapidement à mesure que la demande de systèmes d'IA explose et que les modèles deviennent plus grands et plus complexes, ce qui les rend plus gourmands en énergie. Cette situation a déjà des répercussions négatives, bien que localisées pour l'instant, sur l'environnement des collectivités où sont implantés les centres de données, notamment des pannes d'électricité, des pénuries d'eau et des problèmes de santé. En conséquence, les populations locales sont de plus en plus conscientes des effets négatifs potentiels de l'IA, ce qui peut les inciter à faire pression pour obtenir des changements politiques et législatifs.
8. Dans de nombreux autres secteurs, tels que la biométrie et la surveillance, il convient d'accorder une attention particulière à l'utilisation de l'IA. Compte tenu de l'impact considérable de l'IA sur la vie privée et publique, une réglementation s'impose. Cet impact doit être évalué dès le départ, et des mécanismes de contrôle et de responsabilité sont indispensables pour que l'IA contribue à faire progresser les droits humains et non à les compromettre.

Thème 2 : Lacunes dans les cadres de gouvernance européens et propositions de solutions

9. La section suivante présente les principaux défis d'une gouvernance fondée sur les droits humains, ainsi que des propositions de solutions. Le manque de transparence des systèmes d'IA ne permet pas d'assurer un contrôle humain suffisant. Par ailleurs, l'influence des géants du numérique risque d'entraver les progrès en matière de protection des droits humains en façonnant le discours public sur l'IA, notamment par des campagnes de relations publiques et de marketing qui mettent l'accent sur les capacités (parfois de manière trompeuse), tout en minimisant les risques. Les responsables politiques et le grand public comprennent donc moins bien les effets négatifs de l'IA sur les droits humains. Enfin, les réglementations existantes ne sont actuellement pas suffisamment appliquées, ce qui signifie que la responsabilité et l'obligation de rendre des comptes en cas de préjudices liés à l'IA restent limitées.

10. Dans le contexte géopolitique actuel, les entreprises technologiques sont peu incitées à aligner volontairement leurs pratiques sur l'intérêt public. L'expérience montre que la réglementation peut jouer un rôle crucial pour combler cette lacune. Par exemple, à la suite de l'adoption du règlement sur les services numériques, de nombreuses entreprises ont mis en œuvre des changements structurels. Elles ont, entre autres, renforcé la transparence et les mesures de modération de contenus et procédé à des évaluations des risques.
11. Cependant, à l'heure actuelle, les entreprises suspendent ou annulent leurs efforts de réforme en raison de l'incertitude liée aux changements réglementaires attendus. Dans le secteur privé, le nombre de personnes chargées de la conformité en matière de sécurité et de droits humains a diminué. Face à l'influence croissante des entreprises, qui risque d'affaiblir les protections des droits humains dans la sphère numérique, la réponse la plus efficace de l'UE est une application rigoureuse des réglementations existantes. Il s'agit là d'une revendication majeure de la société civile depuis quelques années.
12. À cet égard, la société civile craint fortement que les garanties et les mécanismes de contrôle ne soient affaiblis alors que l'UE donne la priorité à la compétitivité. Le paquet « Digital Omnibus » récemment proposé par la Commission européenne est présenté comme une initiative visant à rationaliser et à simplifier la législation numérique de l'UE, y compris la mise en œuvre du règlement sur l'IA, et à réduire les charges administratives qui pèsent sur les entreprises. Toutefois, les organisations de la société civile signalent que plusieurs dispositions sont vagues et risquent d'affaiblir les protections des droits humains. Par exemple, en vertu du règlement sur l'intelligence artificielle de l'UE tel qu'il est actuellement en vigueur, les fournisseurs et les déployeurs de systèmes d'IA doivent prendre des mesures pour garantir, dans toute la mesure du possible, que leur personnel possède un niveau suffisant de maîtrise de l'IA. Le « Digital Omnibus » propose de repenser cette disposition comme l'obligation pour les États membres et la Commission d'encourager les fournisseurs et les employeurs à garantir un niveau suffisant de maîtrise de l'IA parmi leur personnel et les autres acteurs concernés. Cette proposition permet aussi, entre autres, de déroger à certaines exigences du règlement sur l'IA applicables aux systèmes d'IA à haut risque. Les organisations de la société civile, soulignent, en outre que les multiples propositions de simplification sont éparpillées entre différents instruments, ce qui entrave leur capacité de s'engager dans le processus de manière opportune et significative.
13. Le débat sur la déréglementation a des répercussions supplémentaires pour les pays candidats à l'UE, y compris ceux des Balkans occidentaux qui sont déjà en train d'aligner leurs cadres juridiques sur la législation de l'UE. Cette préoccupation est double. Premièrement, toute modification ultérieure de la législation européenne devrait également être transposée dans le droit national, ce qui pourrait retarder le processus d'adhésion. Deuxièmement, le programme de simplification risque de laisser entendre à ces pays que l'innovation doit primer sur la protection des droits humains — un message alarmant qui pourrait saper les valeurs fondamentales dans des démocraties déjà fragiles.
14. 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 offre un cadre de référence contraignant axé sur les droits humains et ouvert aux États non membres de l'UE. Véritable pilier de l'engagement de la société civile dans les pays candidats à l'UE, elle est de plus en plus pertinente. C'est pourquoi il est nécessaire d'accélérer sa signature et sa ratification, en particulier parmi les États non membres de l'UE.

15. La société civile souligne l'importance d'agir de concert pour s'opposer efficacement au discours du secteur technologique en faveur de la déréglementation. Une société civile dynamique est essentielle pour assurer un équilibre des pouvoirs face aux évolutions en cours, tant au niveau de la technologie que de sa réglementation. Cependant, le rétrécissement de l'espace civique nuit aux initiatives mises en œuvre pour favoriser le respect des droits humains dans les secteurs de la technologie et de l'IA, tandis que les possibilités pour la société civile de participer de manière significative à la prise de décision et à la surveillance de l'IA restent limitées.
16. La transposition de certaines exigences du règlement de l'UE sur l'intelligence artificielle en normes techniques continue de poser problème, notamment s'agissant de l'obligation de garantir un contrôle humain. La tâche est d'autant plus complexe que les utilisations de l'IA et les profils de risque varient considérablement selon les secteurs, ce qui rend difficile l'élaboration et l'application de normes générales. Le domaine de la normalisation et de la réglementation en matière d'IA en est encore à ses balbutiements et il existe peu de jurisprudence ou de précédents pertinents pour servir de référence.
17. La société civile souligne que toute restriction à l'utilisation de l'IA doit être proportionnée à l'objectif légitime poursuivi et ne doit pas être appliquée de manière généralisée. Par exemple, une interdiction totale des médias sociaux pour les enfants pourrait porter atteinte à leur droit de participer à l'environnement numérique et de s'engager dans la vie civique. En outre, les modalités d'introduction et de mise en œuvre de telles interdictions par les fournisseurs de services dans le cadre de la réglementation numérique de l'UE font actuellement l'objet d'un contrôle limité. Dans le même temps, les autorités devraient exiger des entreprises qu'elles veillent à ce que les produits d'IA destinés aux enfants soient sûrs dès leur conception. Certains algorithmes manipulateurs — en particulier ceux qui exploitent les enfants — devraient être interdits et il devrait incomber aux entreprises de démontrer que leurs produits sont sûrs.
18. Les pouvoirs publics doivent également être prêts à interdire certaines utilisations de la technologie qui portent atteinte aux droits humains ou risquent sérieusement de les violer. Par exemple, les systèmes d'armes dont les effets ne peuvent être suffisamment compris, prévus ou expliqués devraient être interdits, tout comme l'utilisation d'armes autonomes pour sélectionner et attaquer des cibles humaines.
19. La transparence — y compris la transparence du contrôle — est primordiale pour garantir la protection des droits humains et doit inclure l'accès aux données. À l'heure actuelle, il n'existe aucune base juridique ni procédure claire permettant à des chercheurs indépendants ou à des organisations de la société civile d'enquêter sur la manière dont les systèmes d'IA sont utilisés en pratique. Ces acteurs devraient avoir accès aux données, à la documentation et aux outils techniques nécessaires pour assurer un contrôle efficace des systèmes d'IA et pour compléter le suivi institutionnel (avec des garanties appropriées en matière de confidentialité et de sécurité). Cela doit aller de pair avec la garantie que les journalistes d'investigation disposent des ressources nécessaires pour accomplir leur travail. La transparence devrait également être renforcée en ce qui concerne les marchés publics de systèmes d'IA, qui, pour l'instant, restent systématiquement opaques.
20. Le fait que le public comprenne mal comment les données à caractère personnel sont collectées, partagées et traitées par les services en ligne dans le cadre de leur utilisation quotidienne constitue un autre obstacle à la transparence. Des investissements accrus dans la maîtrise du numérique devraient permettre aux utilisateurs de reconnaître les pratiques courantes de suivi, de comprendre

quelles données sont stockées, par qui et à quelles fins, et de connaître leurs droits et les moyens de contrôle dont ils disposent.

21. L'un des moyens de se protéger contre les violations des droits humains liés aux systèmes d'IA consiste à déposer des recours auprès des autorités nationales chargées de la protection des données. Par exemple, les recours déposés par une organisation non gouvernementale dans quatre pays contre Clearview AI — une entreprise qui traitait les données biométriques de particuliers issues d'images faciales — ont conduit les autorités chargées de la protection des données à conclure que ce traitement de données à caractère personnel était dépourvu de fondement juridique. Il reste néanmoins difficile de faire appliquer ces décisions à l'encontre des entreprises qui ne sont pas établies dans l'UE. Les autorités devraient également s'assurer que des voies de recours sont disponibles et facilement accessibles aux personnes.
22. Pour surmonter les difficultés actuelles en matière de gouvernance et d'application de la législation, les institutions européennes et internationales devraient systématiquement demander aux entreprises technologiques de respecter les principes des droits humains et s'efforcer de combler les lacunes des cadres juridiques afin de garantir la responsabilité des entreprises. Les pouvoirs publics ont le devoir de faire respecter la loi, quel que soit le niveau d'investissement du secteur privé dans leur économie nationale. Parallèlement, il convient d'attirer l'attention sur le manque important de ressources auquel font face les organes chargés de l'application de la législation, qui entrave leur capacité à remplir leur mandat. Il en va de même pour les structures nationales des droits de l'homme, y compris les organismes de promotion de l'égalité, qui se voient confier de nouvelles responsabilités en matière de contrôle sans que leurs ressources ne soient augmentées en conséquence. Les autorités doivent veiller à ce que ces organes disposent de ressources humaines, financières et techniques suffisantes pour exercer efficacement leurs fonctions.

Principales conclusions :

- Les systèmes d'IA automatisés devraient faire l'objet d'une réglementation appropriée et d'un contrôle humain effectif, en particulier dans les domaines militaire et de la sécurité, en raison des risques élevés qu'ils représentent pour les droits humains.
- La dégradation de la qualité de l'information et la surabondance d'informations, amplifiées par les systèmes d'IA, transforment les habitudes de consommation et les façons d'interagir avec les contenus et érodent la confiance. L'IA intensifie également la propagation de la désinformation, ce qui est particulièrement préjudiciable à certains groupes de personnes, notamment les enfants, les personnes migrantes, les femmes et les personnes LGBTI.
- L'utilisation de données à caractère personnel par les systèmes d'IA soulève d'importantes questions du point de vue du respect de la vie privée. Les responsables politiques européens devraient promouvoir des modèles de consentement explicite pour la collecte et le traitement des données afin de garantir aux utilisateurs et utilisatrices un plus grand contrôle sur leurs données personnelles. Des garanties contractuelles devraient empêcher toute mise à jour de produits ou de politiques intégrant l'IA sans notification claire ni consentement explicite.
- Les effets de l'IA sur l'environnement et dans le domaine social sont de plus en plus manifestes, notamment dans les chaînes d'approvisionnement mondiales et là où sont construits les centres de données. Ces effets devraient être au cœur des actions de sensibilisation et des débats politiques sur l'IA.

- Les utilisations de l'IA qui portent atteinte aux droits humains ou qui risquent sérieusement de les violer devraient être interdites, et les autorités devraient imposer la sécurité dès la conception pour les produits d'IA destinés aux enfants.
- Il convient de garantir une participation effective de la société civile au processus décisionnel et de lui allouer des ressources suffisantes afin d'assurer un équilibre des pouvoirs en matière de développement et de réglementation de l'IA. Les chercheurs indépendants, les journalistes et les organisations de la société civile devraient avoir accès aux données utilisées par les systèmes d'IA, et la transparence des marchés publics devrait être renforcée.
- Il est essentiel d'assurer une application cohérente et rigoureuse des réglementations européennes existantes en matière de numérique, en allouant des ressources suffisantes aux autorités de contrôle, y compris aux structures nationales des droits de l'homme.
- Il convient d'intensifier les investissements dans la maîtrise du numérique et l'élaboration de normes conformes aux droits humains.

Annexe I

PROGRAMME

GOUVERNANCE DE L'IA GÉNÉRATIVE : COMBLER LES LACUNES POUR PROTÉGER LES DROITS HUMAINS

Consultation avec les organisations de la société civile

Paris, 6-7 novembre 2025

Lieu : Salle 5, Bureaux du Conseil de l'Europe à Paris

55 Avenue Kléber, 75116 Paris

Jeudi 6 novembre 2025

13 h 45 – 14 heures Arrivée et enregistrement

Salle 5, Bureaux du Conseil de l'Europe à Paris

14 heures — 14 h 10 Allocution de bienvenue et introduction par Michael O'Flaherty, Commissaire aux droits de l'homme

14:10 - 14:15 Les travaux du Conseil de l'Europe sur l'IA et les droits humains

Hristijan Koneski, Conseiller auprès du Commissaire aux droits de l'homme

14 h 15 – 14 h 45 Tour de table

- Présentation rapide des participant·es et de leurs travaux

14 h 45 – 15 h 45 Séance 1 : Risques pour les droits humains liés au déploiement de l'intelligence artificielle générative

Cette séance vise à identifier les risques pour les droits humains liés au déploiement de l'IA générative, en mettant l'accent sur les groupes défavorisés et les enfants.

Discussion

15 h 45 – 16 h 05 *Pause-café*

16 h 05 – 17 h 55 Poursuite de la discussion

18 heures *Fin de la première journée*

Vendredi 6 novembre 2025

09 heures — 09 h 05 Introduction

Sandra Veloy Mateu, Conseillère auprès du Commissaire aux droits de l'homme

9 h 05 – 10 h 40 Séance 2 : Lacunes dans les cadres de gouvernance européens et propositions de solutions

Cette séance vise à identifier les principales lacunes de la réglementation européenne actuelle relative au déploiement des technologies d'IA générative et leurs effets sur les droits humains, ainsi qu'à rechercher des solutions pour combler ces lacunes et renforcer la gouvernance de l'IA en conformité avec le droit international en matière de droits humains.

Discussion

- 10:40 - 11:00 Pause-café
- 11:00 - 12:50 Poursuite de la discussion
- 12:50 - 13:00 Remarques finales de Michael O'Flaherty, Commissaire aux droits de l'homme
- 13:00 *Fin de la deuxième journée*

Annexe II

Liste des participant·es

Appelman, Naomi

Çetin, R. Buse

Coppi, Giulio

Dempsey, Mark

Fedchenko, Yevhen

Letouche, Manon

Ristić, Andrijana

Sardeli, Kleanthi

The Racism and Technology Center

AI Forensics

Access Now

ARTICLE 19

Stop Fake

5 Rights Foundation

Share Foundation

NOYB (none of your business)

Conseil de l'Europe

O'Flaherty, Michael

Havula-Lorenzini, Anna

Koneski, Hristijan

Veloy Mateu, Sandra

Commissaire aux droits de l'homme

Assistante

Conseiller auprès du Commissaire

Conseillère auprès du Commissaire