

# L'IA et le droit à l'information

*Les 7 recommandations de RSF  
pour protéger un espace informationnel  
dominé par l'intelligence artificielle*

Pour éviter que l'intelligence artificielle (IA), y compris l'IA générative (GenAI), ne devienne une arme de désinformation massive, RSF appelle à : (1) une régulation renforcée ciblant l'IA et le droit à une information fiable, (2) des politiques publiques ambitieuses soutenant des systèmes d'IA dignes de confiance dans l'écosystème de l'information et (3) une véritable gouvernance institutionnelle dans ce domaine.

Ces recommandations découlent de différents projets et initiatives de RSF ciblant l'IA et l'information : la « [Charte de Paris sur l'IA et le journalisme](#) », qui établit le premier cadre éthique mondial guidant les médias à l'ère de l'IA, le [projet Spinoza](#), un modèle de langage développé par et pour les médias et le rapport « [L'IA comme bien public : garantir un contrôle démocratique de l'IA dans l'espace de l'information](#) » du Forum sur l'information et la démocratie avec la contribution de RSF.

## I. RÉGULATION

**Recommandation 1 : Exiger que les systèmes d'IA donnant accès à l'actualité et à l'information promeuvent des sources fiables et diverses.**

- Classer les systèmes d'intelligence artificielle utilisés pour la diffusion d'informations dans la catégorie des systèmes à haut risque et rendre obligatoire la réalisation d'évaluations indépendantes.
- Exiger que ces systèmes amplifient la visibilité des sources d'information fiables.
- Exiger que ces systèmes adhèrent aux valeurs fondamentales du journalisme et du pluralisme.

**Recommandation 2 : Établir un cadre de responsabilité clair pour combattre la création et la diffusion de "deepfakes" nuisibles.**

- Mettre en place un régime de responsabilité légale transversal assorti de sanctions pénales pour la publication intentionnelle de "deepfakes" préjudiciables.
- Obliger les fournisseurs d'IA à empêcher la production de "deepfakes" nuisibles et réalisés sans le consentement des personnes visées.
- Encourager l'adoption de normes d'authenticité pour les équipements d'enregistrement, les médias et les plateformes numériques.

**Recommandation 3 : Garantir des droits pour les journalistes et les médias face à l'IA.**

- Garantir la transparence et le droit "d'opt-out" pour les médias et les journalistes lorsque leur contenu est utilisé pour entraîner l'IA ou générer du contenu.
- Garantir une rémunération juste et équitable pour l'utilisation du contenu journalistique via une redistribution indépendante et transparente.

## II. POLITIQUES PUBLIQUES

### **Recommandation 4 : Soutenir et financer la sécurité de l'IA et son évaluation indépendante.**

- Exiger un financement proportionnel pour la sécurité de l'IA.
- Soutenir et financer une évaluation indépendante des systèmes d'IA dans le domaine de l'information.
- Mettre en place une plateforme européenne documentant les systèmes d'IA et leurs évaluations.
- Mettre en place un régime de responsabilité pour les développeurs d'IA pour qu'ils répondent des dommages causés dans l'espace informationnel.

### **Recommandation 5 : Reconnaître l'information comme un secteur prioritaire pour le développement de systèmes d'IA dignes de confiance.**

- Investir dans l'IA et les plateformes numériques d'intérêt public.
- Financer le développement d'un modèle de langage fondé sur les données des médias européens.
- Tenir un registre public des données légalement accessibles pour entraîner des modèles d'IA.

## III. GOVERNANCE

### **Recommandation 6 : Mettre fin à la course à l'armement en matière d'IA et promouvoir la coopération internationale pour la gouvernance de l'IA.**

- Renforcer la coopération internationale dans le cadre du partenariat sur l'information et la démocratie.
- Établir des normes et des règles mondiales pour les systèmes d'IA dans l'espace informationnel.
- Impliquer les médias et les journalistes dans la gouvernance internationale de l'IA.

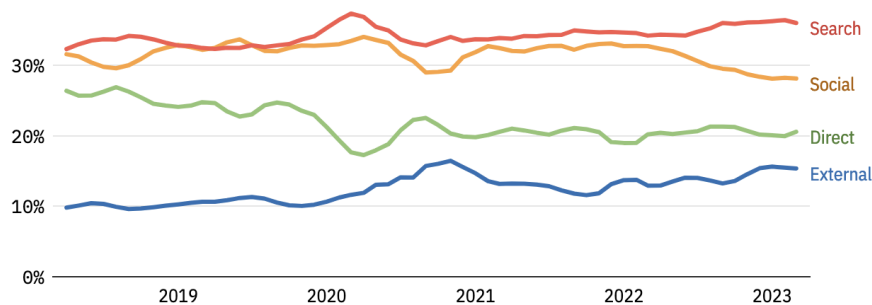
### **Recommandation 7 : Veiller à ce que les capacités de régulation et de supervision de l'IA soient en phase avec la puissance économique et le pouvoir d'influence du secteur.**

- Allouer des moyens humains et matériels de régulation de l'IA proportionnés à la puissance économique du secteur.
- Contrôler le rythme de développement de l'IA pour l'aligner sur les capacités de régulation et de supervision.
- Veiller à ce que les consultations du Bureau de l'IA impliquent principalement la recherche indépendante et des parties prenantes défendant l'intérêt public.

## I. Réguler l'IA pour prévenir les atteintes au droit à l'information fiable

### Recommandation 1 : Exiger que les systèmes d'IA donnant accès à l'actualité et à l'information promeuvent des sources fiables et diverses.

Les moteurs de recherche et les réseaux sociaux sont les principales sources d'audience pour les médias en ligne, représentant [environ deux tiers de leur audience](#). La dépendance des médias à ces plateformes pour atteindre leur public constitue une menace pour l'indépendance et la viabilité du journalisme. Plus précisément, deux entreprises, Google et Meta, occupent une position dominante au niveau mondial dans la distribution des nouvelles et de l'information : **cinq milliards de personnes accèdent chaque jour à l'information par l'intermédiaire de leurs plateformes et de leurs services.**



Source: Press Gazette analysis of Chartbeat data

PressGazette

[Répartition du trafic](#) de 546 sites d'information britanniques et américains par origine. Rouge : moteurs de recherche (~95 % pour Google) ; Orange : médias sociaux (~80 % pour Facebook) ; Vert : connexion directe sur les pages web des médias ; Bleu : trafic provenant d'autres pages web.

La moitié de l'humanité passe [deux heures et demie](#) par jour sur les médias sociaux, où elle consomme majoritairement des contenus recommandés par IA. Les systèmes de recommandation optimisent les revenus publicitaires et le temps passé par les utilisateurs au détriment de la qualité de l'information, amplifiant les [contenus polarisants, nuisibles et sensationnalistes](#). Les dirigeants de ces plateformes disposent d'un pouvoir de décision discrétionnaire sans contre-pouvoir significatif, ce qui leur permet d'orienter la diffusion mondiale de l'information en fonction de leurs intérêts ou de leur idéologie.

### Exemples de préjudices

- Suite à une décision unilatérale de leurs dirigeants, entre fin 2022 et fin 2023, Facebook et X ont respectivement [réduit](#) de 52% et 27% la [visibilité des sites d'information journalistique](#) sur leurs plateformes.
- Sur Tiktok, Dylan Page, un jeune influenceur britannique, voit régulièrement ses contenus davantage recommandés que ceux de la BBC ou du New York Times réunis, même sur des sujets importants comme Gaza.

- Au cours des quatre premiers mois de la pandémie de Covid-19, 34 sites web de désinformation [ont recueilli](#) 80 millions d'interactions sur Facebook, soit plus de dix fois le trafic reçu par le compte de l'OMS sur la même période. Des centaines de milliers de décès [auraient pu être évités](#) grâce à un espace de l'information plus sûr.
- Les plateformes de médias sociaux jouent un rôle important dans l'amplification de la désinformation, souvent avec des conséquences dévastatrices. Elles ont par exemples [déclenché des violences dans le Tigré, en Éthiopie](#), et [alimenté la haine contre les Rohingyas au Myanmar](#).
- Lors de l'élection présidentielle américaine de 2016, l'algorithme de YouTube était [six fois plus susceptible](#) de recommander des vidéos en faveur de Donald Trump que de Hillary Clinton, ce qui a sans doute joué un rôle important dans le résultat de l'élection.
- Lors des élections suisses et allemandes de 2023, il a été [démontré que](#) Microsoft Copilot commettait des erreurs dans 30 % des cas lorsqu'il répondait à des questions factuelles liées aux élections. Les réponses incorrectes étaient souvent attribuées à tort à des sources journalistiques.

### Cadre juridique actuel

La loi sur les services numériques (DSA) exige des très grandes plateformes en ligne (VLOPs), des très grands moteurs de recherche en ligne (VLOSEs) :

- de divulguer les critères utilisés par leurs systèmes de recommandation et les éventuelles possibilités de les modifier (article 27)
- d'offrir au moins un système de recommandation optionnel ne reposant pas sur le profilage (article 38) (N.B. : l'expérience montre que seule une très petite fraction des utilisateurs sélectionne une option alternative lorsque le mode par défaut du système est de maximiser l'engagement, en particulier lorsque l'option est dissimulée).
- d'évaluer et d'atténuer les "risques systémiques" soulevés par leurs services (art. 34 et 35)
- de faire l'objet d'audits externes réguliers visant à évaluer si les risques systémiques identifiés sont correctement surveillés et atténués (article 37)
- de permettre à des chercheurs agréés d'accéder à leurs données (art. 40)

En vertu de la loi sur l'IA :

- Les systèmes d'IA utilisés dans les infrastructures critiques, les services publics ou l'emploi sont considérés comme présentant un risque élevé et, à ce titre, doivent faire l'objet d'une auto-évaluation, d'essais et de mesures d'atténuation des risques ;
- Les fournisseurs de modèles d'IA à usage général (GPAI) présentant des risques systémiques doivent procéder à une auto-évaluation, à des essais et à une atténuation des risques ;
- **Les systèmes d'IA utilisés dans le domaine de l'information ne sont pas considérés comme présentant un risque élevé.**

### Recommandation détaillée

Les systèmes d'IA qui jouent un rôle structurel dans la diffusion des nouvelles et des informations, tels que ceux utilisés par VLOPs, les VLOSEs ou les agents conversationnels répondant à des questions sur les nouvelles et les informations, devraient :

- être considérés comme à haut risque.
- être tenus d'amplifier la visibilité des sources d'information fiables, identifiées comme telles par des normes d'autorégulation telles que la [Journalism Trust Initiative](#) (JTI).

- fournir des garanties solides démontrant leur adhésion aux valeurs fondamentales du journalisme (exactitude, impartialité, absence de préjugé, non-discrimination, responsabilité lorsqu'ils commettent des erreurs) ainsi qu'au pluralisme, lorsqu'ils réutilisent ou retraitent des informations journalistiques. Ces normes doivent être conçues et mises en œuvre par le Bureau de l'IA avec des représentants de l'industrie du journalisme et des médias.
- faire l'objet d'évaluations indépendantes et transversales démontrant leur conformité aux obligations susmentionnées.

## **Recommandation 2 : Établir un cadre de responsabilité clair pour combattre la création et la diffusion des "deepfakes" nuisibles.**

Un an seulement après que les outils d'IA générative sont devenus accessibles au public, le volume d'images créées par l'IA [a dépassé](#) le nombre total de photographies prises par l'humanité au cours des 150 premières années d'existence de la photographie.

Cette explosion de contenus synthétiques réalistes constitue une double menace : d'une part, elle risque d'induire l'opinion publique en erreur en falsifiant des événements ; d'autre part, elle conduit à douter de l'authenticité de tous les contenus, y compris les contenus authentiques.

La loi sur l'IA exige des fournisseurs d'IA qu'ils veillent à ce que les contenus générés par l'IA soient marqués dans un format lisible par la machine afin d'indiquer leur nature artificielle. Il est demandé aux fournisseurs d'IA de divulguer l'origine artificielle des "deepfakes". Cependant, la transparence ne suffit pas : les deepfakes, même lorsqu'ils sont identifiés comme tels, peuvent encore porter préjudice à des personnes, créer de la confusion et polluer des informations factuelles. En outre, les utilisateurs malveillants peuvent facilement contourner les obligations de transparence en utilisant des outils à code source ouvert. Enfin, les outils de détection [ne sont pas fiables](#).

### **Exemples de préjudices**

- Les "deepfakes" sont [largement utilisés](#) lors des élections. Cette technologie peut tromper les électeurs et compromettre la sincérité du scrutin.
- Des milliers de fausses campagnes publicitaires publiées sur les plates-formes de Meta comportaient de [fausses vidéos de journalistes danois](#). La plupart des contenus n'ont pas été supprimés par Meta, bien qu'ils aient été signalés comme frauduleux.
- ElevenLabs permet de [reproduire la voix de n'importe qui](#) sur la base de quelques secondes d'audio, les garde-fous de sécurité étant facilement [contournés](#).
- Meta a récemment [déclaré](#) s'appuyer sur une équipe d'une centaine de "fact-checkers" indépendants pour identifier les "deepfakes". Pour une plateforme comptant 3 milliards d'utilisateurs réguliers, cela fait 1 modérateur pour 30 millions d'utilisateurs, ce qui met en évidence l'inadéquation des moyens déployés avec l'ampleur du problème.

### **Cadre juridique actuel**

En vertu de la loi sur l'IA :

- les fournisseurs de systèmes d'IA sont tenus de veiller à ce que tous les contenus générés par l'IA soient étiquetés d'une manière lisible par les machines ;
- les créateurs de deepfakes doivent afficher de manière transparente la nature synthétique du contenu.

Dans le cadre du DSA, la Commission a publié des [lignes directrices](#) pour l'atténuation des risques systémiques en ligne lors des élections. Il est notamment demandé aux VLOPs et aux VLOSEs d'évaluer et d'atténuer les risques spécifiques liés à l'IA générative, par exemple en étiquetant clairement les "deepfakes".

[La France](#) et [l'Italie](#) ont établi des sanctions pénales pour la création et la diffusion de deepfakes, respectivement, lorsqu'ils sont réalisés sans le consentement des personnes visées et lorsqu'ils causent un préjudice à quelqu'un.

### Recommandation détaillée

- Un régime de responsabilité clair et transversal devrait être mis en place pour combattre la prolifération de "deepfakes" nuisibles, y compris des sanctions pénales pour la publication intentionnelle de "deepfakes" causant un préjudice à autrui ou créés sans le consentement des personnes visées.
- Les fournisseurs d'IA devraient être tenus d'empêcher efficacement la création de deepfakes nuisibles et non consentis. Il existe un droit à la confidentialité de l'image ; toute personne devrait avoir le droit de ne pas se faire voler son apparence.
- L'UE devrait encourager le développement d'une norme d'authenticité pour les fabricants de matériel d'enregistrement (caméras, etc.), et l'adoption rapide de ces normes par les médias et les plateformes numériques.

### Recommandation 3 : Établir des droits pour les médias et les journalistes face à l'IA.

Le rôle croissant des intermédiaires entre les médias et leur public a entraîné une perte importante de revenus publicitaires pour les médias, car une part substantielle de ces revenus est captée par les moteurs de recherche et les médias sociaux qui utilisent un pouvoir de négociation asymétrique. Plus récemment, la majorité des contenus journalistiques mondiaux ont été récupérés par les entreprises d'IA pour entraîner leurs modèles, [sans tenir compte](#) des lois sur les droits d'auteur et des créateurs de contenu.

Étant donné que les agents conversationnels doivent s'appuyer sur des informations à jour pour alimenter leurs réponses, les sociétés d'IA auront toujours besoin de contenus journalistiques. Un nombre croissant d'éditeurs de presse (environ 50 %) [ont bloqué les robots des sociétés d'IA](#) qui accédaient à leurs contenus, tandis que les sociétés d'IA ont commencé à contracter avec certains éditeurs de presse pour obtenir des licences sur leur contenu.

Les chatbots qui utilisent le contenu des médias font peser deux menaces sur la pérennité du journalisme. Tout d'abord, ils risquent de détourner le trafic des médias en utilisant leurs données tout en leur offrant une compensation insuffisante. Deuxièmement, ils risquent d'exercer une influence opaque et arbitraire en choisissant les contenus médiatiques qui font l'objet d'une licence et ceux qui restent invisibles sur leurs systèmes.

### Exemples de préjudices

- L'industrie de l'information subit des [licenciements massifs](#) tandis que les entreprises d'IA connaissent la [croissance la plus rapide](#) de leur histoire, en partie grâce à la valeur créée à partir de données protégées par les droits d'auteur qu'elles n'ont pas payés.

- OpenAI a signé un accord avec quelques éditeurs comme Axel Springer ou Le Monde, risquant de fournir une préférence d'accès aux contenus de ces médias (dont un tabloïd comme Bild) aux plus de 100 millions d'utilisateurs réguliers de ChatGPT.

### Cadre juridique actuel

La législation sur l'IA (AI Act) impose aux systèmes d'IA à usage général (GPAI) (tels que ceux sur lesquels s'appuient certains chatbots) de satisfaire à des exigences de transparence, notamment le respect de la législation européenne sur le droit d'auteur et la publication de résumés détaillés du contenu utilisé pour leur entraînement. Toutefois, le contexte juridique reste peu protecteur des droits d'auteur, avec la doctrine américaine du "fair use" et l'exception européenne relative à l'exploration de textes et de données (Text and Data Mining - TDM). Le cadre juridique actuel ne garantit pas une rémunération équitable ni de conditions de négociation satisfaisantes pour les journalistes et les médias.

### Recommandation détaillée

Il convient d'établir des droits opposables pour les médias, les journalistes et les créateurs de contenu en ce qui concerne l'utilisation de leurs contenus. Ces droits devraient inclure :

- Le droit à la transparence : ils doivent être informés lorsque leur contenu est utilisé pour entraîner l'IA ou créer du contenu généré par l'IA.
- Le droit "d'opt out" : ils devraient pouvoir décider si leur contenu peut ou être utilisé par les systèmes d'IA.
- Le droit à une rémunération équitable : ils doivent recevoir une compensation chaque fois que leur contenu est utilisé pour entraîner ou alimenter un système d'IA. Cela pourrait être facilité par une taxe sur les revenus des entreprises d'IA et/ou en considérant l'utilisation non-consentie de contenus comme une violation des droits d'auteur.

Ces mesures devraient être mises en œuvre par le biais de **négociations collectives avec les éditeurs et les représentants des journalistes**. Les ressources ainsi collectées devraient être distribuées de manière indépendante et transparente pour soutenir l'indépendance éditoriale et la viabilité du journalisme d'intérêt public.



## II. Des politiques publiques en faveur de systèmes d'IA dignes de confiance dans l'écosystème de l'information

### Recommandation 4 : Soutenir et financer la sécurité de l'IA et son évaluation indépendante.

La grande majorité des moyens déployés pour construire et évaluer les systèmes d'IA proviennent du secteur privé, ce qui entraîne des conflits d'intérêts, une sous-estimation ou une omission des risques, des efforts de recherche [fondés sur la recherche de profit](#) plutôt que sur l'intérêt public, des [menaces pour l'évaluation indépendante](#) et un pouvoir de lobbying disproportionné. [Seuls 2 % environ](#) des publications scientifiques sur l'IA au niveau mondial traitent des questions de sécurité. Le déséquilibre en matière de financement de la sécurité des systèmes d'IA est encore pire : il y a un ratio de 1000 pour 1 entre les ressources consacrées à l'augmentation des capacités de l'IA et celles allouées à l'amélioration de la sécurité de l'IA. Les entreprises d'IA [drainent](#) également [les talents des universités](#) et [étendent](#) rapidement leur [influence](#) dans le monde universitaire.

### Exemples de préjudices

- Aujourd'hui, environ [70 % des titulaires d'un doctorat](#) en intelligence artificielle trouvent un emploi dans le secteur privé, contre 20 % il y a vingt ans. L'industrie privée est désormais le principal moteur de la recherche en IA, influençant de manière significative les objectifs pour lesquels l'IA est utilisée, ainsi que son évaluation et sa réglementation.
- Une [enquête menée en 2023](#) a révélé que 70 % des meilleurs experts en IA étaient favorables à accorder davantage de priorité à la sécurité de l'IA.
- Les grandes entreprises technologiques [licencient](#) ou [réduisent](#) régulièrement leurs équipes chargées de l'éthique, en particulier lorsque celles-ci [tirent la sonnette d'alarme](#).

### Législation actuelle

Le [Bureau européen de l'IA](#) a été créé dans le but d'être le centre d'expertise en matière d'IA et la pierre angulaire d'un système européen unifié de gouvernance de l'IA. Il vise à "favoriser une IA digne de confiance dans le marché intérieur" en collaboration avec les acteurs publics et privés concernés. Il jouera un rôle clé dans la mise en œuvre de législation sur l'IA et la préparation des actes délégués, notamment en développant des outils et des méthodologies pour évaluer les capacités et la portée des modèles d'IA à usage général, en classant les modèles présentant des risques systémiques, en élaborant des codes de pratique et en enquêtant sur toute violation des règles.

### Recommandation détaillée

- Exiger un financement proportionnel pour la sécurité : pour chaque dollar investi dans les capacités d'IA, les développeurs d'IA devraient être tenus d'allouer un montant proportionnel aux mesures de sécurité de l'IA et aux cadres de gouvernance.
- Imposer aux modèles d'IA des exigences strictes en matière de sécurité afin de garantir qu'ils respectent les normes éthiques et juridiques reconnues avant d'être diffusés.
- Mettre à disposition du public une plateforme de documentation des systèmes d'IA, rassemblant des informations détaillées sur leurs vulnérabilités, des évaluations indépendantes et des recommandations.

- Adopter des régimes de responsabilité rendant les développeurs d'IA responsables des préjudices causés par leurs systèmes dans l'espace informationnel. Collaborer avec les parties prenantes concernées, y compris les journalistes et les médias, pour déterminer dans quels cas de violation des droits fondamentaux opposer ce régime de responsabilité.

### **Recommandation 5 : Reconnaître l'information comme un secteur prioritaire pour le développement de systèmes d'IA dignes de confiance.**

L'état actuel de l'industrie de l'IA, caractérisé par une poignée d'entreprises privées qui développent et mettent en accès leurs systèmes d'IA à des milliers d'entreprises du secteur de l'information, soulève des inquiétudes quant à l'indépendance du journalisme, à la souveraineté technologique des médias, à la protection des données confidentielles et au respect de la propriété intellectuelle. Les entreprises technologiques occupent déjà une position oligopolistique dans la *diffusion de l'information*. L'IA générative pourrait leur permettre d'étendre leur domination à la *production d'informations*, ce qui constitue un risque important pour l'indépendance, le pluralisme et la viabilité du journalisme.

#### **Exemples de préjudices**

- Jusqu'en 2014, les modèles d'IA les plus importants étaient publiés par des universités. Depuis, l'industrie privée occupe une position dominante. Aujourd'hui, les investissements dans l'IA et le développement de modèles d'IA sont [largement dominés par](#) les grandes entreprises technologiques américaines.
- Les start-ups "européennes" spécialisées dans l'IA entretiennent des [liens étroits](#) avec les entreprises américaines et sont largement influencées par ces dernières.

#### **Législation actuelle**

En janvier 2024, la Commission européenne a lancé un [train de mesures sur l'IA](#) pour soutenir les petites et moyennes entreprises dans le développement d'une IA digne de confiance qui s'aligne sur les valeurs et les réglementations de l'UE. L'initiative "GenAI4EU", qui fait partie de cet ensemble, vise à encourager le développement de nouvelles applications de l'IA dans 14 secteurs industriels européens ainsi que dans le secteur public. Ces secteurs comprennent la robotique, la santé, la biotechnologie, la fabrication, la mobilité, le climat et les mondes virtuels, mais **ils excluent de façon notable le secteur de l'information.**

#### **Recommandation détaillée**

- GenAI4EU devrait reconnaître l'information comme un domaine prioritaire pour développer des outils d'IA et des plateformes numériques d'intérêt public.
- En particulier, l'UE devrait financer le développement d'un modèle de langage basé sur les données des médias européens, intégrant des normes élevées de pluralisme et de fiabilité. Ex : [le projet Spinoza](#).
- L'UE devrait soutenir le mouvement en faveur de l'ouverture des données et tenir un registre public des données légalement accessibles aux systèmes d'IA de formation et d'entraînement, en précisant les conditions dans lesquelles ces données sont accessibles aux développeurs et aux déployeurs d'IA.

### III. Renforcer la gouvernance démocratique mondiale de l'IA

#### Recommandation 6 : Mettre fin à la course à l'armement en matière d'IA et promouvoir la coopération internationale pour la gouvernance de l'IA.

La concurrence entre les grandes entreprises technologiques se répercute au niveau des États, où les gouvernements nationaux, conscients de l'importance stratégique de ces technologies pour le maintien de leur leadership géopolitique, sont incités à promouvoir un cadre réglementaire moins stricte au sein de leurs frontières. Les [récits dominants sur l'IA](#) dans l'UE s'appuient souvent sur de vieux concepts de relations internationales tels que la concurrence entre grands blocs géopolitiques et les courses aux armements. Qu'ils soient justifiés ou non, ces récits sapent les efforts de régulation et créent des prophéties auto-réalisatrices. Le risque que l'UE soit dépassée par les États-Unis, ou que l'Occident soit dépassé par la Chine, est régulièrement utilisé pour justifier la dérégulation du secteur. Pourtant, des technologies mal régulées peuvent aussi avoir — et ont eu — des effets dévastateurs dans les pays dont elles sont issues, comme le démontre le cas des réseaux sociaux. Même dans les secteurs où l'innovation peut effectivement conduire à un avantage géopolitique à court terme (comme les [armes autonomes](#)), une course à l'armement de l'IA est toujours une stratégie collectivement perdante à moyen terme, associée à des menaces existentielles pour nos sociétés.

#### Législation actuelle

En mars 2024, l'Assemblée générale des Nations Unies a adopté par consensus une résolution intitulée "[Saisir les opportunités offertes par les systèmes d'intelligence artificielle sûrs et fiables pour le développement durable](#)" - la toute première à établir un consensus mondial sur la gouvernance de l'IA, encourageant les États membres et de multiples parties prenantes (y compris les médias), "à élaborer et à soutenir des approches et des cadres réglementaires et de gouvernance liés aux systèmes d'intelligence artificielle sûrs et fiables [...], reconnaissant qu'un partenariat et une coopération efficaces entre les gouvernements et les parties prenantes multiples sont nécessaires pour élaborer de telles approches et de tels cadres".

En avril 2024, la France et la Chine ont signé une [déclaration conjointe sur l'intelligence artificielle et la gouvernance mondiale](#), reconnaissant les risques associés à cette technologie et appelant à "renforcer la gouvernance mondiale de l'IA dans le but de promouvoir le développement et l'utilisation pour le bien commun". Plus précisément, "les deux parties s'efforceront également de renforcer la coopération internationale et l'interopérabilité entre les cadres et initiatives de gouvernance de l'IA, en s'appuyant sur les travaux menés au niveau des Nations unies".

#### Recommandation détaillée

Seule une gouvernance mondiale et multilatérale peut garantir la préservation de l'intérêt public, la protection des droits de l'homme et l'innovation au profit de tous, devant la recherche de profits à court terme et d'intérêts nationaux particuliers.

- La gouvernance démocratique de l'espace de l'information et de la communication est une condition préalable à la démocratie elle-même. L'UE doit veiller à ce que les principes démocratiques soient à la base de la gouvernance internationale de l'IA en renforçant la coopération au sein du [Partenariat pour l'information et la démocratie](#).

- L'UE doit s'efforcer d'établir des normes, des principes et des règles au niveau mondial, telles que recommandées ci-dessus, à prendre en compte lors du développement, du déploiement et de l'utilisation des systèmes d'IA dans l'espace informationnel.
- En tant que gardiens essentiels du droit à l'information, les journalistes, les médias et les groupes de soutien au journalisme devraient jouer un rôle actif dans la gouvernance internationale des systèmes d'IA.

### **Recommandation 7 : Veiller à ce que les capacités de régulation et de supervision de l'IA soient en phase avec la puissance économique et le pouvoir d'influence du secteur.**

Il existe un décalage croissant entre l'innovation technologique exponentielle et les cadres normatifs et réglementaires, notamment en ce qui concerne la protection des valeurs démocratiques et des droits fondamentaux. En mars 2023, plus de 1 000 experts du secteur ont signé une lettre ouverte appelant à une pause dans le développement de l'IA afin d'évaluer et d'atténuer les dangers de systèmes tels que le GPT-4. Le développement de l'IA est essentiellement le fait d'entreprises privées à la recherche d'une position dominante sur le marché, qui privilégient souvent la vitesse à la sécurité (voir section 2.1). Leurs produits et services atteignent des centaines de millions de personnes sans décision ni évaluation indépendantes préalables. En outre, les développeurs d'IA ne subissent actuellement que peu ou pas de conséquences pour les dommages causés par leurs modèles, ce qui conduit à un déploiement rapide et potentiellement imprudent.

#### **Exemples de préjudices**

- Les entreprises d'IA exercent un [important pouvoir de lobbying](#) au sein de l'UE. En 2023, [78 % des réunions](#) sur l'IA des hauts fonctionnaires de la Commission européenne ont impliqués des représentants d'entreprises, influençant la régulation pour l'aligner sur leurs intérêts (bien que les responsables des grandes entreprises technologiques aient publiquement [appelé](#) à une régulation mondiale de l'IA).
- Un [sondage réalisé par 2023](#) montre qu'une écrasante majorité d'électeurs américains sont préoccupés par les risques liés à l'IA : 72 % d'entre eux souhaitent ralentir son développement, contre seulement 8 % qui préfèrent l'accélérer. Le sondage indique un manque de confiance dans les dirigeants de la technologie pour autoréguler l'IA, 82 % étant en faveur d'une supervision étatique.

#### **Recommandation détaillée**

- Affecter des ressources à l'évaluation, à la législation et à l'application des réglementations en matière d'IA proportionnellement au pouvoir économique du secteur. Pour ce faire, il convient de fournir un financement et une main-d'œuvre suffisants aux organismes de réglementation en allouant 10 % du budget annuel des entreprises d'IA à la gouvernance et à la surveillance de l'IA au niveau international.
- Contrôler le rythme de développement et de déploiement de l'IA. La vitesse de développement de l'IA doit être limitée par notre capacité à contrôler ces systèmes et à mettre en œuvre des mesures de sécurité et des cadres de gouvernance solides.
- Veiller à ce que les consultations du Bureau de l'IA impliquent principalement des parties prenantes défendant l'intérêt public, telles que des organisations de la société civile et des organismes de recherche publics indépendants.