



Strasbourg, 10 June 2026

CDDEM(2025)18

COMITÉ DIRECTEUR SUR LA DÉMOCRATIE (CDDEM)

IA générative et démocratie

**Étude sur les avantages et les risques de l'utilisation de l'intelligence
artificielle générative dans le débat public et les processus démocratiques**

Préparé par

Prof. Dr. Dr. h.c. Natali Helberger et Prof. Dr. Claes de Vreese, Gionata Bouche, Agustin Ferrari
Braun, Dr. Max van Drunen, Dr. Simon Kruschinski, Dr. Nicolas Mattis, Dr Sophie Morosoli, Dr
Laurens Naudts, Dr. Charis Papaevangelou, Dr. Theresa Seipp, Dr. Fabio Votta, Dr. Teresa Weikmann
(Université d'Amsterdam, AI, Media & Democracy Lab et GESIS - Institut Leibniz des sciences
sociales, Cologne (Dr. Simon Kruschinski))

IA générative et démocratie

Étude sur les avantages et les risques liés à l'utilisation
de l'intelligence artificielle générative dans le débat
public et les processus démocratiques



Table des matières

Résumé	6
1. Introduction	9
2. L'IA Générative et institutions démocratiques	12
2.1 Journalisme	12
2.1.1 Le potentiel transformateur et les applications de l'IA générative	12
2.1.2. Comment la recherche générative remet en question le journalisme.....	16
2.1.3 Point de vue des citoyens sur les risques liés à l'utilisation de l'IA générative dans le journalisme.....	17
2.1.4 La montée de la désinformation alimentée par l'IA	21
2.1.5 Problèmes structurels et démocratiques : dépendance numérique et asymétrie de pouvoir	23
2.1.6 Conclusion	24
2.2 L'IA générative dans les campagnes politiques et les pratiques démocratiques	25
2.2.1 Aperçu du déploiement de l'IA générative dans les campagnes politiques à travers le monde et des dépendances technologiques	26
2.2.2 Activités de campagne assistées par l'IA.....	28
2.2.3 Stratégies de communication assistées par l'IA.....	30
2.2.4 Personnalisation des messages politiques assistée par l'IA	33
2.2.5 L'IA générative pour la recherche d'informations politiques et l'engagement de la société civile	34
2.2.6 Attitudes des citoyens à l'égard de l'utilisation de l'IA générative dans les campagnes politiques	36
2.2.7 Conclusion	39
2.3 Infrastructure	40
2.3.1 La politique des infrastructures.....	41
2.3.2 Contrôle de l'infrastructure nécessaire à l'IA générative	42
2.3.3 Dépendance infrastructurelle et recul démocratique.....	45
2.3.4 Conclusion	46
3. Implications pour les droits humains, la participation politique et les principes démocratiques	46
3.1 Ajouter la dimension des droits humains : l'IA générative du point de vue des droits à l'égalité, à la non-discrimination et à la participation politique	47
3.1.1 Inégalités et implications structurelles	48
3.1.2 Le potentiel du droit de la non-discrimination pour répondre à ces préoccupations	50

3.1.3 Conclusion	52
3.2 IA générative et liberté d'expression	53
3.2.1 Ajouter la dimension des droits humains : l'IA générative du point de vue du droit à la liberté d'expression	53
3.2.2. Le potentiel de la doctrine de la liberté d'expression pour contribuer à la gouvernance démocratique de l'IA	58
4. Cadre juridique du Conseil de l'Europe	60
4.1 Situer l'IA générative dans les recommandations du Conseil de l'Europe	60
4.1.1 Recommandations du Conseil de l'Europe et rôle de l'IA générative dans la société démocratique.....	61
4.1.2 Conclusion	66
4.2 Convention-cadre sur l'Intelligence Artificielle	66
4.2.1 Les droits humains et la démocratie comme point de départ.....	66
4.2.2 Risques pour les droits humains et la démocratie	67
4.2.3 Obligations positives de protéger la réalisation des droits humains et de la démocratie	67
4.2.4 Innovation d'intérêt public.....	68
4.2.5 Démocratiser la gouvernance de l'IA.....	69
4.2.6 Conclusion	71
5. Conclusions générales et recommandations	72

Résumé

Les nouvelles technologies offrent de nouvelles opportunités pour la réalisation des droits fondamentaux. Cela vaut également pour l'IA générative. Les outils de l'IA générative peuvent permettre de nouvelles formes de création et de diffusion de contenus d'actualité et de communication politique, de manière plus accessible, engageante et inclusive, ouvrant ainsi de nouvelles voies pour la réalisation des droits humains tels que la liberté d'expression, la non-discrimination et l'égalité des chances de participation. Cette technologie peut élargir les canaux d'information et d'engagement des électeurs, et offrir de nouveaux moyens de rendre l'information plus pertinente, plus réactive et plus accessible dans différentes langues et différents formats, y compris pour les organisations disposant de ressources limitées. De plus, les fonctionnalités de traduction et d'accessibilité basées sur l'IA créent des opportunités significatives pour atteindre les communautés marginalisées, adapter le contenu à des besoins cognitifs et physiques divers, et renforcer la communication multilingue.

Saisir ces opportunités et créer les conditions propices à l'innovation sociétale nécessitera non seulement de gérer les risques multiples pesant sur la réalisation des droits fondamentaux, mais aussi de créer un environnement institutionnel plus large favorable à l'innovation d'intérêt public. S'appuyant sur des recherches de pointe en sciences de la communication, en journalisme numérique et en droit, le rapport identifie également des préoccupations clés au niveau des citoyens individuels, par exemple celles liées au respect de la vie privée, à l'amplification des préjugés et à la propagation de la désinformation. Il met en évidence les difficultés rencontrées par les institutions démocratiques, notamment les menaces qui pèsent sur la viabilité sociétale et économique du journalisme face à la concurrence des assistants virtuels et des systèmes d'IA des grandes entreprises technologiques, ou à l'utilisation de contenus synthétiques pour intimider et manipuler les électeurs. Il appelle également à prêter attention à la forme et à la structure globales des infrastructures informationnelles sous-jacentes et des dynamiques de pouvoir, qui sont souvent moins visibles pour les décideurs politiques mais qui, en fin de compte, fournissent les conditions matérielles, numériques et organisationnelles dans lesquelles l'information est produite, diffusée et échangée, et constituent la source d'un pouvoir économique et politique considérable.

La principale conclusion de ce rapport est qu'il ne suffit pas de s'assurer que la technologie est sûre et ne présente aucun risque pour les droits humains et les valeurs publiques. L'impact de l'IA générative et son potentiel pour la réalisation des droits fondamentaux dépendent du contexte sociétal, organisationnel et réglementaire dans lequel elle s'inscrit, de la capacité des institutions à faire des choix fondés sur des valeurs, ainsi que de la présence de régulateurs qui encouragent activement l'innovation publique, lorsque les intérêts de la société dans son ensemble l'emportent sur les intérêts purement économiques d'une poignée d'entreprises technologiques. Ce rapport soutient que l'IA générative et ses implications pour les droits humains et la démocratie doivent être considérées dans le contexte des configurations du pouvoir infrastructurel et politique, et de la question de savoir qui est capable et habilité à décider de l'orientation et de la nature de l'innovation.

En mettant l'accent sur les droits humains, la démocratie et l'État de droit, le Conseil de l'Europe est particulièrement bien placé pour adopter une telle perspective plus sociétale sur l'innovation en matière d'IA et pour proposer un contre-discours important face aux discours sur l'innovation technologique davantage axés sur le profit.

À cette fin, le rapport formule un certain nombre de recommandations à l'intention du Conseil de l'Europe :

1. Fournir des orientations et une normalisation de fond

En tant qu'organisation de défense des droits humains, le Conseil de l'Europe est particulièrement bien placé pour élaborer des orientations plus substantielles et identifier, concrétiser et surveiller les risques pour la démocratie, les droits humains et l'égalité politique, qu'il s'agisse de la manière dont l'IA générative est utilisée dans des pratiques démocratiques telles que les élections, des implications plus structurelles pour les institutions démocratiques (telles que la liberté de la presse) ou encore de la gouvernance démocratique de l'IA. Outre l'accent mis sur les risques, le Conseil de l'Europe peut également élaborer des orientations utiles sur la manière dont l'IA peut faire progresser la démocratie et les droits humains, et développer une interprétation fondée sur des valeurs de l'« innovation d'intérêt public » afin d'informer les États, les entreprises et les processus de normalisation

2. Encourager et faciliter l'innovation d'intérêt public

Outre l'approche fondée sur les risques en matière d'IA, qui est également inscrite dans la Convention-cadre du Conseil de l'Europe, le Conseil de l'Europe pourrait élaborer des lignes directrices sur les politiques et les approches de gouvernance qui favorisent et encouragent l'innovation d'intérêt public. Cela pourrait aider à identifier les implications en matière de démocratie et de droits humains des choix en matière de marchés publics, et à développer des discours visant à donner la priorité à l'innovation axée sur les valeurs, aux ensembles de données open source, à la décentralisation et à l'interopérabilité plutôt qu'à la promotion de l'échelle et de l'innovation menées par un nombre restreint d'entreprises technologiques.

3. Renforcer la résilience démocratique et la capacité à gouverner

Le Conseil de l'Europe, et le CDDEM en particulier, pourraient jouer un rôle important dans l'exploration de mécanismes et de solutions réglementaires visant à cartographier et à remédier aux dépendances existantes et à identifier les éléments essentiels à la mission d'un écosystème numérique européen, par exemple en surveillant l'utilisation de l'IA générative lors des élections, en stimulant le développement d'infrastructures d'information d'intérêt public et en renforçant la capacité des États à développer des modes de gouvernance démocratiques et inclusifs de l'IA générative. Outre l'autonomie numérique, donner aux acteurs de la responsabilité publique – tels que la société civile, les organismes de promotion de l'égalité, les universitaires ou les journalistes d'investigation – les moyens de participer à la gouvernance de la transformation numérique constitue une condition essentielle au renforcement de la résilience démocratique. Enfin, la résilience démocratique implique également de permettre un engagement citoyen plus significatif, y compris

de la part de groupes souvent marginalisés, par exemple grâce à la transparence, à l'alphabétisation et à une participation significative à la prise de décision.

Le droit relatif aux droits humains, ainsi que la mise en œuvre de la Convention-cadre du Conseil de l'Europe, exigent non seulement de gérer les implications négatives et les risques liés à l'IA, mais aussi de souligner l'importance de promouvoir activement la contribution positive de l'innovation à la réalisation des droits humains, de la démocratie et de l'État de droit.

1. Introduction¹

Grâce à sa capacité à analyser d'énormes quantités de données, à produire des textes, des images et des contenus audiovisuels de haute qualité, et à ouvrir la voie à de nouvelles formes d'interaction avec les utilisateurs par le biais d'assistants virtuels ou de la recherche générative, l'IA générative a un potentiel énorme pour influencer les institutions et les pratiques démocratiques fondamentales. Cela s'explique par le rôle central que joue la communication dans notre démocratie. La démocratie repose sur la capacité à partager et à accéder à l'information, à exprimer son point de vue et à approuver ou contester celui des autres, à mobiliser, à s'engager, à se rassembler et à appeler à l'action politique. Jusqu'à présent, ce pouvoir de la parole était « ce qui nous distinguait des autres animaux et, jusqu'à présent, des ordinateurs » (Garton Ash, 2016, p. 73) . Avec l'IA générative, ce rôle du langage comme frontière entre l'ordinateur et l'humain s'érode rapidement. L'IA générative, en tant que « synthétiseur de langage », peut directement alimenter toutes ces formes d'interaction communicationnelle auparavant réservées aux humains, et ce faisant, faire ou défaire un discours démocratique, un choix politique autonome ou une sphère publique digne de confiance.

L'IA générative est « une intelligence artificielle (IA) capable de créer du contenu original tel que du texte, des images, des vidéos, des fichiers audio ou du code logiciel en réponse à une demande ou à une requête d'un utilisateur » (Stryker, Cole & Scapicchio, 2025) . On peut citer comme exemples les systèmes ChatGPT, Gemini, Claude ou Grok. Ces systèmes sont alimentés par des modèles linguistiques à grande échelle (LLM, tels que GPT d'Open AI ou LLaMA de Meta), également appelés modèles fondamentaux dans le sens où il s'agit de systèmes d'IA dotés de capacités étendues pouvant être adaptés à toute une série d'objectifs différents. Leur nature polyvalente distingue les systèmes d'IA générative des autres applications d'IA qui sont souvent conçues dans un but précis. Les systèmes d'IA générative deviennent également de plus en plus multimodaux, dans le sens où ils peuvent fonctionner avec et créer du texte, des images, des vidéos ou du son. Une partie du succès et de l'adoption rapide de l'IA générative réside dans sa facilité d'utilisation, qui présente souvent une interface similaire à celles que nous connaissons déjà, par exemple celles des moteurs de recherche. Cependant, en coulisses, l'IA générative fonctionne très différemment d'un moteur de recherche.

Le résultat de l'IA générative est avant tout le fruit d'un calcul statistique qui permet au système de prédire le mot suivant dans une phrase. Il est important de comprendre le caractère statistique de la communication de l'IA générative, car cela signifie que le système ne sait pas et ne réfléchit pas à ce qu'il dit. En conséquence, l'IA générative a été décrite comme une entité qui « assemble au hasard des séquences de formes linguistiques [...] en fonction d'informations probabilistes sur la manière dont elles se combinent, mais sans aucune référence au sens » (Bender et al., 2021, p. 617) . Cela signifie que l'IA générative peut générer des textes, des images, des fichiers audio ou vidéo de haute qualité qui imitent de manière très convaincante le discours produit par les humains, sans aucune compréhension ni aucun sens de la véracité. En raison de la facilité et de la rapidité avec lesquelles les

¹ Nous tenons à remercier les membres du CDDEM, la Secrétaire du Comité, Urška Umek, ainsi que les participants à la table ronde à huis clos intitulée « IA générative et démocratie », organisée le 28 octobre 2025 par le Laboratoire IA, Médias et Démocratie, pour leurs précieux commentaires sur cette étude.

systèmes d'IA générative peuvent produire du contenu, cela abaisse considérablement les seuils pour certaines pratiques communicatives, telles que la personnalisation du contenu afin de l'adapter aux caractéristiques ou aux préférences individuelles des utilisateurs. Cette technologie peut également permettre de nouvelles formes d'engagement (par exemple sous la forme d'assistants virtuels adaptatifs), un niveau plus élevé d'action autonome, ainsi que l'intensification des pratiques de communication. Ces caractéristiques distinguent les technologies basées sur l'IA générative des autres technologies de communication, telles que les systèmes de recommandation conventionnels ou les moteurs de recherche. L'IA générative peut amplifier les pratiques de communication numérique, telles que la diffusion de désinformation synthétique sur les plateformes de médias sociaux existantes.

L'IA générative et ses implications pour la démocratie ne peuvent être considérées indépendamment du contexte plus large dans lequel cette technologie opère. Bien que les discours politiques se concentrent souvent sur la technologie elle-même, les performances de l'IA générative résultent d'un réseau complexe de facteurs et de ressources humains et sociétaux. Il s'agit notamment de la production de contenu par les créateurs et les citoyens, contenu qui est ensuite utilisé pour former les modèles. L'IA générative s'appuie sur le travail des étiqueteurs de données pour préparer le contenu destiné à l'entraînement ou pour informer les pratiques de modération de contenu, sur les red-teamers et les vérificateurs de faits, ainsi que sur l'expertise et les instructions de ceux qui entraînent les modèles. Il faut aussi compter les ressources naturelles nécessaires à la fabrication des puces et au refroidissement des centres de données, qui alimentent les infrastructures informatiques indispensables. Les systèmes et modèles d'IA générative sont également façonnés par les pratiques de gestion internes et les modèles commerciaux, ainsi que par l'engagement d'innombrables personnes qui fournissent les résultats obtenus comme nouvelles données d'entraînement. Enfin, il ne faut pas oublier le rôle des codes éthiques, des directives communautaires et des cadres réglementaires qui orientent le cours de l'innovation et les processus politiques qui sous-tendent ces cadres de gouvernance. En tant que telle, l'IA générative est plus qu'une simple technologie parmi d'autres. Elle s'inscrit dans un processus sociétal d'extraction, de transformation et de redistribution des ressources humaines et naturelles, et en fin de compte, du pouvoir.

Comme pour les réseaux sociaux, le développement de l'IA générative est principalement mené par une poignée de très grandes entreprises technologiques, dont beaucoup ont des liens ou sont identiques aux opérateurs des principales entreprises de réseaux sociaux. Le développement des infrastructures informatiques sous-jacentes est dominé par une concentration de fournisseurs hyperscale (hyper-scalers) très influents, tels qu'Amazon, Microsoft, Google et OpenAI, qui sont fortement incités à utiliser l'IA générative pour consolider davantage leur contrôle sur la fourniture mondiale de services. En outre, la révolution de l'IA générative survient à un moment où les entreprises technologiques jouent également un rôle politique de plus en plus important et influencent les rapports de force géopolitiques. Outre les possibilités offertes par l'IA générative en matière de communication, les nouvelles formes de concentration du pouvoir technologique, économique et communicationnel qui en résultent ont des implications très particulières pour les processus démocratiques et la réalisation des droits humains.

Cette perspective socio-technique plus large orientera également l'analyse présentée dans ce rapport. Adopter une approche socio-technique signifie que nous situons l'IA générative en tant qu'éléments technologiques (par exemple sous la forme d'un agent virtuel ou d'un système de recommandation) au sein d'une chaîne de valeur plus large d'acteurs (y compris les fournisseurs de données, les fournisseurs d'infrastructures cloud, les fournisseurs d'applications, les utilisateurs en aval, etc.) et de pratiques (non) démocratiques. Cela permettra d'avoir une discussion plus nuancée sur les interventions possibles et les points d'intervention, mais aussi d'expliquer comment la dynamique au sein de cette chaîne de valeur affecte en fin de compte la réalisation des droits humains et les processus démocratiques. Nous aborderons en particulier le problème des déséquilibres de pouvoir, de la capture des infrastructures, de la souveraineté numérique et de la manière dont l'IA peut être une technologie de liberté, mais aussi de contrôle autocratique.

Nous le ferons à l'aide de trois études de cas. La première concerne l'utilisation de l'IA générative dans le journalisme. Selon le dernier rapport de Reuters Trends and Prediction intitulé «Trends and Prediction Report (Newman & Cherubini, 2025) », 90 % des rédacteurs en chef, PDG et dirigeants du secteur numérique interrogés dans le monde entier affirment que les salles de rédaction sont en train d'être entièrement ou partiellement transformées par l'IA générative. Les salles de rédaction du monde entier expérimentent l'intégration de l'IA générative tout au long de la chaîne de valeur de l'information, de la recherche et l'analyse des données à la traduction, la transcription et la vérification des faits, en passant par la génération de recommandations personnalisées, de nouvelles formes d'agrégation d'informations et la communication avec le public via des chatbots et des assistants virtuels. Dans cette mesure, l'utilisation de l'IA générative dans le journalisme est instructive, car elle montre comment l'innovation technologique peut affecter le fonctionnement, mais aussi la confiance et la légitimité d'une institution démocratique fondamentale, et être utilisée pour réaliser et faire progresser le droit constitutionnel à la liberté d'expression. Dans le même temps, l'exemple du journalisme est également pertinent pour montrer comment les dynamiques de pouvoir susmentionnées peuvent affecter la durabilité et les conceptions futures du rôle du journalisme dans l'écosystème de l'information.

La deuxième étude de cas porte sur l'utilisation de l'IA générative dans les processus électoraux et les campagnes politiques. Cette étude de cas nous permet de mettre en évidence la manière dont l'IA générative peut façonner les actions de communication entre les citoyens et les partis politiques (par exemple sous la forme de publicité politique personnalisée), donner aux citoyens les moyens de jouer leur rôle démocratique en tant qu'électeurs, ou permettre à des tiers (malveillants) d'utiliser cette technologie à des fins non démocratiques. C'est ici que la capacité de l'IA générative à produire du langage en tant que matière première des pratiques démocratiques, et les implications qui en découlent pour la démocratie, sont les plus prononcées, car les dynamiques sous-jacentes du pouvoir, de l'inclusion et de l'exclusion se manifestent.

La troisième étude de cas porte sur le rôle de l'IA générative dans le cadre de l'infrastructure publique d'information numérique qui sous-tend tous les processus de communication numérique. Si l'utilisation de l'IA générative dans les pratiques communicatives est souvent abordée sous l'angle d'une action et d'une relation très concrètes entre les citoyens et les institutions démocratiques (s'informer ou informer les autres, persuader et s'engager), les implications infrastructurelles sont beaucoup moins visibles, mais ont un impact beaucoup

plus profond et plus généralisé sur les démocraties. Les discussions actuelles sur l'autonomie numérique en sont un bon exemple. Comme la communication numérique dépend des infrastructures technologiques sous-jacentes, mais est également façonnée par celles-ci, la question de savoir qui développe et contrôle ces infrastructures, dans quel but et dans le respect de quelles valeurs, est susceptible d'affecter et de modifier fondamentalement les conditions de la communication publique et de l'exercice des droits humains, au point de mettre en péril l'indépendance des institutions démocratiques telles que la presse et d'élever la concentration du pouvoir économique et politique à un niveau qui n'est plus compatible avec les principes fondamentaux de la démocratie.

L'objectif général de ce rapport est de mettre en évidence les avantages et les risques que l'utilisation de l'IA générative peut présenter pour la démocratie, en mettant particulièrement l'accent sur les médias et les processus électoraux, et de formuler des recommandations sur les priorités à retenir pour le Comité directeur sur la démocratie (CDDEM) du Conseil de l'Europe.

Le rapport s'appuiera sur la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle, les droits humains, la démocratie et l'État de droit (Convention-cadre) ainsi que sur les recommandations et autres instruments du Conseil de l'Europe, mais examinera également les interfaces avec d'autres initiatives régionales/internationales en matière de gouvernance de l'IA générative, en mettant particulièrement l'accent sur l'Union européenne. Le rapport sera axé sur la recherche, en ce sens qu'il utilisera et traduira les conclusions des recherches de pointe dans des domaines tels que les sciences de la communication, le journalisme numérique et le droit, en suggestions de stratégie et de priorisation. Compte tenu de la portée limitée du rapport, l'objectif n'est pas de donner un aperçu complet de l'état de l'art, mais de signaler les tendances et les développements clés. L'objectif du rapport est de fournir un premier aperçu des sujets importants à prendre en considération.

2. L'IA Générative et institutions démocratiques

Après avoir présenté dans l'introduction les grandes lignes de la dynamique sociétale, technologique et économique générale, cette section se concentrera sur la manière dont l'IA générative permet, façonne, modifie ou menace les institutions et les processus démocratiques. Pour ce faire, cette section présentera une synthèse des recherches existantes sur les implications de l'IA générative pour (1) le journalisme, (2) les élections et (3) les infrastructures numériques.

2.1 Journalisme

2.1.1 Le potentiel transformateur et les applications de l'IA générative

Le journalisme est l'un des piliers des démocraties libérales, car il joue un rôle essentiel dans l'information du public, la promotion du débat public et la responsabilisation des dirigeants. En raison de son importance démocratique, le journalisme constitue également un excellent cas d'étude pour évaluer le potentiel transformateur de l'IA (générative) et ses implications

démocratiques et sociétales, d'autant plus que le public, dans différents pays, est plutôt critique à l'égard de l'impact de l'IA générative sur le journalisme (voir Figure 1).

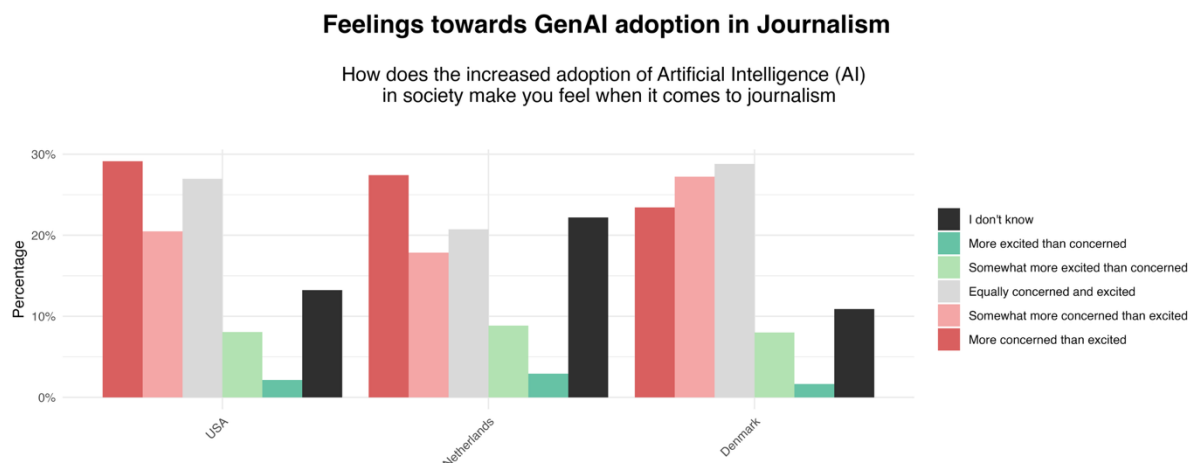


Figure 1. Sentiments à l'égard de l'adoption de l'IA générative dans le journalisme aux États-Unis, aux Pays-Bas et au Danemark. D'après Morosoli et al. (2025).

Un examen des pratiques actuelles dans les salles de rédaction montre que l'IA générative transforme déjà les flux de travail journalistiques. Bien que son utilisation varie d'un organe de presse à l'autre, l'IA générative imprègne de plus en plus l'ensemble de la chaîne de valeur de l'information, de la collecte d'informations et la production de contenu à la distribution potentiellement personnalisée. Les cas d'utilisation concrets comprennent, sans s'y limiter, l'analyse automatisée des tendances lors de la collecte d'informations, la synthèse, la transcription, l'optimisation des titres et même la génération d'articles entiers lors de la production d'informations, ou encore l'optimisation des moteurs de recherche et la personnalisation lors de la distribution d'informations (Cools & Diakopoulos, 2024). En effet, dans une enquête menée en 2024 auprès de professionnels de l'information par Diakopoulos et al. (2024), 70 % des personnes interrogées ont indiqué avoir déjà utilisé des outils d'IA générative pour la production de contenu textuel, et près de 80 % ont déclaré qu'elles aspiraient à le faire à l'avenir (Diakopoulos et al., 2024b).

Interrogés sur leurs motivations pour utiliser l'IA générative, les professionnels de l'information citent souvent sa capacité à améliorer l'efficacité du flux de travail. Selon eux, cela pourrait permettre aux journalistes de se consacrer à des tâches plus importantes, car l'IA générative automatise les aspects les plus fastidieux du travail journalistique et facilite les tâches créatives telles que la recherche d'informations et la génération d'idées (Cools & Diakopoulos, 2024 ; Diakopoulos et al., 2024b). En outre, l'IA générative facilite également la création d'un « journalisme fluide ou modulaire », où le contenu et la présentation s'adaptent aux préférences individuelles des utilisateurs (Cools & Diakopoulos, 2024 ; Jones & Jones, 2019). Entre autres choses, cela pourrait contribuer à rendre l'information plus pertinente, plus diversifiée et plus accessible (Lin & Lewis, 2022), par exemple en créant des résumés compréhensibles ou en générant des versions audio d'articles écrits. Parfois, le contenu synthétique peut également être un moyen pour les citoyens et les journalistes de signaler des problèmes qui, autrement, seraient censurés. Par exemple, au Venezuela, les journalistes se sont tournés vers l'IA générative pour créer des présentateurs journalistiques virtuels afin de protéger leur identité contre les poursuites judiciaires (Phillips & Torres, 2024). Un autre

cas d'utilisation potentiellement intéressant que les organes de presse explorent déjà est celui des agents conversationnels qui peuvent interagir avec les lecteurs et les aider à trouver des informations pertinentes.

En d'autres termes, il y a lieu d'espérer que l'IA générative contribuera à améliorer divers aspects et l'accessibilité du journalisme. Cependant, il est important de noter que ce sentiment n'est pas nécessairement partagé par le public. En effet, les résultats d'une enquête menée en avril 2025 dans six pays (États-Unis, Japon, Afrique du Sud, Brésil, Danemark et Pays-Bas) révèlent une ambivalence et des variations considérables dans la mesure où les citoyens américains, danois et néerlandais perçoivent les avantages des informations hyper-personnalisées (voir Figure 2, graphique B). Le risque le plus répandu dans tous les pays était celui d'une hyper-personnalisation ayant un impact négatif sur la confidentialité des données et celui de faciliter la propagation de fausses informations. Les participants ont vu les principaux avantages de l'hyper-personnalisation dans l'amélioration de l'expérience utilisateur, notamment en fournissant des informations plus pertinentes sur le plan personnel. En outre, la même enquête souligne également un intérêt quelque peu limité pour les informations (co)créées par l'IA dans les deux pays (voir Figure 2, graphique A et section 2.1 b). Dans l'ensemble, les participants étaient les moins disposés à partager ce type de contenu, le Danemark étant le plus sceptique. Dans tous les pays, les individus considéraient également les informations générées par l'IA avec une certaine méfiance, les citoyens néerlandais étant les moins disposés à leur faire confiance. C'est dans le domaine du journalisme basé sur l'IA que nous constatons la plus grande acceptation. Les habitants du Danemark et des États-Unis sont disposés à lire ce type de contenu.

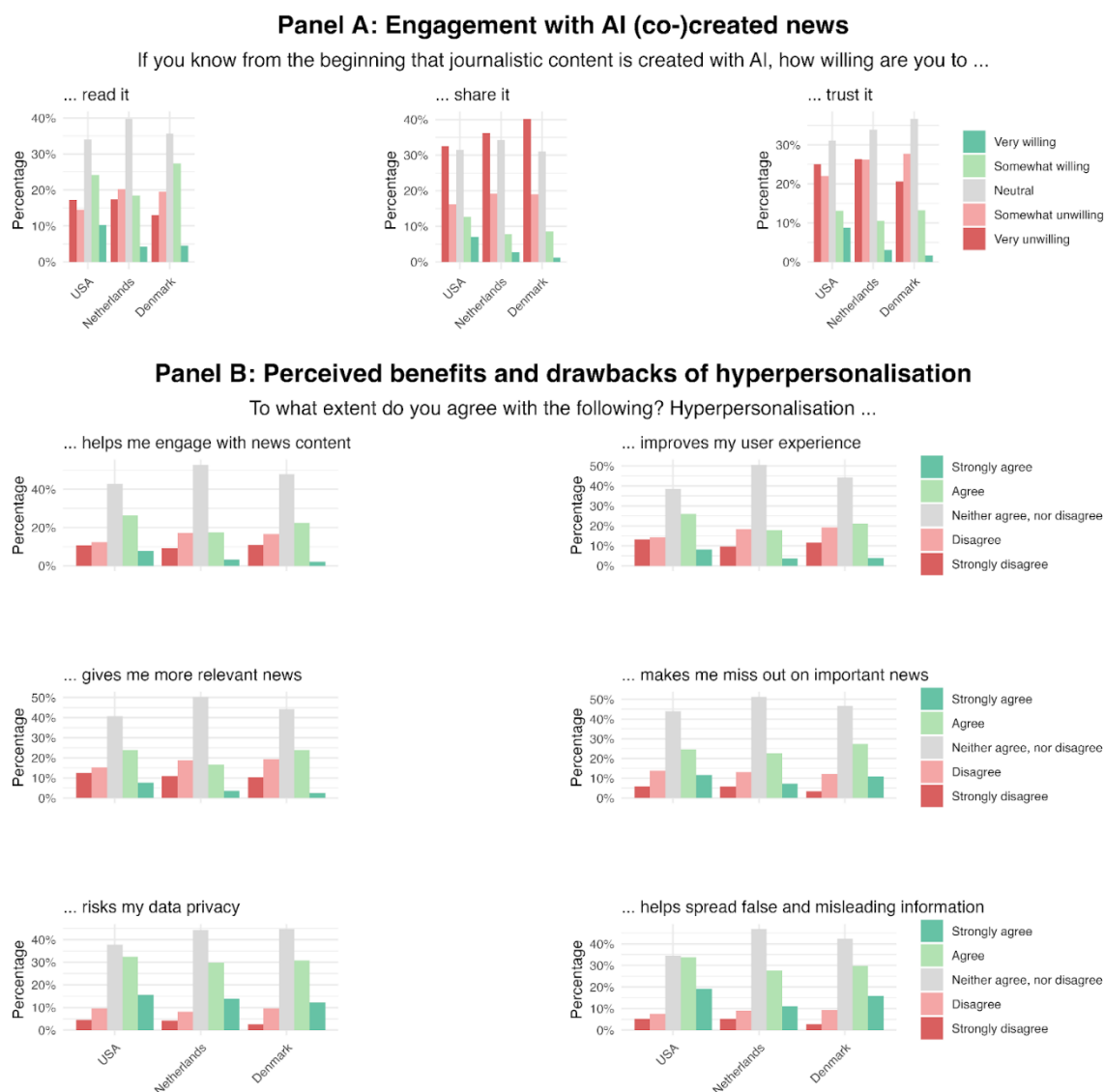


Figure 2. Engagement envers les actualités (co)créées par l'IA (A) et avantages perçus de l'hyper-personnalisation (B) aux États-Unis, aux Pays-Bas et au Danemark. Tiré de (Morosoli et al., 2025)

Cependant, de manière quelque peu paradoxale, nous constatons également qu'un nombre croissant de personnes utilisent l'IA générative pour rechercher des informations et rencontrent fréquemment des fonctionnalités d'IA telles que Google AI Overview dans leur vie quotidienne (Simon et al., 2025). Ici, les entreprises d'IA générative introduisent un nouveau niveau d'intermédiation dans l'accessibilité aux informations grâce à des interfaces de chatbot, des moteurs de réponse et des assistants intelligents intégrés aux systèmes d'exploitation mobiles. Cette évolution permet des formes plus fluides et participatives de consommation de l'information (Caswell, 2024), mais elle a déjà des répercussions économiques et en termes de visibilité sur les éditeurs, en particulier sur les marchés médiatiques anglo-saxons (Maher, 2025). Cela a des implications importantes pour la capacité des médias à exercer le pouvoir de contrôle et de définition de l'agenda qui leur est traditionnellement attribué (voir également la section 2.3).

De plus, la dépendance accrue du public à l'égard des LLM comme points d'accès à l'information constitue également une menace pour la viabilité économique du journalisme. En effet, bien que ces LLM soient souvent entraînés sur du contenu journalistique, ils omettent fréquemment de mentionner les sources d'information de manière adéquate ou déforment factuellement leur contenu. Par exemple, en comparant huit moteurs de recherche IA populaires (dont ChatGPT, DeepSeek, Gemini, Perplexity et d'autres), Jaźwińska et Chandrasekar (2025) constatent que tous échouent fréquemment à retrouver l'article, l'éditeur, le lien ou une combinaison de ceux-ci. Il est important de noter que même les accords de licence de contenu n'ont pas permis aux éditeurs de presse d'avoir la certitude que leur contenu serait correctement cité (Jaźwińska & Chandrasekar, 2025). Une étude récente menée auprès de 22 radiodiffuseurs publics internationaux le confirme, en constatant qu'en moyenne, 45 % des réponses de l'IA présentent au moins un problème important, tel que des hallucinations ou des problèmes de sourcing (EBU, 2025). En matière d'exactitude, cette étude révèle qu'environ la moitié des requêtes testées présentaient des problèmes d'exactitude et qu'environ 20 % des requêtes présentaient des problèmes d'exactitude importants. Étant donné que les LLM sont de plus en plus utilisés comme points d'accès directs à l'information, ces inexactitudes menacent de compromettre davantage la viabilité économique du journalisme en déformant l'origine ou la nature des affirmations journalistiques.

2.1.2. Comment la recherche générative remet en question le journalisme

Les entreprises d'IA générative introduisent un nouveau niveau de contrôle de l'accès à l'information par le biais d'interfaces de chatbot, de moteurs de réponse et d'assistants intelligents intégrés aux systèmes d'exploitation mobiles. La part du trafic transitant par ces canaux augmente régulièrement, les utilisateurs semblant préférer recevoir des réponses synthétiques et toutes faites plutôt que de naviguer à l'aide des hyperliens traditionnels (Reiss et al., 2025 ; Simon et al., 2025). Cette évolution a déjà des répercussions économiques et en termes de visibilité sur les éditeurs, en particulier sur les marchés médiatiques anglo-saxons (Maher, 2025).

La querelle de longue date entre l'industrie de l'édition et les grandes entreprises technologiques américaines a attiré l'attention du public sur la menace existentielle qui pèse sur la viabilité financière du journalisme numérique. Cela a abouti à des interventions législatives dans plusieurs juridictions, imposant une rémunération équitable pour la diffusion de contenus d'information sur les plateformes numériques. L'IA générative entraîne une nouvelle reconfiguration structurelle de l'environnement de l'information en ligne, qui dépasse l'influence autrefois exercée par les moteurs de recherche traditionnels et les plateformes de médias sociaux (voir également la section 2.3.2). Les implications sont à la fois expérientielles, en ce qui concerne la manière dont les citoyens rencontrent et consomment les informations, et politiques, en termes de modification des rapports de force au sein de l'écosystème médiatique. La notion d'« information comme conversation » marque un éloignement de la production artisanale, basée sur des éléments individuels, au profit de formes plus fluides et participatives de consommation de l'information (Caswell, 2024). Le mode de diffusion de l'information à guichet unique, peu contraignant et interactif offert par les systèmes génératifs est de plus en plus attrayant à l'ère de la surcharge d'informations numériques. Contrairement aux intermédiaires précédents, les entreprises d'IA sont désormais directement impliquées dans la production de contenu, s'immisçant davantage dans les fonctions et les valeurs

fondamentales de l'institution journalistique (Guzman & Lewis, 2024 ; Lewis et al., 2025). La dernière étape de ce processus contourne complètement les organes de presse : des plateformes telles que Sora d'OpenAI et Vibes de Meta proposent des flux de contenu généré par l'IA, ce qui pourrait annoncer un avenir où l'espace réservé au contenu produit par les journalistes sur les plateformes ne cessera de se réduire.

La relation entre le secteur de l'information et ces entreprises d'IA est ambiguë. Les éditeurs de presse traditionnels, et en particulier les grands acteurs, exercent toujours une influence considérable sur l'agenda public, une position que les intermédiaires en ligne contestent (Simon, 2024a) . Pourtant, le discours public entourant ces relations évolue, passant d'une confrontation initiale à une collaboration stratégique. Plusieurs éditeurs s'associent désormais à des entreprises d'IA pour former des modèles ou améliorer la production contextuelle grâce à la génération augmentée par la récupération et la recherche en ligne (Tobitt, 2025) . En contrepartie, les entreprises d'IA offrent une rémunération et une visibilité du contenu. Ces collaborations pourraient être considérées comme l'institutionnalisation d'une nouvelle forme de réciprocité entre le journalisme et l'IA. Alors que les systèmes génératifs façonnent la pratique journalistique, les modèles commerciaux et les modes de consommation, le journalisme informe le développement de l'IA en fournissant des données et des bases éthiques pour son utilisation communicative (Guzman & Lewis, 2024, p. 349) .

Cependant, ces tendances comportent des risques latents. En contribuant à améliorer des technologies capables de reproduire et, à terme, de remplacer un nombre croissant de fonctions journalistiques, les éditeurs risquent de compromettre leur propre autonomie et leur viabilité future (Simon, 2024a). Des dépendances systématiques sont susceptibles d'apparaître compte tenu de la puissance financière supérieure des entreprises d'IA générative et de leur contrôle sur les infrastructures, qui vont du matériel informatique aux interfaces utilisateur (Sjøvaag et al., 2024). De plus, comme les modèles génératifs représentent des « mutations épistémiques dans les modes de connaissance computationnels », ils remodelent non seulement la circulation des connaissances, mais aussi les modèles politiques qui la sous-tendent (Amoore et al., 2024). Il convient donc d'accorder une attention particulière aux conséquences normatives de ces mutations, notamment en ce qui concerne la visibilité, la représentation et l'autonomie des utilisateurs. Si certains lecteurs peuvent influencer les résultats de l'IA grâce à l'ingénierie des invites, d'autres, en particulier les utilisateurs moins compétents sur le plan technique ou moins sensibilisés à la politique, peuvent rester confinés dans les cadres informationnels par défaut de ces outils (Gillespie, 2024, p. 3). Les implications pour le pluralisme des médias sont à ce stade théoriques, mais plausibles à long terme : les petits médias et les journalistes indépendants ont peu de chances d'obtenir un traitement ou une rémunération équitable, ce qui aggrave les inégalités en matière d'exposition et d'influence et fausse davantage la dynamique du marché des médias (voir également la section 3.2.1.).

2.1.3 Point de vue des citoyens sur les risques liés à l'utilisation de l'IA générative dans le journalisme

Outre les avantages potentiels de l'IA générative, son intégration rapide dans les processus journalistiques présente des défis importants qui menacent l'intégrité journalistique et la confiance des citoyens dans l'information (Mattis & de Vreese, 2025). L'IA générative n'est pas simplement un nouvel outil qui vient compléter le travail journalistique. Selon la manière

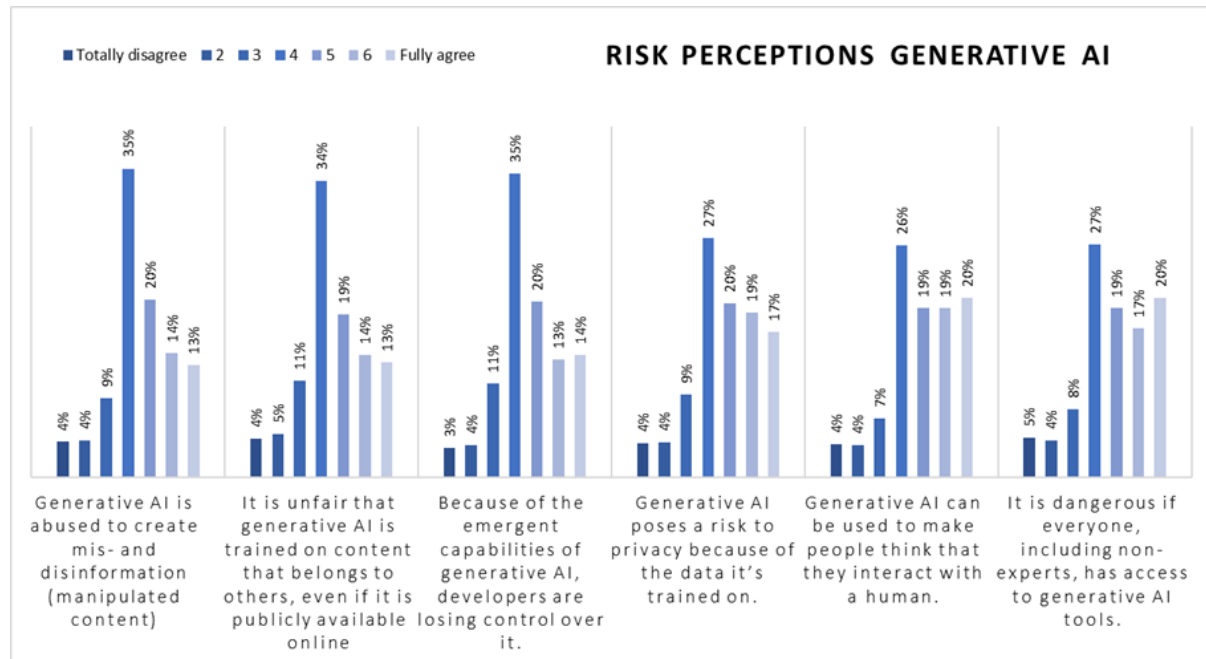
dont elle est utilisée, l' IA générative « intervient dans les processus créatifs fondamentaux du journalisme, remettant en question les normes traditionnelles en matière de paternité, d'originalité et d'identité professionnelle » (Lewis et al., 2025, p. 1). Cela pourrait avoir des implications importantes pour les normes et valeurs journalistiques communes à long terme et pourrait, dans certains cas au moins, nuire à l'intégrité du journalisme, à sa perception par le public et à sa capacité à remplir ses fonctions démocratiques essentielles, telles que la lutte contre la propagation de la désinformation grâce à des reportages précis et de haute qualité (Mattis & de Vreese, 2025).

Et en effet, les citoyens expriment diverses préoccupations et perceptions des risques liés à l'IA générative et à son impact sur leur vie. Les données issues d'un échantillon représentatif de citoyens néerlandais (N = 1545) montrent que les trois principaux risques liés à l'IA générative sont (1) la crainte que l'IA générative puisse être utilisée pour faire croire aux gens qu'ils interagissent avec un être humain ; (2) la perception qu'il est dangereux que tout le monde, y compris les non-experts, ait accès aux outils d' IA générative ; et (3) le fait que l'IA générative présente un risque pour la vie privée en raison des données sur lesquelles elle est entraînée (voir Figure 3). Plus précisément, lorsqu'on les interroge sur les risques liés à l'utilisation de l'IA générative dans le journalisme, les résultats des groupes de discussion avec des citoyens néerlandais indiquent que ces derniers attendent des organes de presse qu'ils fassent preuve de transparence à tout moment quant à leur utilisation de l'IA et souhaitent que ces organes fassent preuve d'un certain degré de responsabilité ou d'imputabilité (Morosoli et al., 2025) . Ces conclusions ne sont pas propres aux Pays-Bas, mais font écho à celles observées, par exemple, dans divers pays africains (Gondwe, 2025) et européens (Kieslich et al., 2024a ; Mitova et al., 2024a ; Monzer et al., 2020). Par exemple, Gondwe (2025) constate que les citoyens du Ghana et du Nigeria considèrent la transparence quant à l'utilisation de l'IA dans la création d'actualités comme un élément crucial pour maintenir la confiance dans les médias. En Europe, Mitova et ses collègues (2024b) concluent que dans quatre pays (Pays-Bas, Pologne, Suisse, Royaume-Uni), le désir de transparence dans le contexte des systèmes de recommandation d'actualités est relativement élevé, d'autant plus que les citoyens sont préoccupés par leur vie privée. À l'aide de scénarios, Kieslich et al. (2024b) ont évalué l'efficacité des obligations de transparence dans le domaine de l'information parmi différentes parties prenantes aux Pays-Bas. Dans l'ensemble, la majorité des répondants parmi les différents groupes de parties prenantes (créateurs de contenu, développeurs de technologies et consommateurs d'actualités) ont accueilli favorablement les obligations de transparence et s'attendaient à un certain niveau d'efficacité, notamment en matière de prévention de la propagation de fausses informations et d'amélioration de la qualité journalistique. Cependant, certains répondants ont exprimé des doutes quant à l'application pratique et à l'impact global des obligations de transparence, soulignant qu'elles pourraient ne pas modifier de manière significative le comportement des consommateurs d'actualités ou des créateurs de contenu.

Compte tenu de ce qui précède, nous observons un décalage entre ce que veulent les citoyens et ce dont ils semblent être informés. Les citoyens ont des attentes fortes en matière de transparence et ont des opinions bien arrêtées sur les risques directement liés à l'IA. Cependant, ils ne semblent pas comprendre les dépendances structurelles plus profondes qui conduiront à des changements institutionnels. Celles-ci restent largement invisibles ou sont considérées comme secondaires. Sans une prise de conscience plus globale de ces

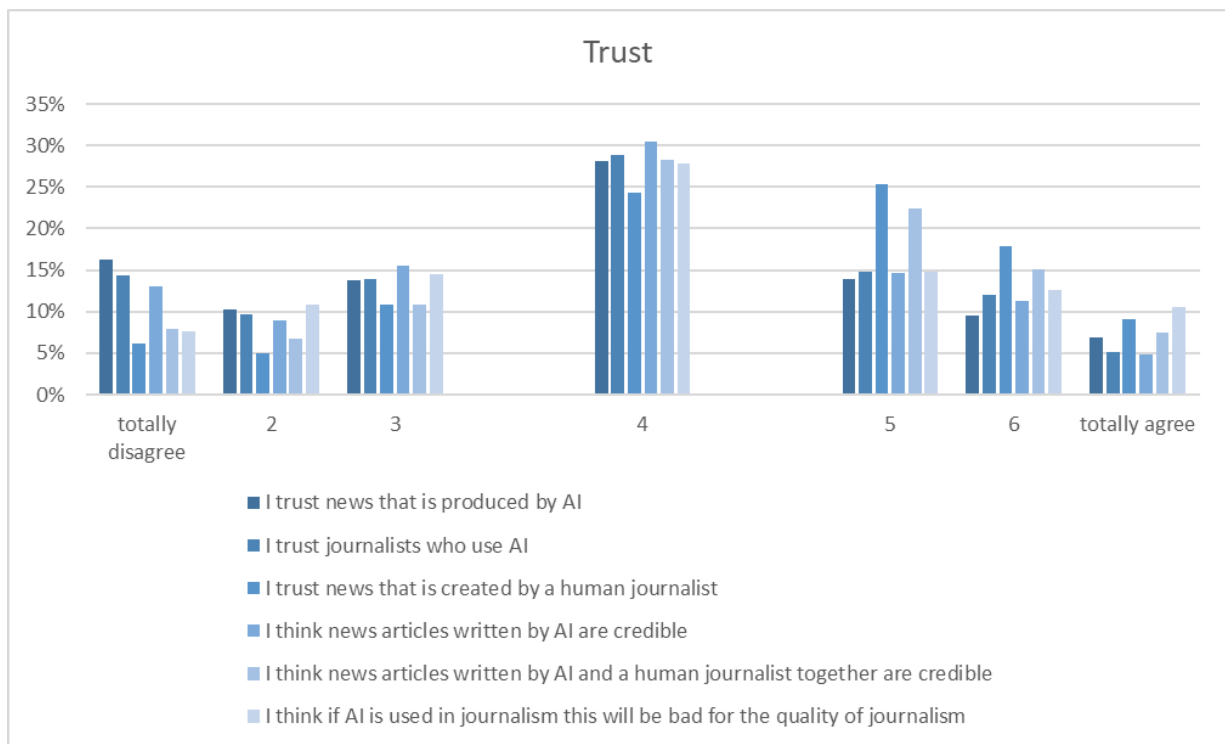
interdépendances, il est difficile pour les décideurs politiques d'élaborer des politiques perçues comme significatives.

Figure 3. Perceptions des risques liés à l'IA générative en %



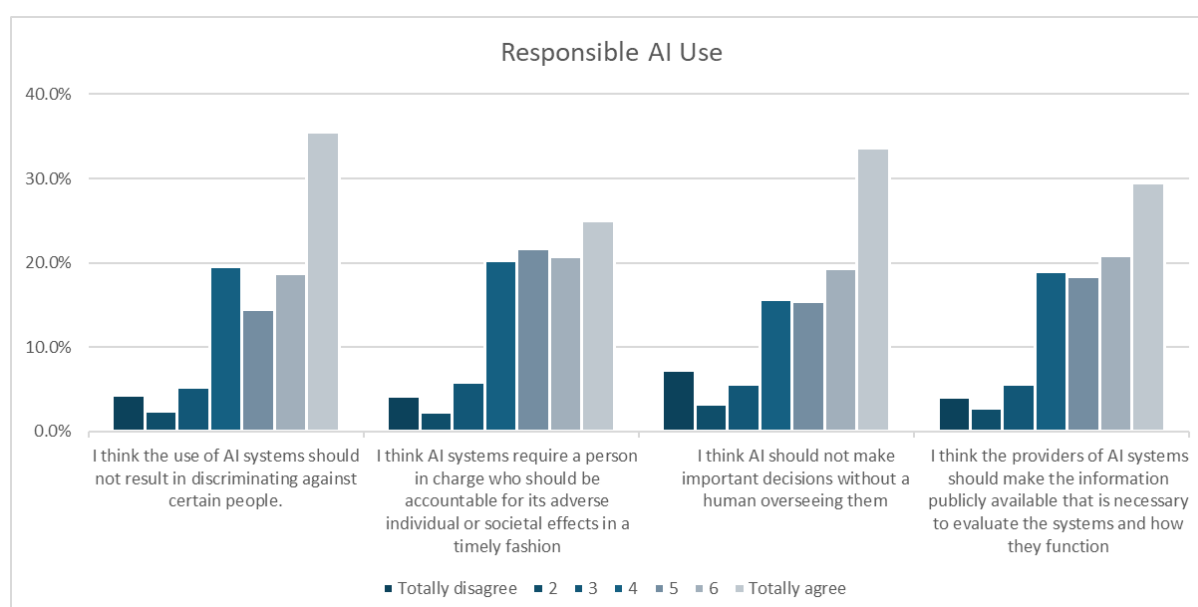
Compte tenu des préoccupations générales concernant l'implication de l'IA générative dans le journalisme, nous avons demandé aux citoyens néerlandais (N = 1 545) interrogés dans le cadre du même sondage que mentionné ci-dessus dans quelle mesure ils faisaient confiance à l'utilisation de l'IA dans la production d'informations. Nous nous sommes particulièrement intéressés au degré de confiance des individus dans les informations générées par l'IA, les informations rédigées par des journalistes humains et les journalistes qui utilisent l'IA dans leur travail de rédaction. Nous leur avons également demandé dans quelle mesure ils jugeaient crédibles les informations produites respectivement. Les résultats révèlent une nette préférence pour les contenus rédigés par des humains par rapport aux contenus générés par l'IA en termes de confiance et de crédibilité (voir Figure 4). La confiance dans les informations générées par l'IA et dans les journalistes utilisant l'IA est relativement faible, tandis que la confiance dans les journalistes humains est plus élevée. Les perceptions de crédibilité suivent un schéma similaire : les articles rédigés par l'IA sont jugés les moins crédibles, suivis par les articles collaboratifs humains-IA, et les articles rédigés par des humains sont jugés les plus crédibles. L'impact perçu de l'IA sur la qualité du journalisme est légèrement supérieur à la neutralité. Ces conclusions ne sont pas propres au cas néerlandais. Dans un rapport récent sur l'IA générative et l'actualité, l'Institut Reuters a constaté que dans six pays (Argentine, Danemark, France, Japon, Royaume-Uni, États-Unis), les citoyens sont les moins à l'aise lorsque les informations sont entièrement générées par l'IA et les plus à l'aise lorsque les informations sont produites par des journalistes humains (Simon et al., 2025).

Figure 4 Niveaux de confiance déclarés en %



Pour (re)construire la confiance sous la forme d'une relation significative entre les citoyens et les médias d'information, une utilisation éthique et responsable de l'IA générative est essentielle. Mais qu'est-ce que les citoyens considèrent comme une utilisation éthique de l'IA ? Nos conclusions dans le contexte néerlandais soulignent qu'une grande majorité s'accorde à dire que l'IA ne devrait pas discriminer certaines personnes, 35,5 % étant tout à fait d'accord ($M = 5,35$), Figure 5. De même, il existe un consensus important sur le fait que les systèmes d'IA doivent être placés sous la responsabilité d'une personne qui doit rendre compte des effets négatifs, et que l'IA ne doit pas prendre de décisions importantes sans supervision humaine, ce qui est également fortement soutenu, avec 33,6 % des personnes interrogées totalement d'accord, ce qui souligne un besoin important de responsabilité. Enfin, il existe un consensus clair sur le fait que les fournisseurs d'IA doivent rendre publiques les informations nécessaires, telles que les données d'entraînement, afin d'évaluer les systèmes, avec 29,5 % des personnes interrogées totalement d'accord. Ces conclusions centrées sur les citoyens concordent avec les perceptions des professionnels des médias. Diakopoulos et al. (2024a) ont interrogé 292 personnes travaillant dans le secteur des médias et ont également constaté que l'absence de supervision humaine, les fausses informations et les préjugés étaient les préoccupations les plus pressantes des professionnels des médias. Dans l'ensemble, ces résultats soulignent l'importance de l'équité, de la transparence et de la responsabilité dans le déploiement des systèmes d'IA auprès des citoyens et des professionnels.

Figure 5. Préoccupations éthiques et utilisation responsable de l'IA



Au vu de l'ensemble des données empiriques, nous préconisons deux mesures pour atténuer l'érosion de la confiance et lutter contre la crainte croissante des fausses informations et de la désinformation générées par l'IA : une transparence significative et une culture de l'IA pour les citoyens et les journalistes. Une transparence significative dans le journalisme basé sur l'IA doit aller au-delà d'une simple mention « IA » (Piasecki et al., 2024). La transparence doit expliquer pourquoi l'IA a été utilisée, ce que le système a fait, quelles données et quels jugements éditoriaux ont façonné le résultat, et quelles incertitudes ou limites subsistent. Outre une transparence utile, des efforts de sensibilisation à l'IA pour les citoyens et les journalistes devraient être développés et mis en œuvre (Deuze & Beckett, 2022 ; Stamboliev, 2023). Ces efforts devraient mettre l'accent sur la capacité des citoyens et des journalistes à remettre en question, à résister et à influencer la manière dont l'IA est conçue et déployée. Il s'agit donc d'aborder les questions d'accès et de relations de pouvoir, et pas seulement d'enseigner l'explicabilité technique (Stamboliev, 2023).

2.1.4 La montée de la désinformation alimentée par l'IA

Une préoccupation centrale est que l'IA générative créera et exposera les citoyens à des informations fausses et trompeuses, qu'ils ne pourront pas reconnaître comme telles, ce qui conduira à une méfiance croissante à l'égard des informations et des processus journalistiques (Morosoli et al., 2025) . Plus précisément, l'essor de l'IA générative a intensifié les inquiétudes quant à son utilisation abusive à des fins de désinformation - informations fausses créées dans l'intention de tromper - ainsi que de mésinformation, c'est-à-dire la diffusion accidentelle de fausses informations. Par exemple, les outils d'IA générative, tels que Sora, peuvent être utilisés pour produire des images représentant des événements qui ne se sont jamais produits, que ce soit sous la forme de deepfakes ou de photographies synthétiques (Weikmann & Lecheler, 2023) (voir également la section 2.2 sur l'utilisation d' IA générative dans les campagnes politiques). Nous constatons ici que le fait d'être trompé par de telles formes de désinformation visuelle peut conduire à une méfiance générale envers les médias visuels, et donc envers tout ce que les citoyens peuvent voir dans les médias

(Weikmann et al., 2025). Pour les pratiques journalistiques, deux préoccupations majeures se posent : (1) la pression accrue pour identifier la désinformation générée par l'IA afin d'éviter de devenir des diffuseurs accidentels de désinformation, et (2) le risque de devenir des producteurs involontaires de désinformation en raison de la dépendance à l'IA pour la recherche ou la création de contenu.

Premièrement, en raison de sa grande similitude avec les informations originales, les médias synthétiques contribuent à une « crise d'authenticité », dans laquelle tant le public que les journalistes ont du mal à faire la distinction entre les contenus créés par des humains et ceux générés par l'IA (Dodds et al., 2025b). Cela soulève la question de savoir si les journalistes ont confiance en leur propre capacité à relever ce défi. Une enquête que nous avons menée auprès d'experts en désinformation de l'Observatoire européen des médias numériques (EDMO) suggère que les journalistes se sentent généralement compétents pour traiter la désinformation à l'ère de l'IA générique. Cependant (voir Figure 6), ils déclarent avoir une confiance nettement inférieure à celle des vérificateurs de faits et un niveau de compétence à peu près équivalent à celui des universitaires qui étudient la question (Weikmann et al. 2026).

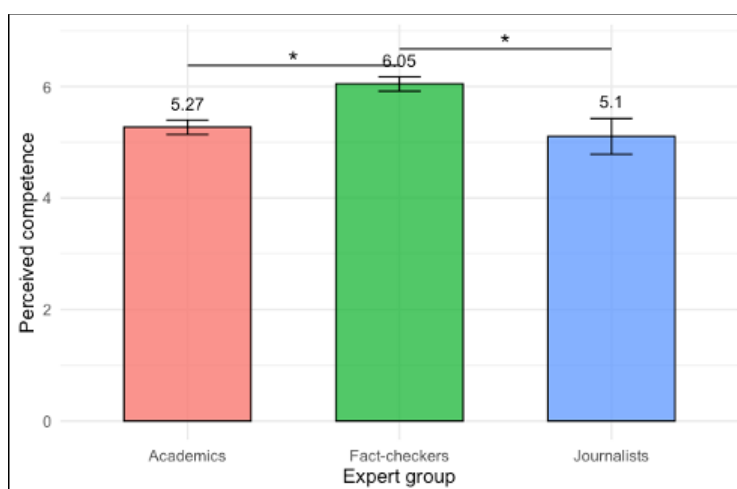


Figure 6. Diagramme à barres représentant la compétence auto-évaluée dans la gestion de la désinformation générée par l'IA parmi les groupes d'experts, y compris les moyennes des groupes sur une échelle de 1 à 7 et les différences significatives (* = $p < .005$).

Deuxièmement, à mesure que les journalistes s'appuient davantage sur l'IA générative dans le processus de production de l'information, les systèmes d'IA peuvent introduire des erreurs, des biais et des « hallucinations » dans le contenu journalistique. Par exemple, un rapport récent de l'Union européenne de radio-télévision (UER, 2025) montre que les assistants IA déforment systématiquement le contenu des informations lorsqu'ils sont utilisés pour la recherche d'informations. Cela peut poser problème non seulement si les citoyens utilisent des assistants IA pour s'informer, mais aussi si les journalistes les utilisent à des fins de recherche. D'une part, les lacunes potentielles des LLM peuvent fournir directement aux utilisateurs des contenus inexacts ou biaisés. D'autre part, la qualité du contenu journalistique peut être affectée s'il est basé sur des contenus erronés provenant de l'IA. Dans le même ordre d'idées, nos recherches montrent que l'utilisation à grande échelle de l'IA générative pour la création de contenu augmente le risque de propagation de fausses informations et de ce qu'on appelle les « AI slop » (contenus de mauvaise qualité produits en masse), qui

peuvent noyer le journalisme crédible (Dodds et al., 2025a). Si rien n'est fait, cela représente une menace directe pour la crédibilité.

2.1.5 Problèmes structurels et démocratiques : dépendance numérique et asymétrie de pouvoir

Les sections précédentes ont démontré qu'avec le déploiement croissant de l'IA générative, une nouvelle transformation numérique du journalisme est en cours. Elles ont également montré qu'à l'heure actuelle, il existe un certain décalage entre ce qui est possible et ce qui est souhaitable du point de vue du public. Si l'éthique professionnelle, la supervision éditoriale et un engagement transparent et honnête envers le public peuvent contribuer à répondre à ces préoccupations, il existe également des limites claires à ce que les organes de presse peuvent faire pour protéger et promouvoir un écosystème d'information sain. Cela s'explique par leur influence limitée sur la manière dont les principaux outils de d'IA générative sont développés, ainsi que par des dépendances structurelles plus profondes et des asymétries de pouvoir.

Ces dépendances structurelles renforcent les déséquilibres de pouvoir, les organes de presse ayant une capacité limitée à influencer, voire à comprendre, les technologies sur lesquelles ils s'appuient (Dodds et al., 2025b). Fonctionnant comme des « boîtes noires », les systèmes d'IA offrent une transparence limitée sur leur fonctionnement interne, leurs biais ou leurs processus décisionnels (Piasecki et al., 2024 ; Simon, 2024b, p. 4). Par conséquent, afin de pouvoir répondre aux questions de confiance et de transparence, les organes de presse s'appuient sur la coopération des fournisseurs de technologies. Dans le même temps, l'IA générative intensifie et renforce encore la concentration du pouvoir technologique, économique et communicationnel (voir également la section 2.1.2). Les organes de presse dépendent de plus en plus d'une poignée de grandes entreprises technologiques, parfois irresponsables, qui non seulement influencent presque toutes les étapes de la chaîne de valeur de l'information, mais stimulent également l'innovation technologique, contrôlent les infrastructures informatiques clés et façonnent les cadres juridiques et éthiques qui régissent le secteur (OSCE, 2025). Ces entreprises ont accumulé des pouvoirs considérables, les médias d'information dépendant d'elles à presque toutes les étapes de la production de l'information, de l'accès et de l'utilisation des outils pertinents à l'infrastructure sous-jacente. Cette dépendance non seulement exacerbe les dépendances existantes et en crée de nouvelles dans la production, mais elle pose également des risques importants pour les droits humains et les valeurs, en particulier la liberté et le pluralisme des médias (Simon, 2024b ; van Druenen, 2025).

Cette dynamique risque d'entraîner une « mainmise sur les infrastructures », les médias d'information devenant incapables de fonctionner de manière durable sans les ressources numériques fournies par des entreprises technologiques externes (Nechushtai, 2018). L'IA générative contribue encore à cette situation, car la dépendance ne se limite plus aux canaux de distribution tels que les plateformes de réseaux sociaux, mais une dépendance fondamentale de nature structurelle, technologique et infrastructurelle apparaît (Braun, 2025). Par exemple, les organes de presse dépendent fortement des services cloud de grandes entreprises technologiques telles qu'Amazon Web Services (AWS) pour héberger leurs données et mener leurs opérations, externalisant ainsi leur infrastructure de base (Braun, 2025). Cela risque d'entraîner des verrous, où les coûts de transfert élevés et l'intégration

technique rendent difficile pour les organes de presse de changer de fournisseur, même si les conditions deviennent défavorables ou ne correspondent plus aux valeurs journalistiques. En conséquence, cet ancrage infrastructurel confère aux entreprises technologiques un pouvoir immense, car elles contrôlent l'architecture sous-jacente sur laquelle repose de plus en plus le journalisme. Bien que le rapport aborde plus en détail la dépendance infrastructurelle dans la section 2.3 ci-après, il est important de noter ici que cette dépendance peut nuire à l'indépendance des médias et compromettre leur liberté et leur pluralisme (van Drunen, 2025).

Ces menaces et ces concentrations de pouvoir reflètent de profonds déséquilibres structurels entre les médias et les géants technologiques. Les obstacles importants au développement technologique en interne confèrent à ces derniers un avantage significatif, tandis que les médias d'information restent limités en termes de données, de puissance de calcul, de capitaux et d'expertise technique (Gerbrandy & Phoa, 2022). Dans ce contexte, la taille et les ressources disponibles d'un organe de presse jouent un rôle crucial dans la détermination de son degré de dépendance, ainsi que de sa capacité à rester indépendant, durable et résilient. Alors que pour les petites rédactions, le recours aux outils des géants de la technologie est souvent une nécessité imposée par des contraintes de ressources, les grands organes de presse adoptent souvent ces outils par choix stratégique, contournant délibérément les alternatives open source. En fait, certaines grandes organisations de presse semblent adopter de plus en plus des caractéristiques et des stratégies de type « plateforme », privilégiant la consolidation des ressources et des médias plutôt que de soutenir une approche plus démocratisée, où les petits médias et les start-ups technologiques ont une chance (Meese & Seipp, 2025).

En conséquence, l'adoption de l'IA (générique) creuse le fossé entre les grands médias disposant de ressources importantes et leurs homologues plus modestes (Eder & Sjøvaag, 2025). Il existe donc un risque tangible que l'IA renforce la position dominante des entreprises médiatiques et technologiques, marginalisant davantage les petits médias et nuisant au pluralisme des médias (Sjøvaag et al., 2024). Ces évolutions entraînent une tendance à la consolidation et favorisent les fusions dans le secteur de l'information afin d'atteindre la taille nécessaire, comme on l'a observé dans les récentes fusions médiatiques. Ces fusions sont souvent présentées comme la seule solution pour éviter la disparition des petits médias et accélérer la formation de « déserts médiatiques » (Blagojev et al., 2023 ; Borchardt, 2022). Si cela peut être vrai dans une certaine mesure, les conséquences à long terme de la consolidation et des fusions croissantes dans le secteur des médias pourraient finalement affaiblir le journalisme dans son ensemble, en particulier le journalisme local et d'investigation, qui jouent un rôle essentiel de garde-fou démocratique.

2.1.6 Conclusion

Il est clair que l'IA générative a un impact sur le journalisme, pilier essentiel des démocraties, tant en termes de production, d'organisation et de contenu que de relation avec le public. En effet, en raison de son importance démocratique, le journalisme constitue une étude de cas essentielle pour évaluer le potentiel de l'IA générative et ses implications démocratiques et sociétales.

D'une manière générale, il est clair qu'au moment de la rédaction de cette étude, la plupart des citoyens considèrent l'IA dans le journalisme avec plus de scepticisme que d'enthousiasme. Il est également clair que les citoyens ont des opinions ambivalentes, avec un décalage entre ce qu'ils veulent et ce qu'ils savent. Les citoyens, par exemple, demandent de la transparence et ont des opinions bien arrêtées sur les risques. Cependant, ils ne semblent pas saisir les dépendances structurelles plus profondes. Celles-ci restent largement invisibles ou sont considérées comme secondaires.

Les opportunités potentielles de l'utilisation de l'IA générative dans le journalisme sont contrebalancées par la méfiance et les inquiétudes relatives du public. Le journalisme a en partie un rôle à jouer pour innover avec le public. De même, les régulateurs ont un rôle à jouer pour garantir la sécurité des technologies, mais dans le même temps, un certain niveau de méfiance et de conscience critique est également essentiel, compte tenu de la crise de confiance plus générale et de la difficulté à distinguer les contenus fiables de ceux qui ne le sont pas.

À un niveau macroéconomique, nous pourrions observer les effets de l'IA générative qui diminuent la confiance dans l'écosystème de l'information en général. C'est un élément à prendre en compte en combinaison avec l'impact de l'IA générative sur d'autres domaines démocratiques, tels que la politique (voir section 2). L'IA générative façonne également de nouveaux modèles économiques axés sur la technologie qui se disputent l'audience et menacent la viabilité économique du journalisme. Cette dynamique rappelle la dépendance du journalisme à l'égard des réseaux sociaux. L'IA générative est donc une technologie qui affecte également le marché journalistique et son modèle économique.

En fin de compte, l'évolution de la production, du contenu et des relations avec le public, ainsi que les nouveaux défis en termes de modèles économiques, de viabilité et de dépendance à l'égard d'entreprises spécifiques, conduisent à une situation qui nécessite des choix politiques et une orientation de l'innovation dans des directions qui favorisent la liberté d'expression et profitent au public. Cependant, tant que les effets à grande échelle sont invisibles, il est peu probable que cette question retienne l'attention politique qu'elle mérite. Il est impératif de relever ces défis de manière rapide et énergique afin d'obtenir des résultats.

2.2 L'IA générative dans les campagnes politiques et les pratiques démocratiques

Grâce à un ensemble croissant d'applications, l'IA générative transforme les campagnes politiques dans le monde entier, comme le montrent les récentes élections en Inde, en Indonésie, en Europe, au Mexique et aux États-Unis (voir Figure 7). Les capacités et la diffusion rapide de l'IA générative ont suscité des débats publics, médiatiques et universitaires sur ses risques et ses opportunités pour les acteurs politiques, les électeurs et les démocraties. Au cours de la super année électorale 2024, l'IA générative n'a peut-être pas été une technologie révolutionnaire qui, à elle seule, a influencé les élections et sapé les processus démocratiques (Kapoor & Narayanan, 2025 ; Simon & Altay, 2025) comme l'avaient prédit des médias tels que le Washington Post (Verma & Zakrzewski, 2024) (voir Figure 7). Cependant, la grande disponibilité de ces outils a fait qu'ils se sont intégrés dans la pratique quotidienne des campagnes, amplifiant principalement les dynamiques existantes et introduisant une nouvelle couche de défis opérationnels et réglementaires.

AI 'resurrects' long dead dictator in murky new era of deepfake electioneering

By Heather Chen, CNN
© 3 min read · Updated 8:20 PM EST, Sun February 11, 2024



An AI-generated depiction of the late dictator Suharto has sparked debate about using AI technology for political gain. CNN/AGG/TA

AI-Generated Video Targets Istanbul Mayor Ahead of Elections

Hamdi Firat Buyuk · Sarajevo · BIRN · January 22, 2024 12:56

A manipulated video of Istanbul's popular opposition mayor praising President Erdogan's party, spread ahead of critical local elections, shows how disinformation is increasingly drawing on the latest technology, such as AI-generated videos.

BBC

AI and deepfakes blur reality in India elections

16 May 2024

Meryl Sebastian
BBC News, Kochi



Fact-checkers say an avatar of Duwaraka was created digitally for a YouTube stream

Trump Posts Fake Video of Himself Flying a 'King Trump' Jet Over Protesters

President Trump shared what appeared to be an A.I.-generated video on social media. It shows Mr. Trump wearing a crown and flying a jet that dumps brown liquid on demonstrators.

Figure 7. Exemples d'utilisation de l'IA à des fins politiques en Indonésie, en Türkiye, en Inde et aux États-Unis. (Pour un aperçu de l'utilisation de l'IA lors des élections de 2024, voir (Elliott, 2025) .

Pour comprendre les implications de l'IA générative dans les campagnes électorales, il faut examiner à la fois la manière dont les acteurs politiques déploient ces technologies et la manière dont les électeurs perçoivent et réagissent au contenu de campagne généré par l'IA. Le potentiel transformateur de l'IA générative s'étend à plusieurs dimensions : elle réduit considérablement les obstacles à la création de contenu, permet une communication stratégique plus sophistiquée, y compris la manipulation visuelle, facilite la personnalisation à grande échelle des messages politiques et élargit les canaux d'information et d'engagement des électeurs. Ces capacités comportent à la fois des opportunités et des risques pour l'intégrité électorale, la participation politique et les processus démocratiques.

Du point de vue des campagnes électorales, cette section examine tout d'abord le déploiement de l'IA dans les campagnes politiques (section 2.2.1) et son potentiel transformateur en matière d'activités opérationnelles (section 2.2.2), de stratégies de communication (2.2.3), de personnalisation (2.2.4) et d'utilisation de l'IA (chatbots) pour l'engagement citoyen et les conseils de vote (2.2.5). Du point de vue des citoyens, nous analysons ensuite comment ceux-ci évaluent l'utilisation de l'IA dans les campagnes politiques (2.2.6) et discutons de nos résultats en tenant compte des défis réglementaires et sociétaux.

2.2.1 Aperçu du déploiement de l'IA générative dans les campagnes politiques à travers le monde et des dépendances technologiques

Le déploiement de l'IA générative dans les campagnes politiques varie considérablement d'un contexte national à l'autre, en fonction de l'interaction de facteurs opérant à trois niveaux

distincts mais interconnectés (Dommett et al., 2024 ; Kruschinski & Votta, à paraître) : Au niveau systémique, les dispositions électorales, les structures des partis et le paysage médiatique d'un pays créent l'environnement stratégique dans lequel les campagnes opèrent et adoptent les technologies d'IA. Les cadres réglementaires, notamment les lois sur la protection des données, les lois sur le financement des campagnes électorales et la législation spécifique à l'IA, établissent des paramètres juridiques variables qui limitent ou permettent le déploiement technologique dans différentes juridictions. Au niveau organisationnel, les décisions individuelles des campagnes électorales concernant la mise en œuvre de l'IA reflètent leurs ressources financières et humaines disponibles, leur sophistication technologique, leurs engagements éthiques et leurs orientations idéologiques. Ces facteurs à plusieurs niveaux interagissent de manière dynamique pour produire des modèles distincts d'adoption de l'IA qui diffèrent d'un pays à l'autre et même entre les campagnes électorales au sein d'un même contexte national.

Les modèles mondiaux d'utilisation de l'IA générative lors des récentes élections montrent comment le contexte façonne l'adoption (voir Elliott, 2025) : aux États-Unis, un environnement réglementaire permissif et des budgets importants ont permis un déploiement précoce et à grande échelle, allant de la génération de messages, d'images et de vidéos aux chatbots, aux appels automatisés et à la sensibilisation des électeurs basée sur les données. L'Inde a utilisé l'IA générative pour gérer la diversité linguistique en créant des supports adaptés à la démographie et des traductions en direct des discours. Au Pakistan, la campagne d'Imran Khan s'est appuyée sur le clonage vocal par IA pour lui permettre de s'adresser à ses partisans depuis sa détention. Dans les systèmes multipartites plus réglementés de l'UE, l'adoption a été plus prudente et expérimentale, bien que des acteurs idéologiques marginaux aient utilisé l'IA pour mener des attaques satiriques ou négatives contre leurs adversaires. Parallèlement, les opérations d'influence étrangères menées par la Russie et la Chine ont utilisé l'IA générative pour produire des contenus sociaux ciblés, des sites d'information parodiques et inonder les discussions de commentaires synthétiques (OpenAI, 2024).

Les facteurs à plusieurs niveaux (système, réglementaire et organisationnel) qui déterminent l'adoption de l'IA générative créent des dépendances à l'égard de la technologie IA qui varient de manière inverse. Les parties et les campagnes disposant de ressources limitées, en particulier dans les systèmes où les limites de dépenses sont strictes, considèrent les outils IA « prêts à l'emploi » accessibles comme un substitut rentable à la main-d'œuvre humaine. Les campagnes bien financées et dotées d'une infrastructure technique établie peuvent quant à elles adopter l'IA générative de manière plus sélective et l'intégrer dans leurs flux de travail existants. Ces effets liés aux ressources sont limités par la réglementation, car des cadres plus stricts en matière de protection des données et de gouvernance de l'IA (par exemple dans l'UE) limitent la manière dont les partis peuvent exploiter les outils d'IA, quel que soit leur budget, tandis que des environnements plus permissifs élargissent l'espace d'adoption tant pour les acteurs disposant de ressources importantes que pour ceux qui en sont dépourvus. Les campagnes politiques se rapprochent en même temps du type de dépendance infrastructurelle observé dans d'autres domaines (voir la section 2.3. pour une discussion plus large sur les infrastructures, et 2.1.5 pour la version structurelle de ce problème dans le journalisme). Les partis considéraient autrefois les technologies numériques comme une infrastructure cyclique qui était principalement activée pendant les périodes électorales. L'essor des campagnes permanentes et la professionnalisation du personnel de campagne brouillent de plus en plus cette distinction (Dommett et al., 2021). Une fois que l'IA

est adoptée pour compenser des ressources limitées, ses gains de productivité peuvent rendre les méthodes traditionnelles économiquement non viables. Cela crée une dynamique d'escalade technologique. Si certains acteurs démontrent que l'IA générique améliore la production de contenu et l'engagement des électeurs, d'autres acteurs sont soumis à une pression concurrentielle pour adopter des outils similaires ou accepter un désavantage stratégique. Sous cette pression, les besoins d'efficacité et de visibilité, y compris la production de contenu synthétique optimisé pour les algorithmes des plateformes, peuvent commencer à l'emporter sur toute considération éthique. La dépendance résulte alors moins des préférences des partis que des possibilités, des coûts et des règles des plateformes qui déterminent ce qui est techniquement viable et visible.

La sous-section suivante montre comment ces modèles dépendants des infrastructures et des ressources se traduisent en tâches de campagne, stratégies de communication et applications d'engagement des électeurs concrètes soutenues par l'IA générative.

2.2.2 Activités de campagne assistées par l'IA

Les technologies l'IA générative peuvent aider les campagnes dans leurs activités opérationnelles (Kruschinski et al., 2025). Les modèles linguistiques tels que ChatGPT, Claude et Gemini peuvent prendre en charge des tâches répétitives, notamment des travaux de recherche, des services de traduction, la synthèse de contenu et l'édition de texte. Ils facilitent également l'élaboration de messages destinés à sensibiliser les électeurs via les canaux médiatiques traditionnels et les plateformes numériques. Ainsi, l'IA générative peut améliorer la productivité des campagnes, ce qui est particulièrement avantageux pour les organisations et les candidats disposant de budgets limités, et peut faciliter l'entrée de nouveaux mouvements politiques dans la sphère publique (Dommett et al., 2024 ; Foos, 2024). Le développement de contenu et la production multimédia représentent des investissements financiers et temporels importants dans les opérations de campagne, qui nécessitent généralement de faire appel à des spécialistes sous contrat offrant des services coûteux. Compte tenu des contraintes de ressources auxquelles sont confrontées la plupart des campagnes, les outils d'IA permettent aux campagnes de déployer leurs ressources humaines, leurs finances et leur temps de manière plus efficace. Cette amélioration de la productivité peut contribuer à uniformiser les règles du jeu dans la compétition électorale en permettant aux acteurs disposant de moins de ressources d'améliorer la sophistication de leur campagne et la qualité de leurs messages (Kruschinski et al., 2025).

L'analyse empirique de notre projet CampAlign Tracker a examiné près de 70 000 publications et publicités sur Facebook et Instagram publiées par plus de 3 000 acteurs politiques, y compris tous les partis majeurs et mineurs et leurs candidats respectifs en Allemagne au cours des quatre semaines précédant le jour du scrutin (Kruschinski & Votta, 2025) (campaigntracker.de). Aux Pays-Bas, nous avons suivi plus de 3 000 comptes appartenant à des partis politiques et des candidats néerlandais, ainsi que des comptes tiers (par exemple, des commentateurs, des influenceurs, des pages de fans, des pages de mêmes) qui ont publié 44 000 contenus sur Facebook, Instagram, TikTok et X, au moment de la rédaction du présent rapport (Votta & Kruschinski, 2025) (campaigntracker.nl). À l'aide d'une méthode de détection semi-automatisée combinant des systèmes commerciaux de détection par IA et une validation humaine qualifiée, nous avons constaté que 1,4 % des contenus publiés lors des élections allemandes comprenaient des contenus visuels générés par IA, avec des taux

similaires aux Pays-Bas (voir Figure 8). Cependant, ce taux d'adoption modeste masque des variations importantes entre les différents types d'acteurs politiques et la portée atteinte par ces visuels générés par IA.

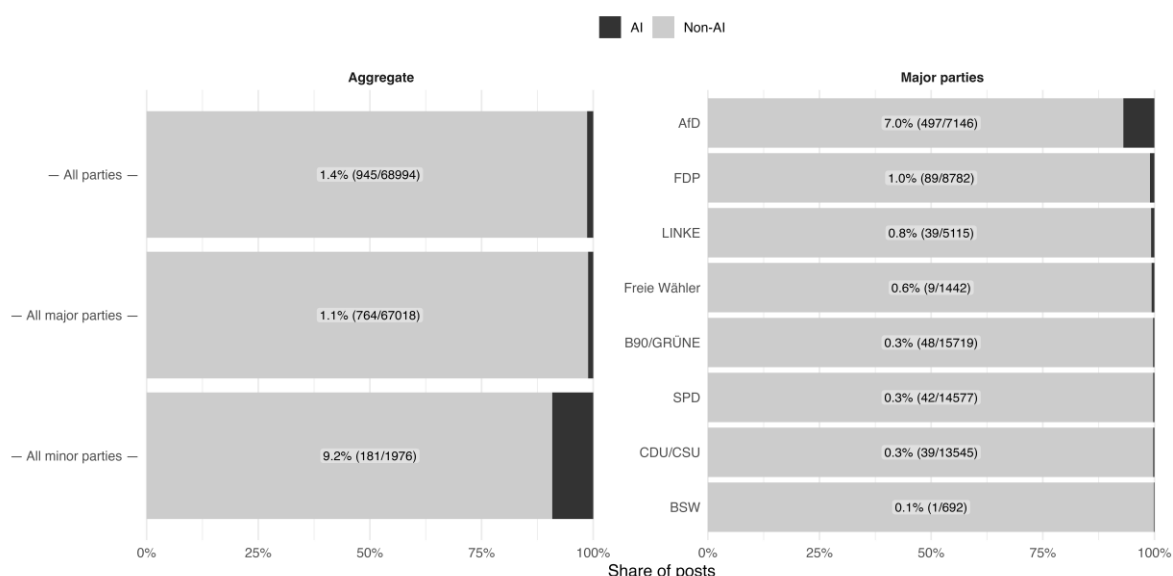


Figure 8. Part des publications visuelles générées par l'IA générative dans la campagne électorale allemande de 2025. Les étiquettes à l'intérieur des barres indiquent le pourcentage de publications générées par l'IA générative et, entre parenthèses, le nombre de publications générées par l'IA générative par rapport au nombre total de publications pour ce groupe. Les petits partis ne sont affichés que dans l'agrégat (Kruschinski & Votta, à paraître) .

Deux modèles distincts d'adoption sont apparus en Allemagne et aux Pays-Bas (voir la Figure 8 pour l'Allemagne). Les petits partis disposant de ressources limitées ont affiché des taux d'utilisation de l'IA plus élevés, probablement parce qu'ils ont tiré parti de cette technologie pour surmonter leurs contraintes en matière de ressources et produire du contenu plus efficacement (Kruschinski & Votta, à paraître ; Votta & Kruschinski, 2025). En outre, les partis soumis à moins de contraintes normatives en matière de déploiement de l'IA, souvent positionnés à l'extrême droite de l'échiquier politique (comme l'AfD en Allemagne ou le PVV aux Pays-Bas), ont affiché des taux d'adoption élevés. Pour ces acteurs, le contenu généré par l'IA servait parfois à provoquer et à choquer le public, testant les limites du discours politique acceptable. En revanche, les partis bien établis et disposant de ressources importantes fonctionnent selon des procédures décisionnelles formalisées et une culture organisationnelle peu encline à prendre des risques, intégrant généralement les nouvelles technologies dans leurs routines existantes et faisant face à des risques plus importants pour leur réputation en cas d'utilisation abusive perçue de technologies controversées. À l'inverse, les partis moins institutionnalisés possèdent une flexibilité organisationnelle et sont soumis à moins de contraintes traditionnelles.

De plus, l'utilisation de l'IA pour générer du contenu politique à grande échelle comporte un risque important de saturation et de pollution de l'information, car même des acteurs individuels de mauvaise foi peuvent désormais produire des volumes de contenu qui auraient été irréalistes auparavant avec des ressources humaines limitées (Simon & Altay, 2025). Le problème crucial ici ne réside pas dans la qualité ou la véracité des textes, images ou vidéos

générés par l'IA, mais plutôt dans leur facilité de production. Cette capacité à produire du contenu à un coût marginal très faible a été qualifiée d'« AI slop » et entrave de plus en plus la capacité des citoyens à distinguer les informations pertinentes du bruit, un problème aggravé par le fait que de nombreuses plateformes et campagnes ne divulguent pas le contenu généré par l'IA en tant que tel (Kruschinski & Votta, à paraître). Seuls 13 % des contenus générés par l'IA comportaient une forme quelconque d'étiquette de divulgation lors des élections fédérales allemandes de 2025. Lorsque cette mention était présente, il s'agissait principalement d'une divulgation volontaire par les partis eux-mêmes plutôt que d'une détection et d'un signalement initiés par la plateforme. Des tendances similaires mais aussi divergentes sont observées dans la campagne électorale néerlandaise : la moitié ou plus du contenu généré par l'IA reste sans mention, mais désormais, la majorité du contenu néerlandais généré par l'IA, s'il est mentionné, l'est par les plateformes plutôt que par les acteurs des partis. Cela marque un changement significatif dans le comportement des plateformes entre les élections fédérales allemandes de février 2025 et les élections parlementaires néerlandaises d'octobre 2025. En conséquence, la communication politique de fond risque d'être noyée, la qualité du discours public peut se détériorer et les électeurs peuvent avoir du mal à s'y retrouver dans leurs processus de décision politique (Jungherr, 2023). En effet, lors de la récente campagne électorale néerlandaise, deux membres du Parlement néerlandais ont systématiquement déployé de grandes quantités de matériel généré par l'IA, y compris des images discriminatoires ciblant les minorités et les femmes via un compte anonyme sur les réseaux sociaux, atteignant ainsi des niveaux d'engagement supérieurs à ceux des politiciens nationaux établis travaillant avec des équipes de campagne complètes (Hofman & Veerbeek, 2024 ; Votta & Kruschinski, 2025). Cela représente le revers de la médaille des avantages potentiels en matière d'égalisation : les outils qui réduisent les obstacles pour les petites campagnes les réduisent également pour les acteurs malveillants qui cherchent à nuire à l'environnement informationnel.

2.2.3 Stratégies de communication assistées par l'IA

Deuxièmement, l'IA offre davantage de possibilités en matière de stratégies de communication (Jungherr, 2023). Les systèmes d'IA visuelle tels que Midjourney, DALL-E, Nano Banana et Sora permettent aux campagnes de générer des images et des contenus vidéo pour diverses applications, allant des affiches de campagne et des publicités télévisées aux sites web et aux flux des réseaux sociaux. Ces outils offrent une grande flexibilité dans l'approche visuelle, allant du réalisme documentaire à l'interprétation artistique, et d'une présentation neutre à des messages suscitant une résonance émotionnelle (Sanseviero et al., 2025). Cette polyvalence permet aux campagnes de présenter les enjeux et les personnalités de manière stratégique en fonction de leurs objectifs de communication, tels que la persuasion, les campagnes négatives, les messages incitant à voter ou les sollicitations de dons (voir Figure 9).



Figure 9. Exemples de différents types d'utilisation de l'IA politique dans la campagne électorale allemande de 2025.

L'IA générative permet également l'utilisation malveillante d'images et de vidéos photoréalistes générées par l'IA, qui peuvent nuire à la fois à l'exactitude des faits et à la confiance politique (Simon & Altay, 2025 ; Weikmann et al., 2024 ; voir également la section 2.1.4). Comme les techniques de deepfake reproduisent désormais les visages, les voix et les gestes avec une grande fidélité, elles créent une « crise d'authenticité » similaire à celle décrite précédemment pour le journalisme dans la section 2.1.4 : dans le contexte électoral, il devient plus difficile de déterminer si un clip apparemment réel est en fait synthétique. Cette incertitude peut être exploitée par des acteurs malveillants pour créer des controverses ou attribuer des déclarations incendiaires à des politiciens, avec des risques en aval tels que des violations des droits de la personnalité et de l'image, la diffusion intentionnelle de fausses informations politiques et l'atteinte à la réputation (Laude & Daum, 2024 ; Weikmann & Lecheler, 2023). Ces capacités sont désormais plus accessibles que jamais, avec une sophistication toujours croissante. Nous avons vu, par exemple, lors de l'élection présidentielle irlandaise, comment des vidéos générées par l'IA, imitant de manière convaincante des diffuseurs et des candidats politiques de confiance, peuvent circuler suffisamment longtemps pour induire les électeurs en erreur et contraindre les plateformes et les régulateurs à réagir (Ryan, 2025). Sur la base de nos recherches CampAlign Tracker, l'analyse du contenu visuel généré par l'IA utilisé en Allemagne et aux Pays-Bas a montré que la grande majorité de celui-ci reposait sur une esthétique photoréaliste plutôt que sur des approches clairement stylisées ou illustratives (voir figure 10 ; Kruschinski & Votta, à paraître ; Votta & Kruschinski, 2025). Cette préférence pour la représentation réaliste suggère des efforts stratégiques visant à exploiter le poids probatoire traditionnellement associé à la photographie, bien que cela augmente simultanément les risques de tromperie et d'érosion de la confiance dans la communication politique visuelle.

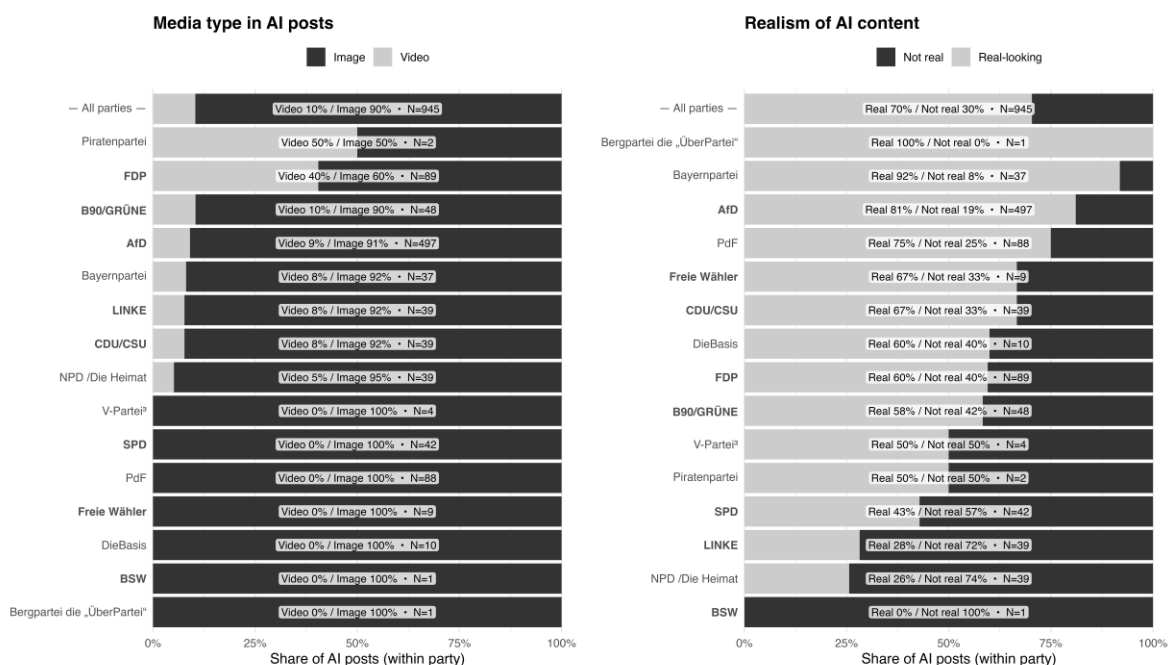


Figure 10. Type de média et réalisme des publications visuelles générées par l'IA dans la campagne électorale allemande de 2025. Les étiquettes à l'intérieur des barres indiquent, pour chaque parti, la part des publications générées par l'IA qui étaient des vidéos par rapport aux images (graphique de gauche) ou qui semblaient réelles par rapport à celles qui ne semblaient pas réelles (graphique de droite), suivies du nombre total de publications générées par l'IA pour ce parti (Kruschinski & Votta, à paraître).

Cependant, le contenu photoréaliste représentait principalement des citoyens ordinaires, des partisans et des membres génériques du public plutôt que des élites politiques, tant en Allemagne qu'aux Pays-Bas. Lorsque des deepfakes de personnalités politiques connues sont apparus, ils représentaient généralement les principaux candidats en Allemagne, notamment Friedrich Merz (CDU/CSU), Alice Weidel (AfD), et aux Pays-Bas, Geert Wilders (PVV) et Frans Timmermans (GroenLinks-PvdA). Un exemple marquant de deepfake de Frans Timmermans le montrait en train d'être arrêté, ce qui a suscité une série de menaces de mort à son encontre et des plaintes pénales contre les auteurs (Feenstra & Sabel, 2025).

Nos conclusions empiriques fournissent des preuves supplémentaires de la manière dont ces dynamiques se manifestent dans la pratique. Les acteurs de droite et d'extrême droite sont apparus comme les principaux producteurs de contenus négatifs générés par l'IA en Allemagne et aux Pays-Bas (Kruschinski & Votta, à paraître ; Votta & Kruschinski, 2025). Ils ont fréquemment utilisé des images générées par l'IA pour critiquer les partis rivaux et les groupes sociaux, y compris les minorités, souvent qualifiés de criminels. Il s'agit là d'une utilisation préoccupante de la technologie à des fins de communication exclusionnaire qui renforce les stéréotypes nuisibles. Les grands partis ont fait preuve d'une plus grande retenue et, lorsqu'ils ont utilisé l'IA à des fins de critique, ils se sont concentrés sur les différences politiques et la responsabilité politique plutôt que sur les attaques personnelles ou la stigmatisation. En outre, l'analyse des modèles d'engagement des contenus générés par l'IA révèle une tendance inquiétante : bien que les visuels générés par l'IA ne constituent qu'une petite fraction du contenu total, ils ont généralement obtenu un engagement médian plus élevé que les publications non générées par l'IA en Allemagne (Kruschinski & Votta, à paraître).

De plus, les visuels générés par l'IA permettent une stratégie de communication plus insidieuse : la simple existence de la technologie deepfake offre une échappatoire pratique aux acteurs politiques pris dans de véritables scandales, qui peuvent désormais prétendre que les enregistrements, images ou vidéos authentiques documentant leur mauvaise conduite sont des fabrications générées par l'IA, exploitant ainsi l'incertitude du public quant à ce qui est réel pour échapper à leur responsabilité. Ce phénomène, connu sous le nom de « *dividende du menteur* », signifie que ceux qui commettent des actes répréhensibles bénéficient paradoxalement de la prolifération des médias synthétiques (Schiff et al., 2025) . De plus, lorsque les citoyens perdent la capacité de différencier de manière fiable les contenus authentiques des contenus fabriqués, un scepticisme généralisé à l'égard des informations politiques et des personnalités publiques peut s'installer (Jungherr, 2023) Weikmann et al., 2024).

2.2.4 Personnalisation des messages politiques assistée par l'IA

L'IA générique permet une personnalisation à plus grande échelle des messages politiques. En traitant des volumes importants d'informations, notamment l'engagement sur les réseaux sociaux, la correspondance avec les électeurs et les données démographiques, les campagnes peuvent identifier l'opinion publique et les priorités des électeurs. Cette capacité d'analyse permet d'élaborer des communications personnalisées qui répondent aux préoccupations, aux préférences et aux positions idéologiques spécifiques des électeurs. Ces messages sur mesure peuvent renforcer la résonance de la communication politique et stimuler l'engagement civique (Foos, 2024). En outre, les fonctionnalités de traduction et d'accessibilité basées sur l'IA peuvent permettre aux campagnes électorales de communiquer plus efficacement avec les communautés marginalisées et les citoyens issus de milieux linguistiques divers. Par exemple, tant dans les médias que dans les campagnes politiques, des outils génératifs pourraient être utilisés pour traduire des contenus afin de tenir compte de la diversité des capacités et des besoins cognitifs et physiques des personnes, jeunes et adultes. Les technologies de reconnaissance vocale, par exemple, pourraient aider à transcrire les programmes audiovisuels pour les personnes malentendantes. En effet, les outils d'IA sont déjà utilisés pour sensibiliser les électeurs multilingues en Inde et aux États-Unis (Bhattacharya, 2024).

Cependant, le revers de la médaille de la sensibilisation adaptative des électeurs est que les mêmes techniques utilisées pour élargir l'inclusion peuvent également servir à diffuser des messages toujours plus granulaires, opaques et chargés d'émotion à des publics étroitement définis, ouvrant la voie à une personnalisation à grande échelle qui brouille la frontière entre persuasion et manipulation. La personnalisation à grande échelle des messages politiques grâce à l'IA introduit des risques de fragmentation sociale, de comportements coordonnés inauthentiques pendant les élections et incite davantage à diffuser des contenus polarisants (Simon & Altay, 2025). Lorsque les systèmes d'IA fournissent de plus en plus d'informations qui renforcent les points de vue préexistants des électeurs, la possibilité d'un dialogue constructif entre les différentes perspectives diminue. Cette dynamique est particulièrement prononcée sur les plateformes de réseaux sociaux, où les systèmes algorithmiques peuvent diriger des messages personnalisés générés par l'IA vers des publics prédisposés à les accepter. Un tel ciblage peut amplifier les points de vue extrêmes et approfondir la polarisation politique (Jungherr, 2023). La capacité de l'IA générative à produire du contenu à grande échelle représente également une menace pour les représentants politiques. Des expériences

sur le terrain ont démontré que les lettres de plaidoyer générées par l'IA étaient presque impossibles à distinguer de la correspondance écrite par des humains, ce qui suggère que des acteurs malveillants pourraient fausser la compréhension qu'ont les législateurs des priorités de leurs électeurs (Kreps & Kriner, 2023). Au-delà de leur influence directe sur les responsables politiques, les systèmes d'IA peuvent façonner les résultats démocratiques en influençant les électeurs grâce à des messages personnalisés et au microciblage (Zuiderveen Borgesius et al., 2018). La convergence entre la génération de contenu à faible coût et la personnalisation basée sur l'IA crée ce que les chercheurs qualifient de « machine à manipuler » potentielle (Simchon et al., 2024), soulevant des questions fondamentales sur le choix démocratique éclairé et l'intégrité électorale lorsque la production à grande échelle basée sur l'IA est combinée à des systèmes de distribution ciblés et algorithmiques qui privilégient les contenus personnalisés pour les publics qui y sont plus sensibles.

2.2.5 L'IA générative pour la recherche d'informations politiques et l'engagement de la société civile

Les systèmes d'IA générative offrent également de nouvelles voies pour l'information des électeurs et l'engagement de la société civile. Alors qu'ils étaient initialement considérés comme indisponibles dans les grands modèles en 2024 (mais voir (Nieuwsuur, 2024), en 2025, ces garde-fous, s'ils ont jamais existé, ont disparu et sont utilisés comme alternatives aux applications de conseil en matière de vote (VAAs) plus établies. Cette évolution présente à la fois des opportunités et des défis. D'une part, les chatbots et les LLM peuvent fournir des services précieux tels que la synthèse de questions politiques complexes, la synthèse des réponses aux consultations et l'accessibilité des informations politiques à différents niveaux de littératie et de profondeur. Cette capacité pourrait idéalement améliorer la réactivité et aider les citoyens à s'engager plus efficacement dans le contenu des politiques. Plusieurs applications de chatbots ont également été testées afin de réduire les obstacles à la connaissance politique en permettant aux citoyens de « discuter » directement avec un manifeste de parti, de poser des questions à des clones numériques de candidats ou de débattre avec un partisan fictif d'un parti d'opposition (Argyle et al., 2023 ; Schiele et al., 2024), voire de réduire les croyances conspirationnistes grâce à des conversations (Costello et al., 2024).

D'autre part, ces chatbots comportent des risques importants, notamment l'attribution erronée des points de vue des partis, la diffusion de fausses informations et l'introduction de bruit dans l'environnement informationnel électoral. Des tests empiriques réalisés par l'Autorité néerlandaise de protection des données (AP) ont révélé un biais systématique dans les recommandations de vote des chatbots. Sur quatre chatbots populaires, plus de 56 % des recommandations favorisaient seulement deux partis (PVV ou GroenLinks-PvdA), un chatbot recommandant ces partis dans plus de 80 % des cas, indépendamment des informations fournies par l'utilisateur. Parallèlement, les autres partis étaient rarement recommandés, même lorsque les positions des utilisateurs correspondaient parfaitement à leurs programmes (Autoriteit Persoonsgegevens, 2025). L'ampleur de l'exposition potentielle à ces risques est préoccupante : une enquête menée auprès de citoyens néerlandais a révélé qu'environ 1 électeur sur 10 déclarait être susceptible de demander des conseils de vote à des chatbots IA, les chiffres les plus élevés étant enregistrés chez les 18-34 ans (17 %) et les 55 ans et plus (6 %) (de León et al., 2025). Pour l'avenir, cela pourrait montrer comment les interfaces

de chat IA pourraient devenir les gardiens de facto de l'information électorale, à l'instar des recherches et des synthèses d'actualités basées sur l'IA (voir également la section 2.1.2).

Au-delà des applications de conseil en matière de vote, l'IA générative offre un potentiel pour la participation citoyenne et l'engagement de la société civile. Les outils d'IA générative peuvent synthétiser des milliers de contributions citoyennes dans le cadre de consultations publiques, rendant la gouvernance participative plus évolutive et accessible (OCDE, 2025). Par exemple, les plateformes de démocratie délibérative peuvent utiliser l'IA pour traiter les contributions des citoyens sur les propositions politiques, identifier les thèmes communs à divers points de vue et faciliter un dialogue structuré entre les participants ayant des perspectives différentes (McKinney, 2024). Ces applications se sont révélées prometteuses pour réduire la polarisation en aidant les citoyens à trouver un terrain d'entente grâce à des discussions médiatisées par l'IA qui mettent en évidence les valeurs communes plutôt que les positions divergentes (Argyle et al., 2023 ; Tsai et al., 2024). Les outils assistés par l'IA permettent aux petites organisations de participer plus efficacement aux débats politiques auparavant dominés par des acteurs disposant de ressources importantes, ce qui pourrait démocratiser l'engagement civique (Bonin et al., 2025 ; Guler et al., 2024). Cependant, comme indiqué dans la section 3.1.1, dans certains cas, le recours à l'IA générative peut appauvrir la diversité et étouffer les voix minoritaires au profit des majorités dominantes ou des discours populaires, y compris ceux qui sont polarisants (voir également : (Lazar & Manuali, 2026). De plus, l'acceptation par le public reste un défi important. Des données expérimentales provenant d'Allemagne révèlent une nette « pénalité IA » : les citoyens se montrent nettement moins disposés à participer à des délibérations facilitées par l'IA qu'à des formats facilités par des humains, la participation chutant de 7,4 points de pourcentage et les notes de qualité perçue baissant de près de 10 points de pourcentage (Jungherr & Rauchfleisch, 2025). Ce scepticisme est particulièrement prononcé chez ceux qui perçoivent des risques plus importants liés à l'IA et des avantages sociétaux moindres, ce qui pourrait créer une nouvelle « fracture délibérative » qui transcende les écarts de participation traditionnels. Enfin, pour les citoyens et les acteurs de la société civile, l'IA générative pourrait contribuer à réduire les obstacles à la communication politique au sein de structures participatives (en ligne), telles que les mécanismes de consultation publique, visant à réduire la distance entre les électeurs et le législateur et à rendre la prise de décision plus démocratique. Cependant, même si l'égalité d'accès à la technologie peut être garantie, il n'est pas certain que, à terme, l'utilisation de l'IA générative dans ces contextes transformera ou, au contraire, aggraverait les inégalités politiques (existantes) (Danaher, 2024). De plus, comme mentionné précédemment, l'IA générative peut être exploitée par des acteurs malveillants pour « générer à grande échelle de faux « sentiments des électeurs » en créant sans effort des messages uniques prenant position sur n'importe quel aspect d'une multitude de questions (Kreps & Kriner, 2023, p. 123).» Il est donc plus difficile pour les institutions publiques de distinguer quels commentaires publics, qu'ils soient coproduits par l'IA ou non, « représentent véritablement les préférences des parties prenantes concernées (Kreps & Kriner, 2023, p. 125) ».

En résumé, à mesure que l'adoption de l'IA se généralise et que les partis reconnaissent les avantages potentiels de son utilisation, le discours démocratique risque de voir les contenus synthétiques remplacer de plus en plus la communication politique de fond, les partis se livrant concurrence non pas par le biais d'arguments politiques, mais par le biais d'appels émotionnels générés par l'IA et d'une négativité stratégique optimisée pour une diffusion virale plutôt que pour une délibération éclairée. Compte tenu de ces opportunités et de ces défis,

les acteurs politiques doivent se demander de manière critique dans quelle mesure l'utilisation stratégique de l'IA générative dans leurs campagnes est justifiable.

Les cas d'utilisation positifs mentionnés tout au long de cette section ne peuvent révéler tout leur potentiel que si les risques liés à la communication politique sont suffisamment pris en compte. En effet, il serait contraire à l'objectif de promotion de l'inclusion et de l'autonomisation des citoyens par le biais de l'IA générative que ces applications ne fassent que reproduire les préjugés sociétaux. Le potentiel démocratisant de l'IA générative s'étend au-delà des campagnes électorales pour toucher la participation démocratique quotidienne. Lorsqu'ils sont déployés de manière responsable, ces outils peuvent réduire les obstacles à l'engagement civique en aidant les citoyens à naviguer dans des documents politiques complexes, à participer à des consultations indépendamment des barrières linguistiques ou éducatives, et à organiser plus efficacement des actions collectives. Il convient toutefois de comprendre que les scénarios susmentionnés ne s'attaquent pas aux inégalités sociales et politiques structurelles sous-jacentes à leur origine et ne doivent donc pas être interprétés comme positionnant l'IA générative comme la solution aux inégalités politiques. Pour accroître véritablement la visibilité politique des communautés marginalisées ou vulnérables, il convient de renforcer la position de la société civile, car les communautés civiles offrent un espace pour exprimer des expériences et des perceptions qui ne font pas partie du discours (politique) dominant et peuvent aider à comprendre et à dénoncer diverses formes d'injustice dans le pouvoir étatique et économique (Young, 2002, p. 211). De plus, comme tout le monde peut utiliser ces outils, y compris les acteurs puissants, les avantages qu'ils offrent aux petits partis et aux mouvements populaires pourraient bientôt s'estomper si leur utilisation générale comme outil de communication devient plus courante.

2.2.6 Attitudes des citoyens à l'égard de l'utilisation de l'IA générative dans les campagnes politiques

La perception et l'évaluation par le public de l'IA générative dans les campagnes politiques influencent à la fois les options stratégiques dont disposent les acteurs politiques et la conception des cadres réglementaires. Cependant, une grande incertitude subsiste quant à l'attitude des citoyens à l'égard de l'utilisation de l'IA dans les élections.

Les enquêtes auprès de la population révèlent que si le terme « IA générative » est largement reconnu, la plupart des gens n'ont pas une compréhension approfondie de cette technologie (Simon et al., 2025), et ce manque de connaissances s'étend à son application dans les campagnes politiques (Kruschinski et al., 2025). En Allemagne, la plupart des personnes interrogées n'ont pas consciemment rencontré de contenu généré par l'IA dans la publicité électorale et restent incertaines quant à savoir si, et comment, les partis utilisent ces outils. Cela suggère que le débat public a devancé la prise de conscience réelle. Des enquêtes transnationales montrent que la perception des risques domine les attitudes à l'égard de l'IA dans les applications politiques (voir figure 11). À l'instar des tendances observées dans le journalisme (voir section 2.1 du rapport), où beaucoup expriment leur scepticisme face aux informations générées par l'IA et craignent une atteinte à la fiabilité et à la transparence (Simon et al., 2025), les électeurs allemands associent les campagnes utilisant l'IA à une baisse de la crédibilité des messages politiques, à la manipulation des électeurs et à des menaces pour la démocratie (voir figure 12 ; (Kruschinski et al., 2025). Tout d'abord,

l'acceptation varie selon l'usage : les fonctions d'assistance technique telles que la correction orthographique ou la traduction, sont largement approuvées. En revanche plus le rôle de l'IA s'étend à la création de contenu, en particulier dans la communication avec les électeurs, plus la résistance s'accroît. L'utilisation de l'IA pour générer des communiqués de presse, des messages sur les réseaux sociaux et surtout des images et des vidéos réalistes suscite les résistances les plus fortes.

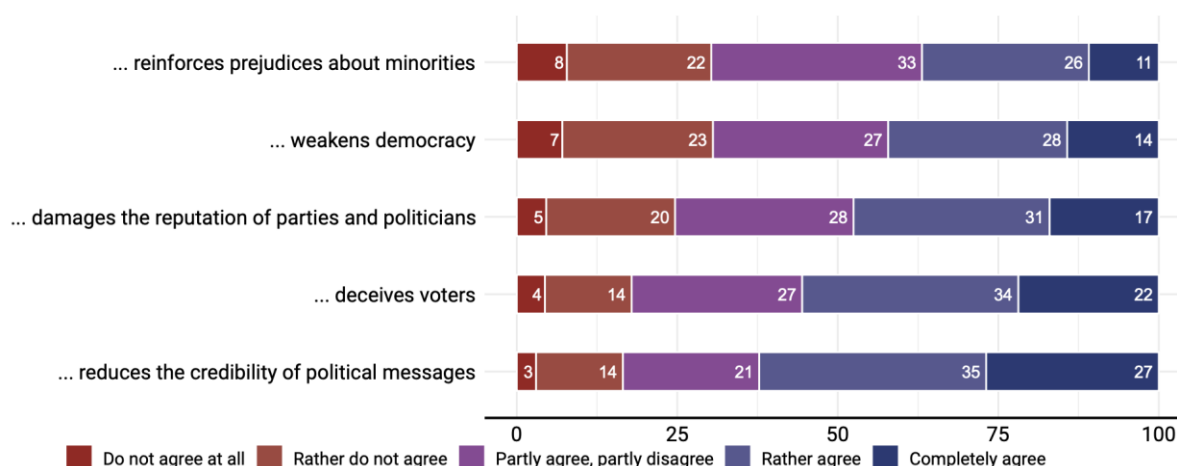


Figure 11 : Perceptions des risques liés aux campagnes électorales basées sur l'IA *Remarque.* N = 1991 ; *question :* « Si les partis et les politiciens utilisent de plus en plus l'IA pour créer des publicités politiques à l'avenir, cela... » Source : (Kruschinski et al., 2025) .

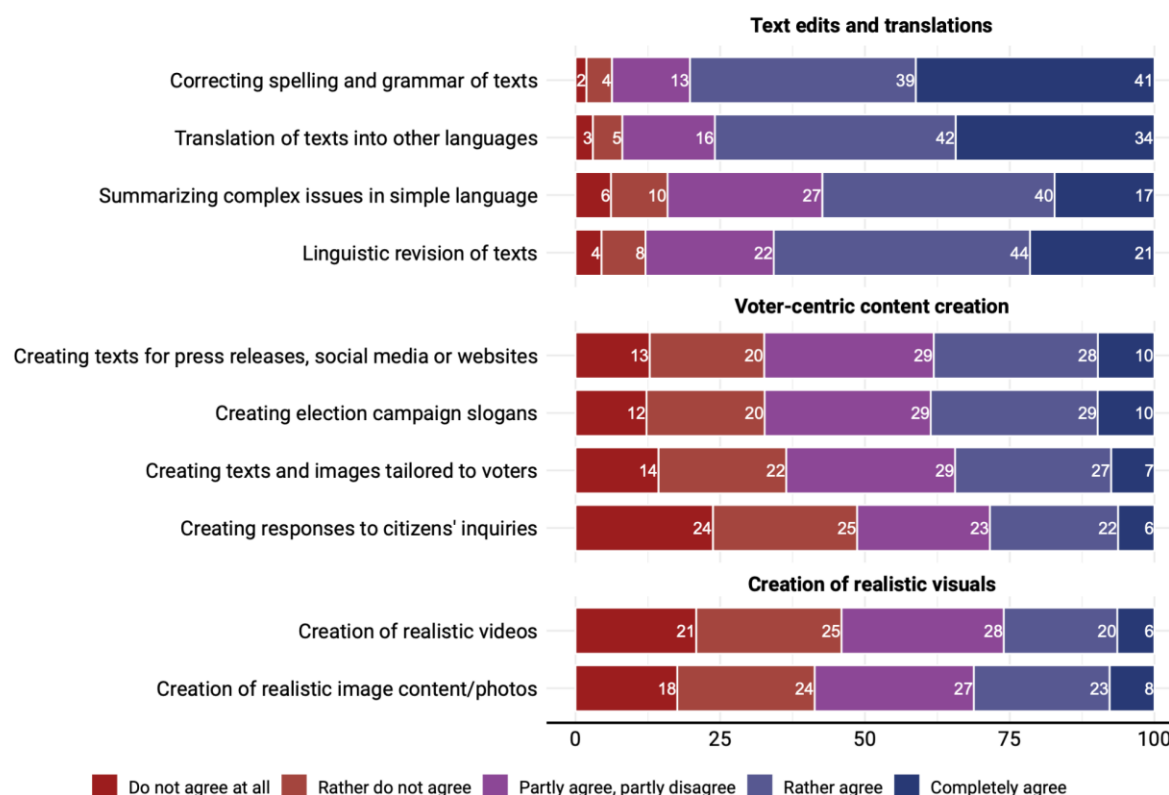


Figure 12 : Acceptation de l'IA dans la communication politique par tâche (en %) N = 1991 ; *question :* « Veuillez indiquer dans quelle mesure vous considérez acceptable l'utilisation de l'IA pour certaines tâches. »

L'analyse des facteurs qui motivent ces attitudes révèle des schémas complexes. Les personnes convaincues de l'importance de leurs actions politiques (forte efficacité politique externe) et celles qui se montrent curieuses vis-à-vis des technologies perçoivent moins de risques et font preuve d'une plus grande acceptation des applications de l'IA. À l'inverse, les personnes plus engagées dans la politique reconnaissent généralement des risques plus importants, probablement parce qu'elles comprennent mieux les processus politiques susceptibles d'être perturbés par l'IA. La connaissance de l'IA produit un paradoxe : les citoyens informés acceptent son utilisation pour des tâches simples, comme la traduction, mais rejettent les applications complexes telles que la génération d'images réalistes, ce qui suggère un jugement discriminatoire plutôt qu'une acceptation ou un rejet global.

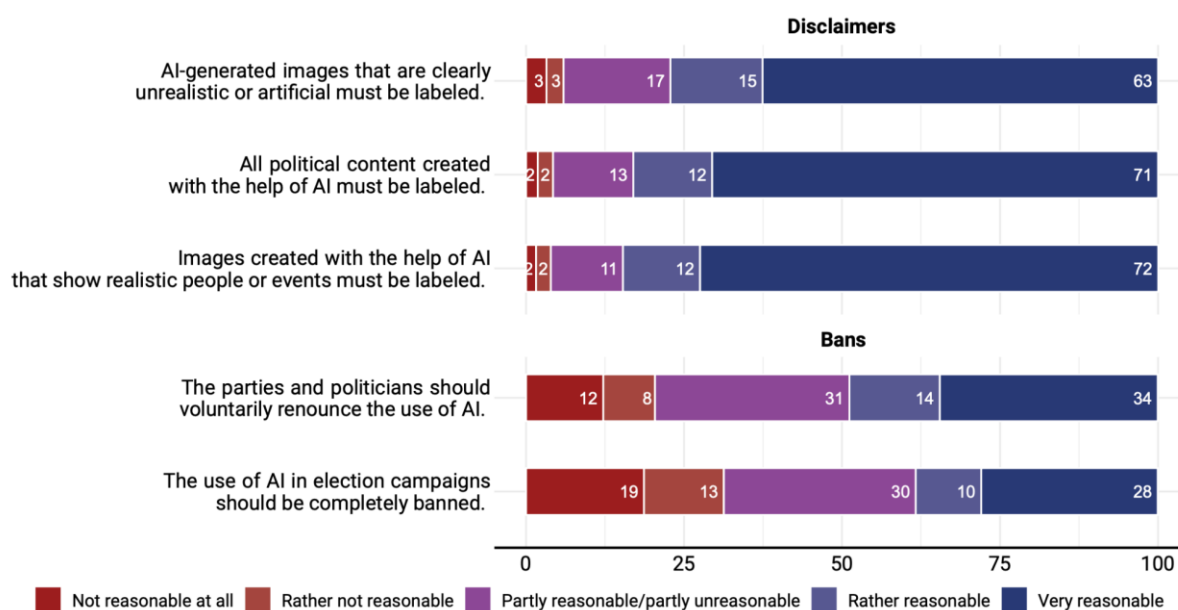


Figure 13 : Opinion publique sur la réglementation de l'utilisation de l'IA dans la communication politique Remarque : N = 1991 ; *question :* « Dans quelle mesure trouvez-vous les propositions suivantes utiles ? »

En matière de réglementation, la transparence obligatoire apparaît comme la principale préférence du public (voir figure 13 ; Kruschinski et al., 2025), ce qui reflète les attentes des citoyens dans le domaine du journalisme (voir 2.1.6). La plupart des Allemands sont favorables à l'obligation d'étiqueter les contenus de campagne générés par l'IA, en particulier les représentations réalistes de personnes ou d'événements, et trouvent l'utilisation de l'IA plus acceptable lorsqu'elle est indiquée de manière transparente. Les mesures plus strictes, telles que l'abstention volontaire ou l'interdiction pure et simple, recueillent moins de soutien. La perception des risques détermine les préférences en matière de réglementation : ceux qui considèrent l'IA comme dangereuse sont favorables à des interventions plus fortes, en particulier à l'interdiction de l'IA. Cependant, les citoyens qui ont une meilleure connaissance de l'IA et qui sont curieux vis-à-vis des technologies soutiennent les mesures de transparence mais s'opposent aux interdictions, préférant disposer d'informations pour se forger leur propre

opinion. Ceux qui ont confiance dans la réactivité du système politique (efficacité externe) s'opposent à la fois aux avertissements et aux interdictions, considérant peut-être l'IA comme moins menaçante lorsqu'ils font confiance aux institutions démocratiques pour s'autoréguler. Ces attitudes hétérogènes, façonnées par la confiance, les connaissances et l'ouverture technologique, révèlent que l'élaboration d'une réglementation légitime en matière d'IA nécessite de tenir compte de la diversité des points de vue plutôt que de rechercher un consensus uniforme.

2.2.7 Conclusion

Dans l'ensemble, les données empiriques actuelles concernant la perspective de campagne suggèrent que l'IA générative est encore utilisée de manière plutôt expérimentale par les partis politiques établis aux Pays-Bas et en Allemagne. En revanche, les partis extrêmes, plus petits et disposant de moins de ressources s'appuient stratégiquement sur l'IA générative, car ils ne fixent pratiquement aucune limite éthique à son utilisation et souhaitent compenser dans une certaine mesure leur désavantage concurrentiel lié aux ressources. À mesure que l'adoption de l'IA se généralise et que les partis reconnaissent les avantages qu'elle présente en termes d'engagement, le discours démocratique risque de voir les contenus synthétiques remplacer de plus en plus la communication politique de fond, les partis se livrant concurrence non pas par le biais d'arguments politiques, mais par le biais d'appels émotionnels générés par l'IA et d'une négativité stratégique optimisée pour une diffusion virale plutôt que pour une délibération éclairée.

Si la démocratisation présente des avantages pour les candidats et les partis disposant de ressources limitées, l'équilibre actuel penche en faveur de risques pour l'intégrité du discours démocratique. Le fait que les partis établis adoptent une approche prudente tandis que les acteurs marginaux expérimentent de manière intensive, suggère qu'il existe une fenêtre d'intervention avant que ces pratiques ne s'ancrent durablement. Des cadres réglementaires complets, assortis de mécanismes d'application solides, sont essentiels pour préserver les bases informationnelles de la prise de décision démocratique, alors que les capacités de l'IA progressent vers des contenus synthétiques de plus en plus difficiles à discerner. Les élections néerlandaises et allemandes de 2025 démontrent que l'autorégulation volontaire s'avère insuffisante, que la gouvernance des plateformes reste faible et que des approches globales, traitant de la transparence, de la responsabilité et des obligations positives de soutenir la participation démocratique doivent être considérées comme prioritaires. Cependant, ces conclusions soulignent également la nature contingente de l'impact politique de l'IA générative : ses effets dépendent moins de la technologie elle-même que des contextes sociaux, organisationnels et réglementaires dans lesquels elle opère.

En ce qui concerne les résultats empiriques relatifs à l'acceptation de l'IA par le public, les sociétés sont confrontées à une question plus large sur les limites entre automatisation et authenticité dans la communication démocratique, à mesure que les campagnes intègrent l'IA générative dans leurs opérations courantes. Les attitudes du public révèlent que les citoyens ont une perception plus forte des risques liés à l'utilisation de l'IA dans les campagnes politiques, mais font la distinction entre les utilisations positives et manipulatrices, avec une acceptation variant de manière systématique : les fonctions d'assistance technique telles que la correction orthographique sont largement approuvées, tandis que la génération d'images et vidéos par l'IA suscite une forte résistance. En outre, la transparence obligatoire pour l'IA

apparaît comme la préférence réglementaire dominante, la plupart des Allemands soutenant les exigences d'étiquetage pour les contenus générés par l'IA plutôt que les interdictions totales. Cependant, la perception des risques, l'acceptation et les préférences réglementaires diffèrent selon les caractéristiques des citoyens. La connaissance de l'IA produit des jugements discriminatoires plutôt que des positions uniformes : les citoyens informés acceptent l'IA pour des tâches simples, mais rejettent les applications complexes telles que la génération de représentations réalistes, ce qui suggère une évaluation sophistiquée des risques spécifiques. L'engagement politique est corrélé à une perception accrue des risques, car les citoyens politiquement actifs comprennent mieux les processus démocratiques que l'IA pourrait perturber. Les attitudes réglementaires se divisent en fonction des connaissances et de l'efficacité politique : ceux qui ont une meilleure connaissance de l'IA privilégient la transparence plutôt que l'interdiction, tandis que ceux qui ont une efficacité politique externe élevée s'opposent à la fois aux obligations de divulgation et aux interdictions.

Ces attitudes hétérogènes et dépendantes des connaissances démontrent qu'une gouvernance légitime de l'IA ne peut pas reposer sur un consensus public uniforme, mais doit tenir compte de perspectives diverses façonnées par la compréhension technologique, la confiance politique et les niveaux d'engagement. Cette acceptation sélective offre à la fois une boussole normative et une feuille de route politique : la transparence, la responsabilité et la supervision humaine restent des conditions préalables essentielles à un déploiement légitime.

Les données empiriques actuelles indiquent donc un double impératif. D'une part, les régulateurs et les plateformes doivent garantir l'intégrité électorale grâce à des régimes de transparence et d'étiquetage applicables. D'autre part, les décideurs politiques, les éducateurs et les acteurs de la société civile doivent cultiver la culture médiatique et l'intelligence artificielle des citoyens, afin de leur permettre de discerner les contenus synthétiques et d'évaluer de manière critique la communication politique. Les campagnes futures combineront probablement de manière toujours plus harmonieuse la créativité humaine et celle des machines, exigeant que les systèmes démocratiques s'adaptent non pas en rejetant l'innovation technologique, mais en l'intégrant dans des garde-fous éthiques et institutionnels solides.

Dans cette optique, l'IA générative ne marque pas le début d'une nouvelle ère dans la communication politique, mais plutôt une nouvelle phase dans la lutte déjà existante qui vise à aligner l'innovation technologique sur la responsabilité démocratique.

2.3 Infrastructure

Le déploiement de l'IA générative dans le journalisme et d'autres secteurs s'appuie sur une infrastructure technologique vaste et multicouche. Il est important de noter que bon nombre de ces couches, notamment la puissance de calcul, le stockage des données, l'accès à des modèles d'IA comportant des milliards de paramètres et les services cloud, sont exploitées par les mêmes entreprises qui détiennent les principaux outils d'IA générative, moteurs de recherche, places de marché et plateformes de médias sociaux mentionnés dans nos études de cas. Ainsi, la concentration de la propriété et du contrôle de notre infrastructure d'information numérique entre les mains d'entreprises technologiques opaques et orientées

vers le profit soulève de sérieuses préoccupations quant à la responsabilité démocratique, à la fiabilité et la qualité de l'environnement de l'information numérique.

2.3.1 La politique des infrastructures

Les infrastructures ont des implications (géo)politiques importantes, notamment en ce qui concerne des sujets critiques tels que la souveraineté (numérique), l'autonomie stratégique et la délibération démocratique. C'est pourquoi il est impératif de prendre conscience de l'importance des infrastructures et de leur politique, en les considérant comme un vecteur de pouvoir. Cependant, le terme « infrastructure » est difficile à définir. Sur le plan conceptuel, elles peuvent être comprises comme « des réseaux construits qui facilitent la circulation des biens, des personnes ou des idées » (Larkin, 2013, p. 328), reliant et organisant diverses dimensions numériques, économiques, sociales et politiques de notre société.

D'un point de vue pratique, d'une part, l'infrastructure fait référence aux substrats matériels et techniques nécessaires à la construction et au fonctionnement de systèmes d'IA à grande échelle, tels que les centres de données, les puces à semi-conducteurs, les câbles sous-marins et les réseaux énergétiques (Mayer, 2023). D'autre part, elles englobent souvent des dimensions plus intangibles, telles que l'architecture logicielle, les spécificités matérielles et les structures organisationnelles qui coordonnent la production, comme les systèmes de gestion de contenu (CMS), les interfaces de programmation d'applications (API) et les kits de développement logiciel (SDK) (Pybus & Côté, 2024). Par extension, les talents, l'expertise technique et les connaissances constituent également des ressources essentielles au développement et le déploiement de ces infrastructures (Rikap, 2021). Le présent rapport adopte une approche synthétique et intégrée qui considère les infrastructures comme les moyens et les conditions de production dans un contexte donné, souvent industriel. En résumé, le terme « infrastructure » englobe un ensemble large et varié de technologies et de conditions facilitant la communication. Concrètement, comme l'ont montré les sections précédentes, alors que les acteurs politiques achètent souvent les outils d'IA fournis par les plateformes en fonction de leurs besoins (2.2.5), la dépendance du journalisme à l'égard des infrastructures d'IA s'étend, par exemple, aux systèmes cloud sur lesquels sont déployés les propres systèmes d'IA des organisations médiatiques (2.1.5).

Les implications de la politique des infrastructures pour la démocratie apparaissent plus clairement lorsqu'on les examine dans le cadre plus large de notre écosystème informationnel, défini comme un « ensemble complexe et interconnecté de systèmes qui [déterminent] qui peut recevoir ou partager quels types d'informations et dans quels contextes » (Mansell et al., 2025, p. 75). Il est essentiel de noter que cet ensemble de systèmes complexes qui composent notre écosystème informationnel n'est pas neutre. Il est au contraire conditionné par l'interconnexion des structures politico-économiques, y compris les configurations de propriété des plateformes (Leerssen, 2025) et l'infrastructure sociotechnique sur laquelle il repose.

Un exemple récent illustre bien ce point : en janvier 2025, juste après la réélection de Donald Trump, Mark Zuckerberg, PDG de Meta Platforms, a annoncé la suppression simultanée de son réseau tiers de vérification des faits aux États-Unis, la réintroduction de contenus politiques dans les recommandations aux utilisateurs ainsi qu'un assouplissement des politiques de contenu risquant d'exacerber les comportements haineux en ligne à l'égard des groupes marginalisés (de Brún & Sawyeddollah, 2025). Ces choix, faits par un fournisseur et

opérateur d'infrastructure dominant, ont une incidence sur les conditions et la qualité des informations dans les flux des utilisateurs, ainsi que sur la manière dont les médias d'information produisent, distribuent et monétisent leur contenu.

Dans ce contexte, les infrastructures – des clouds et serveurs à l'architecture logicielle et aux solutions – apparaissent donc comme l'épine dorsale constitutive de notre écosystème d'information numérique. Elles fournissent les conditions matérielles, numériques et organisationnelles dans lesquelles l'information est produite, diffusée et échangée, et sont la source d'un pouvoir économique et politique considérable. Ensemble, leurs implications pour la démocratie sont fondamentales. Il est donc essentiel de conserver une perspective concrète lorsque l'on aborde la question des infrastructures, en tenant compte des politiques qui les façonnent, y compris les dynamiques de pouvoir dans leurs structures de gouvernance et de propriété.

2.3.2 Contrôle de l'infrastructure nécessaire à l'IA générative

Les entreprises technologiques dominantes sont désormais les mieux placées pour « exploiter les possibilités offertes par les réseaux d'information et de communication numériques et fournir des infrastructures qui facilitent certains types d'interactions » (Cohen, 2019, p. 41) . Le développement et le déploiement de l'IA générative ont été marqués par un paradigme infrastructurel et une économie politique dans lesquels une poignée d'acteurs américains et chinois exercent un contrôle disproportionné sur les technologies critiques (Balayn & Gürses, 2024 ; Whittaker, 2021) . Cette concentration du pouvoir infrastructurel signifie que les valeurs inhérentes à ces systèmes (par exemple, l'efficacité, l'évolutivité, la rentabilité, la surveillance, l'extraction de données, etc.) sont effectivement ancrées dans les services qu'ils prennent en charge. Comme nous l'avons déjà vu dans la section 2.1.5 sur les dépendances structurelles, les grandes entreprises technologiques telles que Microsoft, Amazon et Google ont étendu leur emprise sur le journalisme au-delà des outils et des canaux de distribution, c'est-à-dire l'infrastructure dont dépend largement le journalisme pour sa viabilité et, par conséquent, sa capacité à remplir son rôle dans les sociétés démocratiques.

Ces entreprises technologiques ont systématiquement tiré parti de leur accès et de leur influence sur le capital, l'infrastructure cloud, les données et les talents pour façonner le développement de l'IA générique à leur avantage (Autorité de la concurrence, 2024 ; van der Vlist et al., 2024). En d'autres termes, elles ont utilisé leur pouvoir préexistant sur les infrastructures numériques pour façonner un nouveau terrain de jeu qui, à son tour, leur permet d'approfondir leur contrôle sur ces grands systèmes techniques. Parmi les pratiques courantes, on peut citer l'intégration verticale des technologies clés pour la formation et le déploiement de modèles d'IA et les stratégies d'innovation complémentaires permettant aux acteurs dominants de s'accaparer des marchés spécifiques (Luitse, 2024). Par exemple, les fournisseurs de cloud hyperscale regroupent souvent différents logiciels en un service unifié, forçant ainsi leurs clients à s'engager dans un environnement de production prédéterminé tout en les enfermant (souvent) dans leur infrastructure. Ces exemples illustrent un environnement infrastructurel caractérisé par une concentration du marché, l'absence de réglementation significative et des investissements astronomiques de la part d'un petit nombre d'acteurs dont les objectifs stratégiques généraux sont principalement axés sur le profit (Vipra & Myers West, 2023; Widder & Kim, 2025).

Lorsque les systèmes alimentés par l'IA générique s'intègrent dans notre écosystème informationnel, la question de savoir « qui peut recevoir ou partager quels types d'informations et dans quels contextes » (Mansell et al., 2025, p. 75) est fortement influencée par le paradigme infrastructurel et l'économie politique décrits ci-dessus. En fait, cette imbrication des infrastructures avec l'économie politique du journalisme n'est pas abstraite : elle reconfigure déjà l'environnement médiatique au sens large (Sevignani et al., 2025). Par exemple, les accords de licence avec les fournisseurs d'IA, réduisent le journalisme à des fragments d'informations modulaires prêts à être recombinaés à la demande d'un utilisateur dans un service comme ChatGPT, tandis que des intégrations plus directes réorganisent le fonctionnement des salles de rédaction autour de systèmes tiers. À ce titre, les connaissances sous forme de données, en particulier celles produites par les journalistes, constituent une ressource cruciale pour les fournisseurs d'IA générative, non seulement pour former leurs modèles d'IA, mais aussi pour différencier leurs offres en fournissant des services tels que des résumés d'actualités à jour (par exemple, l'accès aux actualités via ChatGPT ou Google AI Overview). Par conséquent, la recherche et la diffusion d'actualités sont de plus en plus médiatisées par une infrastructure numérique alimentée par l'IA générative. Cela aura très probablement des conséquences sur la visibilité, l'accessibilité et la rentabilité des organes de presse. Par exemple, au moment de la rédaction de cette étude, OpenAI a annoncé sept accords avec des éditeurs de presse dont le siège se trouve en Europe (Italie, France, Allemagne, pays nordiques et Espagne, et deux au Royaume-Uni). Cette approche en matière de conclusion d'accords a des implications pour la fiabilité et la diversité des sources médiatiques sur lesquelles les modèles d'IA peuvent s'appuyer ; la décision des organes de presse de placer leur contenu derrière des paywalls afin d'empêcher les entreprises d'IA de l'utiliser peut également restreindre l'accès du grand public à l'information. L'élaboration et l'application d'accords de licence garantissant que le contenu d'actualité soit accessible aux humains sans être utilisé pour entraîner des modèles d'IA pourraient atténuer ce problème.

Les dépendances infrastructurelles peuvent également compromettre la capacité des médias d'information à exercer leur activité journalistique en dehors des chaînes de valeur technologiques des fournisseurs de l'IA générative (voir à nouveau la section 2.1).

À titre d'illustration, *le Washington Post* (propriété de Nash Holdings, la société d'investissement privée de Jeff Bezos) a développé ArcXP, un CMS destinés aux médias d'information hébergé au sein d'AWS. L'adoption d'ArcXP par des médias d'information tels que *Le Parisien* (Arc XP, 2024) illustre bien la dynamique décrite ci-dessus : une tâche aussi simple que le résumé ou le référencement d'un article passe désormais par la pile cloud d'AWS, intégrant ainsi une gouvernance, des valeurs et une logique organisationnelle externes dans les conditions de travail des journalistes. Ce faisant, ces infrastructures peuvent transformer les salles de rédaction en locataires de leurs systèmes propriétaires, modulariser le travail éditorial, dicter la manière dont la valeur est créée et diffusée et, plus largement, remodeler le rôle essentiel du journalisme dans le bon fonctionnement de l'écosystème de l'information. Comme le résume succinctement Schneiders et Stark : « La valeur et la qualité de l'information dépendent de la qualité du processus de production. Si l'indépendance, la diversité et la pertinence ne sont pas garanties, le journalisme perd ses fonctions – et donc sa *valeur épistémique* pour le public » (Schneiders & Stark, 2025, p. 16) (italique ajouté).

Examinons cela à travers un exemple abordé à la section 2.1 : les accords entre les éditeurs de presse et les entreprises de technologie IA, telles qu'OpenAI, reposent sur la technique en

plein essor de la « génération augmentée par la récupération » (RAG). En substance, un générateur de texte IA générative tel que ChatGPT tente de répondre à une requête utilisateur en utilisant non seulement ses données d'entraînement, mais également un autre ensemble de données (potentiellement plus à jour). C'est ce que fait ChatGPT lorsqu'il répond à une requête demandant des informations sur un sujet et extrait des informations – parfois inexactes, comme le montre la section 2.1(d) – du *Financial Times*, avec lequel OpenAI a signé un accord de plusieurs millions de dollars. De même, l'Associated Press a reconfiguré ses archives pour les rendre « compatibles avec l'IA » afin de garantir que les systèmes d'IA générative puissent utiliser les techniques RAG pour leurs propres offres grand public ou d'entreprise (Helberger et al., 2022) . Le lieu de création de valeur du journalisme est donc élargi et étendu de la diffusion et de la publicité aux chaînes de valeur des entreprises technologiques spécialisées dans l'IA, y compris les conditions d'accès aux API de ces entreprises et aux archives compatibles avec l'IA des éditeurs de presse.

Dans ce contexte, il est essentiel de noter que nous ne parlons pas uniquement de valeur économique, malgré son importance pour la viabilité du journalisme. Nous nous intéressons également à la valeur épistémique et démocratique. C'est ainsi que les citoyens donnent un sens au monde afin de participer aux délibérations démocratiques et à la vie politique en général. Par conséquent, l'exemple de la monétisation des archives des éditeurs, et par extension de la capacité des gens à donner du sens, souligne la valeur transactionnelle que le paradigme (infra)structurel actuel attribue au journalisme. Ce processus s'inscrit dans des écosystèmes propriétaires et payants, soutenus principalement par l'infrastructure informatique en nuage (cloud computing) appartenant aux mêmes entreprises technologiques qui achètent l'accès à ces mines d'informations.

Parallèlement, on observe l'émergence d'initiatives au sein de l'UE visant à expérimenter des modes de gouvernance alternatifs pour l'octroi de licences d'accès aux contenus d'actualité. Par exemple, un consortium d'organisations néerlandaises dirigé par l'Organisation néerlandaise pour la recherche scientifique appliquée (TNO) développe ce qu'il appelle un LLM « souverain » en langue néerlandaise, le GPT-NL (GPT-NL, 2025). Le projet repose sur un modèle de licence pour les éditeurs fondé sur le consentement, une rémunération équitable et la conformité réglementaire, certains grands éditeurs comme DPG Media, Mediahuis et l'agence de presse ANP ayant déjà accepté de concéder sous licence leurs archives. Une telle initiative contraste avec les marchés de licences pilotés par l'industrie et intégrés aux écosystèmes cloud propriétaires mentionnés plus haut. La mise en place et le maintien de modèles compétitifs nécessitent toutefois des ressources financières substantielles, un engagement politique à long terme et l'accès à une infrastructure informatique qui est elle-même fortement contrôlée par quelques entreprises dominantes. De meilleures conditions de licence pourraient ainsi rééquilibrer certains aspects de l'extraction de valeur et renforcer la transparence pour les éditeurs, améliorant potentiellement les conditions d'accès démocratique à l'information, mais elles ne conduisent pas nécessairement à un découplage des écosystèmes européens d'IA par rapport aux dépendances matérielles et cloud concentrées à l'échelle mondiale. En ce sens, la persistance de ces dépendances sous-jacentes met en évidence des problèmes systémiques et structurels plus larges, abordés dans la section suivante.

2.3.3 Dépendance infrastructurelle et recul démocratique

Comme les technologies sont développées et exploitées dans l'opacité, et que les citoyens ne disposent pas d'outils significatifs pour contester les choix et les logiques qui composent leur environnement axé sur les données, l'impact que ces technologies pourraient avoir sur leur environnement numérique et leurs moyens de subsistance est souvent déterminé unilatéralement par d'autres, sans grande transparence ni connaissance approfondie (Naudts, 2024). Cette position de domination technologique a également été identifiée comme un atout stratégique et géopolitique.

L'accent disproportionné mis sur l'échelle qui a caractérisé le développement de l'IA générique a entraîné un enchevêtrement profond des intérêts des entreprises et des États, comme en témoigne explicitement la ruée des leaders technologiques pour courtiser la nouvelle administration américaine en janvier 2025. Bien que privées, les infrastructures informatiques qui sous-tendent le développement des outils d'IA ont été développées en partenariat avec des acteurs étatiques américains et ceux-ci exercent toujours une influence considérable sur elles (Holt, 2024 ; Hu, 2015). Des incidents tels que la suspension des services AWS en Iran (Motamedi, 2019) ou la fermeture du compte Microsoft du procureur de la CPI Karim Khan (Quell, 2025) ont montré que le département d'État américain est prêt à tirer parti du pouvoir infrastructurel des entreprises relevant de sa juridiction pour atteindre des objectifs stratégiques. Comme l'a conclu le gouvernement néerlandais à la suite des suspensions de Khan, la dépendance des institutions publiques à l'égard de ces infrastructures pourrait avoir un impact négatif sur leur autonomie (Veldkamp & van Marum, 2025). L'IA générative renforce encore cette dynamique. D'une part, les acteurs dominants utilisent le développement de l'IA comme prétexte pour étendre davantage leur contrôle sur les infrastructures informatiques (Chander & Gürses, 2024) . D'autre part, les utilisateurs, y compris les institutions publiques, deviennent de plus en plus dépendants de ces infrastructures en intégrant l'IA générative dans leurs activités quotidiennes (Brennan et al., 2025).

Plus précisément, il convient d'être attentif au regroupement et au renforcement des pouvoirs entre les régimes (autoritaires) (États) et les propriétaires de plateformes implantés et opérant sur leurs territoires. Selon Rolf et Schindler, le pouvoir commercial technologique mondial est concentré aux États-Unis et en Chine, où les acteurs numériques agissent comme des partenaires stratégiques en réponse aux tensions économiques et géopolitiques (Rolf & Schindler, 2023). En retour, ces partenaires privilégiés bénéficient d'avantages économiques et réglementaires, ce qui explique également la résistance à laquelle l'Union européenne a été confrontée. Il est important de noter que les plateformes numériques, y compris les propriétaires de modèles à usage général, peuvent être utilisées pour promouvoir et dissimuler certains discours politiques. Par exemple, si des mots tels que « préjugés », « handicap » et « diversité » font l'objet d'un examen minutieux dans la recherche scientifique (Johnson et al., 2025), il n'est pas exclu qu'à l'avenir, des directives soient données à des acteurs privés pour neutraliser, dépolitiser ou censurer ces notions dans l'IA générative. À leur tour, ces interventions pourraient avoir un effet en aval lorsque ces modèles sont utilisés pour générer du contenu journalistique, électoral ou politique. Par exemple, un modèle acheté utilisé pour rationaliser les articles journalistiques pourrait être formé à privilégier délibérément un langage reflétant la binarité des genres afin de rendre invisibles les personnes non binaires. Les systèmes pourraient également rationaliser ces décisions (Kotek et al., 2023) , masquant ainsi la nature discriminatoire ou préjudiciable de ces choix chargés de valeurs.

2.3.4 Conclusion

La concentration des infrastructures ajoute un autre niveau de vulnérabilité pour la démocratie et ses institutions. Cette vulnérabilité peut être difficile à percevoir. Alors que les implications négatives d'un deepfake viral visant un candidat politique sont facilement visibles, les infrastructures fonctionnent en arrière-plan et ont un impact subtil sur la démocratie. Le contrôle des infrastructures ne se traduit pas nécessairement par de la censure ; en contrôlant les technologies et autres ressources qui facilitent la communication, les entreprises technologiques peuvent déterminer le type de communication possible. Ce contrôle se manifeste, par exemple, par le type d'outils d'IA faciles à développer, le type de langage utilisé par défaut par les outils d'IA et la manière dont les outils d'IA peuvent être déployés sur l'infrastructure cloud. Lorsque le contrôle des infrastructures est exercé de manière abusive (par exemple en augmentant les prix ou en coupant l'accès à des infrastructures cloud essentielles), c'est souvent dans une situation où les utilisateurs sont déjà fortement dépendants des infrastructures.

Le concept d'infrastructure est utile pour comprendre comment les entreprises technologiques en sont venues à influencer bon nombre des technologies et des ressources nécessaires à la communication à l'ère numérique, des minéraux aux puces électroniques, en passant par les centres de données et les logiciels. Il permet également de comprendre dans quelle mesure cette infrastructure est exploitée d'une manière contraire aux principes démocratiques. À savoir, par quelques acteurs concentrés, de manière opaque et sans tenir compte des valeurs qui sous-tendent la communication dans un environnement informationnel sain, telles que la diversité, la liberté des médias et la non-discrimination. Cependant, comme le montre clairement la section ci-dessus, le concept d'infrastructure englobe également un large éventail de technologies et de ressources différentes. En se focalisant sur la couche infrastructurelle de l'environnement informationnel, il apparaît clairement que ce à quoi nous assistons n'est pas seulement une concentration du pouvoir entre les mains de quelques acteurs privés – qui sont souvent alliés avec de puissants acteurs étatiques – mais une convergence sans précédent des domaines politique, social, économique et culturel. Les infrastructures qui sous-tendent nos vies de plus en plus numériques sous-tendent également notre environnement informationnel, du journalisme numérique aux campagnes politiques. Cette imbrication profonde a des implications tout aussi profondes sur la manière dont l'information et la valeur qui lui est attribuée sont produites. En ce sens, prendre en compte la couche infrastructurelle signifie également prendre en compte les moyens et les relations de production, y compris la valeur et la qualité de ladite production. En d'autres termes, cette convergence efface – de la manière la plus claire qui soit – les frontières entre valeur économique et valeur démocratique, faisant des infrastructures de notre écosystème informationnel – plutôt que de son contenu – un terrain privilégié pour la contestation du pouvoir.

3. Implications pour les droits humains, la participation politique et les principes démocratiques

Dans les sections précédentes, nous avons montré comment l'IA générative a été déployée, expérimentée ou envisagée pour une utilisation future dans le journalisme et la communication

politique, et comment elle crée à la fois des risques et des opportunités pour la démocratie et les processus démocratiques. Nous avons également discuté de certaines conditions-cadres, notamment le rôle prépondérant et la dépendance à l'égard d'un nombre restreint d'entreprises technologiques qui façonnent la dynamique du marché. L'objectif de cette section est de replacer cette dynamique dans le contexte des cadres normatifs et des instruments existants du Conseil de l'Europe. À cette fin, nous aborderons tout d'abord l'utilisation de l'IA générative dans le journalisme, la communication politique et les infrastructures publique de l'information sous l'angle des humains, en mettant particulièrement l'accent sur les droits à la non-discrimination et à la liberté d'expression. La question centrale à laquelle cette section tente de répondre est de savoir dans quelle mesure la dynamique et les implications identifiées dans la section précédente ont une incidence sur les droits humains, et où nous voyons certains risques, mais aussi des opportunités pour la réalisation des droits humains. Nous nous pencherons ensuite sur l'ensemble des recommandations du Conseil afin de déterminer dans quelle mesure celui-ci a déjà abordé certaines des questions soulevées et proposé des concepts et des solutions sur lesquels s'appuyer. Dans une dernière section, nous nous intéresserons 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 (ci-après « la Convention cadre ») en nous demandant dans quelle mesure celle-ci pourrait constituer un outil permettant de traiter certains des risques et opportunités identifiés.

3.1 Ajouter la dimension des droits humains : l'IA générative du point de vue des droits à l'égalité, à la non-discrimination et à la participation politique

La capacité de participer à la prise de décision publique, et ce dans des conditions d'égalité et de non-discrimination, est au cœur de l'idée démocratique, ainsi que des fondements mêmes du Conseil de l'Europe : « Ses piliers centraux que sont les droits humains, la démocratie et l'État de droit reposent sur la conviction fondamentale que tous les êtres humains sont égaux et dignes en tant qu'agents moraux indépendants » (CoE, 2019, paragraphe 9) . L'inclusivité renvoie à la conviction que les politiques ou les décisions qui ont un impact significatif sur la sphère de vie des personnes ne sont normativement légitimes que lorsque les personnes concernées ont été incluses de manière significative, par la participation ou la représentation, dans le processus décisionnel (Young, 2002, p. 23).

Les technologies génératives influencent la manière dont les personnes sont perçues ou entendues dans la vie sociale et culturelle. Ce faisant, elles peuvent également limiter indûment l'accès et l'exercice des droits civils et politiques *ainsi que* des libertés sociales et économiques de certaines personnes et de certains groupes au profit d'autres. Dans ce contexte, le droit à la non-discrimination – inscrit dans l'article 14 de la Convention européenne des droits de l'homme (ci-après « CEDH » ou « la Convention ») et dans le protocole n° 12 à la Convention – peut jouer un rôle central, car il aide à définir dans quelles conditions les décisions et les politiques qui entraînent (in)directement une jouissance différentielle des droits humains et des libertés associées peuvent être objectivement et raisonnablement justifiées. Toutefois, pour remettre en cause les historiques de désavantage qui sous-tendent le pipeline de l'IA et lutter contre l'impact discriminatoire que les technologies génératives peuvent avoir lors de leur déploiement, les États, les gouvernements et les acteurs privés doivent aller au-delà de l'obligation passive ou négative de ne pas discriminer ou d'empêcher

un traitement différentiel injuste. Comme le soulignent Bartoletti et Xenidis, les parties prenantes doivent plutôt revoir et adapter les mesures existantes à la lumière des changements et des défis technologiques, et adopter des actions et des obligations positives afin de transformer les schémas de discrimination et de tenir compte des différences au sein de la société et tout au long de la chaîne de valeur de l'IA (Bartoletti & Xenidis, 2023, p. 67 ; Fredman, 2016).

3.1.1 Inégalités et implications structurelles

À cet égard, l'idéal de réalisation de l'égalité socio-économique et politique entretient une relation synergique avec le droit à la non-discrimination et la promotion des droits humains. Si les individus ont besoin d'une capacité d'action politique sur leur environnement pour garantir la protection de leurs droits humains, ces mêmes droits garantissent à chacun les opportunités civiles, politiques et socio-économiques nécessaires pour participer à la sphère publique. Dans ce contexte, l'utilisation de l'IA générative dans le journalisme et les campagnes politiques, combinée à une dépendance croissante vis-à-vis des infrastructures, soulève au moins deux préoccupations fondamentales. Premièrement, en raison de leur disponibilité, les technologies génératives risquent de renforcer et d'aggraver les inégalités structurelles, telles que le sexisme et le racisme, dans des domaines de la vie essentiels aux processus démocratiques de délibération et de participation à grande échelle. Deuxièmement, elles risquent d'appauvrir la diversité dans les espaces de débat politique.

Sur le plan structurel, les effets discriminatoires des systèmes d'IA ont été bien documentés, et il a également été démontré que les systèmes d'IA générative reflètent les préjugés sociaux (latents), les formes de discrimination dominantes et les stéréotypes négatifs, y compris les préjugés racistes et sexistes (voir par exemple : Abid et al., 2021 ; Birhane et al., 2021, 2024 ; Kotek et al., 2023 ; Tao et al., 2024) . Ces préjugés peuvent apparaître tout au long du cycle de vie de l'IA, de la collecte des données à leur étiquetage, de la conception au déploiement (Bartoletti & Xenidis, 2023 ; Chen et al., 2021).

La disponibilité généralisée et la facilité d'utilisation de ces systèmes, combinées à la centralisation infrastructurelle du pouvoir en termes de conception, sont susceptibles de renforcer l'institutionnalisation de ces formes d'injustice sociale. Si aucune mesure n'est prise, chaque application en aval dans le domaine du journalisme ou des campagnes politiques, construite sur la base d'un modèle, est susceptible de renforcer les inégalités que ce dernier reflète. Des préjugés systémiques, tels que l'exclusion et la discrimination des groupes historiquement défavorisés et marginalisés, pourraient donc être introduits à plus grande échelle, avec un impact immédiat plus important que celui des systèmes d'IA à usage non général. Dans le même temps, ces applications en aval peuvent également introduire des inégalités (structurelles) si les modèles sont enrichis de données personnalisées biaisées ou utilisés pour renforcer les préjugés et les stéréotypes, comme l'ont démontré les exemples d'utilisation de l'IA générative dans les campagnes politiques (voir section 2.2.3).

Comme l'a souligné la CEDH, les stéréotypes peuvent nuire « au sentiment d'identité d'un groupe et à l'estime de soi et à la confiance en soi de ses membres (*Aksu c. Turquie*, 2012, paragraphe 58) .» Selon Timmer, les stéréotypes doivent être considérés comme « des structures cachées qui sous-tendent l'égalité et la discrimination » (Timmer, 2014) . Les stéréotypes constituent des croyances préconçues et des mythes concernant des groupes

sociaux particuliers qui limitent indûment la mobilité sociale et économique des personnes, y compris leur jouissance des droits humains. Les stratégies de profilage ethnique, par exemple, sont discriminatoires, car elles extrapolent à partir de la perception stéréotypée un lien automatique entre le comportement antisocial et l'origine ethnique d'une personne (voir également *Lingurar c. Roumanie*, 2019, par. 76) . Dans ce contexte, les stéréotypes, les inégalités sociales et la discrimination forment un « cercle vicieux » interdépendant : « les stéréotypes peuvent être une manifestation de la discrimination, mais aussi une rationalisation et une cause de discrimination » (Timmer, 2014, p. 211). Ce faisant, ils contribuent à maintenir les structures sociales et les rapports de force existants. Compte tenu de cette dynamique de renforcement, les traditions, les hypothèses généralisées ou les attitudes dominantes, telles que les stéréotypes raciaux ou les rôles liés au genre, ne suffisent pas à justifier un traitement différencié (voir également *Konstantin Markin c. Russie*, 2012). Pourtant, ce dernier risque de se produire lorsque les systèmes d'IA générique contribuent à modérer les environnements de connaissance des personnes, comme les médias d'information ou les campagnes politiques, et influencent (indirectement) les processus décisionnels démocratiques.

Dans ce contexte, une attention particulière doit également être prêtée à la situation des personnes ou des groupes vulnérables. La vulnérabilité, telle que nous l'entendons ici, est un concept relationnel dans le sens où elle est influencée et façonnée par l'ancrage et la position des personnes dans des relations et des dépendances sociales, économiques et institutionnelles plus larges, y compris les inégalités et les hiérarchies structurelles (Fineman, 2008) . Dans ce contexte, les stéréotypes et la vulnérabilité renforcent les préjugés liés au statut social. D'une part, les groupes vulnérables font souvent l'objet de stéréotypes omniprésents. D'autre part, en reconnaissant le statut vulnérable d'un groupe, nous pouvons identifier les cas où l'application d'hypothèses généralisées est plus susceptible de causer un préjudice (affaire *Carvalho Pinto De Sousa Morais c. Portugal*, 2017 ; Naudts, 2023 ; Timmer, 2014). Dans le même temps, l'utilisation de l'IA générique pour permettre de nouvelles formes d'hyper-personnalisation (voir sections 2.1.1 et 2.2.4) comporte également le risque d'exploiter davantage ou de créer de nouvelles vulnérabilités (Helberger et al., 2022). Dans sa Déclaration sur les capacités manipulatrices des processus algorithmiques, le Comité des Ministres du Conseil de l'Europe avait déjà averti que l'utilisation de moyens informatiques pouvait entraîner « le classement des individus en catégories, renforçant ainsi différentes formes de ségrégation et de discrimination sociales, culturelles, religieuses, juridiques et économiques. Elle facilite également le microciblage des individus sur la base de profils, d'une manière qui peut profondément affecter leur vie ». La limite de la manipulation inacceptable est atteinte, selon le Conseil, lorsque la technologie est utilisée pour « exploiter les vulnérabilités existantes ou les biais cognitifs » (CoE, 2019, paragraphe 9e) .

La deuxième préoccupation majeure du point de vue des droits humains que nous souhaitons souligner concerne les implications potentielles pour l'inclusion et le risque d'uniformisation de la diversité en tant qu'autre principe fondamental des droits humains à la non-discrimination, à la participation politique, mais aussi à la liberté d'expression (voir section 3.2).

Les modèles génératifs pourraient ne pas bien gérer la diversité. Les grands modèles linguistiques, par exemple, risquent de fonctionner comme des machines à consensus participatives imparfaites (Waldo & Boussard, 2024) . Lorsqu'un LLM est interrogé sur des sujets pour lesquels les données en ligne sont limitées, il est plus susceptible de produire des réponses inexactes, mais souvent de manière convaincante ou réaliste. À l'inverse, les LLM

sont plus performants sur des sujets plus populaires, notamment les discussions polarisantes ou celles qui ont fait l'objet d'un consensus. La section 2.2.2 a également montré comment les LLM, en interaction avec la modération algorithmique des contenus, peuvent entraîner la prévalence de certains types de contenus, tels que les contenus émotionnels, par rapport à d'autres. De même, les résultats des systèmes génératifs peuvent refléter les normes culturelles majoritaires ou dominantes. Par exemple, à moins d'être incités à faire autrement, les LLM contemporains ont encore tendance à promouvoir et à reproduire les identités et les récits culturels dominants, « représentant rarement des arrangements et des perspectives moins courants » (Gillespie, 2024). En d'autres termes, la représentation n'est pas une évidence, mais doit être remise en question. Cet appauvrissement de la diversité peut également se produire lorsque les modèles génératifs sont appelés à effectuer des tâches sur de nouvelles données d'entrée. Certaines applications basées sur l'IA promettent une analyse efficace des documents, comme la synthèse des commentaires des citoyens reçus lors de consultations publiques (Lazar & Manuali, 2024 ; Small et al., 2023). Cependant, dans l'exécution de ces tâches également, ces technologies se sont avérées capables de fabriquer des informations, d'introduire des changements (subtils) de sens et de refléter des biais politiques (Lazar & Manuali, 2024) . En outre, Lazar & Manuali soulignent que les LLM « ont du mal à repérer les valeurs aberrantes dans les données qui pourraient contenir des informations potentiellement importantes ». Un LLM pourrait ne pas repérer une idée brillante ou la voix de groupes minoritaires (Lazar & Manuali, 2024, p. 10) .

Comme le note Young : « Un engagement trop fort en faveur du consensus en tant que bien commun peut inciter certains, voire tous, à préconiser de retirer les questions difficiles de la discussion afin de parvenir à un accord et de préserver le bien commun » (Young, 2002, p. 45). Elle poursuit toutefois : « Parfois, ces questions difficiles revêtent une importance capitale pour un groupe qui estime subir une injustice fondamentale, mais elles sont source de désaccord profond car d'autres membres de la société considèrent que la rectification de cette injustice présumée leur coûterait trop cher » (Young, 2002, p. 45). Si la dissidence peut apporter un éclairage intéressant, ces points de vue sont malheureusement noyés dans la masse lorsque la priorité est donnée aux points de vue consensuels ou majoritaires. De même, lorsque des technologies génératives sont utilisées par des acteurs de mauvaise foi pour gonfler artificiellement les préférences des parties prenantes concernant certaines positions politiques, cela réduit la visibilité des expressions politiques authentiques, y compris celles coproduites par l'IA (Kreps & Kriner, 2023, pp. 123–125 ; voir également la section 2.2.5). Que les LLM soient utilisés pour résumer, agréger ou représenter les points de vue des personnes, dans les conditions actuelles, certaines perspectives sont plus susceptibles d'être rendues invisibles que d'être intégrées dans les processus de délibération et de contestation (Mouffe, 1999 ; Young, 2002). Le risque que les systèmes génératifs aplatissent la diversité va également à l'encontre des idéaux de pluralisme défendus par la CEDH, qui a déclaré que la diversité ne devrait jamais être considérée comme une menace, mais plutôt comme un enrichissement de la vision démocratique (*D.H. et autres c. République tchèque*, 2007, paragraphe 176) (voir section 3.2).

3.1.2 Le potentiel du droit de la non-discrimination pour répondre à ces préoccupations

Les dynamiques socio-techniques susmentionnées influencent les résultats produits par l'IA générative. Les résultats produits par l'IA générative peuvent être non inclusifs, carrément dégradants ou abusifs, ou encore représenter de manière inadéquate ou déformée certains

groupes sociaux (voir également (Hacker et al., 2025)). Les chercheurs avertissent toutefois depuis longtemps que la grammaire et le vocabulaire de la législation européenne en matière de non-discrimination risquent de devenir obsolètes à l'ère numérique (Gerards & Xenidis, 2021 ; Gerards & Zuiderveen Borgesius, 2022 ; Naudts, 2023 ; Wachter, 2023 ; Zuiderveen Borgesius et al., 2018). Il n'est donc pas surprenant que les cadres actuels de non-discrimination ne parviennent pas à remédier à l'aggravation des inégalités structurelles et au risque d'invisibilisation et d'exclusion de voix diverses. Dans les paragraphes suivants, nous mettons en évidence certains domaines politiques dans lesquels l'impact démocratique des systèmes génératifs sur le droit à l'égalité et à la non-discrimination peut être amélioré.

Afin de saisir de manière adéquate l'impact négatif que les modèles et systèmes génératifs peuvent avoir sur la capacité des personnes à participer à la société sur un pied d'égalité, toute évaluation de ces technologies doit s'appuyer sur une perspective plus large de la manière dont les éléments sociétaux et technologiques interagissent de manière dynamique pour limiter la jouissance des droits humains par les personnes. En effet, les interventions juridiques ont tendance à cibler des applications et des cas spécifiques de discrimination tout au long de la chaîne de valeur de l'IA, plutôt que ses racines sociétales. Cependant, lorsque l'on se concentre sur un contenu ou une décision en particulier, la nature interconnectée et les propriétés d'amplification uniques des modèles génératifs, et plus encore lorsqu'ils sont à usage général, peuvent être occultées, ignorant ainsi la manière dont les risques pour les droits s'accumulent et s'intensifient au fil du temps (Liu et al., 2018 ; Naudts, 2023) . Pourtant, dans certains cas, lorsqu'il est considéré de manière isolée, un résultat peut être biaisé, voire discriminatoire, mais ne présenter en soi qu'un risque trop faible pour les droits humains globaux des personnes. C'est plutôt l'accumulation de ces types d'injustices à grande échelle, ou la tendance du modèle à les reproduire, qui renforce le discours sociétal général qui, en fin de compte, a un impact sur les libertés et les opportunités globales des personnes (Naudts, 2023 ; Vogiatzoglou, 2025) . De la même manière, les préjudices individualisés causés par les systèmes génératifs résultent d'opérations sur des données au niveau du groupe (Viljoen, 2021). Par exemple, les résultats et l'impact de l'IA générative peuvent être le résultat des nombreuses actions de nombreux individus, y compris des décisions prises à la fois par les concepteurs et les utilisateurs finaux. Ce dernier cas peut se produire lorsque les interactions des citoyens sont utilisées comme données pour alimenter les résultats futurs (Young, 2002).

Bien sûr, cela n'exclut pas la possibilité que certains résultats constituent une violation des droits humains, par exemple lorsqu'ils sont directement insultants, humiliants, nuisibles, haineux ou abusifs envers des individus ou des groupes (voir également (Hacker et al., 2025)). Ces observations illustrent la manière dont les technologies doivent être placées et évaluées dans le cadre d'un réseau complexe et interconnecté d'éléments socio-techniques (Naudts et al., 2025) . Il est donc important que, lorsque des modèles tiers sont utilisés pour éclairer la prise de décision tant privée que publique, les logiques et les interactions qui sous-tendent leur fonctionnement, ainsi que les hypothèses normatives et les conséquences que ces choix impliquent pour les individus et la société dans son ensemble, ne puissent être cachées aux décideurs et aux citoyens. Dans des conditions d'opacité telles, les parties prenantes ne peuvent pas évaluer la proportionnalité de leurs décisions, car les avantages peuvent être projetés, mais les inconvénients potentiels auxquels les utilisateurs concernés pourraient être confrontés restent cachés. Cela mettrait non seulement en péril la légitimité (démocratique) des politiques fondées sur l'IA, mais aussi leur contestation significative par le grand public, y

compris les citoyens, les organismes de promotion de l'égalité et les institutions de défense des droits humains.

Les données qui éclairent les décisions et les interactions de l'IA générative sont variées et, par conséquent, l'injustice (socio-économique) qu'elles introduisent ne peut pas toujours être délimitée par rapport à des groupes sociaux importants ou historiquement défavorisés, mais peut affecter des individus et des groupes dans de nouvelles dimensions, telles que leurs préférences ou d'autres comportements pouvant être surveillés uniquement par le biais de technologies (Gerards & Zuiderveen Borgesius, 2022 ; Naudts, 2023, 2024 ; Wachter, 2023). Dans le cadre de ces discussions, l'accent mis sur un seul axe par les lois européennes en matière de non-discrimination est souvent souligné comme un obstacle. Étant donné que le positionnement des personnes dans les structures de données n'est généralement pas déterminé par un seul point de données, mais plutôt par plusieurs points en constante évolution, dans certains domaines, l'inconvénient subi par les personnes peut ne pas être attribué à une seule caractéristique, mais à plusieurs. Une perspective intersectionnelle, qui reconnaît que les préjudices discriminatoires se forment à la confluence de multiples axes sociaux, pourrait mieux convenir à la complexité et à la subtilité de l'ère numérique (Atrey & Dunne, 2020 ; Bartoletti & Xenidis, 2023 ; Crenshaw, 1989). Bien que le champ d'application de la législation anti-discrimination puisse être étendu pour couvrir de nouvelles formes de discrimination (l'article 14 de la CEDH est un principe ouvert), la question a été soulevée quant aux limites normatives de la législation anti-discrimination et aux types d'injustice qu'elle est censée traiter (Bartoletti & Xenidis, 2023). Toutefois, si l'objectif global est de garantir à chaque personne une chance égale de participer en tant qu'égale sur le plan social et politique aux processus démocratiques, la loi doit prévoir des mécanismes d'intervention lorsque cette ambition globale est menacée. À cet égard, les droits humains pourraient être mis à profit pour abolir les structures d'oppression et de domination fondées sur les données. Dans ce contexte, les actions positives doivent cibler les schémas historiques de discrimination qui, depuis trop longtemps, ont « exclu ou marginalisé la voix et l'influence de certains groupes tout en amplifiant l'influence d'autres groupes » (Young, 2002).

3.1.3 Conclusion

Bien qu'il y ait beaucoup à dire sur les limites et les défis auxquels sont confrontées les lois anti-discrimination, les observations susmentionnées indiquent que les structures décisionnelles qui s'appuient sur les résultats produits par l'IA générative devraient réfléchir à la capacité des infrastructures numériques associées à institutionnaliser et à ancrer davantage les stéréotypes et les vulnérabilités, car ceux-ci aggravent les désavantages socio-économiques subis par les personnes et, par conséquent, l'impact global des décisions politiques envisagées (Helberger et al., 2022 ; Malgieri & Niklas, 2020 ; Naudts, 2024). En outre, nous soulignons l'importance des obligations positives de l'État en matière de promotion de la surveillance (publique) et de la responsabilité, dans le cadre desquelles les risques systémiques pour la démocratie sont également évalués sur la base des dommages cumulatifs que l'IA générative peut produire, et dans le cadre desquelles les acteurs privés, dans certaines circonstances, sont considérés comme des acteurs essentiels dans la formation des structures démocratiques de la société.

3.2 IA générative et liberté d'expression

Dans une démocratie qui fonctionne, les droits à la participation politique et à la non-discrimination sont invariablement liés et interdépendants avec le droit à la liberté d'expression qui, selon la Cour européenne de justice, « constitue l'un des fondements essentiels de toute société démocratique, l'une des conditions fondamentales de son progrès et du développement de chaque individu (paragraphe 61). Ou, pour reprendre les termes de Dworkin, une démocratie équitable exige « que chaque citoyen ait non seulement une voix, mais aussi une parole : une décision prise à la majorité n'est équitable que si chacun a eu la possibilité d'exprimer ses attitudes, ses opinions, ses craintes, ses goûts, ses présupposés, ses préjugés ou ses idéaux, non seulement dans l'espoir d'influencer les autres (même si cet espoir est d'une importance cruciale), mais aussi simplement pour confirmer son statut d'acteur responsable, plutôt que de victime passive, de l'action collective » (Dworkin, 1994) . Dans les paragraphes suivants, nous analyserons certaines des tendances et pratiques décrites à la section 2 sous l'angle de la liberté d'expression.

3.2.1 Ajouter la dimension des droits humains : l'IA générative du point de vue du droit à la liberté d'expression

Le droit à la liberté d'expression prévu à l'article 10 de la CEDH comprend le droit de recevoir, de détenir et de diffuser des informations et des idées, la protection de la liberté d'expression, y compris l'expression politique, contre toute ingérence et toute censure, ainsi que les garanties institutionnelles de la presse et les conditions de son fonctionnement, telles que le pluralisme des médias. Les types de discours protégés par l'article 10 de la CEDH comprennent les textes, mais aussi les images, les vidéos et autres formes d'expression d'idées ou d'informations, quelle que soit la forme sous laquelle elles sont exprimées (Bychawska-Siniarska, 2017). Il est important de noter que, dans le contexte donné, la liberté d'expression peut également s'étendre aux moyens (techniques) de production et de diffusion de l'information (Helberger et al., 2020) , ce qui a pour conséquence que la liberté d'expression peut également s'étendre à l'utilisation de moyens techniques, y compris les outils basés sur l'IA pour la production et la diffusion de contenu. Plus récemment, Bouché et al. ont fait valoir que l'IA générative devrait également être considérée comme une technologie de communication, car les applications d'IA générative peuvent « réduire encore davantage les frontières entre la sphère éditoriale et l'arrière-plan matériel des médias » (Bouché et al., à paraître) .

Opportunités et risques

Comme l'a très clairement démontré la section 2, l'IA générative peut offrir aux citoyens ainsi qu'à d'autres acteurs sociaux un outil puissant pour faire entendre leur voix et exercer ainsi leur droit à la liberté d'expression. Dans un contexte où de plus en plus de citoyens, mais aussi de partis politiques ou de représentants de la société civile, ont accès à de puissants modèles d'IA générative, les obstacles à la création de contenus à diffuser sont considérablement réduits. Les services de traduction et de vérification orthographique sont un exemple évident de la manière dont l'IA générative peut aider les citoyens à faire entendre leur voix et à égaliser les chances de le faire. Il y a ensuite l'utilisation de l'IA générative dans le cadre du journalisme citoyen ou du reportage afin de contourner le contrôle algorithmique des discours sur les plateformes de réseaux sociaux (voir section 2.1). L'IA générative a été

utilisée avec succès comme stratégie numérique pour éviter le silence et retrouver une présence politique. Au Venezuela, par exemple, les journalistes se sont tournés vers l'IA générative pour créer des présentateurs journalistiques virtuels. Ce faisant, les journalistes peuvent garantir la pluralité dans les reportages médiatiques, tout en protégeant leur propre identité contre les poursuites judiciaires (Phillips & Torres, 2024). De nouvelles formes émergentes de journalisme et d'accès du public au journalisme (voir section 2.1), mais aussi l'utilisation de l'IA générative dans les campagnes politiques (sections 2.1 et 2.2) peuvent également faciliter et même démocratiser l'accès à l'information et la participation politique. Si de nouvelles formes plus attrayantes de diffusion de contenus médiatiques invitent des segments de la société qui, jusqu'à présent, avaient tendance à éviter l'actualité ou étaient même exclus (par exemple en raison de barrières linguistiques ou d'un handicap), cela constitue une nouvelle victoire pour la liberté d'expression. Cela suppose bien sûr qu'il existe suffisamment d'espace et d'incitations pour expérimenter et promouvoir l'innovation sociale. En effet, l'inverse peut également être vrai, et l'IA générative peut tout aussi bien être utilisée pour limiter la liberté d'expression, par exemple par la censure, en noyant les voix dans une cacophonie de contenus synthétiques et de désinformation, réduisant ainsi la confiance dans la sphère publique et les institutions chargées de la liberté d'expression. Du point de vue de la liberté d'expression, il est donc également problématique de constater la dissonance apparente entre, d'une part, l'utilisation professionnelle croissante de l'IA générative dans le journalisme et les pratiques politiques et, d'autre part, les nombreuses préoccupations que les citoyens continuent d'avoir en matière de protection de leur vie privée, leur incapacité à distinguer les contenus synthétiques des contenus non synthétiques et leurs craintes d'abus (section 2.1.3). L'autonomisation par l'IA va au-delà de l'enseignement aux citoyens de l'utilisation de l'IA générative, elle implique également de les associer à l'élaboration des solutions qui doivent les aider à faire progresser leurs droits. Cela d'autant plus que l'IA générative remodèle structurellement les pratiques communicatives et la sphère publique, depuis les pratiques communicatives individuelles jusqu'à la forme de l'infrastructure de communication sous-jacente.

L'autre dimension de la liberté d'expression, la liberté de pensée, a jusqu'à présent reçu moins d'attention (O'Callaghan et al., 2024) mais gagnera en importance à mesure que les modèles d'IA générative deviendront plus persuasifs et seront utilisés pour persuader et manipuler. L'utilisation de l'IA générative dans les campagnes politiques en est un exemple, mais aussi dans la publicité commerciale (car la liberté d'expression englobe la communication commerciale). Bien avant que l'IA générative ne soit largement disponible, le Conseil de l'Europe avait déjà mis en garde, dans sa Déclaration sur les capacités manipulatrices des processus algorithmiques, contre le fait que « des niveaux fins, subconscients et personnalisés de persuasion algorithmique peuvent avoir des effets significatifs sur l'autonomie cognitive des individus et leur droit de se forger une opinion et de prendre des décisions indépendantes » (CoE, 2019). Le Conseil a poursuivi en faisant remarquer que non seulement ces processus peuvent affaiblir l'exercice et la jouissance des droits humains individuels, mais qu'ils peuvent également conduire à l'érosion des fondements mêmes du Conseil de l'Europe. Ses piliers centraux que sont les droits humains, la démocratie et l'État de droit reposent sur la conviction fondamentale que tous les êtres humains sont égaux et dignes en tant qu'agents moraux indépendants. Comme le souligne Bayer, les formes de (micro)ciblage assisté par l'IA peuvent causer un double préjudice aux citoyens et aux électeurs, en violant les droits à la vie privée, à l'autonomie individuelle et le droit de ne pas

recevoir d'informations, ainsi qu'en excluant d'autres de l'accès aux contenus médiatiques ou à la communication politique (Bayer, 2020).

Dimension institutionnelle de la liberté d'expression

La citation du Conseil de l'Europe ci-dessus rappelle avec force que le droit à la liberté d'expression n'est pas seulement un droit individuel, mais qu'il revêt également une dimension institutionnelle, car il protège aussi les institutions qui permettent aux citoyens d'exercer leur droit à la liberté d'expression. Il s'agit avant tout du rôle de la presse libre et du journalisme indépendant, mais aussi de la fonction démocratique des partis politiques. Selon la doctrine des droits humains, la presse occupe traditionnellement une place particulière dans une démocratie, car c'est à elle qu'il appartient de diffuser les informations que le public a le droit de recevoir (*Sunday Times c. Royaume-Uni*, 1979). La résilience d'une démocratie est donc inévitablement liée à la santé de l'écosystème médiatique d'une société. Et c'est en raison de ce rôle démocratique que la liberté des médias et des journalistes individuels de diffuser des informations est protégée par la Constitution (Oster, 2015, p. 22). Comme le montrent les sections 2.1.2 et 2.1.5, l'IA générative peut avoir une incidence non seulement sur des processus journalistiques concrets, tels que la recherche d'un article, la rédaction de titres ou même la production. Elle peut également avoir des implications pour la presse en tant qu'institution, par exemple si sa prolifération affecte la confiance et la viabilité démocratique de la presse, ou si les pratiques des développeurs de technologies affectent sa viabilité économique. Du point de vue de la théorie de la liberté d'expression, il apparaît clairement que les défis posés à la viabilité économique du journalisme, en particulier des médias indépendants et de petite taille, ne sont pas seulement un problème de politique industrielle, mais aussi un signal d'alarme annonçant (une nouvelle) crise démocratique imminente.

En effet, outre la diffusion d'informations que le public a le droit de recevoir, la mise à disposition d'une sphère publique et le contrôle des pouvoirs publics sont souvent décrits comme d'autres fonctions démocratiques essentielles (Wetenschappelijke Raad voor Regeringsbeleid (WRR), 2024). Ces deux fonctions sont sous pression, et encore plus sous l'impact de l'IA générative. En ce qui concerne la capacité à contrôler le pouvoir, l'IA générative et sa capacité à traiter des quantités inimaginables de données offrent aux journalistes d'investigation un outil puissant pour de nouvelles formes de journalisme assisté par les données et l'IA. Les exemples abondent, comme les récents prix Pulitzer décernés à des journalistes qui ont utilisé l'IA générative dans leurs reportages de guerre. Pourtant, l'utilisation de l'IA générative dans le journalisme est une arme à double tranchant, car elle s'accompagne, comme nous l'avons démontré dans la section 2.1.5, d'une profonde dépendance structurelle vis-à-vis d'une poignée d'entreprises technologiques et des infrastructures informatiques qu'elles contrôlent. Le rôle du journalisme, qui consiste à demander des comptes au pouvoir, s'étend aux entreprises qui exercent un pouvoir économique et politique de plus en plus important dans notre démocratie. Et pourtant, il existe un réel danger que la dépendance technique se traduise par une réduction de l'autonomie et de l'indépendance de la presse.

Les différentes façons dont les entreprises technologiques favorisent mais aussi menacent cette sphère publique ont été largement discutées dans plusieurs recommandations du Conseil de l'Europe (voir également la section 4.1). C'est là que la dynamique de concentration du pouvoir économique, technologique, politique et culturel est la plus visible,

une dynamique qui pourrait être encore renforcée par l'essor de l'IA générative et les processus d'intégration verticale qui en résultent. Comme le montre la section 2.1.5, les acteurs qui contrôlent certaines des plateformes de médias sociaux les plus influentes au monde sont également les principaux moteurs du développement des principaux systèmes d'IA générative et des LLM.

Les efforts politiques et juridiques visant à remédier à ces déséquilibres de pouvoir et à ces dépendances n'ont jusqu'à présent apporté que peu de solutions significatives. Les plus grands groupes de médias ont pris leurs propres mesures de protection, soit en concluant des accords de licence avec des entreprises d'IA, comme l'ont fait Springer et OpenAI, soit en intentant des poursuites judiciaires pour défendre leurs droits d'auteur, comme dans le cas du *New York Times* contre OpenAI et Microsoft.² Cependant, ces stratégies ne restent accessibles qu'aux acteurs disposant de ressources importantes, laissant les petits médias sans moyens de protection comparables. Cela met une nouvelle fois en évidence les problèmes structurels plus profonds expliqués précédemment. De plus, les réglementations obsolètes en matière de concentration des médias et l'approbation des fusions dans ce secteur, souvent justifiées par la nécessité d'empêcher la disparition des petits médias, pourraient finalement s'avérer préjudiciables à la préservation du pluralisme des médias à long terme (Seipp et al., 2024). Dans ce contexte, les organismes de normalisation et de défense des droits humains tels que le Conseil de l'Europe jouent un rôle crucial dans l'élaboration de cadres qui soutiennent le journalisme dans le contexte de ces changements technologiques et économiques. À l'avenir, les politiques devraient se concentrer sur la promotion de la collaboration plutôt que sur la consolidation, par exemple en finançant publiquement des infrastructures numériques partagées, en permettant la négociation collective entre les organes de presse, en particulier locaux, en établissant des normes inclusives pour l'achat de technologies et en soutenant les start-ups technologiques d'intérêt public qui proposent des solutions innovantes et adaptées au journalisme par le biais de crédits d'impôt.

Devoirs et responsabilités

Il est important de reconnaître que le manque d'autonomie numérique des médias résulte en partie de choix très stratégiques de leur part de s'associer aux géants de la technologie. Du point de vue de la théorie des droits humains, la décision des médias d'utiliser s'étend à la liberté de décider quelles technologies utiliser lorsque les médias exercent leur fonction démocratique (Helberger et al., 2020 ; Oster, 2015, p. 77) . Dans cette mesure, la décision d'utiliser Facebook ou YouTube comme plateforme de distribution centrale, ou de construire de futurs pipelines de traitement de l'information sur les produits et infrastructures d'OpenAI ou de Microsoft, fait partie de la liberté des médias de décider comment informer et transmettre l'information. D'autre part, les choix effectués par les médias peuvent avoir des implications qui ne concernent pas seulement chaque média individuellement, mais aussi la forme et l'ouverture de notre infrastructure commune d'information publique. Cela soulève la question de savoir si, pour décider de conclure ou non des accords avec Open AI & Co, la protection constitutionnelle des médias nécessite également de repenser les devoirs et responsabilités

² https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/openai-partners-news-publisher-axel-springer-meta-faces-claims-authors-infringing-copyright-2023-12-22_en ; <https://harvardlawreview.org/blog/2024/04/nyt-v-openai-the-timess-about-face/>

existants de la presse. Traditionnellement, ces responsabilités se concentrent sur le résultat de l'activité journalistique, par exemple le fait que le contenu soit produit conformément à l'éthique du journalisme (Bychawska-Siniarska, 2017 ; Wetenschappelijke Raad voor Regeringsbeleid (WRR), 2024). À l'avenir, on pourrait faire valoir que les devoirs et les responsabilités des médias devraient aller au-delà de la production responsable de contenu et inclure également l'utilisation responsable de la technologie et des choix technologiques responsables. Ailleurs, nous avons fait valoir que les devoirs et les responsabilités peuvent également inclure l'obligation de développer une éthique journalistique pour l'utilisation de l'IA (Bastian et al., 2021) , ce qui est de plus en plus le cas (Becker et al., 2023 ; Cools & Diakopoulos, 2024) . Mais les devoirs et les responsabilités peuvent également être considérés comme impliquant que les choix technologiques effectués par les médias contribuent à créer une infrastructure d'information publique qui respecte les valeurs démocratiques fondamentales (telles que le respect des droits humains), mais aussi des intérêts sociétaux plus larges, par exemple en tenant compte des besoins d'information d'une population diversifiée, de la protection de l'environnement ou en s'abstenant de toute exploitation abusive des auteurs, des médias, des travailleurs et de la planète.

Pluralisme et diversité

Le pluralisme est un concept central dans la théorie de la liberté d'expression et de la démocratie (« Il ne peut y avoir de démocratie sans pluralisme », *Centro Europa 7 S.R.L. et Di Stefano c. Italie*, 2012, p. 129). C'est également un concept central pour examiner les implications de l'IA générative pour la démocratie. Au niveau des modèles d'IA générative, l'aplatissement de la diversité et l'incapacité à représenter la diversité culturelle ou linguistique tant dans les données d'entraînement que dans le modèle affectent non seulement la qualité des résultats des modèles d'IA générative, mais risquent également d'ancrer des biais structurels tout au long de la chaîne de valeur de l'IA générative.

Au niveau des systèmes et des applications de IA générative, comme l'hyper-personnalisation des informations ou les messages politiques ciblés, la diversité est potentiellement menacée par la réduction de l'exposition à la diversité, en particulier si les citoyens se voient présenter une réponse unique et convaincante sans avoir besoin d'en chercher d'autres (voir section 2.1), par l'invisibilisation d'autres perspectives ou par des réponses biaisées (voir section 2.2.5 démontrant comment les chatbots peuvent favoriser uniquement une petite sélection de partis politiques).

Enfin, au niveau structurel, l'émergence, respectivement le renforcement des rôles de gardien et la concentration du pouvoir économique, politique et technologique entre les mains de quelques-uns sont en conflit direct avec les principes fondamentaux de la démocratie. Ou, pour reprendre les termes d'Edwin Baker : « [Un] pays n'est démocratique que dans la mesure où les médias, ainsi que les élections, sont structurellement égalitaires et politiquement importants. La meilleure interprétation institutionnelle de cette vision démocratique de la sphère publique est, selon moi, une répartition égalitaire du contrôle... Il s'agit d'un principe de répartition démocratique du pouvoir de communication – une affirmation selon laquelle la démocratie implique une dispersion aussi large que possible du pouvoir dans le discours public » (Baker, 2006) .

3.2.2. Le potentiel de la doctrine de la liberté d'expression pour contribuer à la gouvernance démocratique de l'IA

Création d'un cadre conceptuel d'évaluation des risques

Le droit à la liberté d'expression peut servir de référence importante pour identifier les cas dans lesquels les applications et les systèmes d'IA générative sont bénéfiques ou problématiques, à savoir lorsque leur développement et leur déploiement compromettent les conditions nécessaires à l'exercice du droit à la liberté d'expression. Le droit à la liberté d'expression peut également aider à expliquer pourquoi certaines applications de l'IA générative méritent davantage d'attention réglementaire que d'autres, à savoir dans la mesure où elles sont plus proches du cœur même de la réalisation du droit à la liberté d'expression. Par exemple, la protection constitutionnelle de la liberté d'expression s'étend également aux propos qui « choquent ou offensent » (Handyside, paragraphe 49). En ce sens, même l'utilisation de l'IA générative pour créer des contenus polarisants, voire choquants (voir section 2.2), peut toujours relever de la protection du droit à la liberté d'expression, car elle peut s'inscrire dans le cadre du discours démocratique. La limite du discours inadmissible est franchie lorsqu'il met en péril la tolérance et le respect de la dignité égale de tous les êtres humains, qui sont, selon la Cour, le fondement d'une société démocratique et pluraliste. Selon la CEDH, « il peut être nécessaire dans les sociétés démocratiques de sanctionner, voire d'empêcher, toutes les formes d'expression qui propagent, incitent, encouragent ou justifient la haine fondée sur l'intolérance » (*Féret c. Belgique*, 2009, paragraphe 64), paragraphe 64). Il s'agit là d'un exemple illustrant comment l'article 10 de la CEDH peut aider à identifier les formes d'utilisation de l'IA générative qui ne sont pas acceptables sur le plan social et qui nécessitent une action politique précisément parce qu'elles sont contraires aux principes fondamentaux de la démocratie.

Le Conseil de l'Europe a déjà franchi une étape importante vers la concrétisation et la systématisation des implications de l'IA générative sur la liberté d'expression dans sa Note d'Orientation sur l'IA générative et la liberté d'expression. La Note examine les implications de l'IA générative sur la liberté d'expression à chaque niveau de la pile d'IA générative. Cela inclut la couche fondamentale et la manière dont des informations biaisées mais aussi inexactes peuvent affecter le droit de recevoir des informations ou mettre en danger la diversité linguistique ; la couche « outils » et le potentiel de modération et de filtrage des contenus pour supprimer et censurer la liberté d'expression, ainsi que la couche « produits », où des applications telles que les chatbots ou les agents IA peuvent faciliter l'accès à l'information, mais aussi entraîner une manipulation, un renforcement des stéréotypes ou des inexactitudes, ainsi que des implications structurelles, par exemple en omettant d'afficher les sources et en s'appropriant des contenus médiatiques pour développer des passerelles d'information concurrentes. Les tentatives de systématisation et d'illustration des défis importants qui pèsent sur le droit à la liberté d'expression, tant au niveau individuel qu'institutionnel, peuvent éclairer les pratiques d'évaluation des risques, mais aussi donner aux universitaires et à la société civile les moyens d'auditer les LLM existants (voir par exemple l'exemple intéressant de (Calvet-Bademunt et al., 2025, qui examine les lois nationales et les pratiques des entreprises régissant l'IA afin de vérifier leur compatibilité avec la liberté d'expression). Cela pourrait également permettre la création d'un cadre de surveillance et de signalement efficace, soit au sein du Conseil, soit en habilitant et en mettant en place des réseaux de surveillance

composés d'universitaires, de membres de la société civile, de vérificateurs de faits et de régulateurs.

De la même manière, la théorie de la liberté d'expression peut nous éclairer sur les utilisations responsables/d'intérêt public de l'IA générative, par exemple lorsque celle-ci est utilisée pour donner aux citoyens les moyens de diffuser des informations, d'atteindre des groupes de la société qui sont généralement exclus ou moins susceptibles d'avoir accès à un journalisme indépendant ou à des informations politiques, ou pour promouvoir l'accessibilité de l'information, car toutes ces actions peuvent créer un environnement propice à la réalisation de la liberté d'expression et pourraient donc également relever des obligations de protection positive des États membres.

Mettre en avant la perspective de l'innovation d'intérêt public

Pour les gouvernements, la liberté des médias impliquait jusqu'à présent à la fois l'obligation de s'abstenir de toute ingérence et de toute censure (l'IA générative crée de nouveaux outils pour y parvenir) et l'obligation positive de créer les conditions permettant aux médias d'exercer leur liberté et aux citoyens de bénéficier de la liberté d'expression (*Dink c. Turquie*, 2010). On peut soutenir que cette dernière dimension – les obligations positives – devient de plus en plus importante en raison des immenses asymétries de pouvoir entre les citoyens et certaines plateformes et entreprises technologiques, ainsi que du potentiel transformateur de l'IA générative. Au cours de la dernière décennie et à diverses occasions, le Conseil de l'Europe a souligné et expliqué le rôle des États dans la création d'un environnement informationnel favorable.

Traditionnellement, ces « obligations positives » des États en matière de protection et de promotion de la liberté des médias allaient de la mise en place d'un paysage médiatique résilient et pluraliste à l'établissement de privilèges journalistiques tels que l'accès aux informations détenues par les pouvoirs publics ou la protection des sources, en passant par le soutien financier et non financier aux journalistes et la protection de la sécurité des journalistes et des citoyens afin qu'ils puissent exprimer leurs idées et leurs opinions sans crainte de répression (Pentney, 2024). Compte tenu du potentiel de l'IA générative pour contribuer également à la réalisation des droits à la liberté d'expression ou à la transformation sociétale de la presse, ces obligations positives pourraient également inclure des mesures visant à encourager et à permettre l'autonomie numérique et les infrastructures alternatives, ou à inciter à l'innovation médiatique orientée vers l'intérêt public, par exemple par le biais de programmes de financement, d'incitations fiscales, de bacs à sable, de mandats de service public étendus et de la concrétisation du rôle et de la mission des médias de service public, ou d'autres formes de réglementation fondée sur des incitations. Comme l'affirment Bouché et al., « en reconnaissant le rôle des infrastructures médiatiques (technologiques) et du droit dans la formation de l'opinion publique, les attentes idéales concernant le rôle de l'État dans la protection du pluralisme devraient tenir compte de la création de conditions juridiques et infrastructurelles qui soutiennent la multiplicité des « publics » et leurs différentes aspirations » (Bouché et al., à paraître).

Outre les États, les plateformes et les entreprises technologiques elles-mêmes entrent en scène en tant que destinataires des obligations en matière de droits humains. Alors que traditionnellement, les droits humains s'appliquaient dans le cadre des relations entre le

gouvernement et les citoyens, les recommandations plus récentes du Conseil de l'Europe incluent de plus en plus les grands fournisseurs de technologies, et en particulier les plateformes de médias sociaux, dans le cercle des entités tenues de respecter et de protéger les droits humains des citoyens, y compris la liberté d'expression. L'approche fondée sur les risques dans la Convention-cadre en est un exemple, mais comme nous le verrons plus loin (section 4.2), la Convention-cadre est plus ambitieuse et souligne explicitement la nécessité de réaliser le potentiel des technologies d'IA pour créer de la valeur pour la société en promouvant et en faisant progresser la réalisation des droits humains et des principes démocratiques.

4. Cadre juridique du Conseil de l'Europe

4.1 Situer l'IA générative dans les recommandations du Conseil de l'Europe

Les recommandations du Conseil de l'Europe ont une longue histoire d'adaptation des normes relatives aux droits humains afin de tenir compte des évolutions technologiques. Certains instruments se concentrent sur des technologies spécifiques, comme les recommandations sur les intermédiaires ou les systèmes algorithmiques (CoE, 2018, 2020). Plus généralement, les recommandations analysent le rôle de la technologie dans le cadre d'une analyse globale d'une question spécifique, telle que les discours de haine ou les élections (CoE, 2022a, 2022b).

Les recommandations du Conseil de l'Europe n'ont pas encore abordé directement l'impact de l'IA générative sur la démocratie, bien qu'une note d'orientation ait été élaborée en 2025 sur les implications de l'IA générative pour la liberté d'expression. Les recommandations se sont plutôt concentrées sur la technologie utilisée pour diffuser l'information. Par exemple, les recommandations ont analysé les responsabilités des intermédiaires en matière de suppression des discours haineux diffusés par leurs systèmes (CoE, 2022b, paragraphe 30), de garantie du respect de la vie privée des utilisateurs par leurs technologies de publicité ciblée (CoE, 2022a, paragraphe 5), et de promotion d'une information plus diversifiée et d'un journalisme de haute qualité (CoE, 2018, paragraphe 2.5, 2022a, 2022b).

Les recommandations ont analysé les précurseurs et les technologies étroitement liés à l'IA générative. Par exemple, en 2017, le Comité des ministres a adopté une recommandation sur les mégadonnées au service de la culture, du savoir et de la démocratie (CoE, 2017). Les grands volumes de données sont, bien sûr, une ressource clé qui permet les progrès actuels de l'IA générative, et la recommandation appelait déjà les États membres à améliorer la culture numérique des utilisateurs et à mettre en place des processus multipartites pour orienter l'avenir du Big Data dans le respect des droits humains. Cependant, la recommandation se concentrait sur l'archivage et l'amélioration de l'accessibilité des données culturelles, ce qui n'a qu'une pertinence limitée pour les défis et les opportunités actuellement créés par l'IA générative. La recommandation sur l'impact des systèmes algorithmiques sur les droits humains est sans doute celle qui se rapproche le plus d'une approche directe de l'IA générative. Sa définition large des systèmes algorithmiques (qui apparaît également dans d'autres recommandations (CoE, 2022a, paragraphe 4, 2022c, paragraphe 7) les définit comme « des applications qui, utilisant souvent des techniques d'optimisation mathématique,

effectuent une ou plusieurs tâches telles que la collecte, la combinaison, le nettoyage, le tri, le classement et l'inférence de données, ainsi que la sélection, la hiérarchisation, la formulation de recommandations et la prise de décision » (CoE, 2020, paragraphe 2). Bien que cette définition ne fasse pas directement référence aux algorithmes qui génèrent des données, elle englobe bon nombre des étapes (collecte, combinaison, déduction) de l'IA générative. Plus important encore, le contenu même de la recommandation souligne l'importance de l'infrastructure sous-jacente et des ensembles de données utilisés dans les systèmes algorithmiques (CoE, 2020, paragraphe 2.2). Comme le montrent les sections ci-dessus, ces technologies continuent de jouer un rôle clé dans le contexte de l'IA générative.

En résumé, si l'IA générative est une technologie nouvelle pour les recommandations du Conseil de l'Europe, celles-ci ont une longue histoire d'adaptation des normes relatives aux droits humains aux changements technologiques et ont commencé à montrer comment le faire avec les technologies liées à l'IA générative. La section suivante s'appuiera sur ces enseignements et sur l'analyse précédente des risques et des opportunités de l'IA générative pour voir ce que nous pouvons apprendre des recommandations du Conseil de l'Europe sur le rôle que devrait jouer l'IA générative dans une société démocratique.

4.1.1 Recommandations du Conseil de l'Europe et rôle de l'IA générative dans la société démocratique

Contenus préjudiciables

L'un des principaux défis posés par l'IA générative est qu'elle peut être utilisée pour créer de faux messages discriminatoires à l'égard de groupes sociaux spécifiques (voir section 3.1). Comme l'ont expliqué en détail les sections consacrées à l'utilisation de l'IA générative en politique et dans le journalisme, l'IA générative réduit considérablement les obstacles à la production de contenus, y compris de contenus discriminatoires ou haineux. Par exemple, la section 2.2 a montré comment les acteurs politiques (en particulier ceux qui ont déjà moins de préoccupations normatives concernant l'utilisation de l'IA) peuvent utiliser cette technologie pour produire rapidement et facilement des contenus organiques plus extrêmes. Dans certains cas, ces contenus peuvent constituer des discours haineux illégaux. Si les contenus moins extrêmes n'atteignent pas nécessairement le niveau de toxicité requis pour être considérés comme des discours haineux ou du harcèlement illégaux, ils peuvent néanmoins avoir un impact négatif sur le droit à la non-discrimination et nuire à la tolérance (voir également Hacker et al. 2024).

La recommandation sur le discours de haine reconnaissait déjà le danger d'une exposition plus large aux discours haineux diffusés par des intermédiaires, soulignant « que la disponibilité persistante du discours de haine en ligne exacerbe son impact » (CoE, 2022b). Pour remédier à ce problème, la recommandation a souligné l'importance de garantir des procédures de retrait rapides et des exigences de diligence raisonnable pour les intermédiaires. Ces procédures sont sans doute encore plus importantes dans la mesure où l'IA générative facilite la production à grande échelle de discours haineux qui peuvent être diffusés par des intermédiaires et amplifiés par des algorithmes (voir la section sur les campagnes politiques). Cependant, l'argument selon lequel les entreprises technologiques ont la responsabilité de veiller à ce que leurs services ne soient pas utilisés de manière abusive s'applique également aux entreprises dont les systèmes d'IA sont utilisés pour

générer des discours haineux. Cet argument est sans doute encore plus pertinent dans ce contexte, étant donné que les entreprises d'IA générative ne se contentent pas de diffuser les discours haineux téléchargés par leurs utilisateurs, mais contrôlent la manière dont ce contenu est produit grâce à la conception de leurs modèles. Combinées, la recommandation sur le discours de haine et les recommandations qui traitent plus largement des obligations des entreprises technologiques en matière de droits humains fournissent de nombreux points de départ pour mettre en œuvre cette responsabilité (CoE, 2018b, 2020, 2022a, 2022e). Par exemple, les entreprises d'IA générative pourraient intégrer les droits humains dans leurs politiques de contenu (à l'instar des politiques de modération de contenu des intermédiaires), embaucher du personnel ayant une expertise suffisante dans les normes pertinentes en matière de droits humains, tenir compte du contexte juridique, local, culturel, sociopolitique et historique, et mettre en place des processus de diligence raisonnable qui garantissent la sécurité de leurs systèmes. Dans tous les cas, la réponse aux discours haineux doit être adaptée à leur gravité et tenir compte de la liberté d'expression.

L'IA générative peut également amplifier involontairement les préjugés sociaux lorsqu'elle est utilisée par des acteurs qui n'ont pas l'intention de produire des contenus préjudiciables. Par exemple, comme le montre la section 2.1, les journalistes qui utilisent l'IA générative pour collecter des données ou rédiger des textes peuvent involontairement adopter les stéréotypes intégrés dans ses données d'entraînement. Dans ce contexte, la recommandation sur le discours de haine souligne que les fonctionnaires, les partis politiques et les médias doivent éviter de diffuser le discours de haine ou des expressions qui encouragent l'intolérance (CoE, 2022b, paragraphes 28 et 41). La réglementation de ces expressions est délicate, car la liberté d'expression des médias et des acteurs politiques, y compris de manière offensive, est fortement protégée par l'article 10 de la CEDH (voir section 3.2). Toutefois, comme indiqué précédemment, les entreprises technologiques ont l'obligation d'assurer la sécurité de leurs systèmes, ce qui inclut sans doute la responsabilité de veiller à ce que leurs systèmes n'incitent pas leurs utilisateurs à produire involontairement des contenus préjudiciables. En outre, les initiatives d'éducation qui alertent les utilisateurs sur le danger que l'IA générative recrée des stéréotypes préjudiciables peuvent prévenir une utilisation abusive involontaire de l'IA générative. Enfin, la recommandation sur le discours de haine souligne l'importance des contre-discours. Les acteurs politiques peuvent condamner ouvertement l'utilisation de l'IA pour produire des discours haineux, tandis que les médias peuvent faire contrepoids en fournissant des informations fiables et diversifiées.

Manipulation

L'IA générative peut également nuire à la capacité des individus à prendre des décisions éclairées. Comme l'ont montré les sections précédentes, l'IA générative peut être utilisée pour manipuler les choix politiques des individus, en créant par exemple des messages adaptés aux vulnérabilités et aux caractéristiques spécifiques d'un individu. Les recommandations existantes du Conseil de l'Europe, en particulier la Recommandation sur la communication électorale et la couverture médiatique des campagnes électorales, la Déclaration sur les capacités de manipulation des processus algorithmiques et les Lignes directrices relatives à la protection des personnes à l'égard du traitement des données à caractère personnel dans le cadre des campagnes politiques, fournissent déjà des orientations utiles sur les cas dans lesquels une telle communication politique personnalisée doit être considérée comme inadmissible (CoE, 2019, 2021a, 2022c). Dans ce contexte, l'un des principaux

enseignements à tirer pour les régulateurs est de fonder toute adaptation de ces normes sur la manière dont l'IA générative est utilisée dans la pratique à des fins de manipulation hyperpersonnalisée. Lors de la précédente itération de ce débat, les craintes concernant le potentiel des publicités politiques microciblées hautement efficaces n'ont pas été confirmées par les données disponibles sur leur utilisation et leur impact réels. Adopter une approche similaire à l'égard de l'IA générative détournerait l'attention des régulateurs de risques sans doute plus pressants.

L'IA générative peut également nuire à la capacité des individus à s'informer. Comme le montre la section 2.2, en réduisant les obstacles à la création de contenu, l'IA générative permet aux acteurs de produire de grands volumes d'informations de mauvaise qualité (également appelées « slop ») ainsi que des informations inexactes (également appelées « hallucinations »). Du point de vue des recommandations du Conseil de l'Europe, il est utile de distinguer deux implications différentes de l'abaissement des barrières à la production de contenu. Premièrement, l'IA générative peut être intentionnellement détournée de son usage pour nuire au débat démocratique en permettant aux acteurs de produire plus facilement de grandes quantités de désinformation. La recommandation sur la communication électorale a déjà souligné le danger que représentent les bots et les faux comptes utilisés pour diffuser de la désinformation pendant les campagnes. Pour remédier à ces problèmes, la recommandation a souligné l'importance de surveiller les campagnes de désinformation et a appelé les intermédiaires à garantir l'intégrité de leurs services tout en tenant compte de la liberté d'expression (CoE, 2023a, paragraphe 21). L'IA générative renforce l'importance de telles mesures (2023a, par. 21) ; conformément à l'analyse des contenus préjudiciables ci-dessus, elle exige sans doute également des fournisseurs de modèles d'IA générative qu'ils garantissent la sécurité de leur technologie.

Deuxièmement, l'IA générative facilite la production de grands volumes d'informations de mauvaise qualité. Le problème dans ce contexte n'est pas tant que les individus se fient à des informations fausses. C'est qu'en facilitant la production d'informations qui semblent intéressantes mais qui ont peu de valeur informative, l'IA générative rend beaucoup plus difficile pour les individus de trouver (et par conséquent pour les médias de produire et de financer) des informations de haute qualité. Concrètement, le journalisme d'investigation coûte beaucoup de temps et d'argent. La création d'une vidéo attrayante à l'aide de l'IA générative n'en coûte pas autant. De plus, l'IA générative permet aux entreprises spécialisées dans l'IA de produire elles-mêmes des résumés d'actualités. Même lorsque ces résumés renvoient aux articles d'actualité sous-jacents, rien ne garantit que les utilisateurs suivront ces liens vers la source originale (et ils ont d'ailleurs peu d'intérêt à le faire). Et si les entreprises d'IA ont commencé à obtenir des licences pour utiliser le contenu des médias d'information dans leurs modèles d'IA, ces accords n'ont profité qu'à quelques éditeurs et leurs conditions ne sont pas accessibles au public.

Les recommandations précédentes du Conseil de l'Europe ont déjà largement mis en garde contre les effets négatifs du « rythme effréné du partage d'informations en ligne » facilité par les intermédiaires, soulignant qu'il compromette financièrement d'un journalisme de qualité, la profondeur et l'exactitude des reportages, ainsi que l'accès du public à des informations fiables (CoE, 2018a, 2022d). Malheureusement, aucune solution ne se dégage. Si le contrôle exercé par les entreprises technologiques sur l'attention (et les revenus qui en découlent) est largement reconnu comme problématique, la manière d'organiser cette attention et ces

revenus des médias reste controversée. Cela vaut tout particulièrement pour les mécanismes qui obligerait les entreprises technologiques à garantir la visibilité du contenu des médias sur leurs plateformes (CoE, 2021b ; OSCE, 2025). Néanmoins, la recommandation relative à la création d'un environnement favorable au journalisme de qualité met en avant plusieurs outils politiques qui pourraient être plus facilement adaptés pour améliorer les relations entre les entreprises d'IA et les médias, notamment la nécessité de mettre à jour la législation sur la concentration des médias (par exemple, pour permettre aux médias de négocier collectivement avec les entreprises d'IA) et de garantir le partage des données (afin que les médias sachent comment leur contenu est utilisé et monétisé par les entreprises d'IA) (CoE, 2022d). Au-delà du rétablissement des relations entre les médias et les entreprises technologiques, les États peuvent mobiliser les mécanismes déjà existants pour financer et produire du journalisme, tels que les avantages fiscaux, le financement direct du journalisme d'investigation et le soutien aux médias de service public (CoE, 2022d, paragraphe 1.3). Dans de nombreux pays, ces mécanismes sont déjà disponibles et peuvent donc être mis en œuvre relativement rapidement. Ils présentent en outre l'avantage de consolider la relation directe entre les médias et leur public, plutôt que d'accroître leur visibilité sur les services contrôlés par les entreprises d'IA.

Infrastructure

L'infrastructure est un aspect essentiel mais souvent invisible de l'IA générative, comme l'a montré en détail la section 2.c. Le contrôle de l'infrastructure technologique nécessaire au déploiement et au développement de l'IA générative est aujourd'hui concentré entre les mains de quelques entreprises technologiques. Cela confère à ces acteurs le pouvoir de façonner la manière dont l'IA générative peut être utilisée et développée, voire de restreindre l'accès à des infrastructures critiques. Jusqu'à présent, l'infrastructure technologique n'a occupé qu'une place secondaire dans les recommandations du Conseil de l'Europe, étant traitée dans des sous-sections (CoE, 2022e, paragraphes 2.1-2.7) et de manière ponctuelle (CoE, 2020, paragraphe 2.3, 2022a, paragraphe 11). Toutefois, à partir de ces recommandations et des travaux antérieurs sur la propriété des médias et la neutralité du réseau (CoE, 2016, 2018a), il est déjà possible de dégager plusieurs principes qui devraient guider toute future réglementation des infrastructures.

Les recommandations du Conseil de l'Europe soulignent l'importance d'empêcher le contrôle centralisé des infrastructures que nous observons actuellement dans le contexte de l'IA générative. Une telle centralisation peut limiter la capacité des États à remplir leurs obligations en matière de droits humains (CoE, 2020, paragraphe 2.3) et favoriser l'émergence de pouvoirs excessifs sur l'opinion publique (CoE, 2022e, paragraphe 2.4). Afin de limiter ce contrôle centralisé, les normes soulignent l'importance d'une application et de modifications efficaces (ex ante) de la législation en matière de concurrence et de concentration des médias (CoE, 2018a, 2022e, paragraphe 2.4, 2022a, paragraphe 11). Par ailleurs, les recommandations soulignent l'importance de soutenir l'innovation menant à des infrastructures alternatives (CoE, 2020, paragraphe 2.3), par exemple grâce au soutien de l'État aux médias de service public ou à des investissements directs réalisés par les médias eux-mêmes (CoE, 2022d, paragraphe 2.4.2, 2022e, paragraphe 2.6).

En outre, les États peuvent exiger que les infrastructures soient conçues de manière ouvertes, interopérables et à permettre la portabilité des données. De telles mesures peuvent empêcher

le verrouillage des utilisateurs, qui non seulement freine la concurrence, mais les rend également plus vulnérables aux hausses de prix ou aux abus de pouvoir des fournisseurs d'infrastructures (CoE, 2022e, paragraphe 2.3). Enfin, lorsque les utilisateurs dépendent d'un fournisseur d'infrastructure, les États peuvent intervenir pour garantir des conditions d'accès équitables, raisonnables et non discriminatoires (CoE, 2022a, paragraphe 11, 2022e, paragraphe 2.2). De telles mesures peuvent empêcher (ou au moins corriger) les représailles des fournisseurs d'infrastructures à l'encontre d'acteurs spécifiques en raison de leurs opinions politiques ; plus largement, elles contribuent également à garantir que l'infrastructure nécessaire pour utiliser l'IA générative afin de contribuer au débat public soit accessible à tous (CoE, 2016).

Opportunités offertes par l'IA générative

Les recommandations du Conseil de l'Europe rappelle régulièrement l'importance d'aller au-delà de la simple gestion des risques posés par les nouvelles technologies et de créer activement les conditions permettant à la technologie de soutenir la démocratie. Elles ont par exemple mis en évidence le potentiel des technologies numériques pour accroître la portée et la diversité des informations auxquelles les individus sont exposés et pour créer de nouveaux réseaux de solidarité (CoE, 2018a, 2022e), et soulignent la nécessité pour les États d'instaurer un « environnement sûr et propice au débat public et à la communication », ainsi qu'un environnement favorable à un journalisme de qualité (CoE, 2022b, 2022d).

L'un des principaux moyens pour les États de concrétiser ces obligations positives consiste à faciliter les opportunités offertes par l'IA générative. Une telle approche ne dissuadera pas les acteurs qui cherchent délibérément à utiliser cette technologie pour porter atteinte aux valeurs démocratiques. Cependant, elle peut empêcher l'utilisation irresponsable et involontaire de l'IA générative et faciliter la tâche des acteurs (tels que les médias et les partis politiques) qui souhaitent l'utiliser d'une manière qui soutient la démocratie. Un exemple qui a été mentionné à plusieurs reprises dans ce rapport est que l'IA générative peut être utilisée pour rendre l'information plus accessible en l'adaptant aux besoins spécifiques ou aux handicaps cognitifs/physiques d'une personne.

Les États pourraient faciliter l'utilisation responsable de l'IA générative de plusieurs manières. L'une d'elles consiste à publier des orientations concrètes sur l'impact de l'IA générative sur les responsabilités des médias et sur la manière dont les médias peuvent déployer l'IA générative de manière responsable (voir notamment (CoE, 2023b)). Une autre approche consiste à stimuler l'innovation qui exploite le potentiel positif de l'IA générative (CoE, 2023b). Conformément à la section précédente, une telle stratégie devrait également tenir compte de l'infrastructure sous-jacente sur laquelle repose l'IA générative, car celle-ci conditionne le type d'innovation possible. Bien sûr, la création des conditions d'une démocratie saine ne se limite pas à encourager une meilleure utilisation de l'IA générative. Au contraire, étant donné les difficultés que présente l'IA générative pour les individus qui souhaitent distinguer les contenus authentiques des contenus non authentiques, il est sans doute d'autant plus important d'investir dans des mesures traditionnelles visant à préserver la démocratie. Dans ce contexte, il convient de souligner le rôle essentiel des médias dans la fourniture d'informations fiables et de haute qualité. Ces informations peuvent constituer un contrepoids important aux grandes quantités d'informations de mauvaise qualité ou inexactes qui peuvent être générées par l'IA.

4.1.2 Conclusion

Les normes du Conseil de l'Europe constituent sans doute un outil politique particulièrement utile pour relever les défis posés par l'IA. Comme nous l'avons vu plus haut, le principal défi que pose l'IA générative à la démocratie ne réside sans doute pas dans les nouveaux contenus à fort impact tels que les deepfakes et les discours de haine synthétiques illégaux. Il s'agit plutôt (1) des volumes importants de contenus de mauvaise qualité, polarisants et discriminatoires que l'IA générative permet de produire, et (2) de l'expansion du pouvoir des entreprises technologiques dans la production de contenus et les infrastructures sous-jacentes, et de la perte correspondante de contrôle des médias sur l'innovation technologique, le financement et l'accès aux audiences.

Les recommandations du Conseil de l'Europe peuvent contribuer à répondre à ces défis en proposant une vision plus positive, qui va au-delà de la simple tentative d'atténuer les excès de l'IA générative et clarifie à quoi pourrait ressembler une IA favorable à la démocratie. Elles peuvent également indiquer les mesures que les différentes parties prenantes (y compris les États, les médias et les entreprises d'IA elles-mêmes) peuvent prendre pour concrétiser cette vision. Proposer une telle vision est utile en soi, car elle permet de lutter contre la perception de l'IA comme un vecteur d'innovation et de croissance économique qui ne doit pas être freiné par des réglementations visant à limiter son impact sur la société démocratique. Mais cela peut également ancrer les efforts en cours pour rendre l'IA plus responsable dans les normes européennes en matière de droits humains. Sans ces orientations, le risque est grand que ces efforts privilégient les intérêts économiques à court terme au détriment des principes fondamentaux des droits humains, ou qu'ils ne reflètent pas pleinement les subtilités des droits humains européens telles que définis dans les normes du Conseil de l'Europe.

4.2 Convention-cadre sur l'Intelligence Artificielle

La Convention-cadre du Conseil de l'Europe sur l'IA est le tout premier traité international juridiquement contraignant pour les Parties visant à garantir que les systèmes d'IA respectent les droits humains, la démocratie et l'État de droit. À l'instar de la loi européenne sur l'IA et d'autres cadres comparables (comme la loi brésilienne sur l'IA en cours d'élaboration), l'objectif de la convention-cadre est de créer un cadre réglementaire horizontal applicable à toute une série de secteurs afin de garantir un développement et un déploiement responsables de l'IA. Il existe toutefois un certain nombre de différences importantes qui pourraient faire de la convention-cadre un instrument puissant pour faire progresser les droits humains, la démocratie et l'État de droit à l'ère de l'IA générative.

4.2.1 Les droits humains et la démocratie comme point de départ

La promotion de la démocratie est l'objectif central de la convention-cadre. Elle s'inscrit ainsi dans la longue tradition du Conseil de l'Europe en tant qu'organisation de défense des droits humains. Le Conseil de l'Europe écrit ailleurs : « Le lien entre les droits humains et la démocratie est profond et réciproque : chacun dépend en quelque sorte de l'autre et est incomplet sans l'autre. »³ Cela distingue la Convention-cadre d'autres cadres de gouvernance de l'IA qui mettent l'accent sur les aspects économiques et réglementaires du marché (voir

³ <https://www.coe.int/en/web/compass/democracy>, consulté le 12 novembre 2025.

également le paragraphe 5 de l'exposé des motifs de la Convention-cadre. La loi européenne sur l'IA en est un exemple : fondée sur la réglementation en matière de responsabilité du fait des produits, la loi sur l'IA vise également à protéger la démocratie et les droits humains, mais ce n'est là qu'un objectif parmi d'autres. En outre, la loi sur l'IA doit trouver un équilibre entre la protection des droits humains et la démocratie d'une part, et d'autres objectifs tels que la réalisation du marché intérieur et l'innovation d'autre part. En pratique, cela fait de la protection des droits humains et de la démocratie l'un des nombreux objectifs à concilier, la rendant ainsi vulnérable aux pressions économiques et géopolitiques. En outre, le droit de la responsabilité du fait des produits n'a pas à ce jour de tradition en matière de protection des droits humains ou de la démocratie.

4.2.2 Risques pour les droits humains et la démocratie

À l'instar de la loi européenne sur l'IA et d'autres approches comparables en matière de gouvernance, la Convention-cadre adopte une approche fondée sur les risques. Conformément à l'article 16 de la Convention-cadre, les États membres doivent mettre en place, adopter ou maintenir un cadre de gestion des risques, en mettant explicitement l'accent sur les droits humains, la démocratie et l'État de droit. On peut citer comme exemples l'utilisation de l'IA générative pour inciter à la haine ou pour maintenir et renforcer les stéréotypes et les formes structurelles de discrimination. Dans la jurisprudence de la CEDH, qui se développe sur les questions liées à l'IA et reste limitée en ce qui concerne l'IA générative, ainsi que dans les recommandations du Conseil de l'Europe, on trouve quelques orientations générales lorsque l'IA et l'IA générative entrent en conflit avec les droits humains et la démocratie. Un exemple de lignes directrices plus spécifiques liées à l'IA est la Note d'orientation sur les implications de l'IA générative pour la liberté d'expression élaborée par le Comité directeur sur les médias et la société de l'information en décembre 2025 (Conseil de l'Europe, 2025). Cette Note d'orientation systématise les défis spécifiques que les systèmes d'IA et l'IA générative peuvent poser à la liberté d'expression. Par conséquent, il serait utile de s'appuyer sur les orientations spécifiques à l'IA pour d'autres droits humains et principes démocratiques (y compris les situations où l'application de l'IA générative est incompatible avec les droits humains ou les principes démocratiques) et de les développer, ce qui permettrait au Conseil de l'Europe de tirer parti de son expertise considérable et de son mandat spécifique en matière de droits humains. Le développement de la méthodologie HUDERIA, un outil conçu spécifiquement pour identifier et atténuer les risques pesant sur la protection et la promotion des droits humains, de la démocratie et de l'État de droit, ainsi que le modèle HUDERIA proposant des mesures concrètes pour mettre en œuvre cette méthodologie, constituent des avancées importantes dans cette direction.⁴

4.2.3 Obligations positives de protéger la réalisation des droits humains et de la démocratie

Contrairement à d'autres instruments de gouvernance, tels que le Règlement européen sur l'IA, la Convention-cadre ne se contente pas d'aborder les risques pour les droits humains et la démocratie. L'article 10 de la Convention-cadre invite les parties à s'assurer que l'IA respecte non seulement l'égalité, mais aussi à adopter des mesures visant à surmonter les

⁴ Pour plus d'informations sur la méthodologie et le modèle, cliquez ici : <https://rm.coe.int/huderia-methodology-and-model-adopted-provisional-version-2026/48802ac001>

inégalités afin d'obtenir des résultats équitables et justes. Comme indiqué à la section 3.2, il est essentiel de s'attaquer aux causes sociétales sous-jacentes des pratiques discriminatoires, stéréotypées ou exclusives pour lutter contre les effets discriminatoires de l'IA générative, précisément en raison de la manière dont la technologie influence et est influencée par les dynamiques sociétales et les données sous-jacentes. Dans le même ordre d'idées, l'article 7 de la Convention-cadre exige que des mesures soient prises pour respecter et promouvoir la dignité humaine et l'autonomie individuelle, ce qui est essentiel pour lutter contre les effets manipulateurs des services (hyper)personnalisés. Après tout, ces effets ne peuvent être considérés indépendamment des modèles commerciaux sous-jacents qui sont optimisés pour exploiter les vulnérabilités individuelles et réduire les utilisateurs à des sources rentables de données et de revenus publicitaires. Cela correspond à l'approche de la Convention-cadre qui consiste non seulement à se concentrer sur les systèmes d'IA au sens technique strict, mais aussi à situer ces systèmes dans leur contexte socio-économique plus large (voir préambule).

Comme le montre le présent rapport, le développement et le déploiement de l'IA générative peuvent profiter aux institutions démocratiques telles que le journalisme et la presse indépendante, mais ils peuvent également en compromettre l'intégrité et l'indépendance. La concentration du contrôle sur les éléments clés de l'infrastructure de l'IA générative, la dépendance des institutions journalistiques à l'égard d'une poignée d'entreprises commerciales à but lucratif et les importantes asymétries de pouvoir peuvent non seulement remettre en cause le journalisme ou la communication politique en tant que pratiques et institutions démocratiques, mais aussi perturber l'équilibre des pouvoirs dans une démocratie. L'importance d'empêcher le contrôle centralisé des infrastructures de communication critiques a déjà été soulignée par le Conseil de l'Europe dans ses recommandations (voir section 4.1). Par conséquent, l'élaboration de stratégies visant à réduire cette dépendance, par exemple en encourageant les alternatives, mais aussi en appelant à la responsabilité de toutes les parties (y compris les entreprises technologiques et les institutions journalistiques), doit être un point important à prendre en compte dans la mise en œuvre des obligations découlant de la présente convention.

4.2.4 Innovation d'intérêt public

Une autre différence importante réside dans l'approche de la Convention-cadre, qui ne se limite pas à analyser les implications négatives et les risques potentiels de l'IA, mais aussi à souligner l'importance de promouvoir activement la contribution positive de l'innovation à la réalisation des droits humains, de la démocratie et de l'État de droit (article 13 et l'exposé des motifs, paragraphe 90).

En d'autres termes, la philosophie qui sous-tend la Convention-cadre ne repose pas seulement sur la volonté de protéger les citoyens et la société contre une IA nuisible ou préjudiciable, mais aussi sur celle de promouvoir l'utilisation de l'IA pour faire progresser les droits humains et la démocratie. Par exemple, l'exposé des motifs de la Convention-cadre explique que les rédacteurs ont explicitement inclus des références à « la dignité humaine et l'autonomie individuelle en tant que valeurs et principes fondamentaux essentiels à la pleine réalisation des droits de l'homme, de la démocratie et de l'État de droit », ainsi qu'à la mise en œuvre de mesures visant à surmonter les inégalités structurelles et historiques qui peuvent alimenter et causer des biais algorithmiques (exposé des motifs, paragraphe 77). Si l'objectif

des approches fondées sur les risques est de protéger la société contre une IA préjudiciable, une approche plus axée sur les valeurs, comme celle de la Convention-cadre, peut créer de puissantes incitations à favoriser le développement d'une IA d'intérêt public. En mettant l'accent sur les droits humains, en fournissant des orientations concrètes (sur ce que ces droits impliquent en relation avec l'IA (générique) et, plus généralement, sur l'expérience et l'expertise du Conseil de l'Europe dans l'analyse des droits humains à la lumière des évolutions technologiques, la Convention-cadre pourrait également constituer une base importante pour l'établissement de normes en matière de droits humains et la fourniture d'orientations substantielles (également, par exemple, dans le contexte de l'application du Règlement européen sur l'IA, qui adopte une approche plus procédurale).

Les mesures positives visant à faire progresser et à promouvoir l'innovation d'intérêt public peuvent prendre diverses formes, allant de la promotion de stratégies de co-conception inclusives, à l'adoption de mesures d'équité et à la réalisation d'évaluations d'impact sur l'égalité, en passant par le renforcement du rôle de la société civile afin de mieux permettre la mobilisation politique et la diffusion des connaissances ou les lignes directrices en matière de marchés publics. Dans ce contexte, les évaluations réalisées lors de la conception, du développement ou de l'acquisition de technologies, y compris les évaluations des risques, devraient être rendues publiques afin de favoriser la responsabilité publique et la contestabilité démocratique. Définir précisément ce qu'implique l'innovation « d'intérêt public » pourrait être un moyen important d'aider les États membres et les acteurs privés à stimuler la demande et à adopter des systèmes d'IA qui non seulement « ne nuisent pas », mais qui sont également dédiés à la promotion des droits humains et de la démocratie. Plus concrètement, faciliter la recherche et l'accès à des solutions techniques conformes aux normes en matière de droits humains, comme les outils Radar de l'UNESCO⁵, peut être un moyen important d'aider les petits fournisseurs et les journalistes indépendants en particulier à contribuer à l'innovation d'intérêt public. La promotion de l'innovation d'intérêt public constituerait également un contre-discours important et indispensable à d'autres conceptions plus technologiques et économiques de l'innovation qui dominent actuellement une grande partie du débat mondial sur l'IA.

4.2.5 Démocratiser la gouvernance de l'IA

L'IA a été qualifiée de « technologie systémique » (Wetenschappelijke Raad voor Regeringsbeleid (WRR), 2021) afin de souligner l'impact systémique potentiellement profond et étendu de l'IA (et de l'IA générative) sur une grande partie de la société et des processus sociétaux, que ce soit dans le domaine du travail, du commerce, de l'industrie, de la politique, de la santé, de la justice, du journalisme, de l'éducation et dans de nombreux autres secteurs. Compte tenu du potentiel transformateur de l'IA, il est important que celle-ci non seulement respecte et mais aussi fasse progresser les droits humains, la démocratie et l'État de droit. En outre, la gouvernance de l'IA doit être plus démocratique et inclusive. Cela est important à la fois en soi (pour garantir que les personnes concernées par une technologie qui transforme la société aient leur mot à dire sur la manière dont elle est gouvernée), mais aussi pour améliorer la gouvernance démocratique et la responsabilité du développement et du

⁵ <https://radar.ircai.org/en/about/>

déploiement de l'IA générative (en intégrant les points de vue des personnes directement concernées).

L'élaboration de la loi européenne sur l'IA est un exemple typique qui montre comment, en raison précisément du potentiel transformateur de cette technologie, le processus législatif peut être caractérisé par d'immenses asymétries de pouvoir, le lobbying de l'industrie et la difficulté d'atteindre et d'inclure un groupe plus large de parties prenantes sociales et, surtout, les citoyens concernés dans le processus décisionnel, malgré toutes les bonnes intentions. L'élaboration d'approches de gouvernance inclusives, participatives et garantissant une implication significative, y compris des groupes ayant moins de pouvoir de négociation ou traditionnellement sous-représentés dans ces débats, sera primordiale pour une gouvernance démocratique de l'IA. La Convention-cadre pourrait constituer un point de départ important pour élaborer et mettre en œuvre des principes de gouvernance plus démocratique de l'IA.

Selon l'article 19 de la Convention-cadre, « Chaque Partie cherche à veiller à ce que les questions importantes soulevées par les systèmes d'intelligence artificielle soient, le cas échéant, dûment examinées dans le cadre de débats publics et de consultations multipartites, en tenant compte des incidences sociales, économiques, juridiques, éthiques, environnementales et des autres implications pertinentes ». Cette disposition est un élément important d'une approche plus démocratique de la gouvernance de l'IA, car elle exige des États membres qu'ils créent des forums et des processus permettant d'inclure les citoyens et les autres parties prenantes de la société non seulement dans la gouvernance de la technologie de l'IA au sens strict, mais aussi, et surtout, dans les implications sociétales plus larges des choix de gouvernance. À l'avenir, le défi sera de rendre ces processus inclusifs, significatifs et efficaces dans la pratique.

Dans ce contexte, les États membres devraient promouvoir une participation significative des citoyens, sous la forme de mécanismes de participation, de représentation et de contestation, tout au long du cycle de vie de l'IA générative, afin de garantir une IA véritablement centrée sur l'humain. Cela pourrait également impliquer de doter les organismes chargés de l'égalité, les médiateurs en matière de discrimination ou les institutions nationales des droits humains des ressources, du mandat et des pouvoirs nécessaires pour identifier, suivre et évaluer l'impact de l'IA générative sur la société, notamment en leur donnant accès aux documents nécessaires relatifs aux données, aux choix de conception et aux structures organisationnelles entourant le développement et le déploiement des technologies (Bartoletti & Xenidis, 2023, p. 70), un mandat qui fait actuellement défaut dans leur boîte à outils. Au-delà des mesures axées sur la technologie, les campagnes de sensibilisation au numérique peuvent également jouer un rôle essentiel. Dans ce contexte, il convient de noter que la Cour, dans sa jurisprudence antérieure, a examiné l'impact d'une éducation médiocre, susceptible de toucher davantage les communautés défavorisées, sur la capacité des personnes à peser le pour et le contre d'une proposition ou d'une initiative politique particulière (*D.* 2007, paragraphe 203).

Rendre la gouvernance de l'IA plus démocratique implique également s'attaquer aux asymétries de pouvoir existantes qui amplifient certaines voix ou certains discours au détriment d'autres. Dans ce contexte, les universitaires et la société civile jouent un rôle essentiel de contre-pouvoir : ils aident les individus et les communautés à exprimer des expériences et des perceptions qui échappent souvent au discours politique et économique

dominant et, grâce à des efforts collectifs, peuvent contribuer à dénoncer les injustices commises tant par l'État que par les acteurs économiques (Young, 2002, p. 155). Pour remplir ce rôle, il est toutefois indispensable que les universitaires et la société civile s'interrogent de manière critique sur leur capacité à prendre réellement en compte des points de vue de ceux qui sont affectés par l'IA générative mais ont peu d'influence sur la manière dont elle est régie.

Au-delà de la création de processus et de procédures visant à intégrer dès le départ les acteurs sociaux dans la gouvernance de l'IA, la mise en place d'un cadre de responsabilité efficace et de droits concrets permettant aux parties concernées de contester certains développements ou déploiements de l'IA contraire à leurs droits humains ou l'État de droit constitue une boucle de rétroaction sociale importante et est un élément indispensable du contrôle démocratique. Encore une fois, et contrairement à d'autres cadres réglementaires, la Convention-cadre fournit des bases importantes. Elle impose notamment aux parties de garantir la responsabilité et l'obligation de rendre compte des effets négatifs sur les droits humains, la démocratie et l'État de droit (art. 9 de la Convention-cadre). Elle oblige également les États à assurer la transparence, l'accès aux données et la possibilité effective pour les personnes affectées de déposer un recours (art. 14 de la Convention-cadre) et d'offrir des droits concrets et des garanties procédurales aux personnes affectées (art. 15 de la Convention-cadre).

Enfin, la capacité des gouvernements et des régulateurs supranationaux à réguler l'IA (et les intérêts économiques et politiques qui la sous-tendent) dépendra de leur volonté politique et de leur capacité à gouverner. Ces deux dimensions sont soumises à rudes épreuves. Au cours des dix dernières années, les relations entre les régulateurs et les entreprises technologiques ont considérablement changé en raison de la croissance du pouvoir politique et économique de certaines des principales entreprises technologiques, du degré de dépendance de la société à l'égard de leurs technologies, du niveau d'investissement dans les infrastructures et de la perte de contrôle sur l'environnement de production, ainsi que de l'érosion de l'autonomie numérique publique. Cette dynamique affecte directement la capacité des gouvernements nationaux à exercer leur rôle, comme le montrent une fois encore les débats actuels au sein de l'UE, où certaines entreprises technologiques font pression sur la Commission européenne pour suspendre, voire abroger, des règles déjà en vigueur. Par conséquent, préserver l'autonomie numérique des États, protéger leur marge de manœuvre politique et réduire leur dépendance stratégique vis-à-vis des grandes entreprises technologiques constitue un pilier supplémentaire, et incontournable, d'une gouvernance démocratique de l'IA.

4.2.6 Conclusion

Le Comité directeur sur la démocratie (CDDEM) du Conseil de l'Europe pourrait jouer un rôle important en proposant un contre-discours axé sur les droits humains et la démocratie face aux discours dominants sur l'innovation, la rapidité et l'échelle, et la Convention-cadre pourrait constituer un instrument important permettant aux États membres de collaborer avec le Comité en vue de mettre en place des garanties structurelles et individuelles afin de promouvoir une approche démocratique et sociétale sur l'IA générative et sa gouvernance. Cela signifie également que la convention-cadre sur l'IA attribue un rôle plus important aux États pour qu'ils traitent de manière proactive les risques et les impacts sur les droits humains,

la démocratie et l'État de droit, y compris lorsqu'ils émanent d'acteurs privés. Pour ce faire, un cadre solide est nécessaire pour concrétiser la gouvernance démocratique de l'IA, fondé sur le droit des droits humains et rappelant que les droits humains comprennent non seulement l'obligation de s'abstenir de toute ingérence, mais aussi de créer les conditions permettant aux individus d'exercer pleinement leurs droits fondamentaux.

5. Conclusions générales et recommandations

Le présent rapport a mis en lumière plusieurs façons majeures dont l'IA générative transforme non seulement les pratiques démocratiques, qu'il s'agisse de l'information des citoyens, les campagnes politiques ou la participation de la société civile et des acteurs sociaux à la prise de décision, mais aussi les institutions démocratiques telles que le journalisme ainsi que l'infrastructure plus large de l'information publique. Il a mis en évidence certaines opportunités et donné des exemples de pratiques qui sont contraires, voire nuisibles, à la démocratie et à la réalisation des droits humains, tels que le droit à la non-discrimination ou la liberté d'expression. Enfin, il a souligné que l'IA générative et ses conséquences pour les droits humains et la démocratie doivent être analysées dans le contexte des configurations du pouvoir infrastructurel et politique, ainsi qu'au regard de ceux qui ont la capacité et la légitimité de définir l'orientation et la nature de l'innovation.

L'analyse de certaines des implications les plus marquantes de l'IA générative sur les droits humains, des recommandations existantes et de la Convention-cadre montre clairement que le défi est double : il s'agit certes de gérer les risques, mais aussi, et surtout, de saisir les opportunités qu'offre l'IA générative pour faire progresser les droits humains et les pratiques et institutions démocratiques. Compte tenu du potentiel transformateur de l'IA et du fait que ce potentiel dépend de manière substantielle des ressources sociétales, qu'il s'agisse de main-d'œuvre, de contenu, de données environnementales ou personnelles, la notion d'« innovation » appliquée à l'IA générative doit être interprétée au sens large : celui d'une innovation d'intérêt public où les intérêts de la société dans son ensemble l'emportent sur les intérêts purement économiques d'une poignée d'entreprises technologiques. Axée sur les droits humains, la démocratie et l'État de droit, la Convention-cadre du Conseil de l'Europe pourrait constituer un point de départ important pour concrétiser cette perspective plus sociétale de l'innovation en matière d'IA et pour garantir qu'elle profite à la société dans son ensemble.

Dans ce contexte, le Conseil de l'Europe, et dans le cadre de son mandat, le CDDEM, pourraient jouer un rôle important en aidant à concrétiser le rôle et la mise en œuvre de la Convention-cadre pour la démocratie, et en définissant les conditions permettant de tirer parti des opportunités démocratiques offertes par l'IA générative, tout en renforçant et en protégeant les institutions, les processus et les cadres de gouvernance démocratiques. Dans la suite de ce document, et sur la base de notre analyse, nous formulons trois recommandations concernant les priorités politiques possibles pour l'Organisation et ses organes.

Orientations substantielles

Contrairement à d'autres instruments internationaux de gouvernance, la Convention-cadre du Conseil de l'Europe se concentre spécifiquement sur les risques et les opportunités que présente l'IA pour la démocratie, les droits humains et l'État de droit. Cet ancrage protège au moins en partie la Convention-cadre contre les pressions visant à abaisser ses normes au profit d'une innovation purement technique et axée sur le profit. Compte tenu de son mandat en tant qu'organisation de défense des droits humains, de sa grande expertise et de ses nombreux travaux (par exemple sous forme de recommandations pertinentes et de la jurisprudence de la Cour européenne des droits de l'homme), le Conseil de l'Europe est particulièrement bien placé pour élaborer des orientations substantielles permettant d'identifier les cas où l'IA générative met en danger les droits humains et la démocratie, voire constitue une menace inacceptable du point de vue des droits humains. Dans la présente étude, nous avons notamment mis en évidence le danger d'un appauvrissement de la diversité, d'inégalités structurelles, de manipulation, d'atteinte à la confiance et aux conditions pratiques et institutionnelles nécessaires à la réalisation de la liberté d'expression. Ces risques sont encore amplifiés par des dépendances infrastructurelles et les asymétries de pouvoir considérables, ainsi que par le rôle prépondérant joué par des acteurs privés à but lucratif qui ne sont pas soumis aux obligations en matière de droits fondamentaux ni aux garanties de l'État de droit lorsqu'il s'agit de décider quels risques sont (in)acceptables pour la société.

Grâce à la méthodologie et au modèle HUDERIA, ainsi qu'à des recommandations telles que la Note d'orientation sur l'IA générative et la liberté d'expression, le Conseil de l'Europe a déjà fait des progrès importants pour fournir des orientations plus substantielles. Le Conseil de l'Europe pourrait jouer un rôle clé en identifiant, précisant et suivant les risques pour la démocratie et l'égalité politique, que ce soit dans la manière dont l'IA générative est utilisée dans les pratiques démocratiques telles que les élections, dans les implications plus structurelles pour les institutions démocratiques (telles que la liberté de la presse) ou dans la gouvernance démocratique de l'IA.

Outre des orientations substantielles sur les cas où l'IA générative est en contradiction avec les droits humains et les principes démocratiques, des orientations sur les cas et les moyens par lesquels l'IA peut faire progresser la démocratie et les droits humains sont également indispensables. En particulier, il existe actuellement peu d'orientations sur la signification exacte de l'innovation « d'intérêt public » ou sur les valeurs à privilégier dans les politiques de marchés publics.

Outre des orientations substantielles, le Conseil de l'Europe pourrait également jouer un rôle dans la normalisation et aider les organismes de défense des droits humains et les gouvernements des États membres à s'accorder sur des normes en matière d'innovation d'intérêt public et de marchés publics.

Encourager et faciliter l'innovation d'intérêt public

Dans les récents débats politiques et réglementaires autour de l'IA générative, l'accent a été mis sur la gestion des risques et sur la nécessité de rendre ces technologies plus sûres et plus fiables, tout en restant économiquement viables. En revanche, les approches de gouvernance qui encouragent et favorisent l'innovation d'intérêt public, dans le sens du

développement et du déploiement de l'IA générative pour faire progresser les droits humains et les principes démocratiques, ont reçu relativement moins d'attention. Une contribution importante de la Convention-cadre qui la distingue des autres cadres réglementaires est d'aller au-delà d'une approche purement fondée sur les risques et d'exiger des États membres qu'ils encouragent l'utilisation de l'IA pour faire progresser les droits humains et la démocratie. Cela est conforme à la doctrine existante des droits humains et aux obligations positives des États de créer les conditions permettant aux citoyens de bénéficier des droits humains.

Le Conseil de l'Europe pourrait jouer un rôle important en rendant plus visibles les implications démocratiques relatives aux droits humains des décisions en matière de marchés publics, et en aidant à identifier les innovateurs d'intérêt public, les obstacles à ce type d'innovation et les moyens nécessaires pour les surmonter. Il pourrait étudier des mécanismes et des solutions réglementaires visant à encourager et à soutenir le développement d'alternatives résilientes d'intérêt public, ainsi que l'investissement dans des institutions de confiance de haute qualité, telles que les médias locaux, les médias de service public et les médias qui mettent l'accent sur la qualité, l'ouverture et les normes d'intégrité, qu'ils prennent la forme de médias traditionnels ou de nouveaux types de médias rendus possibles par l'IA. Une fois encore, il pourrait également apporter un contre-discours important aux discours technologiques dominants, en privilégiant l'innovation axée sur la valeur, les ensembles de données l'open source, la décentralisation et l'interopérabilité plutôt que la promotion de l'échelle et de l'innovation impulsées par un nombre concentré d'hyper-scalers.

Renforcer la résilience démocratique et la capacité à gouverner

La démocratie, implique aussi la capacité d'une société à déterminer elle-même le rôle de la technologie dans la société et les valeurs auxquelles elle adhère. Face à l'intégration croissante d'une poignée de solutions technologiques propriétaires et axées sur le profit dans les institutions démocratiques, aux dépendances qui en résultent et à l'influence politique grandissante des entreprises technologiques, nous avons distingué cinq dimensions de l'autodétermination démocratique et de la gouvernance de la technologie :

- l'utilisation de l'IA générative dans les élections et les processus de prise de décision démocratique, ainsi que ses implications en matière d'inclusion, d'égalité des chances de participation et de jouissance des libertés politiques ;
- l'intégration dans les infrastructures démocratiques, et la mesure dans laquelle l'IA générative peut contribuer à la qualité de la sphère publique et à la résilience des institutions démocratiques, ou les mettre en péril ;
- la sensibilité et la réactivité des entreprises technologiques et des utilisateurs en aval de l'IA générative face à la responsabilité démocratique, à l'observabilité et au contrôle ;
- la capacité des gouvernements et des autorités de régulation à prendre des décisions de manière autonome ;
- l'inclusivité et la représentativité de la prise de décision démocratique, ainsi que la capacité à atténuer les asymétries de pouvoir et la vulnérabilité numérique.

Le suivi de l'impact de l'IA générative sur ces cinq dimensions de la prise de décision démocratique et l'aide apportée aux gouvernements pour élaborer une stratégie visant à préserver leur autonomie numérique et à promouvoir la résilience démocratique pourraient devenir un axe prioritaire du Conseil de l'Europe et du CDDEM.

Concrètement, cela pourrait inclure le suivi et l'élaboration de lignes directrices pour une utilisation responsable de l'IA générative lors des élections, la cartographie des dépendances existantes, ainsi que l'exploration de mécanismes et de solutions réglementaires visant à encourager le développement de solutions infrastructurelles basées en Europe, offrant une alternative aux écosystèmes des hyper-scalers. Il pourrait s'agir notamment de cartographier les dépendances existantes, et d'accompagner les gouvernements dans leurs procédures de marchés publics pour limiter les dépendances potentielles vis-à-vis des acteurs dominants, tels que les grandes entreprises technologiques étrangères. D'autres mesures similaires pourraient consister à reconsidérer le mandat des médias d'intérêt public afin d'y inclure le développement d'éléments essentiels à la mission de l'écosystème numérique européen, et à travailler avec les États membres à l'élaboration de dispositions favorables au développement d'alternatives locales et communautaires aux produits des grandes entreprises technologiques.

Une autre approche pourrait consister à aider les États membres à mettre en place des moyens de renforcer le pouvoir des acteurs chargés de veiller à la responsabilité publique, tels que les universitaires, la société civile et les journalistes d'investigation. Cela pourrait se faire, par exemple, en encourageant les collaborations, en créant des boucles de rétroaction entre les acteurs chargés de la responsabilité publique, les entreprises et les régulateurs, ou encore en mettant en place des stratégies financières et. Dans ce contexte, il convient de réaffirmer la nécessité urgente de la transparence, tant au niveau individuel qu'institutionnel et sociétal. Comme l'a démontré la première partie de l'étude, le manque de transparence et d'observabilité constitue un obstacle majeur à la responsabilité démocratique et va à l'encontre des attentes et des besoins des citoyens qui doivent naviguer dans l'environnement numérique. Cela nous amène à notre recommandation finale :

Jusqu'à présent, une grande partie du débat sur la gouvernance de l'IA s'est déroulée au-dessus de la tête des citoyens que ce soit concernant le développement de solutions et de pratiques basées sur l'IA générative ou l'élaboration des cadres de gouvernance. Pour rendre cette gouvernance de l'IA générative plus démocratique, il est nécessaire de développer des moyens de promouvoir une participation citoyenne significative (participation, représentation et contestation) tout au long du cycle de vie de l'IA générative, en accordant une attention particulière aux inégalités structurelles auxquelles sont confrontées les personnes concernées et les groupes (historiquement) défavorisés. La Convention-cadre comprend déjà un ensemble d'obligations importantes dans ce domaine, allant de la prise en compte des points de vue des personnes dont les droits peuvent être affectés dans les cadres de gestion des risques et des impacts (art. 16 (2)(c) de la Convention-cadre), à l'exigence d'organiser des consultations publiques (art. 19 de la Convention-cadre) et à l'importance de la culture numérique (art. 20 de la Convention-cadre), de la transparence (art. 14 (2) de la Convention-cadre) et des droits concrets et des garanties procédurales pour les citoyens afin d'assurer la contestabilité (art. 15 de la Convention-cadre). Compte tenu des intérêts et des droits concrets des citoyens dans le développement des infrastructures numériques européennes, l'utilisation de l'IA générative dans les processus décisionnels démocratiques et la qualité de la prise de

décision démocratique concernant l'IA générative pourraient constituer une autre contribution importante du CDDEM.

Références

- Abid, A., Farooqi, M., & Zou, J. (2021). Persistent Anti-Muslim Bias in Large Language Models. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, AIES '21*, 298–306. <https://doi.org/10.1145/3461702.3462624>
- Amoore, L., Campolo, A., Jacobsen, B., & Rella, L. (2024). A world model: On the political logics of generative AI. *Political Geography*, 113, 103134. <https://doi.org/10.1016/j.polgeo.2024.103134>
- Arc XP. (2024). Incorporating AI in the Newsroom: Le Parisien's Journey to Innovation and Efficiency. *Arc XP*. <https://arcmarketing-arc-marketing-prod-staging.web.arc-cdn.net/2024/07/22/incorporating-ai-in-the-newsroom-le-parisiens-journey-to-innovation-and-efficiency/>
- Argyle, L. P., Bail, C. A., Busby, E. C., Gubler, J. R., Howe, T., Rytting, C., Sorensen, T., & Wingate, D. (2023). Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences*, 120(41), e2311627120. <https://doi.org/10.1073/pnas.2311627120>
- Atrey, S., & Dunne, P. (Eds). (2020). *Intersectionality and Human Rights Law*. Hart Publishing. <https://doi.org/10.5040/9781509935321>
- Autoriteit Persoonsgegevens. (2025, October 21). *AP waarschuwt: Chatbots geven vertekend stemadvies | Autoriteit Persoonsgegevens*. <https://www.autoriteitpersoonsgegevens.nl/actueel/ap-waarschuwt-chatbots-geven-vertekend-stemadvies>
- Autorité de la concurrence. (2024). *Avis relatif au fonctionnement concurrentiel du secteur de l'intelligence artificielle générative* (Nos 24-A-05). <https://www.autoritedelaconcurrence.fr/fr/avis/relatif-au-fonctionnement-concurrentiel-du-secteur-de-lintelligence-artificielle-generative>
- Baker, E. (2006). *Media Concentration and Democracy: Why Ownership Matters*. Cambridge University Press.
- Balayn, A., & Gürses, S. (2024, September 30). Misguided: AI regulation needs a shift in focus | Internet Policy Review. *Internet Policy Review*. <https://policyreview.info/articles/news/misguided-ai-regulation-needs-shift/1796>
- Bartoletti, I., & Xenidis, R. (2023). *Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination* (p. 94). Council of Europe. <https://edoc.coe.int/fr/intelligence-artificielle/11649-study-on-the-impact-of-artificial-intelligence-systems-their-potential-for-promoting-equality-including-gender-equality-and-the-risks-they-may-cause-in-relation-to-non-discrimination.html>
- Bastian, M., Helberger, N., & Makhortykh, M. (2021). Safeguarding the Journalistic DNA: Attitudes towards the Role of Professional Values in Algorithmic News Recommender Designs. *Digital Journalism*, 9(6), 835–863. <https://doi.org/10.1080/21670811.2021.1912622>
- Bayer, J. (2020). Double harm to voters: Data-driven micro-targeting and democratic public discourse. *Internet Policy Review*, 9(1). <https://policyreview.info/articles/analysis/double-harm-voters-data-driven-micro-targeting-and-democratic-public-discourse>

- Becker, K. B., Simon, F. M., & Crum, C. (2023). *Policies in Parallel? A Comparative Study of Journalistic AI Policies in 52 Global News Organisations* [SocArXiv].
<https://osf.io/preprints/socarxiv/c4af9>
- Beckett, C., & Yaseen, M. (2023). *Generating Change. A Global Survey of what news organisations are doing with AI*. Journalism AI.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.
<https://doi.org/10.1145/3442188.3445922>
- Bhattacharya, A. (2024, September 19). Political campaigns embrace AI to reach voters across language barriers. *Rest of World*. <https://restofworld.org/2024/aapi-victory-alliance-ai-voter-outreach/>
- Birhane, A., Dehdashtian, S., Prabhu, V., & Boddeti, V. (2024). The Dark Side of Dataset Scaling: Evaluating Racial Classification in Multimodal Models. *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1229–1244.
<https://doi.org/10.1145/3630106.3658968>
- Birhane, A., Prabhu, V. U., & Kahembwe, E. (2021). *Multimodal datasets: Misogyny, pornography, and malignant stereotypes* (arXiv:2110.01963). arXiv.
<https://doi.org/10.48550/arXiv.2110.01963>
- Blagojev, T., Da Costa Leite Borges, D., Brogi, E., Kermer, J. E., Trevisan, M., & Verza, S. (2023). *News Deserts in Europe: Assessing Risks for Local and Community Media in the 27 EU Member States* (EUI, RSC, Preliminary Report). Centre for Media Pluralism and Media Freedom (CMPF).
- Bonin, A.-K., Pecher, M., Peter, T., Schönacher, N., & Wettinger, A. (2025). Between Authenticity and Automation: Nonprofit-Organizations Navigating AI-driven Communication. In A. Godulla, F. Decker, L. Hartmann, C. Pankoke, M. Pecher, S. Pütter, & C. Santangelo (Eds), *Strategic Communication in Disruptive Times: How Sociopolitical Polarization, Virtual Media and AI Reshape Organizational Communication* (pp. 93–113).
- Borchardt, A. (2022). Go, Robots, Go! The Value and Challenges of Artificial Intelligence for Local Journalism. *Digital Journalism*, 10(10), 1919–1924.
<https://doi.org/10.1080/21670811.2022.2149584>
- Bouché, G., Helberger, N., de Rijke, M., & van Drunen, M. Z. (2025). LLMs as media technology: Implications under article 10 ECHR. *International Journal of Law and Information Technology*, 33, eaaf027. <https://doi.org/10.1093/ijlit/eaaf027>
- Braun, A. F. (2025, April 8). Independent Media has an Infrastructural Problem. *Tech Policy Press*. <https://techpolicy.press/independent-media-has-an-infrastructural-problem>
- Brennan, K., Kak, A., & Myers West, S. (2025). *Artificial Power: 2025 Landscape Report*. AI Now Institute. <https://ainowinstitute.org/publications/research/ai-now-2025-landscape-report>
- Bychawska-Siniarska, D. (2017). *Protecting the right to Freedom of Expression under the European Convention on Human Rights—A handbook for legal practioners*. Council of Europe.
- Calvet-Bademunt, J., Mchangama, J., & Anzabi, I. (2025). *That Violates My Policies> AI Laws, Chatbots, and The Future of Expression*. The Future of Speech.
<https://futurefreespeech.org/wp-content/uploads/2025/10/AI-Report-2025-Full-Report.pdf>

- Caswell, D. (2024). Audiences, automation, and AI: From structured news to language models. *AI Magazine*, 45(2), 174–186. <https://doi.org/10.1002/aaai.12168>
- Chander, S., & Gürses, S. (2024, October 15). From Infrastructural Power to Redistribution: How the EU's Digital Agenda Cements Securitization and Computational Infrastructures (and How We Build Otherwise). *AI Now Institute*. <https://ainowinstitute.org/publications/from-infrastructural-power-to-redistribution-how-the-eus-digital-agenda-cements-securitization-and-computational-infrastructures-and-how-we-build-otherwise>
- Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. (2021). *Bias and Debias in Recommender System: A Survey and Future Directions* (arXiv:2010.03240). arXiv. <https://doi.org/10.48550/arXiv.2010.03240>
- Cohen, J. E. (2019). *Between truth and power: The legal constructions of informational capitalism*. Oxford University Press.
- Cools, H., & de Vreese, C. H. (2025). From Automation to Transformation with AI-Tools: Exploring the Professional Norms and the Perceptions of Responsible AI in a News Organization. *Digital Journalism*, 1–20. <https://doi.org/10.1080/21670811.2025.2505982>
- Cools, H., & Diakopoulos, N. (2024). Uses of Generative AI in the Newsroom: Mapping Journalists' Perceptions of Perils and Possibilities. *Journalism Practice*, 1–19. <https://doi.org/10.1080/17512786.2024.2394558>
- Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385(6714), eadq1814. <https://doi.org/10.1126/science.adq1814>
- Crenshaw, K. (1989). *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*. 31.
- Danaher, J. (2024). Generative AI and the future of equality norms. *Cognition*, 251, 105906. <https://doi.org/10.1016/j.cognition.2024.105906>
- de Brún, P., & Sawyeddollah, M. (2025, February 17). *Meta's New Content Policies Risk Fueling More Mass Violence And Genocide*. Tech Policy Press. <https://techpolicy.press/metas-new-content-policies-risk-fueling-more-mass-violence-and-genocide>
- de León, E., Votta, F., Araujo, T., & de Vreese, C. (2025, October 29). 1 in 10 Dutch citizens are likely to ask AI for election advice. This is why they shouldn't. *University of Amsterdam / AlgoSoc*. <https://www.uva.nl/en/shared-content/faculteiten/en/faculteit-der-maatschappij-en-gedragswetenschappen/news/2025/10/dont-ask-ai-for-election-advice.html>
- Deuze, M., & Beckett, C. (2022). Imagination, Algorithms and News: Developing AI Literacy for Journalism. *Digital Journalism*, 10(10), 1913–1918. <https://doi.org/10.1080/21670811.2022.2119152>
- Diakopoulos, N., Helberger, N., Cools, H., Li, C., Kung, E., & Rinehart, A. (2024a). *Generative AI in Journalism: The Evolution of Newswork and Ethics in a Generative Information Ecosystem*.
- Diakopoulos, N., Helberger, N., Cools, H., Li, C., Kung, E., & Rinehart, A. (2024b). *Generative AI in Journalism: The Evolution of Newswork and Ethics in a Generative Information Ecosystem*.
- Dodds, T., Vandendaele, A., Simon, F. M., Helberger, N., Resendez, V., & Yeung, W. N. (2025b). Knowledge Silos as a Barrier to Responsible AI Practices in Journalism?

- Exploratory Evidence from Four Dutch News Organisations. *Journalism Studies*, 26(6), 740–758. <https://doi.org/10.1080/1461670X.2025.2463589>
- Dodds, T., Zamith, R., & Lewis, S. C. (2025a). The AI turn in journalism: Disruption, adaptation, and democratic futures. *Journalism*, 14648849251343518. <https://doi.org/10.1177/14648849251343518>
- Dommett, K., Kefford, G., Kruschinski, S., Dommett, K., Kefford, G., & Kruschinski, S. (2024). *Data-Driven Campaigning and Political Parties: Five Advanced Democracies Compared*. Oxford University Press.
- Dommett, K., Kefford, G., & Power, S. (2021). The digital ecosystem: The new politics of party organization in parliamentary democracies. *Party Politics*, 27(5), 847–857. <https://doi.org/10.1177/1354068820907667>
- Dworkin, R. (1994). A new map of censorship. *Index on Censorship*, 23(1–2), 9–15. <https://doi.org/10.1080/03064229408535633>
- EBU. (2025). *News Integrity in AI Assistants. An international PSM study*.
- Eder, M., & Sjøvaag, H. (2025). Falling behind the adoption curve: Local journalism's struggle for innovation in the AI transformation. *Journal of Media Business Studies*, 22(4), 325–343. <https://doi.org/10.1080/16522354.2025.2473301>
- Elliott, V. (2025). The WIRED AI Elections Project. *Wired*. <https://www.wired.com/story/generative-ai-global-elections/>
- Feenstra, W., & Sabel, P. (2025, October 27). Twee PVV-Kamerleden vallen met nepbeelden anoniem Timmermans aan, GroenLinks-PvdA doet aangifte. *de Volkskrant*. <https://www.volkskrant.nl/politiek/twee-pvv-kamerleden-vallen-met-nepbeelden-anoniem-timmermans-aan-groenlinks-pvda-doet-aangifte~b6958ea1/>
- Fineman, M. A. (Ed.). (2008). The vulnerable subject: Anchoring equality in the human condition. In *Transcending the Boundaries of Law* (0 edn, pp. 177–191). Routledge-Cavendish. <https://doi.org/10.4324/9780203848531-26>
- Foos, F. (2024). The Use of AI by Election Campaigns. *LSE Public Policy Review*, 3(3), 8. <https://doi.org/10.31389/lseppr.112>
- Garton Ash, T. (2016). *Free Speech: Ten Principles for a Connected World*. Yale University Press. (1227501). <https://research.ebsco.com/linkprocessor/plink?id=5a1596fe-8094-3709-af3b-f82edf3fbb49>
- Gerards, J., & Xenidis, R. (2021). *Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non discrimination law*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2838/544956>
- Gerards, J., & Zuiderveen Borgesius, F. (2022). Protected Grounds and the System of Non-Discrimination Law in the Context of Algorithmic Decision-Making and Artificial Intelligence. *Colorado Technology Law Journal*, 20, 56.
- Gillespie, T. (2024). Generative AI and the politics of visibility. *Big Data & Society*, 11(2), 20539517241252131. <https://doi.org/10.1177/20539517241252131>
- Gondwe, G. (2025). Perceptions of AI-driven news among contemporary audiences: A study of trust, engagement, and impact. *AI & SOCIETY*, 1–12. <https://doi.org/10.1007/s00146-025-02294-x>
- Guler, G., Remfry, E., Kherroubi Garcia, I., Barrow, N., Duarte, T., & Ibison, Y. (2024). *Grassroots and non-profit perspectives on generative AI*. Joseph Rowntree Foundation. <https://www.jrf.org.uk/ai-for-public-good/grassroots-and-non-profit-perspectives-on-generative-ai>
- Guzman, A. L., & Lewis, S. C. (2024). What Generative AI Means for the Media Industries, and Why it Matters to Study the Collective Consequences for Advertising,

- Journalism, and Public Relations. *Emerging Media*, 2(3), 347–355.
<https://doi.org/10.1177/27523543241289239>
- Hacker, P., Zuiderveen Borgesius, F., Mittelstadt, B. D., & Wachter, S. (2025). *Generative discrimination: What happens when generative AI exhibits bias, and what can be done about it* (1st edn). Oxford University Press.
<https://ora.ox.ac.uk/objects/uuid:cf87e259-368f-440a-a2df-5285433ffcb1>
- Helberger, N., Sax, M., Strycharz, J., & Micklitz, H.-W. (2022). Choice Architectures in the Digital Economy: Towards a New Understanding of Digital Vulnerability. *Journal of Consumer Policy*, 45(2), 175–200. <https://doi.org/10.1007/s10603-021-09500-5>
- Helberger, N., van Drunen, M., Eskens, S., Bastian, M., & Möller, J. (2020). A freedom of expression perspective on AI in the media: With a special focus on editorial decision making on social media platforms and in the news media. *European Journal of Law and Technology*, 11(30).
- Hofman, E., & Veerbeek, J. (2024, December 11). Verantwoording bij het data-onderzoek naar de AI-beelden van radicaal-rechts. *De Groene Amsterdammer*.
<https://www.groene.nl/artikel/data-onderzoek-naar-de-ai-beelden-van-radicaal-rechts>
- Holt, J. (2024). *Cloud Policy: A History of Regulating Pipelines, Platforms, and Data*. The MIT Press. <https://doi.org/10.7551/mitpress/14683.001.0001>
- Hu, T.-H. (2015). *A Prehistory of the Cloud*. The MIT Press.
<https://doi.org/10.7551/mitpress/9780262029513.001.0001>
- Jaźwińska, K., & Chandrasekar, A. (2025, March). AI Search Has A Citation Problem. *Columbia Journalism Review*. https://www.cjr.org/tow_center/we-compared-eight-ai-search-engines-theyre-all-bad-at-citing-news.php
- Jones, R., & Jones, B. (2019). Atomising the News: The (In)Flexibility of Structured Journalism. *Digital Journalism*, 7(8), 1157–1179.
<https://doi.org/10.1080/21670811.2019.1609372>
- Jungherr, A. (2023). Artificial Intelligence and Democracy: A Conceptual Framework. *Social Media + Society*, 9(3), 20563051231186353.
<https://doi.org/10.1177/20563051231186353>
- Jungherr, A., & Rauchfleisch, A. (2025). Artificial Intelligence in deliberation: The AI penalty and the emergence of a new deliberative divide. *Government Information Quarterly*, 42(4), 102079. <https://doi.org/10.1016/j.giq.2025.102079>
- Kapoor, S., & Narayanan, A. (2025, April 15). We Looked at 78 Election Deepfakes. Political Misinformation is not an AI Problem. *AI as Normal Technology*.
<https://www.normaltech.ai/p/we-looked-at-78-election-deepfakes>
- Kieslich, K., Diakopoulos, N., & Helberger, N. (2024a). Anticipating impacts: Using large-scale scenario-writing to explore diverse implications of generative AI in the news environment. *AI and Ethics*. <https://doi.org/10.1007/s43681-024-00497-4>
- Kieslich, K., Diakopoulos, N., & Helberger, N. (2024b). Anticipating impacts: Using large-scale scenario-writing to explore diverse implications of generative AI in the news environment. *AI and Ethics*. <https://doi.org/10.1007/s43681-024-00497-4>
- Kotek, H., Dockum, R., & Sun, D. Q. (2023). Gender bias and stereotypes in Large Language Models. *Proceedings of The ACM Collective Intelligence Conference*, 12–24. <https://doi.org/10.1145/3582269.3615599>
- Kreps, S., & Kriner, D. (2023). How AI Threatens Democracy. *Journal of Democracy*, 34(4), 122–131.
- Kruschinski, S., Jost, P., Fecher, H., & Scherer, T. (2025). *Künstliche Intelligenz in politischen Kampagnen*. Otto Brenner Stiftung.

- Kruschinski, S., & Votta, F. (2025). *CampAlignTracker.de: Detection and Analysis of Visual GenAI in the 2025 German Federal Election*. <https://doi.org/10.17605/OSF.IO/CZEB6>
- Kruschinski, S., & Votta, F. (forthcoming). Efficiency Tool or Manipulation Machine? The Use, Transparency, and Communication Strategies of Visual Generative AI on Social Media in the 2025 German Federal Election. In S. Geise & M.-M. Leidecker-Sandmann (Eds.), *Die Medien im Wahlkampf—Analysen zur Bundestagswahl 2025*. Wiesbaden: Springer VS.
- Larkin, B. (2013). The Politics and Poetics of Infrastructure. *Annual Review of Anthropology*, 42(Volume 42, 2013), 327–343. <https://doi.org/10.1146/annurev-anthro-092412-155522>
- Laude, L., & Daum, A. (2024). *KI als neues Wahlkampfinstrument*. <https://doi.org/10.59704/9e96c5f69bbfa94c>
- Lazar, S., & Manuali, L. (2026). Using LLMs to Enhance Democracy. *Minds and Machines*, 36(1), 12. <https://doi.org/10.1007/s11023-026-09767-y>
- Leerssen, P. (2025). From Murdoch to Musk: Platform ownership and the political economy of online content governance. *Platforms & Society*, 2, 29768624251386260. <https://doi.org/10.1177/29768624251386260>
- Lewis, S. C., Guzman, A. L., Schmidt, T. R., & Lin, B. (2025). Generative AI and its disruptive challenge to journalism: An institutional analysis. *Communication and Change*, 1(1), 9. <https://doi.org/10.1007/s44382-025-00008-x>
- Lin, B., & Lewis, S. C. (2022). The One Thing Journalistic AI Just Might Do for Democracy. *Digital Journalism*, 10(10), 1627–1649. <https://doi.org/10.1080/21670811.2022.2084131>
- Liu, L. T., Dean, S., Rolf, E., Simchowitz, M., & Hardt, M. (2018). Delayed Impact of Fair Machine Learning. *arXiv:1803.04383 [Cs, Stat]*. <http://arxiv.org/abs/1803.04383>
- Luitse, D. (2024). Platform power in AI: The evolution of cloud infrastructures in the political economy of artificial intelligence. *Internet Policy Review*, 13(2). <https://policyreview.info/articles/analysis/platform-power-ai-evolution-cloud-infrastructures>
- Maher, B. (2025, February 21). ChatGPT referrals to top publishers up eight times in six months but still negligible. *Press Gazette*. <https://pressgazette.co.uk/platforms/chatgpt-perplexity-news-ai-referral-search/>
- Malgieri, G., & Niklas, J. (2020). Vulnerable data subjects. *Computer Law & Security Review*, 37, 105415. <https://doi.org/10.1016/j.clsr.2020.105415>
- Mansell, R., Durach, F., Kettemann, M., Lenoir, T., Procter, R., Tripathi, G., & Tucker, E. (2025). Information ecosystem and troubled democracy. *OID*. <https://observatory.informationdemocracy.org/report/information-ecosystem-and-troubled-democracy/>
- Mattis, N., & de Vreese, C. (2025). Breaking the News? Generative AI's Impact on Journalism and Its Implications for Disinformation. *International Journal of Communication*, 19, 1–25.
- Mayer, V. (2023). When do we go from here? Data center infrastructure labor, jobs, and work in economic development time and temporalities. *New Media & Society*, 25(2), 307–323. <https://doi.org/10.1177/14614448221149947>
- McKinney, S. (2024). Integrating Artificial Intelligence into Citizens' Assemblies: Benefits, Concerns and Future Pathways. *Journal of Deliberative Democracy*, 20(1). <https://doi.org/10.16997/jdd.1556>

- Meese, J., & Seipp, T. (2025). The Press as Platform: Institutional Isomorphism and the Strategic Adoption of Platform Logics. *Digital Journalism*, 1–17. <https://doi.org/10.1080/21670811.2025.2573074>
- Mitova, E., Blassnig, S., Strikovic, E., Urman, A., De Vreese, C., & Esser, F. (2024a). Exploring users' desire for transparency and control in news recommender systems: A five-nation study. *Journalism*, 25(10), 2001–2021. <https://doi.org/10.1177/14648849231222099>
- Mitova, E., Blassnig, S., Strikovic, E., Urman, A., De Vreese, C., & Esser, F. (2024b). Exploring users' desire for transparency and control in news recommender systems: A five-nation study. *Journalism*, 25(10), 2001–2021. <https://doi.org/10.1177/14648849231222099>
- Monzer, C., Moeller, J., Helberger, N., & Eskens, S. (2020). User Perspectives on the News Personalisation Process: Agency, Trust and Utility as Building Blocks. *Digital Journalism*, 8(9), 1142–1162. <https://doi.org/10.1080/21670811.2020.1773291>
- Morosoli, S., Naudts, L., Cools, H., Venkatraj, K., Helberger, N., & De Vreese, C. (2025). Public accountability and regulatory expectations for AI in journalism: Qualitative evidence from focus groups with Dutch citizens. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-025-02591-5>
- Motamedi, M. (2019, October 2). Locked out: Why is Amazon blocking Iranians from its services? *Al Jazeera*. <https://www.aljazeera.com/economy/2019/10/2/locked-out-why-is-amazon-blocking-iranians-from-its-services>
- Mouffe, C. (1999). Deliberative Democracy or Agonistic Pluralism? *Social Research*, 66(3), 745–758.
- Naudts, L. (2023). *Fair or Unfair Differentiation? Reconsidering the Concept of Equality for the Regulation of Algorithmically Guided Decision-Making (Doctoral Dissertation)*. KU Leuven.
- Naudts, L. (2024). The Digital Faces of Oppression and Domination: A Relational and Egalitarian Perspective on the Data-driven Society and its Regulation. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, FAccT '24*, 701–712. <https://doi.org/10.1145/3630106.3658934>
- Naudts, L., Helberger, N., Sax, M., & Veale, M. (2024). Toward Constructive Optimisation: A new perspective on the regulation of recommender systems and the rights of users and society. In N. Helberger, B. Kas, H.-W. Micklitz, M. Namysłowska, L. Naudts, M. Sax, & M. Veale (Eds), *Digital Fairness for Consumers* (pp. 25–68). BEUC - The European Consumer Organisation.
- Naudts, L., Helberger, N., Veale, M., & Sax, M. (2025). A Right to Constructive Optimization: A Public Interest Approach to Recommender Systems in the Digital Services Act. *Journal of Consumer Policy*. <https://doi.org/10.1007/s10603-025-09586-1>
- Nechushtai, E. (2018). Could digital platforms capture the media through infrastructure? *Journalism*, 19(8), 1043–1058. <https://doi.org/10.1177/1464884917725163>
- Newman, N., & Cherubini, F. (2025). *Journalism, media, and technology trends and predictions 2025*. Reuters Institute.
- Nieuwsuur. (2024, August 5). *Chatbots mogen niets zeggen over verkiezingen, maar dat lukt niet altijd*. <https://nos.nl/nieuwsuur/artikel/2531896-chatbots-mogen-niets-zeggen-over-verkiezingen-maar-dat-lukt-niet-altijd>
- O'Callaghan, P., Cronin, O., Kelly, B. D., Shiner, B., Walmsley, J., & McCarthy-Jones, S. (2024). The right to freedom of thought: An interdisciplinary analysis of the UN

- special rapporteur's report on freedom of thought. *The International Journal of Human Rights*, 28(1), 1–23. <https://doi.org/10.1080/13642987.2023.2227100>
- OCDE. (2025). *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions*. OECD Publishing. <https://doi.org/10.1787/795de142-en>
- OpenAI. (2024, March 13). Disrupting malicious uses of AI by state-affiliated threat actors. *OpenAI*. <https://openai.com/index/disrupting-malicious-uses-of-ai-by-state-affiliated-threat-actors/>
- OSCE. (2025). *Safeguarding Media Freedom in the Age of Big Tech Platforms and AI*. <https://www.osce.org/representative-on-freedom-of-media/598525>
- Oster, J. (2015). *Media Freedom as a Fundamental Right*. Cambridge University Press.
- Pentney, K. (2024). States' positive obligation to create a favourable environment for participation in public debate: A principle in search of a practical effect? *Journal of Media Law*. (world). <https://www.tandfonline.com/doi/abs/10.1080/17577632.2024.2326278>
- Phillips, T., & Torres, P. (2024, August 27). 'Being on camera is no longer sensible': Persecuted Venezuelan journalists turn to AI. *The Guardian*. <https://www.theguardian.com/world/article/2024/aug/27/venezuela-journalists-nicolas-maduro-artificial-intelligence-media-election>
- Piasecki, S., Morosoli, S., Helberger, N., & Naudts, L. (2024). AI-generated journalism: Do the transparency provisions in the AI Act give news readers what they hope for? *Internet Policy Review*, 13(4). <https://doi.org/10.14763/2024.4.1810>
- Pybus, J., & Côté, M. (2024). Super SDKs: Tracking personal data and platform monopolies in the mobile. *Big Data & Society*, 11(1), 20539517241231270. <https://doi.org/10.1177/20539517241231270>
- Quell, M. (2025, May 15). Trump's sanctions on ICC prosecutor have halted tribunal's work. *AP News*. <https://apnews.com/article/icc-trump-sanctions-karim-khan-court-a4b4c02751ab84c09718b1b95cbd5db3>
- Reiss, M. V., Knor, E. L., Stöwing, E., Merten, L., & Möller, J. (2025). Zwischen Neugier und Skepsis: Nutzung und Wahrnehmung generativer KI zur Informationssuche in Deutschland. *Arbeitspapiere des Hans-Bredow-Instituts*, 76, 39 S. <https://doi.org/10.21241/SSOAR.100907>
- Rikap, C. (2021). *Capitalism, Power and Innovation: Intellectual Monopoly Capitalism Uncovered* (1st edn). Routledge. <https://doi.org/10.4324/9780429341489>
- Rolf, S., & Schindler, S. (2023). The US–China rivalry and the emergence of state platform capitalism. *Environment and Planning A: Economy and Space*, 55(5), 1255–1280. <https://doi.org/10.1177/0308518X221146545>
- Ryan, Ó. (2025, October 22). Deepfake AI video depicting Catherine Connolly quitting presidential race removed by Meta. *The Irish Times*. <https://www.irishtimes.com/politics/2025/10/22/meta-removes-ai-video-purporting-to-show-catherine-connolly-quitting-presidential-race/>
- Sanseviero, O., Cuenca, P., Passos, A., & Whitaker, J. (2025). *Hands-On Generative AI with Transformers and Diffusion Models*. <https://www.oreilly.com/library/view/hands-on-generative-ai/9781098149239/>
- Schibstedt. (2024). *Editorial media as defenders of democracies*. <https://cdn.schibstedmedia.com/wp-content/uploads/2024/11/25125955/Editorial-media-as-defenders-of-democracies.pdf>

- Schiele, M., Gittmann, Y., Ilchmann, S., Gojsalić, A., Jurinčić, D., & Klempt, P. (2024). *Voting Advice Applications: Implementation of RAG-supported LLMs*. Preprints. <https://doi.org/10.36227/techrxiv.172115156.64500701/v1>
- Schiff, K. J., Schiff, D. S., & Bueno, N. S. (2025). The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability? *American Political Science Review*, 119(1), 71–90. <https://doi.org/10.1017/S0003055423001454>
- Schneiders, P., & Stark, B. (2025). Ensuring News Quality in Platformized News Ecosystems: Shortcomings and Recommendations for an Epistemic Governance. *Media and Communication*, 13, 10042. <https://doi.org/10.17645/mac.10042>
- Seipp, T. J., Helberger, N., De Vreese, C., & Ausloos, J. (2024). Between the cracks: Blind spots in regulating media concentration and platform dependence in the EU. *Internet Policy Review*, 13(4). <https://doi.org/10.14763/2024.4.1813>
- Sevignani, S., Theine, H., & Tröger, M. (2025). Unpacking Property: Media, Ownership, and Power in Transformation| Toward Media Environment Capture: A Theoretical Contribution on the Influence of Big Tech on News Media. *International Journal of Communication*, 19, 21–21.
- Shrishak, K., & Pénicaud, S. (2025). *European Policy Guidelines on AI and Algorithm-Driven Discrimination for Equality Bodies and Other National Human Rights Structures* (p. 14). Council of Europe. <https://rm.coe.int/policy-guidelines-eng-web/48802a38a5>
- Simchon, A., Edwards, M., & Lewandowsky, S. (2024). The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS Nexus*, 3(2), pgae035. <https://doi.org/10.1093/pnasnexus/pgae035>
- Simon, F. M. (2024a). Escape Me If You Can: How AI Reshapes News Organisations' Dependency on Platform Companies. *Digital Journalism*, 12(2), 149–170. <https://doi.org/10.1080/21670811.2023.2287464>
- Simon, F. M. (2024b). Escape Me If You Can: How AI Reshapes News Organisations' Dependency on Platform Companies. *Digital Journalism*, 12(2), 149–170. <https://doi.org/10.1080/21670811.2023.2287464>
- Simon, F. M. (2024b). *Artificial Intelligence in the News: How AI Retools, Rationalizes, and Reshapes Journalism and the Public Arena*. <https://doi.org/10.7916/ncm5-3v06>
- Simon, F. M., & Altay, S. (2025). Don't panic (yet): Assessing the evidence and discourse around generative AI and elections. *Knight First Amendment Institute at Columbia University*. <https://ora.ox.ac.uk/objects/uuid:ea305c74-27f2-41cd-a6f5-010dd90e60eb>
- Simon, F. M., Nielsen, R. K., & Fletcher, R. (2025). *Generative AI and news report 2025: How people think about AI's role in journalism and society*. Reuters Institute for the Study of Journalism. <https://doi.org/10.60625/RISJ-5BJV-YT69>
- Sjøvaag, H., Ferrer-Conill, Raul, & Olsen, R. Kr. (2024). Capture Beyond the Platforms: The Material and Infrastructural Conditions for Digital Journalism. *Digital Journalism*, 0(0), 1–20. <https://doi.org/10.1080/21670811.2024.2377078>
- Small, C. T., Vendrov, I., Durmus, E., Homaei, H., Barry, E., Cornebise, J., Suzman, T., Ganguli, D., & Megill, C. (2023). *Opportunities and Risks of LLMs for Scalable Deliberation with Polis* (arXiv:2306.11932). arXiv. <https://doi.org/10.48550/arXiv.2306.11932>
- Stamboliev, E. (2023). Proposing a Postcritical AI Literacy: Why We Should Worry Less about Algorithmic Transparency and More about Citizen Empowerment. *Media Theory*, 7(1), 201–232. <https://doi.org/https://doi.org/10.70064/mt.v7i1.890>
- Stryker, Cole, C., & Scapicchio, M. (2025). What is generative AI. *IBM.Com*. <https://www.ibm.com/think/topics/generative-ai>

- Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus*, 3(9), pgae346. <https://doi.org/10.1093/pnasnexus/pgae346>
- Timmer, A. (2014). *Strengthening the Equality Analysis of the European Court of Human Rights: The Potential of the Concepts of Stereotyping and Vulnerability*. Ghent University.
- Tobitt, C. (2025, November 4). Who's suing AI and who's signing: Getty signs with Perplexity, Penske Media sues Google over AI Overviews. *Press Gazette*. <https://pressgazette.co.uk/platforms/news-publisher-ai-deals-lawsuits-openai-google/>
- Tsai, L. L., Pentland, A. S., Braley, A., Chen, N. L., Ramón Enríquez, J., & Reuel, A. (2024). Generative AI for Pro-Democracy Platforms. *An MIT Exploration of Generative AI*. <https://mitgovlab.org/research/generative-ai-for-pro-democracy-platforms/>
- van der Vlist, F., Helmond, A., & Ferrari, F. (2024). Big AI: Cloud infrastructure dependence and the industrialisation of artificial intelligence. *Big Data & Society*, 11(1). <https://doi.org/10.1177/20539517241232630>
- van Drunen, M. Z. (2025). Safeguarding media freedom from infrastructural reliance on AI companies: The role of EU law. *Telecommunications Policy*, 49(7), 102990. <https://doi.org/10.1016/j.telpol.2025.102990>
- Veldkamp, C., & van Marum, E. (2025). *Beantwoording vragen over het bericht 'Wake-up call: Microsoft sluit e-mail ICC zonder pardon af.'* [Kamerstuk]. Ministerie van Algemene Zaken. <https://www.rijksoverheid.nl/documenten/kamerstukken/2025/07/01/beantwoording-vragen-over-het-bericht-wake-up-call-microsoft-sluit-e-mail-icc-zonder-pardon-af>
- Verma, P., & Zakrzewski, C. (2024, April 23). AI deepfakes threaten to upend global elections. No one can stop them. *The Washington Post*. <https://www.washingtonpost.com/technology/2024/04/23/ai-deepfake-election-2024-us-india/>
- Viljoen, S. (2021). A Relational Theory of Data Governance. *The Yale Law Journal*, 82.
- Vipra, J., & Myers West, S. (2023, September 27). Computational Power and AI. *AI Now Institute*. <https://ainowinstitute.org/publications/compute-and-ai>
- Votta, F., & Kruschinski, S. (2025). CampAlignTracker.nl: Detection and Analysis of Visual GenAI in the 2025 Dutch Parliamentary Election. *OSF*. <https://osf.io/sjc2b>
- Wachter, S. (2023). The Theory of Artificial Immutability: Protecting Algorithmic Groups Under Anti-Discrimination Law. *Tulane Law Review*, 97(2), 149.
- Waldo, J., & Boussard, S. (2024). GPTs and Hallucination: Why do large language models hallucinate? *Queue*, 22(4), 19–33. <https://doi.org/10.1145/3688007>
- Weikmann, T., Greber, H., & Nikolaou, A. (2025). After Deception: How Falling for a Deepfake Affects the Way We See, Hear, and Experience Media. *The International Journal of Press/Politics*, 30(1), 187–210. <https://doi.org/10.1177/19401612241233539>
- Weikmann, T., & Lecheler, S. (2023). Visual disinformation in a digital age: A literature synthesis and research agenda. *New Media & Society*, 25(12), 3696–3713. <https://doi.org/10.1177/14614448221141648>
- Weikmann, T., Wouters, F., Tulin, M., Hameleers, M., de Vreese, C. H., Zarouali, B., & Opgenhaffen, M. (under review). *On the same page? Experts are mostly, but not always aligned in their perceptions of self-competence, risks and responsibilities surrounding disinformation in times of generative AI*.

- Wetenschappelijke Raad voor Regeringsbeleid (WRR). (2021). *Opgave AI. De nieuwe systeemtechnologie*. <https://www.wrr.nl/publicaties/rapporten/2021/11/11/opgave-ai-de-nieuwe-systeemtechnologie>
- Wetenschappelijke Raad voor Regeringsbeleid (WRR). (2024). *Aandacht voor media. Naar nieuwe waarborgen voor hun democratische functies*. WRR.
- Whittaker, M. (2021). The steep cost of capture. *Interactions*, 28(6), 50–55. <https://doi.org/10.1145/3488666>
- Widder, D. G., & Kim, N. (2025). *How Big Cloud becomes Bigger: Scrutinizing Google, Microsoft, and Amazon's investments* (SSRN Scholarly Paper No. 5377426). Social Science Research Network. <https://doi.org/10.2139/ssrn.5377426>
- Xenidis, R. (2025). *Legal Protection Against Algorithmic Discrimination in Europe: Current Frameworks and Remaining Gaps* (p. 73). Council of Europe. <https://rm.coe.int/legal-protection-eng-web/48802a38d7>
- Young, I. M. (2002). *Inclusion and Democracy* (online edition). Oxford University Press. <https://doi.org/10.1093/0198297556.001.0001>
- Zuiderveen Borgesius, F. J., Möller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., Bodo, B., & De Vreese, C. (2018). Online Political Microtargeting: Promises and Threats for Democracy. *Utrecht Law Review*, 14(1), 82. <https://doi.org/10.18352/ulr.420>

Instruments du Conseil de l'Europe

Traités

- CoE (1950). Convention de sauvegarde des droits de l'homme et des libertés fondamentales (STE n° 005) <https://www.coe.int/fr/web/conventions/full-list?module=treaty-detail&treatynum=005>
- CoE (1981 et 2018). Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (STE n° 108) <https://www.coe.int/fr/web/conventions/full-list?module=treaty-detail&treatynum=108> et Protocole amendant la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (CETS n° 223) <https://www.coe.int/fr/web/conventions/full-list?module=treaty-detail&treatynum=223>
- CoE (2024). Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (STE n° 225) <https://www.coe.int/fr/web/conventions/full-list?module=treaty-detail&treatynum=225>

Autres instruments

- CoE. (2016). Recommandation CM/Rec(2016)1 du Comité des ministres aux États membres sur la protection et la promotion du droit à la liberté d'expression et du droit à la vie privée dans le contexte de la neutralité du réseau <https://search.coe.int/cm?i=09125948801fe6c6>
- CoE. (2017). Recommandation CM/Rec(2017)8 du Comité des Ministres aux États membres sur le Big Data au service de la culture, de l'alphabétisation et de la démocratie <https://search.coe.int/cm?i=091259488022cd3d>

- CoE. (2018a). Recommandation CM/Rec(2018)1 du Comité des ministres aux États membres sur le pluralisme des médias et la transparence de la propriété des médias
<https://search.coe.int/cm?i=091259488022ff43>
- CoE. (2018b). Recommandation CM/Rec(2018)2 du Comité des ministres aux États membres sur les rôles et responsabilités des intermédiaires de l'internet
<https://search.coe.int/cm?i=091259488022ff4e>
- CoE. (2019). Déclaration du Comité des ministres sur les capacités de manipulation des processus algorithmiques (adoptée par les Délégués des ministres le 13 février 2019)
<https://search.coe.int/cm?i=0912594880233b7e>
- CoE. (2020). Recommandation CM/Rec(2020)1 du Comité des ministres aux États membres sur les incidences des systèmes algorithmiques sur les droits de l'homme
<https://search.coe.int/cm?i=09125948802662bf>
- CoE. (2021a). Lignes directrices sur la protection des personnes à l'égard du traitement des données à caractère personnel par et pour les campagnes politiques. Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel.
https://edoc.coe.int/fr/module/ec_addformat/download?cle=7d787bb950167c81630d2f1256eec848&k=0774034411ed423d22776a377fd233b1
- CoE. (2021b). Note d'orientation sur la hiérarchisation des contenus d'intérêt public en ligne. Comité directeur sur les médias et la société de l'information.
<https://rm.coe.int/cdmsi-2021-009-note-dorientation-sur-la-hierarchisation-des-contenus-d/1680a524c5>
- CoE. (2022a). Recommandation CM/Rec(2022)11 du Comité des ministres aux États membres sur les principes de gouvernance des médias et de la communication
<https://search.coe.int/cm?i=09125948802680f6>
- CoE. (2022b). Recommandation CM/Rec(2022)16 du Comité des ministres aux États membres sur la lutte contre les discours de haine
<https://search.coe.int/cm?i=0912594880268339>
- CoE. (2022c). Recommandation CM/Rec(2022)12 du Comité des ministres aux États membres sur la communication électorale et la couverture médiatique des campagnes électorales
<https://search.coe.int/cm?i=09125948802680fa>
- CoE. (2022d). Recommandation CM/Rec(2022)4 du Comité des ministres aux États membres sur la promotion d'un environnement favorable au journalisme de qualité à l'ère numérique
<https://search.coe.int/cm?i=0912594880267f96>
- CoE. (2022e). Recommandation CM/Rec(2022)13 du Comité des ministres aux États membres sur les impacts des technologies numériques sur la liberté d'expression
<https://search.coe.int/cm?i=09125948802680f8>
- CoE. (2023a). Note d'orientation sur la lutte contre la propagation de la désinformation et des fausses informations en ligne par le biais de la vérification des faits et de solutions de conception des plateformes respectueuses des droits de l'homme. Comité directeur sur les médias et la société de l'information.
<https://rm.coe.int/cdmsi-2023-15-note-d-orientation-msi-inf/1680add25f>

- CoE. (2023b). Lignes directrices sur la mise en œuvre responsable des systèmes d'intelligence artificielle dans le journalisme. Comité directeur sur les médias et la société de l'information.
https://edoc.coe.int/fr/module/ec_addformat/download?cle=63f35343d6f9681cad0598452a17f0df&k=0774034411ed423d22776a377fd233b1
- CoE. (2025). *Note d'orientation sur les implications de l'IA générative pour la liberté d'expression* (CDMSI(2025)15rev).
<https://rm.coe.int/cdmsi-2025-15rev-note-dorientation-sur-les-implications-de-lintelligen/488029df81>

Jurisprudence de la Cour européenne des droits de l'homme

- Aksu c. Turquie, Nos. 4149/04, 41029/04 (CEDH [GC] 15 mars 2012).
<https://hudoc.echr.coe.int/eng?i=001-109578>
- Carvalho Pinto de Sousa Morais c. Portugal, No. 17484/15 (CEDH 25 juillet 2017).
<https://hudoc.echr.coe.int/eng?i=001-176013>
- Centro Europa 7 S.r.l. et Di Stefano c. Italie, No. 38433/09 (CEDH 7 juin 2012).
<https://hudoc.echr.coe.int/eng?i=001-111396>
- D.H. et autres c. République tchèque, No. 57325/00 (CEDH [GC] 13 novembre 2007).
<https://hudoc.echr.coe.int/fre?i=001-83258>
- Dink c. Turquie, No. 2668/07 (CEDH 14 septembre 2010)
<https://hudoc.echr.coe.int/fre?i=001-100383>
- Féret c. Belgique, No. 15615/07 (CEDH 16 juillet 2009).
<https://hudoc.echr.coe.int/eng?i=001-93626>
- Handyside c. Royaume Uni, No. 5493/72 (CEDH 7 décembre 1976)
<https://hudoc.echr.coe.int/fre?i=001-62057>
- Konstantin Markin v. Russia, No. 30078/06 (ECtHR [GC] 22 March 2012).
<https://hudoc.echr.coe.int/fre?i=001-109871>
- Lingurar c. Roumanie, No. 48474/14 (CEDH 16 avril 2019).
<https://hudoc.echr.coe.int/eng?i=001-192466>
- Sunday Times c. Royaume Uni, No. 6538/74 (CEDH 26 avril 1979).
<https://hudoc.echr.coe.int/fre?i=001-62140>