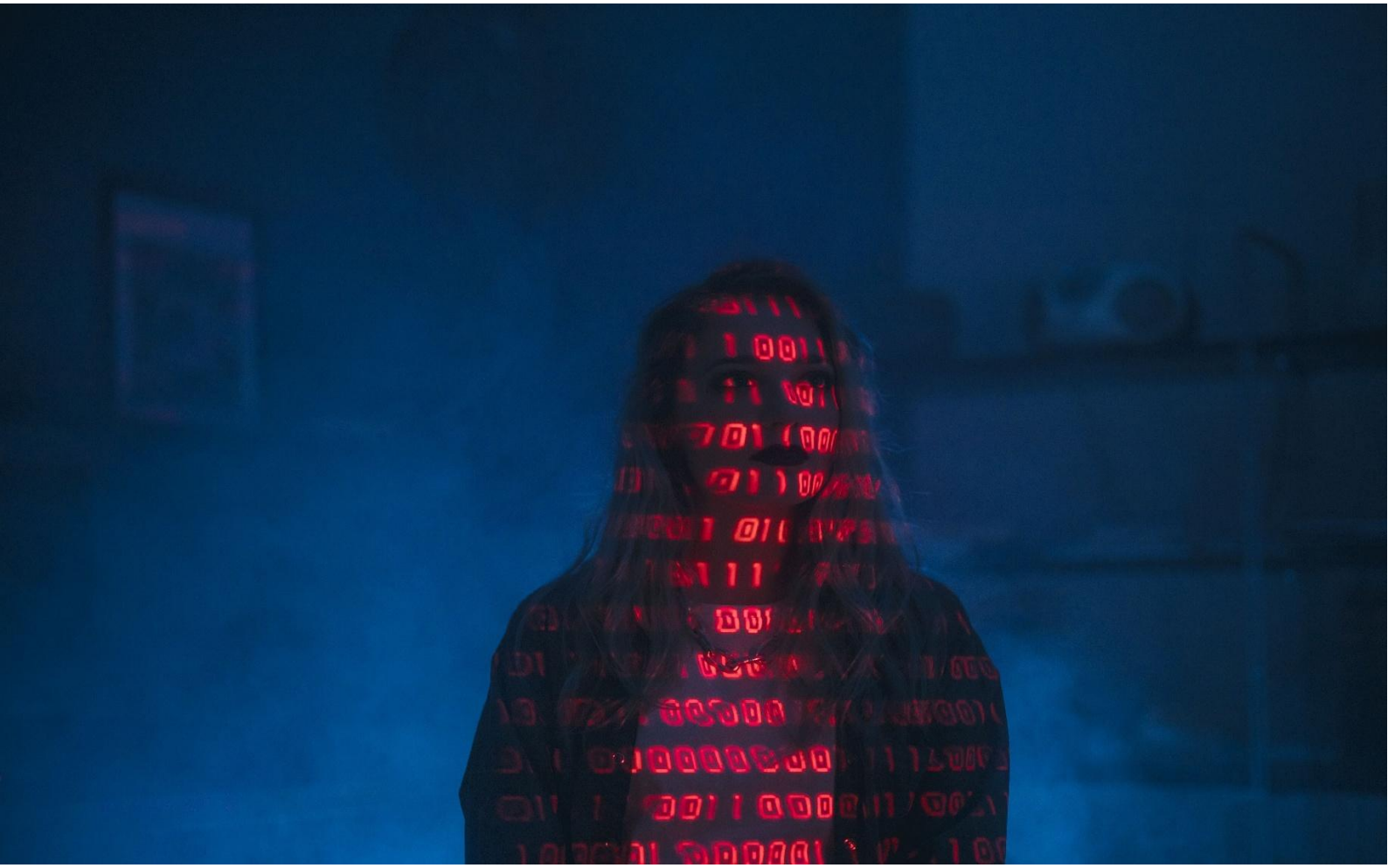
CNCDH
COMMISSION NATIONALE CONSULTATIVE DES DROITS DE L'HOMME
RÉPUBLIQUE FRANÇAISE

AVIS
A – 2026 – 06

Avis pour une conception des services numériques respectueuse des droits humains

25 juin 2026



L'*Avis pour une conception des services numériques respectueuse des droits humains* a été adopté lors de l'Assemblée plénière du 25 juin 2026 (adoption à l'unanimité).

Résumé

Cet *Avis* s'inscrit dans la lignée des travaux de la CNCDH relatifs aux enjeux pour la protection des droits humains soulevés par les nouvelles technologies. Toutefois, la réflexion qui y est élaborée constitue un tournant dans l'approche du sujet, et ce à plusieurs titres. En premier lieu, la CNCDH fait le choix d'une approche holistique plutôt que sectorielle, en considérant que le modèle dominant de conception des services numériques constitue une menace pour les droits humains de chaque personne (quel que soit notamment son âge, son genre, son origine…). Cette approche laisse toutefois la place à la prise en compte de facteurs de vulnérabilité, qui, en particulier lorsqu'ils sont croisés, augmentent les risques de subir des atteintes aux droits humains. En second lieu, plutôt que d'adopter une approche curative, la CNCDH promeut une approche préventive, visant non pas seulement à modérer le contenu, mais à encadrer l'architecture même des services numériques, pour une protection intransigeante des droits humains en ligne.

L'*Avis* propose tout d'abord un état des lieux des atteintes aux droits humains engendrées par la conception des services numériques. Leur modèle économique étant fondé sur la maximisation de l'engagement, la conception de ces services repose sur l'exploitation des vulnérabilités humaines afin de favoriser l'hyperconnexion. En usant de ressorts cognitifs, la conception provoque des comportements addictifs, alimentés notamment par l'hyperpersonnalisation du contenu ainsi que par l'anthropomorphisation des technologies émergentes, en particulier des systèmes d'intelligence artificielle. La conception des services numériques entraîne ainsi des risques tant individuels (contacts malveillants, exposition à des contenus dangereux…) qu'interpersonnels (isolement social, propagation de messages de haine…) et systémiques (dégradation de la santé mentale, désinformation, risques pour la démocratie…). Le panorama ainsi dressé par la CNCDH permet de dénoncer un modèle toxique de conception des services numériques, attentatoire aux droits humains.

Forte de ce constat, la CNCDH propose, dans la deuxième partie de l'*Avis*, plusieurs modalités d'action. Les pistes envisagées visent à imposer une conception des services numériques protectrice des droits humains, qui mette un terme aux dynamiques addictives et à la logique délétère de l'économie de l'attention. En particulier, la CNCDH propose, sur le modèle du droit européen de la consommation, d'interdire certaines pratiques de conception toxiques, et d'en inscrire d'autres au sein d'une liste de pratiques présumées comme telles jusqu'à preuve du contraire. Dans la lignée des propositions du Parlement européen, elle se prononce également en faveur de la reconnaissance d'un droit de ne pas être dérangé en ligne. Les propositions formulées par la CNCDH ont également à cœur de favoriser l'autonomisation des utilisateurs et utilisatrices des services numériques. Celle-ci repose sur une série d'actions aux niveaux individuel et collectif : conférer une liberté de choix en ligne, éduquer aux enjeux du numérique et encourager l'action de la société civile.

Enfin, la CNCDH considère que la prise de conscience de la toxicité des modalités dominantes de conception des services numériques doit mener vers la pleine responsabilisation des fournisseurs de ces services. En conséquence, d'une part, le cadre règlementaire doit être renforcé, en refusant de céder à la logique de dérégulation actuellement à l'œuvre. D'autre part, au regard des conséquences particulièrement délétères entraînées par les modalités de conception toxique, le cadre légal doit être complété afin de permettre l'engagement de la responsabilité des fournisseurs de services numériques. La CNCDH propose ainsi la création d'un nouveau régime de responsabilité du fait de la conception des services numériques.

La CNCDH estime que seule une approche à la fois holistique et préventive permettra d'assurer la protection efficace des droits humains en ligne, et ainsi de répondre aux enjeux majeurs qui résultent de l'utilisation des réseaux sociaux et du déploiement massif des systèmes d'intelligence artificielle, pour le respect des droits fondamentaux, ainsi que pour la préservation de l'État de droit et de la démocratie.

Table des matières

Introduction

1. Les services numériques occupent une place incontournable dans nos vies. Ces services désignent l'ensemble des solutions et des plateformes qui offrent différentes fonctionnalités et proposent des contenus à leurs utilisateurs, au moyen d'une mise à disposition via internet ou un réseau numérique. Parmi ces services, les réseaux sociaux en particulier se sont imposés comme des espaces majeurs de notre vie sociale. En 2025, près d'un internaute sur deux les consulte chaque jour[1]. Leur rôle dépasse désormais la simple mise en relation pour recouvrir l'information et le divertissement, mais aussi la diffusion des discours publics et la participation citoyenne. Les réseaux sociaux offrent ainsi des espaces « *de socialisation, de diffusion d'idées, de partage de connaissances et de soutien* »[2]. Plus récemment, les outils dits d'« intelligence artificielle » (IA), et en particulier les services reposant sur des grands modèles de langage (*large language models*, LLM), ont connu un essor fulgurant. Capables de produire du texte, de l'image ou du son à partir de requêtes formulées en langage naturel, ces IA génératives s'inscrivent dans un mouvement plus large d'automatisation de la communication et de la production de contenus. Leur diffusion rapide et massive, tant dans la sphère privée que professionnelle, leur utilisation intensive à tous les niveaux et en particulier leur intégration aux réseaux sociaux, le renouvellement technologique très rapide (quelque mois) et le remplacement incessant de modèles par des modèles plus performants, bouleversent nos habitudes à peine acquises et prennent de vitesse l'utilisateur et le régulateur.

2. Or, l'ensemble de ces plateformes, et en particulier les réseaux sociaux et systèmes d'IA, ne sont pas des environnements neutres : leur architecture technique, leurs choix de conception et le modèle économique sous-jacent participent activement à la manière dont les interactions se produisent et dont les contenus circulent. L'organisation des interfaces, les systèmes de recommandation et les algorithmes de classement captent l'attention et hiérarchisent les discours, au risque d'entraîner polarisation du débat, diffusion accélérée de contenus haineux et/ou violents, désinformation, ou encore atteintes à la dignité des personnes.

3. En conséquence, les plateformes numériques et la façon dont elles sont conçues sont susceptibles d'emporter des conséquences délétères pour la santé et le bien-être de leurs utilisateurs. Parmi les risques identifiés figurent celui de développer un comportement addictif, des troubles du comportement alimentaire, des pensées dépressives et/ou suicidaires ou encore d'être confronté à des contenus violents, haineux, pornographiques et/ou illicites. Une étude publiée en octobre 2025 révèle qu'en France, un usage excessif des réseaux sociaux conduirait à 600 000 cas de dépression supplémentaires, et 800 décès par suicide pour les jeunes nés entre 1990 et 2021[3]. Bien que chacun soit potentiellement concerné, les inquiétudes se sont essentiellement concentrées sur les enfants et les adolescents. En effet, ce public est particulièrement vulnérable en raison non seulement de sa présence massive sur les réseaux sociaux et du temps quotidien consacré à leur consultation, mais également des répercussions que cette activité entraîne sur son développement cognitif. En effet, ainsi que l'indique l'Agence nationale de sécurité sanitaire (Anses), « *l'adolescence constitue une période particulière de vulnérabilité aux stratégies de captation de l'attention, en raison de*

[1] Centre de Recherche pour l'Étude et l'Observation des Conditions de Vie (Crédoc), Baromètre du numérique 2026, février 2026.

[2] Anses, *Avis relatif aux « effets de l'usage des réseaux sociaux numériques sur la santé des adolescents »*, saisine n° « 2019-SA-0152 », décembre 2025, p. 17.

3 N. Hoertel *et al.*, « *Impact of excessive social media use on adolescent depression and its consequences in France: An individual-based microsimulation model* », *PLOS Medicine*, 2025.

capacités encore limitées de régulation émotionnelle et comportementale par rapport à l'âge adulte »[4]. Or, selon une enquête publiée par l'Arcom en 2025, 99 % des enfants de onze à dix-sept ans utilisent au moins une plateforme en ligne, dont 83 % se connectent quotidiennement à une très grande plateforme[5]. Parmi ces jeunes, 88 % se considèrent exposés à des risques graves en ligne.

4. Comme les développements au sein du présent *Avis* permettront de le souligner, ces technologies entraînent des effets différenciés en fonction d'un certain nombre de facteurs. Le Comité des ministres du Conseil de l'Europe identifie parmi les personnes les plus vulnérables « *les femmes et les filles, les enfants, les personnes en situation de vulnérabilité et exposés à la discrimination, notamment les personnes handicapées, les minorités ethniques, linguistiques et religieuses nationales, les communautés LGBTI, ainsi que les migrants et les personnes issues de l'immigration. Toute personne perçue comme appartenant à ces groupes peut être confrontée à des abus ciblés, y compris de nature intersectionnelle, à une discrimination structurelle ou à une exclusion algorithmique, ce qui limite leur capacité à exercer leurs droits en ligne* »[6]. En premier lieu, les filles et les femmes sont exposées de manière systémique aux cyberviolences. En France, 82 % des victimes de violence en ligne sont des femmes[7]. De même, les personnes appartenant à la communauté LGBTQI+ (qui sont 85 % à avoir subi des cyberviolences) et les personnes racisées (71 %) sont également particulièrement visées[8]. Les personnes défavorisées sur le plan socio-économique figurent également parmi les plus vulnérables, dans la mesure où les personnes à bas revenus indiquent faire plus fréquemment face à des contenus violents, injurieux, ou encore liés aux troubles alimentaires[9]. En effet, la surexposition aux écrans est deux fois plus fréquente au sein des foyers les moins favorisés, ce qui représente un facteur supplémentaire d'accroissement des inégalités[10]. Enfin, les risques sont multipliés lorsqu'une ou plusieurs de ces caractéristiques se croisent[11].

5. Les outils d'IA amplifient encore ces menaces, en rendant possible la génération automatisée et à grande échelle de contenus parfois difficiles à distinguer de la production

[4] Anses, 2025, *op. cit.*, p. 9.

[5] Arcom, « Protection des mineurs en ligne : quels risques ? Quelles protections ? », 25 septembre 2025. Les « *très grandes plateformes* » (ou VLOP pour *Very Large Online Platforms*) correspondent aux services visés par le Règlement sur les services numériques. L'étude concerne YouTube, Snapchat, TikTok, Instagram, Pinterest, X.

[6] Comité des ministres du Conseil de l'Europe, « Exposé des motifs relatif à la Recommandation CM/Rec(2026)4 du Comité des Ministres aux États membres sur la sécurité et l'autonomisation en ligne des utilisateurs et des créateurs de contenu », 8 avril 2026.

[7] Féministes contre le cyberharcèlement, Point de contact et Stop Fisha, « Grande enquêtes sur les cyberviolences sexistes et sexuelles. Synthèse », juin 2026. Voir, pour une étude spécifique aux 18-24 ans : A. Hadj Larbi (SSMSI), M. Rakotobe et B. Traore (Depp) pour l'Insee, Cyberviolences dans les établissements scolaires et dans la société, 14 octobre 2025 : « *Les femmes majeures déclarent davantage être victimes de cyberviolence : 3,8 % contre 2,6 % des hommes. Au collège, 31 % des filles déclarent des cyberviolences, contre 26 % des garçons* ».

[8] Féministes contre le cyberharcèlement avec Ipsos, « Cyberviolence et cyberharcèlement : le vécu des victimes », 2022. Voir également Comité des ministres du Conseil de l'Europe, 2026, *op. cit.*

[9] Crédoc, 2025, *op. cit.*, p. 23.

[10] L. Poncet *et al.*, « *Sociodemographic and behavioural factors of adherence to the no-screen guideline for toddlers among parents from the French nationwide Elfe birth cohort* », *International Journal of Behavioural Nutrition and Physical Activity*, août 2022, cité par C. Bousquet-Bénard et A. Pascal, (mission « Enfants et écrans »), « À la recherche du temps perdu », 30 avril 2024.

[11] Féministes contre le cyberharcèlement, Point de Contact et Stop Fisha, 2026, *op. cit*. : « *79 % des victimes [de cyberviolences] déclarent être exposées à au moins une forme de discrimination et plus de la moitié en cumulent deux ou plus*. »

humaine (on parle alors d'« *hypertrucage* », ou plus communément de *deepfake* en anglais). Dans le cadre de son avis sur la protection de l'intimité des jeunes en ligne, la CNCDH a dénoncé les risques entraînés par ces technologies, en particulier lorsqu'elles sont utilisées pour générer des images à caractère sexuel[12]. La combinaison de ces outils avec les réseaux sociaux entraîne le risque de créer un environnement informationnel délétère : des messages produits par IA, personnalisés pour toucher des publics spécifiques, peuvent être produits et diffusés massivement, et bénéficier des mécanismes de recommandation propres aux plateformes. Dans ce contexte, la frontière entre communication et manipulation devient poreuse, fragilisant les conditions d'un débat démocratique éclairé et ouvrant la voie à de nouvelles formes d'ingérence. Les réseaux sociaux développent par ailleurs désormais des IA incluses directement dans leur plateforme et accessibles par les utilisateurs sans que ceux-ci l'aient sollicité, à l'instar de MetaAI (Facebook, Instagram, WhatsApp) ou de Grok (X). À cela s'ajoute le développement des « compagnons IA », des intelligences artificielles conçues pour développer des interactions sociales et émotionnelles avec un utilisateur, qui augmentent encore les risques liés à un anthropomorphisme excessif déjà soulignés par la CNCDH au sein de son avis sur l'impact de l'intelligence artificielle sur les droits fondamentaux. Ainsi, les risques engendrés par l'utilisation des services numériques se mesurent à la fois aux niveaux individuels, interpersonnels et systémiques – c'est-à-dire qu'ils peuvent entraîner, par des réactions en chaîne, des effets négatifs considérables sur l'ensemble de la société.

6. La discussion s'est longtemps focalisée sur les usages et les comportements des utilisateurs, ainsi que sur la régulation *a posteriori* des contenus, notamment au travers des possibilités de les « signaler » auprès d'équipes de modération. Cependant, depuis plusieurs années, des voix s'élèvent en faveur d'un déplacement du regard[13] : au lieu de se limiter à sanctionner ou modérer ce qui circule, en se concentrant sur le contenu, il s'agirait d'interroger les conditions mêmes de production et de circulation de celui-ci, autrement dit la *conception* même (ou le *design*) des services numériques. En effet, « *nombreux sont désormais les rapports dénonçant leurs possibles effets délétères résultant de choix directement réalisés par des fournisseurs de services dont le modèle repose sur l'économie de l'attention* »[14]. Selon le Comité directeur sur les médias et la société de l'information (CDMSI) du Conseil de l'Europe, la « conception » se réfère à « *toutes les grandes décisions qui structurent le fonctionnement d'une plateforme numérique et déterminent la manière dont elle est perçue par les utilisateurs* »[15]. Cette définition englobe également « *les moyens techniques permettant aux plateformes de mettre en œuvre, maintenir et mettre à jour leurs architectures et interfaces* »[16].

[12] CNCDH *Avis sur la protection de l'intimité des jeunes en ligne*, A-2025-1, 23 janvier 2025.

[13] Voir notamment : Panoptykon Foundation, People vs Big Techs, « *Safe by Default. Moving away from engagement-based ranking towards safe, rights-respecting, and human centric recommender systems* », 7 mars 2024 ; mission « Enfants et écrans », *op. cit.* ; Assemblée nationale, *Rapport fait au nom de la Commission d'enquête sur les effets psychologiques de TikTok sur les mineurs*, clôturée le 4 septembre 2025, « Avant-propos » ; F. Forestier, M. Khamassi, S. Broadbent, C. Zolynski, *Pour une nouvelle culture de l'attention*, Odile Jacob, 2024.

[14] C. Zolynski (dir.), « Étude des dispositifs légaux relatifs à l'usage des réseaux sociaux par les mineurs. Analyse du cadre juridique existant, des enjeux et des évolutions en cours » novembre 2025, p. 140.

[15] Comité directeur sur les médias et la société de l'information du Conseil de l'Europe, « Note d'orientation sur la lutte contre la propagation de la mésinformation et de la désinformation en ligne par le biais de la vérification des faits et de la conception de plateformes dans le respect des droits de l'homme », CDMSI(2023)015, 12 décembre 2023, p. 3

[16] *Ibid.*

Cette perspective invite alors à considérer la responsabilité des plateformes en leur qualité non seulement de diffuseurs, mais aussi de concepteurs d'environnements numériques[17].

7. Face aux dérives des acteurs du numérique, depuis plusieurs années, l'Union européenne a durci la régulation des pratiques en ligne et renforcé la protection accordée aux citoyennes et citoyens européens. En 2018, le Règlement général sur la protection des données (RGPD)[18] a révisé la directive de 1995[19] et posé les bases de la protection de l'individu et de ses données personnelles face aux grandes plateformes gérées par des entreprises multinationales. En 2022, l'adoption du Règlement sur les services numériques[20] (RSN, ou *Digital Services Act*, DSA) a permis de renforcer la lutte contre les contenus illicites et la désinformation en ligne, et amélioré la transparence et la responsabilité des plateformes. Il a renforcé les obligations de modération sur les réseaux sociaux en imposant aux grandes plateformes[21] une obligation de retrait ou de blocage d'accès des contenus illicites dont elles ont connaissance (que ceux-ci soient haineux, pédocriminels, terroristes…). Le texte a également permis aux régulateurs d'avoir un droit de regard sur la façon dont fonctionnent les algorithmes, ainsi que la manière dont les décisions de retrait de contenus sont prises, et dont les publicitaires ciblent les usagers. Le règlement interdit également le profilage des mineurs à des fins publicitaires[22]. En juillet 2025, en application de l'article 28 § 4 du RSN, la Commission a publié des lignes directrices aux fins d'aider les fournisseurs de grandes plateformes accessibles aux mineurs à satisfaire à leur obligation de « *garantir un niveau élevé de protection de la vie privée, de sûreté et de sécurité des mineurs sur leur service* »[23]. Par ailleurs, en 2024, le Règlement sur l'intelligence artificielle[24] (RIA) a posé un cadre de régulation en classant les systèmes d'IA selon leur niveau de risque et en prévoyant des obligations de transparence, de sécurité et de respect des droits fondamentaux[25]. La Commission européenne prépare actuellement un projet de Règlement sur l'équité numérique (REN, ou *Digital Fairness Act*, DFA)[26]. Le texte devrait concerner les schémas déceptifs (*deceptive* ou *dark patterns*)[27], la personnalisation des contenus, ainsi que les contrats et le marketing d'influence. L'objectif est essentiellement de protéger les consommateurs contre les

[17] Voir en ce sens : CJUE, 16 juin 2026, *WebGroup Czech Republic et autres*, aff. C-188-24, et *Coyote c. France*, aff. C-190-24.

[18] Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 *relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données*.

[19] Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, *relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données*.

[20] Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 *relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE*.

[21] La majorité de ces mesures s'appliquent seulement aux plateformes qui ont plus de 45 millions d'usagers dans l'Union européenne (notamment Facebook, YouTube, X/Twitter et TikTok).

[22] Article 28 § 2 du RSN.

[23] Commission européenne, « La Commission publie des lignes directrices sur la protection des mineurs », 14 juillet 2025.

[24] Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 *établissant des règles harmonisées concernant l'intelligence artificielle*.

[25] CNCDH, A-2025-1, *op. cit.*

[26] Voir la consultation publique préalable. Voir également le site du Parlement européen qui récapitule les projets de texte.

[27] Les schémas déceptifs sont des astuces utilisées dans les sites Web et les applications qui poussent l'utilisateur à faire des choses qu'il n'avait pas l'intention de faire, comme acheter ou s'inscrire à quelque chose. Parmi les exemples, on peut citer la fausse urgence (par exemple, les faux compteurs à rebours) pour pousser l'utilisateur à agir, la publicité déguisée et la manipulation émotionnelle.

pratiques commerciales déloyales. Enfin, la Commission prépare également un paquet législatif applicable au domaine numérique [28]. Publiées le 19 novembre 2025 par la Commission européenne, les propositions d'« omnibus numériques » en matière d'intelligence artificielle et de protection des données personnelles visent à modifier les règles encadrant les activités numériques, au nom de la compétitivité des entreprises européennes. Ces propositions suscitent néanmoins des inquiétudes quant à un potentiel processus de dérégulation.

8. Sur le plan national, la législation concerne essentiellement la régulation du contenu. La loi sur la majorité numérique[29] prévoit que les utilisateurs doivent être âgés d'au moins 15 ans pour s'inscrire sur les plateformes de réseaux sociaux, sauf si leurs parents ou les titulaires de la responsabilité parentale ont donné leur consentement. La loi précise également que les plateformes de réseaux sociaux doivent utiliser des systèmes de vérification technique tels que spécifiés par l'Autorité de régulation de la communication audiovisuelle et numérique (Arcom). Les règles ont été établies pour s'appliquer aux réseaux sociaux qui exercent leur activité en France. Toutefois, en raison de son empiètement sur les compétences de l'Union européenne en la matière, la loi n'a pas été mise en application. En 2024, la loi visant à sécuriser et à réguler l'espace numérique [30] (dite « loi SREN ») a interdit l'affichage de contenus pornographiques sans que la majorité des utilisateurs n'ait été contrôlée par un processus en « double anonymat »[31]. Début 2026, une proposition de loi visant à protéger les mineurs des risques auxquels les expose l'utilisation des réseaux sociaux a été adoptée en première lecture par l'Assemblée nationale et le Sénat[32]. Elle vise notamment à interdire aux personnes mineures de moins de quinze ans l'accès à certains réseaux sociaux dès lors qu'ils représentent un danger pour eux. À l'heure de l'adoption de cet *Avis*, une commission mixte paritaire devait encore se réunir concernant ce texte.

---

[28] Proposition de Règlement du Parlement européen et du Conseil modifiant les règlements (UE) 2016/679, (UE) 2018/1724, (UE) 2018/1725 et (UE) 2023/2854 ainsi que les directives 2002/58/CE, (UE) 2022/2555 et (UE) 2022/2557 en ce qui concerne la simplification du cadre législatif numérique, et abrogeant les règlements (UE) 2018/1807, (UE) 2019/1150 et (UE) 2022/868 ainsi que la directive (UE) 2019/1024 (règlement omnibus numérique) et Proposition de règlement du Parlement européen et du Conseil modifiant les règlements (UE) 2024/1689 et (UE) 2018/1139 en ce qui concerne la simplification de la mise en œuvre des règles harmonisées concernant l'intelligence artificielle (train de mesures omnibus numérique sur l'IA).

[29] Loi n° 2023-566 du 7 juillet 2023 *visant à instaurer une majorité numérique et à lutter contre la haine en ligne*, JORF du 8 juillet 2023.

[30] Loi n° 2024-449 du 21 mai 2024 *visant à sécuriser et à réguler l'espace numérique*, JORF du 22 mai 2024.

[31] La vérification d'âge « en double anonymat » repose sur le fait que l'opérateur fournissant l'attestation de majorité ne sait pas ce pour quoi cette attestation va être utilisée, permettant de respecter le principe de l'anonymat en ligne. Un arrêté du 26 février 2025 oblige 17 sites pornographiques à contrôler l'âge de leurs utilisateurs pour empêcher les mineurs d'accéder à leurs contenus. Par sa décision adoptée en référé le 15 juillet 2025, le Conseil d'État a confirmé la conformité de cette mesure au droit de l'Union européenne (CE, 15 juillet 2025, n° 505472). La loi confie également à l'Arcom un pouvoir de blocage administratif des services de communication au public en ligne ayant une responsabilité éditoriale et des services de plateforme de partage de vidéos diffusant des contenus à caractère pornographique. Par ailleurs, la loi charge également l'Arcom d'établir des exigences techniques contraignantes (« référentiel ») pour les systèmes de vérification de l'âge auxquelles doivent se conformer les sites web qui proposent des contenus pornographiques. Le référentiel de l'Arcom a été adopté le 8 octobre 2024. Voir : Arcom, *Référentiel technique sur la vérification de l'âge*…, 8 octobre 2024.

[32] Proposition de loi visant à protéger les mineurs des risques auxquels les expose l'utilisation des réseaux sociaux, n° 2107, déposée à l'Assemblée nationale le mardi 18 novembre 2025.

9. Par ailleurs, à l'échelle régionale ou internationale, plusieurs institutions plaident pour l'adoption d'une perspective centrée sur les droits humains dans le cadre de la régulation des services numériques, en particulier le Conseil de l'Europe. C'est notamment le sens de la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'Homme, la démocratie et l'État de droit, adoptée le 5 septembre 2024[33]. Au jour de l'adoption du présent *Avis*, la Convention n'était pas encore entrée en vigueur.

10. Dans la continuité de ses travaux sur la haine en ligne[34], sur l'intelligence artificielle[35] et sur la protection de l'intimité des jeunes en ligne[36], la CNCDH s'est auto-saisie de la thématique de la conception des services numériques. Ce sujet soulève des enjeux de proportionnalité entre le respect de la liberté d'expression, de la liberté d'information, de la vie privée ou encore de la liberté d'entreprendre. Il pose des questions juridiques, éthiques et politiques majeures concernant la place des droits fondamentaux dans la conception des architectures numériques. La CNCDH souhaite adopter une approche systémique du sujet, c'est-à-dire qui ne se concentre pas uniquement sur les comportements et les risques au niveau individuel. En effet, d'une part, les risques systémiques et sociétaux, tels que la mésinformation et la désinformation, la haine en ligne, le cyberharcèlement de masse ou encore la radicalisation des contenus doivent être pris en compte. D'autre part, plutôt que de rejeter la responsabilité des comportements en ligne sur les utilisateurs et utilisatrices ou sur leurs parents, il convient d'interroger la responsabilité des fournisseurs de services numériques qui conçoivent les plateformes sur lesquelles les contenus sont émis et partagés. Enfin, au regard des risques posés par les services numériques, la CNCDH se positionne en faveur de l'autonomisation et la mise en pouvoir d'agir des utilisateurs et utilisatrices en général, et ce sans se limiter à la protection des personnes mineures.

11. Dans le présent *Avis*, la CNCDH s'attachera, premièrement, à dresser un panorama des atteintes aux droits humains causées par le modèle dominant de conception des services numériques (I). Elle proposera ensuite des modalités d'action afin d'intégrer la protection des droits humains dès la conception des services (II). La liste de l'ensemble des recommandations adressées aux autorités européennes et nationales sera enfin abordée en troisième partie (III).

[33] Conseil de l'Europe, Convention-cadre sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit, ouverte à la signature le 5 septembre 2024. Voir également le Rapport explicatif accompagnant la Convention.

[34] CNCDH, *Avis sur la lutte contre la haine en ligne*, 8 juillet 2021, A-2021-9.

[35] CNCDH, *Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux*, A-2022-6, 7 avril 2022.

[36] CNCDH, A-2025-1, *op. cit.*

I. Un modèle dominant de conception des services numériques attentatoire aux droits humains

12. La CNCDH prend note de la typologie des risques en ligne adoptée par la Commission européenne en matière de protection des personnes mineures sur internet, dite typologie des « 5 C »[37]. Celle-ci repose sur le constat que les enfants sont exposés à différentes catégories de risques :

- **risques d'exposition à certains contenus** (*content*) : confrontation à des contenus haineux ou illicites ; désinformation… ;
- **risques liés au comportement** (*conduct*) : envoi ou publication de contenus haineux, violents ou pornographiques, participation à des défis dangereux… ;
- **risques liés au contact** : confrontation à des adultes malveillants, exposition à des actes de pédocriminalité… ;
- **risques pour les consommateurs** (*consumers*) : risques liés au marketing ; risques de profilage commercial ; risques financiers (arnaques, fraudes, dépense de sommes importantes) ; risques liés à la sécurité (achat et consommation de drogues, de médicaments, d'alcool…) ;
- **risques transversaux** (*cross-cutting*) : risques liés aux technologies avancées (agents conversationnels fondés sur l'IA susceptibles de fournir des informations préjudiciables…) ; risques pour la santé et le bien-être mental, émotionnel ou physique (dépendance, dépression, troubles anxieux, perturbations des rythmes du sommeil, isolement social…) ; risques en matière de protection de la vie privée et des données (accès à des informations sur les personnes mineures, fonctions de géolocalisation…).

13. La CNCDH considère que la pertinence de cette typologie ne se limite pas aux risques rencontrés par les personnes mineures, mais s'étend à ceux auxquels tous les utilisateurs et utilisatrices en ligne sont confrontés. Ainsi, la suite des développements s'appuiera en partie sur les risques identifiés au sein de cette typologie. Seront toutefois distingués les risques comportementaux fondés sur l'exploitation des vulnérabilités (A), les risques individuels liés aux contenus dangereux auxquels les utilisateurs et utilisatrices se trouvent exposés en ligne (B) et enfin les risques systémiques entraînés par la conception des services numériques (C). Chacun de ces risques est entendu comme découlant de la conception même des services.

[37] Commission européenne, « Lignes directrices concernant des mesures visant à garantir un niveau élevé de protection de la vie privée, de sûreté et de sécurité des mineurs en ligne », conformément à l'article 28, paragraphe 4, du règlement (UE) 2022/2065, Bruxelles, 7 octobre 2025, C(2025) 6826 final. Cette typologie des risques repose sur les travaux de l'OCDE et du milieu académique. Voir notamment : OCDE, *Children in the digital environment – Revised typology of risks*, 2021 ; S. Livingstone et M. Stoilova, *The 4Cs: Classifying Online Risk, Short Report Series on Key Topics*, 2021.

A. Une conception fondée sur l'exploitation de nos vulnérabilités

14. Les modèles économiques de la plupart des services numériques reposent sur ce qu'il est désormais convenu d'appeler « *l'économie de l'attention* »[38]. En effet, si les services numériques paraissent la plupart du temps « gratuits », leur rentabilité dépend le plus souvent de la collecte et de l'exploitation des données des utilisatrices et utilisateurs[39]. Ainsi, afin de maximiser leur profit, certains médias sociaux et fournisseurs de systèmes d'IA ont intérêt à prolonger la connexion le plus longtemps possible, dans l'optique de collecter un maximum de données et notamment de les utiliser à des fins de profilage, de vendre des espaces publicitaires personnalisés, ou encore d'inciter l'utilisateur à s'abonner à une version payante du service. Les travaux de recherche menés sur le sujet, ainsi que les auditions conduites par la CNCDH, l'amènent à constater que la conception des services numériques est fondée sur l'exploitation des vulnérabilités humaines. Cette exploitation repose sur une conception addictive (1), et s'appuie sur des mécanismes d'hyperpersonnalisation (2) ainsi que des technologies anthropomorphiques (3).

1. La conception addictive des services numériques

15. Afin de maximiser l'engagement des utilisateurs et utilisatrices et de maintenir leur connexion, plusieurs services en ligne s'appuient sur un certain nombre de paramètres qui visent à susciter la connexion dans la durée et à « maximiser l'engagement » de l'utilisateur. Ces paramètres reposent sur différents ressorts cognitifs, parmi lesquels :

- **la suppression des moments naturels d'arrêt** : par exemple, les notifications *push,* le défilement infini du fil d'actualité, la lecture automatique des vidéos, qui entraînent un comportement quasi-hypnotique[40] ;
- **la sollicitation d'une interaction de la part de l'utilisateur** : par exemple, le fait de « tirer » sur la page pour en rafraîchir le contenu, les notifications de « recapture » de l'intérêt une fois que l'utilisateur a quitté la plateforme ;
- **la pression sociale** et la « peur de ne pas en être » (*fear of missing out*, FOMO en anglais) : par exemple, les *likes*, les commentaires, l'affichage du nombre d'abonnés sur les réseaux sociaux, ou encore les fonctions « *vu* », la mention ou le symbole « *est en train d'écrire…* », ou les notifications artificielles (soit le fait d'afficher le symbole d'une notification alors que rien de nouveau n'est apparu sur la page) ;
- **l'usage de techniques de jeu** (*gamification*) : il s'agit de prévoir des fonctionnalités ludiques pour que l'usage s'apparente à un jeu, ce qui recouvre notamment le fait de donner des récompenses variables et aléatoires, sous forme de *likes* ou de commentaires, ou d'assigner chaque jour des objectifs à atteindre (nombre de *likes* ou

[38] Voir notamment : S. Abiteboul et J. Cattan, « L'économie de l'attention », in *Nous sommes les réseaux sociaux*, p. 45-49 ; Direction générale du Trésor, « L'économie de l'attention à l'ère du numérique », n°369, septembre 2025.

[39] Conseil national du numérique, « Votre attention, s'il-vous-plaît ! Quels leviers face à l'économie de l'attention ? », janvier 2022.

[40] Ce mécanisme peut donner lieu au phénomène de *doomscrolling*, soit le fait ou le sentiment de « *ne pas pouvoir s'empêcher de faire défiler indéfiniment des contenus multimédias anxiogènes* ». Voir : M. Dupont, « Le "*doomscrolling*", ou l'ascenseur émotionnel sans fin des réseaux sociaux », *Le Monde*, 20 avril 2022.

d'abonnés …), d'octroyer des récompenses en cas de connexion prolongée ou de connexions en « séries », ou encore d'organiser des compétitions entre utilisateurs[41] ;

- **le regroupement de fonctionnalités** (*bundling*) : il s'agit de « *fonctionnalités pratiques qui ne sont pas liées à l'objectif principal d'une application* », comme les services de messagerie intégrés dans une application de type réseau social, l'intégration d'un assistant IA au sein de l'application (comme l'assistant IA dans WhatsApp), ou encore le déploiement d'une place de marché au sein du service (comme le TikTok Shop ou le Snap Store)[42].

16. Ainsi, la « conception addictive » désigne l'ensemble des « *fonctionnalités d'interface, des systèmes de recommandation et d'autres aspects des services numériques qui sont délibérément optimisés pour encourager une utilisation compulsive ou un engagement excessif* »[43]. Il est important de garder à l'esprit que ces mécanismes ne sont pas des défauts isolés, mais des stratégies marchandes conscientes, qui visent à maximiser l'engagement et la captation de données, souvent renforcées par le profilage comportemental, qui porte notamment atteinte aux droits au respect de la vie privée et à la protection des données. La conception addictive résulte de l'effet cumulatif de ces mécanismes, qui altèrent la capacité des utilisateurs à contrôler le temps passé sur le service et leur attention en exploitant leurs vulnérabilités comportementales et émotionnelles[44].

17. Les pratiques de conception addictive ont de graves répercussions sur la jouissance des droits fondamentaux, notamment le respect de la dignité humaine, le bien-être mental et la participation démocratique, avec des conséquences particulièrement graves pour les enfants, les utilisateurs marginalisés et les personnes en situation de précarité. En ce qui concerne les personnes mineures, les conclusions de la Commission d'enquête parlementaire sur les effets psychologiques de TikTok sur les mineurs [45] dénoncent les « *effets dévastateurs* » de l'application sur la santé mentale des jeunes. Les utilisateurs ignorent souvent le danger de ces mécanismes, et ne disposent pas de véritables alternatives. Ces asymétries d'information, ainsi que la « rétention » dans laquelle les utilisateurs sont placés, portent atteinte à leur autonomie et peuvent entraîner des préjudices financiers, en particulier

[41] Voir à ce sujet : A. Delaporte et S. Vojetta, « Influence et réseaux sociaux. Face aux nouveaux défis, structurer la filière de création, outiller l'État et mieux protéger », rapport remis à Monsieur le Premier ministre et Madame la ministre déléguée chargée de l'Intelligence artificielle et du Numérique, janvier 2026. Le rapport évoque notamment la pratique des *live matches*, qui permettent à des influenceurs de s'affronter, le vainqueur étant celui ou celle qui obtiendra le plus de points ou cadeaux de sa communauté. Le rapport indique que « *Cette incitation au don est accentuée par des effets visuels et sonores semblables à ceux des casinos, rendant la pratique extrêmement addictive pour les utilisateurs, en plus du sentiment exacerbé d'appartenance à la communauté de l'influenceur renforcé par des mentions régulières du nom des donneurs (ou des meilleurs donneurs) par leur "idole"* », p. 24.

[42] Pôle d'Expertise de la Régulation Numérique (PEReN), « Éclairage n° 10… Une exploration des fonctionnalités engageantes des plateformes numériques », décembre 2025.

[43] EDRi, *DFA Background Paper*, 24 octobre 2025, p. 31.

[44] Ainsi que l'a fait remarquer le Parlement européen : « *certaines entreprises technologiques se servent de la conception et des fonctionnalités du système pour tirer profit des vulnérabilités des utilisateurs et des consommateurs, dans le but de capter leur attention et d'augmenter le temps qu'ils passent sur les plateformes numériques* ». Voir : Parlement européen, Résolution du 12 décembre 2023 sur la conception addictive des services en ligne et la protection des consommateurs sur le marché unique de l'UE (2023/2043(INI)) (C/2024/4164).

[45] Assemblée nationale, *2025, op. cit.*

dans les contextes de vulnérabilité où des déclencheurs psychologiques sont utilisés pour inciter à des dépenses répétées, par le biais de techniques telles que les *loot boxes*[46], les offres « ludifiées », ou la monétisation.

18. La Commission européenne a initié plusieurs procédures à l'encontre des plateformes. Au mois de février 2026, la Commission a publié ses conclusions préliminaires dans lesquelles elle estime que la conception addictive de TikTok est contraire à la législation sur les services numériques. Elle indique notamment que TikTok n'a pas évalué de manière adéquate la manière dont ces caractéristiques addictives pourraient nuire au bien-être physique et mental de ses utilisateurs, et précise que ces conséquences concernent en particulier les mineurs ainsi que certains majeurs vulnérables[47]. En ce qui concerne Meta, la Commission poursuit actuellement son enquête à propos des obligations d'évaluation et d'atténuation des risques découlant de la conception des interfaces en ligne de Facebook et d'Instagram, « *qui peuvent exploiter les vulnérabilités et l'inexpérience des mineurs, conduisant à un comportement addictif et renforçant les effets dits de "trou de lapin"* » (voir *infra*). La CNCDH note avec intérêt que, dans le cadre de cette enquête, la Commission européenne s'est notamment appuyée sur les enquêtes menées par les organisations de la société civile (voir *infra*).

19. Au fil de ses recherches et des auditions menées, et notamment dans le cadre de l'audition de Meta, Tiktok, Snapchat, Google et YouTube, la CNCDH a pu prendre connaissance des mesures adoptées par les plateformes pour atténuer ces risques, tels les outils de gestion du temps d'écran et les outils de contrôle parental. Toutefois, la CNCDH considère que ces mesures sont insuffisantes à permettre une conception saine des réseaux sociaux et à réduire suffisamment les risques encourus. De même, la Commission européenne a conclu à titre préliminaire que TikTok semblait ne pas mettre en œuvre de mesures raisonnables, proportionnées et efficaces pour atténuer les risques découlant de sa conception addictive. En conséquence, la Commission européenne invite TikTok à modifier la « *conception de base* » de son service, en « *désactivant des fonctionnalités addictives clés telles que le* scroll *infini* », ou « *en mettant en œuvre des "interruptions de temps d'écran" efficaces, y compris pendant la nuit* ». Enfin, la CNCDH rappelle que le financement d'actions de prévention ou de partenariat avec des associations de défense des droits humains ne saurait exonérer les plateformes de leur responsabilité directe concernant la conception de leurs services et les algorithmes utilisés.

20. Par ailleurs, il est important de noter que s'ajoutent désormais aux réseaux sociaux des fonctionnalités de « commerce social ». Le commerce social (ou *s-commerce* en anglais) correspond à une combinaison des technologies des médias sociaux et des fonctionnalités commerciales, afin de faciliter les interactions en ligne et de concentrer l'acquisition de produits et de services en un seul endroit[48], ce qui bouleverse le commerce électronique (*e-*

[46] Les *loot boxes*, ou « coffres à butin », sont des formes de récompenses que l'on peut obtenir sur les réseaux sociaux gamifiés.

[47] Commission européenne, « La Commission conclut à titre préliminaire que la conception addictive de TikTok est contraire à la législation sur les services numériques », 6 février 2026.

[48] T. P. Liang et E. Turban, « *Introduction to the Special Issue Social Commerce: A research Framework for social commerce* », *International Journal of Electronic Commerce*, 16(2), 2011, pp. 5-13.

commerce)[49]. Dans ce contexte, la CNCDH tient à alerter le public quant aux risques émergents entraînés par ces fonctionnalités, et en particulier en termes d'achats compulsifs et de risques d'achat de produits dangereux ou illicites. En effet, ces plateformes peuvent amplifier la conception addictive déjà reprochée à certaines places de marché en ligne, comme en atteste la procédure ouverte par la Commission européenne à l'encontre de Shein. La Commission reproche à la fois à l'entreprise l'insuffisance des « *systèmes mis en place pour limiter la vente de produits illicites dans l'Union européenne, y compris les contenus susceptibles de constituer du matériel pédopornographique, tels que les poupées sexuelles ressemblant à des enfants* », et « *la conception addictive du service, y compris le fait de donner aux consommateurs des points ou de récompenses pour leur engagement* »[50].

2. Les algorithmes de recommandation et l'hyperpersonnalisation du contenu

21. Les auditions menées par la CNCDH ainsi que les recherches effectuées dans le cadre de cet avis ont permis de souligner que la conception des algorithmes sur lesquels reposent les services numériques entraîne des risques pour le bien-être et la santé de leurs utilisateurs. En effet, les algorithmes des plateformes reposent majoritairement sur le principe de l'hyperpersonnalisation : l'analyse des données d'utilisation, couplée à des technologies d'intelligence artificielle, permet de « proposer » un contenu hautement personnalisé. Ainsi, chacune des actions en ligne, et la collecte de données comportementales en résultant (temps passé à regarder une vidéo, *like*, partage, commentaire…), sont utilisées pour nourrir l'algorithme et influencer le contenu qui est mis en avant. En conséquence, le contenu disponible repose davantage sur le suivi du comportement et la déduction artificielle plutôt que sur les choix effectués en connaissance de cause par l'utilisateur.

22. En raison d'une conception fondée sur l'hyperpersonnalisation, les utilisateurs et utilisatrices des réseaux sociaux sont susceptibles d'être exposés, parfois à répétition, à des contenus inadaptés, tels des contenus extrêmement violents, promouvant des comportements alimentaires dangereux ou encore encourageant à l'automutilation voire au suicide. En octobre 2025, Amnesty International France a publié les résultats d'une enquête mettant en lumière les « *risques systémiques que [*TikTok*] fait courir aux enfants et aux jeunes* »[51]. Le rapport pointe en particulier le phénomène de « *rabbit hole* » (terrier de lapin), qui correspond au fait d'être entraîné dans une spirale de contenus similaires et de plus en plus intenses. Cet effet découle notamment du fil « Pour toi » prévu par l'application, qui propose automatiquement du contenu personnalisé en fonction des goûts et intérêts de l'utilisateur, anticipés par l'algorithme. Amnesty montre notamment qu'au bout de trois à quatre heures d'utilisation, les comptes créés pour les besoins de l'enquête et ayant manifesté un intérêt pour un contenu « *triste* » se sont vu proposer des vidéos donnant « *une vision romanesque du suicide* » ou montrant des jeunes faisant part de leur intention de mettre fin à leurs jours, comprenant des informations sur les méthodes de suicide. De même, l'organisation états-unienne *Center for*

[49] N. Hajli, « *A social commerce investigation of the role of trust in a social networking site on purchase intentions* », *Journal of Business Research,* Vol. 71, 2016, pp. 133-144 ; A. Delaporte et S. Vojetta, 2026, *op. cit.*

[50] Commission européenne, « La Commission ouvre une enquête sur Shein au titre du règlement sur les services numériques », communiqué de presse, 17 février 2026.

[51] Amnesty International France, « Entraîné·e·s dans le "*rabbit hole*" : De nouvelles preuves montrent les risques de TikTok pour la santé mentale des enfants », 20 octobre 2025. Voir également l'enquête menée par des journalistes de Ouest France : E. Benech, Y. Qi et T. Launay, « Nous avons laissé un robot parcourir TikTok pendant 100 heures, voici sa descente aux enfers », *Ouest-France*, 19 avril 2026.

countering digital hate (CCDH) a souligné la façon dont l'algorithme de TikTok met en avant le contenu promouvant les troubles du comportement alimentaire, voire encourageant au suicide[52]. Par ailleurs, cette hyperpersonnalisation est un élément de marketing, qui permet de proposer des publicités supposément alignées avec les goûts de l'utilisateur ou de l'utilisatrice. Cette logique d'hyperpersonnalisation est également appliquée aux agents conversationnels fondés sur l'IA, qui enregistrent dans leur « mémoire » les informations dévoilées au fil des conservations, et ce afin d'adapter la réponse apportée au plus près des besoins anticipés de l'utilisateur ou de l'utilisatrice.

23. Plusieurs travaux ont à présent mis en évidence le fait que la conception des services numériques joue un rôle majeur dans les risques rencontrés sur ces services. Un rapport publié en 2024 par la Panoptykon Foundation relève notamment que la présentation du contenu basé sur l'engagement « *amplifie de manière disproportionnée les contenus de mauvaise qualité, trompeurs ou sensationnels qui suscitent une forte réaction émotionnelle chez les utilisateurs plutôt que de viser à leur apporter une valeur réelle* »[53]. Une autre étude montre que, sur Facebook, le contenu « *borderline* », c'est-à-dire celui qui est le plus proche de la violation des conditions générales d'utilisation, obtient un engagement plus élevé, et donc une plus grande amplification par les systèmes de recommandation[54]. Afin de remporter la compétition concurrentielle à laquelle les services numériques se livrent, certains sont tentés d'accentuer ce phénomène, afin que le contenu le plus sensationnel soit mis en avant[55].

24. La CNCDH a pris bonne note des possibilités fournies par les plateformes de configurer l'algorithme de leur application de façon à ne pas subir cette hyperpersonnalisation, en choisissant par exemple de ne pas recevoir de recommandations (sur YouTube notamment), de visualiser uniquement les publications issues des comptes suivis (sur Instagram) ou encore de recevoir un fil d'actualité chronologique. Toutefois, elle considère que ces fonctionnalités sont trop peu accessibles, ou conduisent, par le *design* adopté, à décourager l'utilisateur de faire un choix alternatif. En outre, elle remarque que le *design* même de la plateforme peut aboutir à la dissimulation de ces fonctionnalités. C'est en effet le sens de la décision rendue par un tribunal d'Amsterdam, dans une affaire opposant l'association Bits of Freedom à Meta. La juridiction a conclu qu'en recourant à des stratagèmes pour dissimuler le fil d'actualité non personnalisé derrière un logo, et en imposant la présentation systématique du fil d'actualité personnalisé par Meta en dépit du choix contraire de l'utilisateur, la plateforme n'avait pas respecté le RSN. Elle a affirmé qu'une « *option de choix non persistante dans un système de recommandation va à l'encontre de l'objectif du RSN, qui est de garantir aux utilisateurs une véritable autonomie, la liberté de choix et le contrôle sur la manière dont l'information leur est présentée* ». Le juge a également conclu que la conception même des plateformes par Meta

[52] Center for Countering Digital Health (CCDH), « *Deadly by Design. TikTok pushes harmful content promoting eating disorders and self-harm into young users' feeds* », 15 décembre 2022 ; CCDH, « *YouTube's Anorexia Algorithm. How YouTube recommends eating disorder videos to young girls in the EU* », 3 février 2025.

[53] Panoptykon Foundation, People vs BigTechs, « *Safe by Default. Moving away from engagement-based ranking towards safe, rights-respecting, and human centric recommender systems* », mars 2024.

[54] People vs BigTechs, « *Fixing Recommender Systems: From Identification of Risk Factors to Meaningful Transparency and Mitigation* », 23 août 2023.

[55] Voir BBC Two, « *Inside the Rage* Machine », 25 mars 2026. Ce reportage montre comment TikTok a sciemment manipulé son algorithme afin de favoriser la diffusion du contenu « *borderline* ».

constituait « *une atteinte significative à l'autonomie des utilisateurs de Facebook et d'Instagram* »[56].

3. L'anthropomorphisation des systèmes d'intelligence artificielle

25. Les systèmes d'intelligence artificielle (SIA) sont des technologies émergentes qui sont encore peu régulées. L'article 3 du règlement sur l'IA (RIA) définit un SIA comme « *un système automatisé conçu pour fonctionner à différents niveaux d'autonomie, qui peut faire preuve d'une capacité d'adaptation […] et qui […] déduit, à partir des données d'entrée qu'il reçoit, la manière de générer des résultats tels que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels* ».

26. Les risques nouveaux présentés par ces technologies sont accrus par leurs caractéristiques anthropomorphiques, c'est-à-dire l'ensemble des techniques permettant de faire apparaître l'agent conversationnel comme « plus humain ». Il peut s'agir de lui donner un prénom et de lui faire utiliser la première personne pour s'exprimer, de lui faire employer un langage informel ou humoristique, de ralentir le rythme de ses réponses pour imiter une réflexion humaine, de concevoir une voix tremblante ou hésitante pour imiter celle d'une personne réelle, ou encore d'adopter des caractéristiques visuelles ressemblant à celles des êtres humains… Dans le cadre d'un accompagnement en santé mentale, la conception anthropomorphique de ces *chatbots* permet d'accroître la satisfaction des utilisateurs et leur intention de réutiliser le service[57].

27. Or, les compagnons IA poussent cette anthropomorphisation à l'extrême, en cumulant à ces caractéristiques un haut degré de personnalisation de la « relation » à l'utilisateur, et en adaptant leurs réponses aux informations connues de la personne qui utilise le service. Par exemple, en se fondant sur le texte rédigé par l'utilisateur, l'expression de l'outil peut s'adapter pour refléter une personnalité similaire à la sienne, ce qui aura pour effet d'augmenter son engagement[58]. Par ailleurs, ces outils sont paramétrés pour faire preuve de « flagornerie » (*sycophancy*), c'est-à-dire pour valider, donner raison, voire flatter l'utilisateur de sorte qu'ils soutiennent les propos et jugements que défend ce dernier, quelle que soit leur nature, afin de maintenir la « conversation », ou des interactions entre l'utilisateur et le système.

B. Une conception génératrice de risques individuels

28. Au-delà du caractère addictif qui en découle, la conception des services numériques expose leur public à un certain nombre de risques. La CNCDH souligne que, parmi les utilisateurs et utilisatrices de ces services, et quand bien même la conception des services numériques repose sur l'exploitation de l'ensemble des vulnérabilités humaines, certains groupes sont *particulièrement* vulnérables. S'agissant des personnes mineures, la CNCDH a pris connaissance des résultats de l'enquête menée par l'Arcom, publiés au mois de

[56] Tribunal d'Amsterdam (Chambre de droit privé, Cour civile de première instance), *Bits of Freedom v Meta*, 2 octobre 2025. Le texte original du jugement peut être consulté en ligne. Une traduction en anglais est également disponible.

[57] G. Park, S. Lee, J. Chung, « *Do anthropomorphic chatbots increase counseling satisfaction and reuse intention? The moderated mediation of social rapport and social anxiety* », *Cyberpsychology, Behavior and Social Networking*, Vol. 26(5), 2023, pp.357-365.

[58] M. Shumanov et L. Johnson, « *Making conversations with chatbots more personalized* », *Computers in Human Behavior*, Vol. 117, 2021.

septembre 2025[59]. L'étude permet notamment de constater que 99 % des mineurs de 11 à 17 ans utilisent au moins une plateforme en ligne, parmi lesquels 83 % se connectent quotidiennement à une grande plateforme[60]. L'utilisation des réseaux sociaux est de plus en plus précoce : 22 % des enfants de 11 ans indiquent avoir utilisé pour la première fois les réseaux sociaux avant leur dixième anniversaire, contre seulement 4 % des jeunes de 17 ans. Les mineurs eux-mêmes identifient six risques majeurs : l'hyperconnexion, l'exposition aux contenus choquants, les défis dangereux, le cyberharcèlement, les adultes mal intentionnés et les arnaques. Par ailleurs, l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (Anses) a publié un rapport sur les risques sanitaires liés à l'usage des réseaux sociaux par les mineurs[61]. L'étude juridique qui l'accompagne identifie elle aussi plusieurs sources de risques : le cyberharcèlement, la diffusion non consentie d'images intimes, la sextorsion, les *deepfakes* et l'atteinte à l'intimité, l'exposition à des contenus violents et à caractère pornographique, l'incitation à des conduites à risque, et enfin la modification de la perception de soi[62].

29. Dans le cadre du panorama des risques présenté au sein de cet avis, la CNCDH étudiera en particulier : les risques de contacts (1) et de contenus (2) dangereux, ainsi que les risques particuliers posés par les systèmes d'intelligence artificielle (3).

1. Les risques de contacts dangereux

30. L'un des premiers risques induits par les services numériques concerne les « contacts » qu'une personne peut y avoir avec une autre personne mal intentionnée. Cela intervient typiquement lorsqu'une personne mineure entre en contact avec une personne majeure dangereuse (cyberharcèlement, sollicitations sexuelles, prédation pédocriminelle, *grooming*[63], sextorsion[64]…)[65]. Plusieurs travaux ont désormais mis en évidence le rôle que la conception des services peut jouer dans la génération de risques liés au contact. Par exemple, ainsi que le souligne l'étude juridique qui accompagne l'avis de l'Anses de novembre 2025[66], lorsque les comptes utilisateur sont paramétrés en « mode public » par défaut et qu'ils sont recommandés à des personnes se trouvant hors de la sphère de contacts, ceci accroît les risques que des personnes mal intentionnées puissent cibler ou contacter des personnes mineures à des fins de chantage sexuel[67]. Le même constat ressort du verdict d'un tribunal civil de Santa Fe, au Nouveau-Mexique, dans une affaire visant Meta. Le 24 mars 2026, un

[59] Arcom, 2025, *op. cit.*

[60] Le terme utilisé est « VLOP », pour *Very large online platform* (très grande plateforme en ligne). Ce terme recouvre YouTube, Snapchat, TikTok, Instagram, Pinterest et X (anciennement Twitter).

[61] Anses, 2025*, op. cit.*

[62] C. Zolynski, 2025, *op. cit.*

[63] Le *grooming* correspond à la sollicitation en ligne d'enfants à des fins sexuelles. Voir : Agence des droits fondamentaux de l'Union européenne (FRA), *Manuel de droit européen en matière de droits de l'enfant*, 2015, p. 119.

[64] La sextorsion correspond au fait de soutirer de l'argent à une victime en effectuant du chantage à caractère sexuel. Voir : Cybermalveillance.gouv, « Que faire en cas de sextorsion ? », 2024.

[65] B. O'Neill, « *The influence of social media on the development of children and young people* », étude commandée par le comité CULT du Parlement européen, 2023.

[66] C. Zolynski, *2025, op. cit.*, p. 139.

[67] Knight-Georgetown Institute & Panoptykon Foundation, *European Board for Digital Services and European Commission Report on Systemic Risks and Mitigations under the Digital Services Act*, 7 avril 2025, p. 3.

jury a condamné l'entreprise pour manquements graves à la protection des enfants s'agissant des risques d'abus sexuels, de racolage en ligne et de traite des êtres humains sur Instagram et Facebook[68]. Le jury a en effet considéré que l'entreprise avait conscience des risques générés par son algorithme. Le procureur a notamment dénoncé les choix opérés par Meta qui, privilégiant la maximisation de l'engagement et des revenus publicitaires, avait menti aux utilisateurs et à leurs parents sur la protection offerte aux personnes mineures[69].

31. Plusieurs procédures sont par ailleurs en cours, aux niveaux national et européen, à l'encontre des grandes plateformes, en matière de surexposition à des contacts dangereux. La Commission européenne a ainsi conclu de façon préliminaire à la violation du RSN par Meta[70]. Cette décision repose sur le constat que l'entreprise n'a pas identifié, évalué et atténué avec suffisamment de diligence les risques liés à l'accès à ses services par des mineurs de moins de treize ans. La Commission enjoint ainsi Facebook et Instagram à modifier leurs méthodes d'évaluation des risques et à renforcer leurs mesures de prévention, de détection et de retrait des mineurs de moins de treize ans de leur service. Par ailleurs, la Commission considère que Meta doit « *lutter efficacement contre les risques que les mineurs de moins de treize ans pourraient rencontrer sur les plateformes et les atténuer* », ce qui correspond à l'exigence de « *garantir un niveau élevé de confidentialité, de sûreté et de sécurité pour les mineurs* »[71]. La Commission européenne a également ouvert une procédure formelle à l'encontre de Snapchat[72]. Cette procédure est fondée sur des accusations d'exposition des personnes mineures à des tentatives de pédopiégeage et de recrutement à des fins criminelles, ainsi qu'à des informations sur la vente de produits illégaux (drogues, vapotage ou alcool). À ce propos, les risques concernant les mineurs sont également soulevés par les technologies de géolocalisation ou par l'activation automatique du son ou de l'image qui sont intégrées au sein d'appareils tels que les montres, les écouteurs ou encore les lunettes connectés. Dès lors, la CNCDH insiste sur la nécessité d'étendre l'approche « *by design* », qui consiste à protéger les droits humains dès la conception, aux objets connectés.

2. Les risques de contenus dangereux

32. Un autre risque accentué par la conception des services numériques concerne le contenu qui y est proposé. Selon le *Baromètre du numérique 2026*, « *près des deux tiers des utilisateurs de réseaux sociaux et de plateformes de partage de vidéos déclarent être exposés souvent ou de temps en temps à au moins un type de contenus inappropriés ou inadaptés, qu'ils soient mensongers, haineux ou injurieux, violents, pornographiques ou en lien avec des conduites à risque (troubles alimentaires, suicide, automutilation)* ». En particulier, les jeunes de 18 à 24 ans, les 25-39 ans, les hommes (69 %) et les personnes disposant d'un bas niveau

[68] *State of New Mexico v. Meta Platforms*, 24 mars 2026.

[69] Par ailleurs, en France, le 13 avril 2026, la Haute-Commissaire à l'Enfance a saisi l'Arcom et la Commission européenne après une enquête révélant plus de 350 annonces à caractère proxénète sur TikTok.

[70] Commission européenne, « La Commission conclut à titre préliminaire que Meta a enfreint la législation sur les services numériques pour ne pas avoir empêché des mineurs de moins de 13 ans d'utiliser Instagram et Facebook », communiqué de presse, 29 avril 2026.

[71] *Ibid*.

[72] Commission européenne, « La Commission enquête sur le respect par Snapchat des règles de protection de l'enfance en vertu de la législation sur les services numériques », communiqué de presse, 26 mars 2026.

de revenus se disent « *significativement plus exposés* »[73]. À ce propos, la CNCDH souligne les constats dressés notamment par l'Arcom concernant l'exposition des personnes mineures aux contenus pornographiques. En 2023, l'agence constatait que, chaque mois, « *2,3 millions de mineurs fréquentent des sites pornographiques* », un chiffre en croissance rapide et constante. Dès 12 ans, plus de la moitié des garçons se rend en moyenne chaque mois sur ces sites[74]. Ces constats soulèvent des interrogations quant au respect par les plateformes de leurs obligations en matière de contrôle de l'âge, ainsi que d'avertissement clair et accessible des utilisateurs et utilisatrices quant aux contenus illégaux figurant sur leurs plateformes. En effet, les enquêtes ouvertes par la Commission européenne en la matière indiquent que la conception des interfaces peut être trompeuse et manipulatrice et ne pas informer correctement les utilisateurs et utilisatrices[75].

33. Par ailleurs, la notion de « contenus » figurant sur les plateformes numériques concerne également les produits qui y sont mis en vente. De ce point de vue, la conception des services numériques peut également amplifier le risque que des produits illégaux soient diffusés. C'est en effet parce que l'entreprise n'avait pas suffisamment anticipé ni pris en compte ce risque dans son rapport d'évaluation des risques systémiques que la Commission européenne a infligé à la plateforme Temu une amende de 200 millions d'euros[76].

3. Les risques spécifiques présentés par les systèmes d'intelligence artificielle

34. Les inquiétudes au sujet des contenus dangereux sont renouvelées par les systèmes d'IA (SIA). Dans le cadre du présent *Avis*, le propos se concentrera sur les risques présentés par les intelligences artificielles génératives (IAG) et les compagnons IA. Les premières sont des modèles d'IA à usage général qui permettent la production flexible de contenus tels que du texte, de l'audio, des images ou de la vidéo[77]. Les seconds sont des « *systèmes d'IA dotés d'interfaces en langage naturel, produisant des réponses personnalisées et anthropomorphiques* »[78].

35. S'agissant en premier lieu des intelligences artificielles génératives, la CNCDH rappelle les constats formulés au sein de son *Avis sur la protection de l'intimité des jeunes en ligne*[79]. Sans garde-fous appropriés, ces technologies sont capables de générer du contenu hautement problématique, voire illégal, et de le diffuser massivement. En particulier, elle note que ces technologies sont des vecteurs de violences contre les femmes et les filles, en étant

[73] Crédoc, 2026, o*p. cit.*, p. 21.

[74] Arcom, « La fréquentation des sites "adultes" par des mineurs », mai 2023. Les résultats de l'enquête indiquent en outre qu'en moyenne, 12% de l'audience des sites adultes est réalisée par les mineurs.

[75] Voir notamment : Commission européenne, « La Commission enquête sur le respect par Snapchat des règles en matière de protection des mineurs prévues par le règlement sur les services numériques », 26 mars 2026 ; « La Commission ouvre des enquêtes au titre du règlement sur les services numériques afin de protéger les mineurs des contenus pornographiques », communiqué de presse, 27 mai 2025.

[76] Commission européenne, « DSA : la Commission inflige une amende de 200 millions d'euros à Temu », communiqué de presse, 28 mai 2026.

[77] Considérant 99 du Règlement sur l'intelligence artificielle.

[78] Observatoire de l'Intelligence artificielle de Paris 1 Panthéon-Sorbonne, « IA compagnon, une amie qui nous veut du bien ? », colloque organisé en collaboration avec le Département de recherche en droit de l'immatériel (DReDIS) de l'Institut de recherche juridique de la Sorbonne (IRJS) et la CNCDH, 19 janvier 2026.

[79] CNCDH, A-2025-1, *op. cit.*

notamment utilisées pour générer des images hyperréalistes (*deepfakes*) à caractère sexuel. Celles-ci prennent parfois la forme d'applications qui « dénudent » numériquement les femmes et les enfants, créant ainsi des images à caractère pornographique qui sont ensuite diffusées massivement sur internet. La CNCDH souligne en outre que les technologies d'IA embarquées sur les plateformes de réseaux sociaux démultiplient les risques, dans la mesure où l'image générée par l'IA est ensuite partagée sur la plateforme, dont l'algorithme promeut le contenu produit par cette technologie, entraînant des flots de *deepfakes*. Ainsi, l'introduction et le « débridage » de la fonctionnalité Grok du réseau X (anciennement Twitter), en décembre 2025, a conduit à la diffusion massive de contenus illicites, qu'il s'agisse d'images à caractère sexuel non consenties ou à caractère pédocriminel[80]. En outre, l'étude conduite par *AI Forensics* a notamment conclu que cette IA était également utilisée pour générer du contenu négationniste et nazi, des représentations de Hitler ou encore des messages de soutien à l'organisation État islamique[81]. La Commission européenne a annoncé avoir lancé une nouvelle enquête sur les systèmes de recommandation de Grok et de X au titre du RSN[82].

36. En second lieu, les compagnons IA sont essentiellement des agents conversationnels dotés d'une dimension fortement personnalisée, car ils sont « *conçus pour simuler des interactions interpersonnelles ou émotionnelles durables* »[83]. Ils peuvent être indépendants ou bien intégrés à d'autres outils, comme au sein des réseaux sociaux, de jeux vidéo ou de jouets connectés. Une étude parue en mai 2026 indique qu'en France, près de 90 % des jeunes de 11 à 25 ans utilisent une IA conversationnelle, et que près de la moitié d'entre elles et eux y livrent des informations intimes[84]. Les risques présentés par ces technologies sont multiples[85]. Ils ont notamment été référencés par la Commission européenne dans le cadre des *Lignes directrices sur l'intelligence artificielle et les personnes vulnérables*[86]. En ce qui concerne les enfants, les lignes directrices considèrent que les interactions avec ce type de système peuvent entraver leur développement social et émotionnel normal, leurs relations avec d'autres personnes humaines et leurs compétences socio-émotionnelles telles que l'empathie, la régulation émotionnelle, la compréhension sociale et l'adaptabilité. Cela pourrait contribuer à façonner leurs valeurs, leurs croyances et leurs comportements de manière potentiellement préjudiciable. Ces craintes sont corroborées par des travaux de recherche qui soulignent le risque d'isolement de l'être humain, amplifié par le fait que les modèles reposent sur

[80] Arcom, Communiqué relatif aux contenus intimes non consentis générés par Grok sur la plateforme X, 15 janvier 2026.
[81] M. Tual, « IA Grok : la moitié des images générées pendant les fêtes représentaient des personnes partiellement dénudées », 5 janvier 2026.
[82] Commission européenne, « La Commission enquête sur les systèmes de recommandation de Grok et de X au titre du règlement sur les services numériques », communiqué de presse, 26 janvier 2026.
[83] Observatoire de l'Intelligence artificielle de Paris 1 Panthéon-Sorbonne, *op. cit*.
[84] Cnil, « IA conversationnelle et santé mentale des jeunes : résultats de l'enquête européenne », 5 mai 2026.
[85] Voir notamment : K. Favro et C. Zolynski, « Quelle régulation pour les IA compagnons ? (Il est temps d'agir) », *Dalloz IP/IT* 2026, p. 72.
[86] Commission européenne, « Lignes directrices concernant des mesures visant à garantir un niveau élevé de protection de la vie privée, de sûreté et de sécurité des mineurs en ligne », conformément à l'article 28, paragraphe 4, du règlement (UE) 2022/2065, C(2025) 6826 final, 7 octobre 2025.

l'engagement des utilisateurs[87] et par la tendance des machines à la *sycophancy*[88] (voir *supra*), qui provoquerait l'érosion de la capacité de réflexion et d'empathie envers les autres[89].

C. Une conception génératrice de risques systémiques

37. Au-delà des risques individuels auxquels les services numériques donnent lieu, certains risques systémiques doivent également être envisagés. En particulier, les services numériques entraînent des risques concernant la santé publique (1), la lutte contre la haine et les discriminations (2) et menaçant le système démocratique (3).

1. Des risques sanitaires : services numériques et santé publique

38. La conception actuelle des services numériques entraîne également des risques en termes de santé publique. En effet, les personnes qui ont un usage important des médias numériques « *sont deux fois plus susceptibles d'avoir des problèmes de santé mentale, y compris des facteurs de risque de suicide et d'automutilation* »[90]. De ce point de vue, les risques sont particulièrement élevés en ce qui concerne la jeunesse. En effet, plusieurs risques sanitaires sont identifiés chez les adolescentes et les adolescents : l'altération du sommeil, la dévalorisation de soi, l'incitation aux comportements à risque, l'exposition aux cyberviolences[91]… Ainsi, aux États-Unis, Meta et YouTube ont été condamnées en raison de la conception addictive de leurs services, au regard des conséquences que celle-ci avait entraînées sur la santé mentale d'une requérante[92]. Cette dernière défendait l'idée selon laquelle les services en question étaient aussi addictifs que des pratiques telles que la cigarette et les casinos, et devraient en conséquence faire l'objet d'avertissements et de réglementation aussi restrictives.

39. Par ailleurs, en raison des risques auxquels elles font face et des violences qu'elles subissent, les filles sont plus exposées que les garçons à l'altération de leur bien-être et de leur santé mentale[93]. Elles sont en effet « *plus de deux fois plus susceptibles que les garçons de présenter des niveaux cliniquement significatifs de symptômes dépressif* »[94]. En outre, « *les adolescents LGBTQIA+, les adolescents souffrant de certains troubles psychiatriques,*

---

[87] La plupart des compagnons IA posent systématiquement une question de « relance » après avoir répondu à une requête, invitant à poursuivre la conversation.

[88] Dans le domaine de l'IA générative, le terme de *sycophancy* désigne la tendance des systèmes d'IA à générer des réponses qui s'alignent sur les préférences ou les préjugés perçus de l'utilisateur, en se montrant souvent excessivement conciliants ou en fournissant des commentaires trop positifs. L'IA peut ainsi privilégier la génération de réponses agréables au détriment de la fourniture d'informations précises ou critiques, ce qui peut conduire à des résultats trompeurs ou déséquilibrés. Voir : Conseil des barreaux européens, *Guide on the use of generative AI for lawyers*, 2 octobre 2025.

[89] W. Hartzog et J. Silbey, « *How AI destroys institutions* », Research Paper n° 5870623, *UC Law Journal*, vol. 77, 2026, à paraître.

[90] Parlement européen, *op. cit.,* p.3

[91] Anses, 2025, o*p. cit.*

[92] Los Angeles County Superior Court, *P. F., et al. (K.G.M.) v. Meta Platforms, Inc., et al.,* 25 mars 2026.

[93] Anses, 2025, o*p. cit.*, p. 9. L'Anses explique ce phénomène par le fait que les filles utilisent plus les réseaux sociaux que les garçons, qu'elles utilisent davantage des réseaux sociaux visuels, qu'elles subissent davantage de pressions liées aux stéréotypes de genre, qu'elles subissent davantage de cyberharcèlement et qu'elles y accordent plus d'importance, en faisant preuve d'un engagement émotionnel plus marqué.

[94] Parlement européen, *op. cit.,* p. 4.

notamment dépressifs ou de troubles neurodéveloppementaux (…) sont particulièrement impactés »[95].

40. En outre, l'usage des réseaux sociaux peut entraîner des conséquences importantes sur le sommeil, et, par ricochet, sur la santé physique et mentale. En effet, selon l'Anses, les contenus consultés et les interactions en ligne stimulent l'éveil physiologique, cognitif et émotionnel, entravant ainsi le processus d'endormissement et de maintien du sommeil. Son rapport indique en particulier, que, chez l'adolescent, un sommeil insuffisant « *entraîne somnolence diurne, irritabilité et tristesse* »[96]. L'agence rappelle qu'une réduction chronique du temps de sommeil augmente le risque d'apparition de maladies chroniques, et nuisent particulièrement à la santé mentale. De même, la cybervictimation, qui est plus forte chez certains publics (comme les femmes et les filles ainsi que les personnes minorisées), est associée à une augmentation des symptômes dépressifs[97].

41. Certains risques émanent tout particulièrement des technologies les plus récentes. Les résultats d'une enquête parue en janvier 2026 indiquent qu'au sein des quatre pays étudiés (Allemagne, France, Irlande, Suède), près de la moitié des jeunes déclare utiliser l'IA pour parler de sujets intimes ou personnels. En France, 64 % des jeunes considèrent l'IA « *comme un conseiller de vie* »[98]. Au sein de la population étudiée, deux tiers des jeunes de 11 à 25 ans présentent des signes de troubles anxieux. De nouveau, on constate des effets genrés de ce phénomène, puisque les filles déclarent se sentir moins bien que la moyenne.

42. Les interactions entre les êtres humains et les compagnons IA pourraient conduire au phénomène dit d'« ***extraction de données par l'intimité*** »[99]. En effet, les utilisateurs étant moins conscients des informations personnelles qu'ils divulguent en raison de la nature émotionnellement engageante de l'interaction, ils seraient davantage disposés à partager certaines informations intimes. Ils pourraient ainsi être amenés à révéler des informations intimes sur eux-mêmes et leurs proches, allant des problèmes de santé mentale, à leur orientation et pratiques sexuelles. Cette caractéristique particulière de la conception des compagnons IA pose des risques spécifiques à l'heure où les jeunes, en particulier, utilisent les compagnons IA pour résoudre des problèmes interpersonnels. En effet, une étude récente montre que la conception « sycophantique » des outils d'IA réduit les intentions « prosociales » et favorise la dépendance [100] . D'autres travaux indiquent que cette caractéristique représente un danger particulier pour les utilisateurs souffrant déjà de troubles psychiatriques[101]. Aux États-Unis, le suicide de Sewell Garcia, jeune homme de quatorze ans qui avait une utilisation compulsive de Character AI, ou encore celui d'Adam Raise qui avait

[95] *Ibid*., p. 11.
[96] Anses, 2025, *op. cit.,* p. 13.
[97] *Ibid*., p. 14.
[98] Cnil, « Les jeunes européens et l'IA conversationnelle », janvier 2026.
[99] V. Bernardo, European Data Protection Supervisor, « *AI Companions* », n.d.
[100] M. Cheng *et al*., « *Sycophantic AI decreases prosocial intentions and promotes dependence* », *Science*, Vol. 391, Iss. 6792, 2026.
[101] M. Naddaf, « *AI chatbots are sycophants - researchers say it's harming science* », *Nature,* Vol. 647, 2025, pp. 13-14.

échangé avec ChatGPT d'Open AI illustrent de façon dramatique les conséquences que peuvent entraîner ces « relations » nouvelles[102].

43. Enfin, les conséquences engendrées par les médias sociaux, et par les compagnons IA en particulier, sont à placer dans le contexte du délabrement actuel de la santé mentale et de celle des jeunes en particulier, mais également de la crise de l'hôpital public. La CNCDH exprime son inquiétude quant au cercle vicieux qui semble s'installer, entre la baisse de la fréquentation des psychologues, la diminution des actions des pouvoirs publics pour permettre l'accès à des soins en santé mentale, la pénurie de professionnels de santé mentale, le recours croissant aux outils d'IA à des fins de conseils psychologiques, et la hausse des troubles de santé mentale, ici encore en particulier chez les jeunes. En effet, en conséquence d'une exposition massive aux écrans, on observe de moindres performances cognitives (langagières, socio-relationnelles et attentionnelles) et scolaires. On observe également une corrélation entre l'intensité de l'usage du numérique et les symptômes de trouble de l'attention, d'hyperactivité et d'impulsivité[103]. La CNCDH rappelle que la santé mentale, « grande cause nationale » des années 2025 et 2026, doit susciter des actions structurelles et urgentes.

44. Plus généralement, la CNCDH souligne que l'usage des services numériques entraîne d'importantes conséquences en matière environnementale, qui ont elles-mêmes des répercussions sur les plans sanitaire et social. En effet, la fabrication des appareils numériques repose sur l'exploitation de ressources naturelles rares, tandis que le fonctionnement des services numériques repose sur des technologies hautement consommatrices en énergie[104]. Ainsi que le souligne l'Anses, les réseaux sociaux numériques, « *en encourageant la connexion permanente, amplifient la consommation énergétique des infrastructures numériques et aggravent ainsi leur impact environnemental* » [105]. La conception de l'infrastructure et du logiciel d'un système numérique, en particulier l'apprentissage et la formation des SIA, soulèvent les mêmes enjeux[106].

2. La promotion d'un environnement haineux : violences de genre et discriminations

45. Ainsi que la CNCDH l'avait relevé dans son *Avis sur la protection de l'intimité des jeunes en ligne*[107], et ainsi que cela a été confirmé par les auditions menées dans le cadre du travail préparatoire à ce présent *Avis*, la majorité des actes malveillants en ligne visent les femmes et les filles[108]. En effet, les services numériques fonctionnent souvent comme un reflet,

---

[102] É. Viniacourt, « Aux États-Unis, une mère accuse une IA d'avoir poussé son fils de 14 ans au suicide », *Libération*, 24 octobre 2024.

[103] C. Schwarzer *et al.*, « *Association of media use and early childhood development: cross-sectional findings from: the LIFE Child Study* », *Pediatric Research*, Vol. 91, 2022, pp. 247–253 ; R.M. Silva Santos *et al.*, « *The association between screen time and attention in children : A Systematic Review* », *Developmental Neuropsychology*, Vol. 47, 2022, pp. 175-192 ; C.K. Ra *et al.*, « *Association of digital media use with subsequent symptoms of attention-deficit / hyperactivity disorder among adolescents* », *JAMA*, 2018, pp. 255-263. Ainsi, un tiers des adolescents utilisent des écrans (principalement des réseaux sociaux) après minuit, et plus d'un tiers des filles de 11 à 15 ans se sentent « *accro* » à certains réseaux sociaux. Voir : *U.S. Surgeon General*, *Social Media and Youth Mental Health*, 2023.

[104] Voir : Arcom, Arcep, Ademe, « Étude de l'impact environnemental des usages audiovisuels en France », 2022.

[105] Anses, 2025, o*p. cit.*, p. 17.

[106] Arcep, « Intelligence artificielle générative : quels défis environnementaux ? », mai 2026.

[107] CNCDH, A-2025-1, o*p. cit.*

[108] *Ibid.*, p. 5.

voire un miroir grossissant[109] de la société. Les violences vécues sur internet s'inscrivent alors dans le *continuum* des violences de genre que les femmes et les filles subissent toute leur vie. L'association Féministes contre le cyberharcèlement indique ainsi que « *84 % des victimes de violences en ligne sont des femmes* »[110]. En outre, plusieurs travaux montrent que les femmes et les filles partageant des caractéristiques se situant à l'intersection de divers motifs de discrimination sont surexposées au risque de subir des violences. A titre d'exemple, une récente enquête menée par le Département pour la science, l'innovation et la technologie (DSIT) du gouvernement britannique a relevé que 75 % des femmes s'identifiant comme LBTQ+ avaient déclaré avoir subi une forme de violence en ligne, contre 37 % des femmes qui ne partagent pas cette caractéristique[111].

46. Ces violences peuvent prendre différentes formes, et doivent être sans cesse réévaluées à l'aune des « progrès » technologiques. En effet, les systèmes d'IA générative renouvellent et amplifient les risques auxquels les femmes et les filles sont exposées, en permettant notamment de générer des contenus hypertruqués à caractère sexuel, et/ou en proposant des outils de « nudification » artificiel. Un rapport d'enquête publié par l'ONG AI Forensics en janvier 2026 révèle les conséquences de l'insuffisance des garde-fous encadrant le fonctionnement de Grok, l'agent IA intégré à X. Le rapport indique notamment que, du 25 décembre 2025 au 1er janvier 2026, 53 % de 20 000 images générées par la plateforme étudiées représentaient des personnes dénudées ou très peu vêtues, dont 80 % étaient des femmes[112], tandis que 2 % des images représentaient des enfants. Une mise à jour de l'enquête a indiqué qu'au 14 janvier 2026, le nombre d'images à caractère sexiste ne représentaient plus que 10 % du total[113].

47. La CNCDH considère que, loin de se limiter à des situations individuelles, les dangers présentés par ces constats relèvent du défi de société. Non seulement les expériences de violence peuvent être traumatisantes pour les femmes et les filles qui en sont victimes, mais en outre, en conséquence de la violence qu'elles y trouvent, celles-ci renoncent parfois à s'exprimer sur internet. Or, il apparaît que les services numériques ne jouent pas un rôle passif de simple mise à disposition de contenus sexistes et violents : ils « organisent leur visibilité »[114]. Alors que les recherches sur le sujet sont encore peu nombreuses, le rapport issu de l'enquête du DSIT précitée a récemment établi que la conception des services numériques constitue un « *vecteur structurel* » de violences en ligne à l'égard des femmes et des filles[115]. Il souligne que ces violences procèdent, pour une part déterminante, de choix architecturaux opérés en amont par les concepteurs de services numériques. En effet, ces choix de conception (paramètres de confidentialité, systèmes de messagerie directe, mécanismes de recommandation ou processus de signalement) structurent l'environnement

[109] Plan International, « *Free to be online?* », 2020. L'enquête révèle notamment que 50 % des femmes et des filles ont déclaré être davantage victimes de harcèlement en ligne que de harcèlement de rue.
[110] L. Salmona, Table ronde organisée au Sénat sur la thématique des cyberviolences et de la lutte contre la haine en ligne, 30 avril 2026.
[111] UK Government – DSIT, « *Platform design and the risk of online violence against women and girls* », 6 février 2025, p. 18.
[112] AI Forensics, « *Grok Generating Flood of Sexualized Images of Women and Minors* », 5 janvier 2026.
[113] AI Forensics, « *AI-Generated Image Abuse: An Update on Grok Unleashed* », 20 janvier 2026.
[114] S. Benoualid, Table ronde organisée au Sénat, *supra* note 39.
[115] UK Government – DSTI, 2025, *op. cit.*

dans lequel les violences s'exercent. Le rapport souligne notamment que, lorsque les paramètres de confidentialité sont définis par défaut sur la visibilité maximale, certaines utilisatrices, et notamment les plus jeunes d'entre elles, s'y retrouvent exposées sans l'avoir voulu, et parfois sans même en avoir pleinement conscience. Par ailleurs, l'un des défauts de conception les plus documentés réside dans le fait que les fonctionnalités de protection (le blocage de compte, le masquage de contenu, le signalement d'un utilisateur ou d'un propos…) sont quasi-exclusivement « réactifs » et reposent donc sur l'initiative de la personne ciblée. Ce mode de fonctionnement fait reposer sur les victimes la charge de leur propre sécurité, au lieu de responsabiliser les concepteurs des services. Cette logique se révèle particulièrement problématique, à deux titres : d'une part, elle est inefficace lorsque des femmes sont exposées à des volumes massifs de messages violents (ce que l'on appelle le « cyberharcèlement en meute »), rendant tout signalement individuel vain ; d'autre part, elle pourrait constituer une « double peine », en contraignant les victimes à documenter les violences subies à de nombreuses reprises et indirectement à revivre ces violences.

48. La CNCDH estime ainsi que ce modèle est incompatible avec une approche centrée sur les droits fondamentaux. Enfin, elle constate que les services numériques ne collectent peu ou pas de données ventilées par sexe concernant les violences en ligne. Ce manque prive les pouvoirs publics ainsi que les acteurs de la société civile des connaissances nécessaires et des preuves utiles à l'élaboration de politiques ciblée et efficaces.

49. Enfin, la conception des services numériques favorisant à la fois le caractère extrême des publications, en accentuant la diffusion des contenus les plus choquants (voir *supra*), ainsi que la création de « chambres d'écho », qui plongent les individus dans des espaces clos qui ne laissent pas place à la contradiction, un risque systémique particulièrement inquiétant résulte de la radicalisation de la pensée à laquelle conduit cette polarisation. En effet, les développements ci-dessus ont indiqué que certaines catégories d'utilisateurs et d'utilisatrices sont particulièrement concernées par les risques posés par les services numériques. En conséquence du sexisme structurel sur les plateformes, de nombreuses femmes ont ainsi choisi de ne plus s'exprimer en ligne, dans une forme d'autocensure inquiétante pour la pluralité des opinions et le droit à la liberté d'expression. Ont également été évoqués les radicalisations masculinistes, ainsi que les propos racistes, négationnistes et antisémites diffusés et amplifiés par les outils d'IA sur les plateformes. Ainsi que le souligne le Comité des ministres du Conseil de l'Europe, « *les femmes et les filles, les enfants et les personnes en situation de vulnérabilité ou exposées à la discrimination rencontrent dans l'environnement en ligne des risques spécifiques et accrus, dont le ciblage fondé sur l'identité et des obstacles intersectionnels, au plein exercice de leurs droits humains* »[116]. La CNCDH craint que la dynamique enclenchée favorise et accentue les discriminations et les actes de haine. De manière générale, la CNCDH rappelle que la vulnérabilité est une notion avant tout situationnelle : elle ne réside pas dans des traits individuels, mais émerge de l'interaction entre l'individu présentant certaines caractéristiques et des environnements malveillants. Ce constat appelle des actions urgentes et structurantes visant l'architecture des services numériques.

3. Des risques pour la démocratie : désinformation et menaces d'ingérence

50. Le Comité des ministres du Conseil de l'Europe définit la désinformation comme « *les informations dont on peut vérifier qu'elles sont fausses, inexactes ou de nature à induire en*

[116] Comité des ministres du Conseil de l'Europe, « Recommandation sur la sécurité et l'autonomisation en ligne des utilisateurs et des créateurs de contenu », CM/Rec(2026)4, 8 avril 2026.

erreur, créées et diffusées dans l'intention délibérée de causer un préjudice ou d'obtenir un avantage politique ou économique en trompant le public »[117]. Ainsi que le relève le Comité, la désinformation « *est devenue une préoccupation majeure dans les démocraties du monde entier et fait partie intégrante de l'ère numérique* »[118].

51. La CNCDH estime que les choix de conception des plateformes peuvent favoriser les conditions permettant l'émergence d'un débat public sain et constructif, et garantir un environnement favorable à la liberté d'expression. À l'inverse, ces choix peuvent également enfermer les utilisateurs dans des contenus peu fiables ou manifestement faux, sans possibilité de les contester. En effet, le rapport issu de la Commission d'enquête parlementaire sur les effets psychologiques de TikTok souligne notamment que le phénomène « *d'ultra-personnalisation* » expose les utilisateurs et utilisatrices au risque de se retrouver enfermés et intellectuellement isolés à cause d'une sélection algorithmique effectuée à leur insu, entraînant une réduction de la diversité des informations auxquelles ils et elles ont accès[119]. Le rapport souligne, en outre, que l'utilisateur peut se retrouver confronté à des « *chambres d'écho* » au sein desquelles certaines informations, idées ou croyances sont amplifiées et renforcées, et rarement remises en question. À ce propos, le rapport du Haut Conseil à l'Égalité 2026 consacré au masculinisme indique que la désinformation constitue « *un levier d'action majeur* » pour les mouvements masculinistes[120]. Ces derniers déploient des stratégies « *visant à essentialiser les questions de genre et à renforcer une vision strictement binaire des rôles sociaux* », afin de consolider les stéréotypes en présentant les différences de genre comme « *naturelles et immuables* ». La radicalisation masculiniste, de par sa propagation via les médias sociaux, soulève ainsi des enjeux de sécurité nationale. Dans ce contexte, la CNCDH soutient la création d'un Observatoire national du masculinisme et des radicalisations sexistes, confié au HCE, qui aurait notamment pour mission d'engager un dialogue structuré avec les plateformes numériques[121].

52. En outre, la CNCDH considère que « *la quantité croissante de désinformation générée et diffusée à l'aide d'outils de l'intelligence artificielle fait peser des menaces particulières sur le dialogue démocratique* »[122]. La CNCDH souligne le rôle essentiel que jouent les services numériques dans l'exercice du droit à la liberté d'expression et du droit à la liberté d'information. Toutefois, ainsi que l'a relevé la Commission européenne, elle note également que les systèmes de recommandation sont capables de façonner l'environnement informationnel et l'opinion publique[123]. En conséquence, elle considère qu'il est de la

[117] Cdmsi, « *Note d'orientation sur la lutte contre la propagation de la mésinformation et de la désinformation en ligne par le biais de la vérification des faits et de la conception de plateformes dans le respect des droits de l'homme* », CDMSI(2023)015, 12 décembre 2023.

[118] Cdmsi, « Projet d'exposé des motifs de la Note d'orientation sur la lutte contre la propagation de la mésinformation et de la désinformation en ligne par le biais de la vérification des faits et de la conception de plateformes dans le respect des droits de l'homme », CDMSI(2023)016, 12 décembre 2023.

[119] Assemblée nationale, 2025, *op. cit.*, p.68.

[120] Haut Conseil à l'Égalité entre les femmes et les hommes (HCE), « Rapport 2026 sur l'état du sexisme en France : la menace masculiniste », p. 73.

[121] *Ibid.*, p. 79.

[122] Comité directeur sur les médias et la société de l'information du Conseil de l'Europe, CDMSI(2023)015, 2023, o*p. cit.* p. 2.

[123] Commission européenne, Lignes directrices de la Commission à l'intention des fournisseurs de très grandes plateformes en ligne et de très grands moteurs de recherche en ligne sur l'atténuation des

responsabilité des plateformes, en particulier celles qui concentrent un nombre significatif d'utilisateurs et d'utilisatrices, de promouvoir un environnement en ligne favorable à l'exercice de ces droits. À ce titre, elle signale que les actions de vérification des faits réalisées par la communauté, disponibles sur certaines plateformes, réduisent la diffusion des publications trompeuses[124]. Elle encourage ainsi les plateformes à mettre en place des outils de ce type, qui ont l'avantage de conférer une certaine autonomie aux utilisateurs et utilisatrices. En outre, la CNCDH insiste sur la nécessité, rappelée par le Comité des ministres du Conseil de l'Europe, pour les plateformes de « *s'attacher à comprendre les enjeux de sécurité en ligne spécifiques au contexte local, y compris les risques liés au genre* »[125]. Dans ce sens, elle incite les plateformes à veiller à recruter un nombre suffisant de membres du personnel informés des contextes politique, culturel, social et linguistique des territoires sur lesquels ils et elles sont amenés à intervenir, et à entraîner leurs outils automatisés dans le même esprit.

53. Par ailleurs, les médias sociaux et les services numériques en général peuvent également être le théâtre de campagnes délibérées de désinformation. Des travaux de recherche indiquent que la conception des plateformes peut contribuer à la circulation et à l'amplification de ces fausses informations[126]. Dans le même registre, l'étude conduite par AI Forensics au sujet des « *dérives* » de la génération de contenus hypertruqués par Grok. Cela produit un brouillage des repères entre le vrai et le faux, l'authentique et le falsifié, qui affecte la qualité du débat public[127].

54. Enfin, les algorithmes peuvent également être manipulés par les plateformes elles-mêmes. Par exemple, certains propos peuvent être minimisés, en étant secrètement déclassés voire déréférencés (c'est ce que l'on appelle le *shadow ban*). Plusieurs travaux indiquent que la voix des femmes et des filles, de même que celles des personnes LGBTQI+, est réduite au silence sur certaines grandes plateformes[128]. À l'inverse, d'autres sujets peuvent être artificiellement promus afin d'influencer les utilisateurs. À ce sujet, la section de lutte contre la cybercriminalité du parquet de Paris a ouvert une enquête à l'encontre de la plateforme X, à la suite de signalements dénonçant des algorithmes biaisés dans son fonctionnement. Elle soupçonne la plateforme d'avoir modifié l'affichage de ses contenus « *à des fins d'ingérence étrangère* », en mettant en avant des thématiques favorables à l'extrême

risques systémiques pour les processus électoraux, présentés en vertu de l'article 35 § 3 du règlement (UE) 2022/2065, Section 3.2.1., d).

[124] Y. Chuai *et al*., « *Community-based fact-checking reduces the spread of misleading posts on X (formerly Twitter)* », *Nature Communications*, Vol. 17, 2026.

[125] Comité des ministres du Conseil de l'Europe, CM(Rec)(2026)4(2), *op. cit.*, § 29.

[126] P.M. Krafft et J. Donovan, « *Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign* », *Political Communication*, Vol. 37, Iss. 2, 2020, pp. 194-214. Les auteurs montrent notamment qu'un facteur important dans la propagation de campagnes de désinformation est la décontextualisation du contenu à mesure qu'il circule d'une plateforme à l'autre. Ainsi, mettre en place des protocoles facilitant le suivi des contenus qui migrent d'une plateforme à l'autre pourrait atténuer cet effet de « filtrage ».

[127] AI Forensics, « *Grok Generating Flood of Sexualized Images of Women and Minors* », 5 janvier 2026.

[128] Voir notamment : D. Delmonaco, S. Mayworm, H. Thach, J. Guberman, A. Augusta, O.L. Haimson, « *What are you doing, TikTok" : How Marginalized Social Media Users Perceive, Theorize, and "Prove" Shadowbanning* », *University of Michigan Library*, avril 2024.

droite[129]. Cette enquête a été élargie « *à la suite d'autres signalements dénonçant le fonctionnement de Grok sur la plateforme X, ayant conduit à la diffusion de contenus négationnistes et de* deepfakes *à caractère sexuel* »[130]. La CNCDH s'inquiète tout particulièrement de ces développements au regard des risques que peuvent entraîner les outils tels que Grok, de même que les autres IA adossées aux réseaux sociaux, dans la diffusion d'informations biaisées[131]. À ce propos, elle remarque que la modération du contenu déléguée à des outils d'IA peut accentuer les biais constatés, dans la mesure où les systèmes algorithmiques sont confrontés à plusieurs difficultés dans la modération des contenus haineux, dont la prise en compte du contexte socio-politique ou des spécificités langagières. En effet, l'autoentraînement des systèmes d'IA et la baisse de l'intervention humaine constatée au sein des équipes de modération peuvent conduire à des décisions inadaptées, et ainsi porter atteinte à la liberté d'expression et à la liberté d'information.

55. Dans ce contexte, la CNCDH considère dès lors que la conception même des outils de modération doit être fondée sur les droits humains. Ainsi, ces outils doivent être intégrés à l'algorithme de recommandation, de sorte que les propos identifiés comme étant de nature violente et/ou discriminatoire ne puissent pas être promus par l'algorithme. Par ailleurs, la CNCDH rappelle l'importance de la modération humaine afin d'assurer un retrait approprié des contenus[132]. En outre, elle souligne le caractère décisif de la vérification des faits dans la lutte contre la désinformation.

[129] D. Leloup et M. Untersinger, « L'enquête sur des soupçons d'ingérences étrangères sur X en France se rapproche d'Elon Musk », *Le Monde*, 11 juillet 2025.

[130] FranceInfo/AFP, « Les locaux français de X perquisitionnés par la section de lutte contre la cybercriminalité du parquet de Paris », 3 février 2026.

[131] Voir : O. Clairouin, « Grok, l'IA d'Elon Musk, est avant tout une redoutable machine à désinformer », *Le Monde*, 21 novembre 2025.

[132] Voir CNCDH, A-2021-9, *op. cit.*, § 43.

II. Assurer la protection des droits humains dès la conception des services numériques

56. La CNCDH partage les considérations du Comité des ministres du Conseil de l'Europe, selon lequel « *les plateformes en ligne devraient être conçues de façon à garantir le plus haut niveau de protection des droits humains* »[133]. Pour ce faire, la CNCDH produit un certain nombre de recommandations visant remettre en cause le modèle toxique de conception des services numériques (A), à garantir l'autonomisation des utilisateurs et utilisatrices en ligne (B), et responsabiliser les fournisseurs de services numériques (C).

A. Remettre en cause un modèle toxique de conception des services numériques

57. La CNCDH considère qu'une nouvelle méthodologie de conception doit prévaloir afin de garantir la protection des droits humains. Celle-ci doit permettre d'adopter une conception éthique des services numériques : assurant le respect, dès la conception, des droits humains des utilisateurs et utilisatrices (1), renonçant à la conception addictive (2) et à la maximisation de l'engagement (3).

1. Imposer une conception protectrice des droits humains

58. La CNCDH considère que les risques identifiés au sein de la première partie du présent *Avis* résultent de choix de conception délibérés. En conséquence, les mécanismes de modération *a posteriori* apparaissent insuffisants pour apporter une réponse appropriée aux risques engendrés. En France, les débats autour de la question se sont longtemps concentrés autour des dangers particuliers que ces outils génèrent pour les mineurs. La réponse actuellement privilégiée consiste à interdire l'accès aux réseaux sociaux autres que les services de messagerie aux mineurs de moins de quinze ans et à instaurer un « *couvre-feu numérique* » de 22h à 8h[134]. Toutefois, tout en constatant l'urgence d'agir pour la protection des utilisateurs et utilisatrices et en particulier celle des plus jeunes, la CNCDH estime que ces débats ne doivent pas occulter la nécessité d'agir sur la conception des plateformes, seule à même de garantir le respect des droits humains en ligne pour tous et toutes. La CNCDH relève à ce sujet que le Haut-Commissaire de l'ONU aux droits de l'Homme estime que « *les États doivent exiger des entreprises technologiques qu'elles intègrent la sécurité dès la conception de leurs plateformes, au lieu de faire porter le fardeau aux parents et aux enfants* »[135].

59. La prise de conscience du caractère attentatoire aux droits humains de la conception actuelle des services numériques doit conduire à un changement de paradigme. Celui-ci doit intégrer la protection des droits humains dès leur conception. Une grande part de cet effort consiste à assurer la sécurité, la sûreté et le respect de la vie privée dès la conception *(« by design »)*, c'est-à-dire sans que l'utilisateur ou l'utilisatrice ait à intervenir activement pour

[133] Comité des ministres du Conseil de l'Europe, Recommandation du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme (adoptée par le Comité des Ministres le 8 avril 2020, lors de la 1373e réunion des Délégués des Ministres.

[134] Voir Assemblée nationale, *2025, op. cit.*, « Conclusions », desquelles a été tirée la proposition de loi n° 2107 précitée.

[135] ONU Info, « Protection des enfants en ligne : l'ONU plaide pour une approche globale au-delà des restrictions d'accès », 20 mai 2026.

assurer sa propre protection. Afin de promouvoir une conception des services numériques respectueuse des droits humains, le *Center for Countering Digital Hate* (CCDH), une organisation états-unienne, a développé un cadre permettant de concevoir des plateformes plus vertueuses. Intitulé « STAR Framework », ce cadre repose sur quatre principes : *safety by design, transparency, accountability, responsibility* [136]. La CNCDH considère qu'il conviendrait de s'inspirer de ce cadre afin d'interdire les conceptions attentatoires aux droits humains. En premier lieu, la « *sécurité dès la conception* » implique deux catégories d'actions. Les entreprises doivent tout d'abord anticiper les risques et faire preuve de proactivité afin de garantir que leurs produits et services ne présentent aucun danger pour le public. Ensuite, les principes de la sécurité dès la conception reposent sur une approche systémique préventive face aux risques. Cela implique d'intégrer les considérations de sécurité à travers des évaluations des risques et des décisions prises dès la conception, et lors du déploiement et de la modification des produits et services. En deuxième lieu, en matière de transparence, trois domaines particuliers mériteraient de faire preuve de publications de la part des plateformes : les algorithmes, le respect des règles et les aspects économiques, notamment les gains tirés du traitement des données. Par ailleurs, la CNCDH recommande d'étudier la pertinence de l'approche « *by design* » au domaine des objets connectés, dont elle a relevé les risques particuliers en termes notamment de contacts dangereux[137].

2. Mettre un terme à la conception addictive

60. La CNCDH estime que la conception addictive des services numériques doit être remise en cause. En 2023, le Parlement européen a adopté une résolution invitant la Commission européenne à envisager l'adoption d'une législation relative à la conception addictive[138]. En considérant que « *la dépendance liée à l'utilisation de l'internet peut avoir des effets secondaires similaires aux dépendances liées aux substances, y compris avec des preuves de tolérance et de rechute* », le Parlement propose ainsi une « *conception éthique des services en ligne* ». Celle-ci repose notamment sur la création d'un « *droit numérique de ne pas être dérangé* ». Le Parlement établit en outre une liste de bonnes pratiques, parmi lesquelles celles « *consistant à "réfléchir avant de partager", à désactiver toutes les notifications par défaut, à formuler des recommandations en ligne plus neutres, telles que celles fondées sur l'ordre chronologique ou sur un contrôle accru de l'utilisateur, à choisir d'emblée entre des applications en couleur et des applications en niveaux de gris, ou à émettre des avertissements lorsque les utilisateurs ont passé plus de 15 ou 30 minutes sur un service spécifique ou à verrouiller automatiquement certains services après une durée d'utilisation prédéfinie* […] » [139]. L'organisation EDRi, une fédération paneuropéenne d'associations œuvrant pour la protection des droits humains sur internet, se prononce également en faveur de l'établissement d'une liste noire de pratiques bannies[140]. Ces pratiques pourraient inclure le défilement infini, la lecture automatique des vidéos, ainsi que tout mécanisme qui incite à la connexion prolongée ou suscite des interactions continues.

[136] CCDH, « *Star Framework: CCDH's global standard for regulating Social Media Companies* », 2022.
[137] Voir *supra*, I.B.1.
[138] Parlement européen, Résolution du 12 décembre 2023, *op. cit.*
[139] *Ibid.*, p. 8.
[140] EDRi, 2025, *op. cit.*, p. 31.

3. Mettre un terme au modèle reposant sur la maximisation de l'engagement

61. En second lieu, la CNCDH considère que les systèmes de recommandation fondés sur la manipulation des émotions des utilisateurs et des utilisatrices représentent un risque important pour la santé de chacune et chacun. En outre, ils soulèvent des enjeux de taille en ce qui concerne le débat démocratique. Si l'article 38 du RSN impose aux plateformes de conférer la possibilité de choisir un système de recommandation qui ne soit pas basé sur le profilage, les obligations qui en résultent ne sont pas suffisamment précises. En effet, les systèmes alternatifs sont souvent difficilement accessibles et peu compréhensibles, ce qui détourne les utilisateurs et utilisatrices de leur usage, et autorise les plateformes à prétendre que leur public préfère le classement optimisé fondé sur la personnalisation. Par ailleurs, les alternatives prévues ne sont parfois pas persistantes, de sorte qu'un paramétrage ne sera reflété que sur un temps court, et pas de façon permanente sur l'application. Il convient ainsi dans un premier temps d'assurer que les fils les plus neutres soient activés par défaut, et de garantir la visibilité et l'accessibilité des options de paramétrage. Ensuite, la possibilité de choisir des systèmes alternatifs est également essentielle. Or, le texte laisse à la plateforme la liberté de proposer à ses utilisateurs toute option alternative, sans garantie que celle-ci réponde réellement à leurs intérêts propres.

62. L'une des solutions fréquemment mises en avant comme alternative aux fils d'actualité qui maximisent l'engagement concerne les fils d'actualité reposant sur une présentation chronologique du contenu, qui seraient plus simples et « neutres » que les fils personnalisés[141]. Certes, les fils d'actualité chronologiques peuvent limiter l'engagement de l'utilisateur. Néanmoins, le faible intérêt que leur portent les utilisateurs pourrait les inciter à retourner vers des fils d'actualité personnalisés, lorsque ces derniers sont disponibles[142]. Cela pourrait conduire à imposer ce type de présentation à l'ensemble des plateformes concernées. Or, à l'heure actuelle, bien que les systèmes de recommandation alternatifs soient techniquement disponibles, ils sont souvent dissimulés dans des paramètres difficilement accessibles, en particulier pour un public jeune.

63. Par ailleurs, les flux chronologiques peuvent modifier la composition des contenus recommandés de manière inattendue, en accroissant l'exposition relative d'un utilisateur aux contenus abusifs, en réduisant la visibilité des contenus provenant des comptes de son réseau social et en amplifiant la prévalence des contenus politiques et peu fiables[143]. Ainsi, il convient d'inscrire cette présentation chronologique du contenu dans une conception globalement plus saine, qui lutte également contre la promotion du contenu sensationnaliste. Enfin, les flux chronologiques peuvent également créer un « biais de récence », qui encourage les comportements de publication de type spam[144]. Ici encore, il conviendrait donc de prévoir des mécanismes qui permettent d'éviter ce type de détournement. En somme, différentes

---

[141] *Ibid.*, p. 8.

[142] J. Bandy et T. Lazovich, « *Exposure to Marginally Abusive Content on Twitter* », *Proceedings of the international Conference on Web and Social Media, Vol. 17,* pp. 24-33, 2023 ; A. Guess *et al.* « *How Do Social Media Feed Algorithms Affect Attitudes and Behavior in an Election Campaign?* », *Science*, pp. 398-404, 2023 ; A. Moehring, « *Personalization, Engagement, and Content Quality on Social Media: An evaluation of Reddit's News Feed* », 2024.

[143] *Ibid*.

[144] P. Bengani, « *What's Right and What's Wrong with Optimizing for Engagement* », *Medium*, 27 avril 2022.

possibilités de paramétrage des fils d'actualité devaient être offertes afin de privilégier l'intérêt déclaré des utilisateurs du service[145].

64. Au-delà, afin de limiter le caractère conflictuel des interactions sur les réseaux sociaux, la pratique du « *bridging* », qui consiste à créer du lien entre les utilisateurs et utilisatrices, repose sur des recommandations algorithmiques qui priorisent les publications promouvant le dialogue ou les émotions positives[146]. Par exemple, la pratique de certaines plateformes permettant de rédiger des « notes » en lien avec les publications répond à l'objectif de contextualiser le contenu plutôt que de cliver les utilisateurs et utilisatrices. Une autre pratique consiste à proposer d'attribuer une note au système de recommandation, ou encore de hiérarchiser les contenus qu'ils et elles préfèrent voir.

65. De manière générale, les recherches récentes soulignent l'importance d'autoriser la personnalisation des systèmes de recommandation par les utilisateurs et utilisatrices, même, voire surtout, lorsque ces systèmes ne sont pas fondés sur l'engagement supposé[147]. En conséquence, le pluralisme des systèmes de recommandation apparaît essentiel au maintien d'un environnement informationnel sain (voir *infra*).

B. Garantir l'autonomisation des utilisateurs et des utilisatrices

66. Selon le Comité des ministres du Conseil de l'Europe, l'autonomisation des utilisateurs et des utilisatrices désigne « *les mesures visant à améliorer la compréhension des utilisateurs, leur aptitude à poser des choix éclairés et leur contrôle des effets des technologies numériques sur leurs droits, notamment en encourageant l'éducation aux médias et à l'information, les possibilités pour les utilisateurs d'exercer leurs droits et les moyens d'action collective* »[148].

67. La CNCDH constate que les pratiques dominantes en matière de conception des services numériques contraignent l'autonomie des utilisateurs et des utilisatrices, notamment en les manipulant, en captant leur attention et en favorisant leur surexposition à certains types de contenus ou à des fausses informations. En effet, la conception addictive et la personnalisation abusive des plateformes sont de nature à interférer avec la liberté des utilisateurs et utilisatrices, et avec le respect du droit à la non-discrimination, à la liberté d'information et à la liberté d'expression. L'autonomisation des utilisateurs et utilisatrices doit leur permettre de faire efficacement valoir leurs droits humains et conduire à la réduction des inégalités résultant des différents usages des services numériques. La CNCDH est donc d'avis qu'une régulation des services numériques doit avoir pour objectif de renforcer l'autonomie de chacune et chacun en ligne. Cet objectif appelle des actions à plusieurs niveaux : il convient de conférer une liberté de choix dans l'utilisation des services (1), d'éduquer aux enjeux du numérique (2) et de soutenir l'action de la société civile (3).

1. Conférer une liberté de choix

68. La mise en évidence des risques engendrés par la conception actuelle de la majorité des services numériques en matière d'atteintes aux droits fondamentaux des utilisateurs et

[145] Knight-Georgetown Institute, « *Better Feeds: Algorithms That Put People First* », p. 20. Par exemple, les fonctionnalités permettant de « voir plus » ou « voir moins » tel ou tel type de contenu devraient être généralisées.
[146] J. Stray, « *Designing recommender systems to depolarize* », *First Monday, 2022*
[147] Knight-Georgetown Institute, *ibid.*, p. 22.
[148] Comité des ministres du Conseil de l'Europe, Recommandation CM/Rec(2026)4, *op. cit.*, §11.

utilisatrices implique de soulever la question des alternatives. En effet, face à des produits délétères, il est important que les utilisateurs et utilisatrices retrouvent une liberté de choix et d'action, en particulier lorsque les conséquences d'un usage dérégulé peuvent être si importantes sur les plans individuels, interpersonnels, sociétaux et systémiques. Le rapport issu de la commission parlementaire sur les effets psychologiques de TikTok préconise notamment d'introduire une obligation de « *pluralisme algorithmique* ». Cette notion recouvre, d'une part, la capacité de permettre aux utilisateurs de paramétrer leurs systèmes de recommandation et de modération. Ainsi, il pourrait être envisageable de conférer à l'utilisateur et à l'utilisatrice la possibilité de choisir un fil d'actualité qui ne soit pas fondé sur la collecte de ses données comportementales. De même, il pourrait être permis de choisir un algorithme qui ne diffuse que du contenu en noir et blanc, qui priorise la nuance et la contextualisation plutôt que la confrontation et le sensationnel, ou encore qui s'adapte aux préférences spécifiquement indiquées. D'autre part, le concept de pluralisme algorithmique repose sur la capacité des tiers à proposer des fonctionnalités complémentaires à celles proposées par la plateforme propriétaire du réseau social[149]. La CNCDH exprime le grand intérêt qu'elle porte au principe de pluralisme algorithmique afin de redonner un pouvoir d'agir aux utilisateurs et utilisatrices des services numériques. En particulier, elle considère que le pluralisme des systèmes de recommandation est essentiel à une information de qualité. Elle soutient en outre l'interopérabilité des réseaux sociaux afin de favoriser la liberté de choix des utilisateurs et utilisatrices et, plus généralement, un espace informationnel de qualité[150].

69. En somme, la CNCDH considère qu'une conception protectrice des droits humains ne doit reposer ni sur l'exploitation des vulnérabilités individuelles et collectives, ni sur la maximisation de l'engagement. Elle ne doit pas conduire à manipuler ou à tromper les utilisateurs et utilisatrices des services numériques. Elle ne doit pas non plus chercher à créer de dépendance, et doit garantir que les utilisateurs et utilisatrices ont pleinement le contrôle et peuvent agir de manière consciente et éclairée en ligne. Par ailleurs, la liberté d'action de l'utilisateur, nécessaire au respect de ses droits, repose également sur la désactivation, par défaut, des systèmes d'IA intégrées aux plateformes ainsi que des systèmes de recommandation reposant sur les signaux implicites fondés sur l'engagement, pour privilégier les systèmes reposant uniquement sur les signaux explicites fournis par l'utilisateur, en se fondant sur les données résultant des intérêts déclarés par l'utilisateur lors de la définition de son profil, et ses retours d'information. Ces préférences doivent directement influencer les systèmes de recommandations, en fournissant à l'utilisateur la possibilité de signaler à la plateforme si ses préférences sont insuffisamment prises en compte.

2. Éduquer aux enjeux du numérique

70. La CNCDH rappelle le rôle primordial de l'éducation, et souligne l'importance d'éduquer aux usages du numérique. Elle préconise des actions d'éducation aux enjeux du numérique, de sensibilisation aux risques entraînés par les services numériques à tout âge, et de formation aux bonnes pratiques. Au regard de l'ampleur qu'ont pris les risques soulevés par les services numériques, non seulement au niveau individuel mais aussi en termes de rapports interindividuels et sociaux, elle considère que les actions menées ne doivent pas se limiter à

[149] Voir à ce sujet la tribune « Pour le pluralisme algorithmique ! » publiée dans *Le Monde* le 25 septembre 2024 par une soixantaine de personnalités, associations et entreprises.

[150] K. Szymielewicz, « *Reclaiming the Algorithm: A call for social media interoperability* », *DSA Observatory*, 23 février 2026.

des actions de sensibilisation ponctuelles, mais s'inscrire dans un programme structurel au long cours. Ces actions doivent être assurées par des intervenants humains, en s'appuyant au besoin sur la société civile. Elles doivent concerner, en premier lieu, les personnels enseignant, les enfants et leurs parents, ainsi que les personnels médicaux et paramédicaux, avant d'être étendues au reste de la population.

71. Les mesures de sensibilisation et de prévention doivent porter sur les différents usages permis par les services numériques et intégrer les dernières évolutions techniques en la matière, c'est-à-dire notamment sensibiliser aux enjeux soulevés par les outils d'IA, y compris sur les plans social et environnemental. L'éducation aux services numériques doit également intégrer une dimension critique quant aux risques entraînés par ces technologies. Elle doit permettre d'éveiller l'esprit critique des utilisateurs et utilisatrices, afin notamment de lutter contre la désinformation. Ces actions doivent ainsi nécessairement s'inscrire dans un cadre de réflexion autour de l'usage du numérique, notamment à l'école.

72. Les mesures d'éducation adoptées doivent également porter sur l'éducation aux droits humains et l'exercice des droits et libertés en ligne, en intégrant les pratiques de paramétrage. À ce titre, les mesures d'éducation aux services numériques doivent notamment s'étendre aux actions entreprises dans le cadre de l'éducation à la vie affective, relationnelle et sexuelle (ÉVARS), qui, ainsi que la CNCDH l'a rappelé à de nombreuses reprises, constituent un maillon essentiel dans la lutte contre les violences sur internet. En effet, l'exposition majeure à la pornographie (40 % des enfants y ayant été confrontés avant la fin de l'école primaire, et 100 % avant la fin du collège) entraîne notamment une altération de la construction de la sexualité, du rapport à l'autre et de la relation amoureuse[151]. Par ailleurs, l'exposition croissante aux discours masculinistes via les plateformes, qui vont parfois jusqu'à la promotion de comportements violents, entraîne des inquiétudes pour le développement des jeunes garçons. En conséquence, les mesures d'éducation au numérique doivent être globales et intégrées en tant que telles aux parcours et aux programmes scolaires. Leur déploiement doit donner lieu à la formation des personnels enseignants. L'approche privilégiée doit aboutir à un *continuum* de l'éducation aux droits humains, en ligne et hors ligne.

73. Par ailleurs, la CNCDH rappelle que l'éducation et la sensibilisation au numérique doivent être adaptées au public visé[152]. Elle encourage donc les autorités, associations et plateformes à s'adapter aux codes des différents publics, notamment des mineurs, dans le cadre de partenariats avec des professionnels ou des organismes expérimentés dans de telles interventions, par exemple en réalisant des vidéos en direct. Cela peut aussi prendre la forme de jeux interactifs, de quiz ou de mises en situation. Il convient également de promouvoir la concertation avec des spécialistes du graphisme et, par exemple, des pédopsychiatres pour s'assurer d'une communication effective du message en cause.

74. Enfin, la CNCDH rappelle que les actions envisagées doivent être élaborées en co-construction avec les publics et organisations associatives concernées, qui ont développé une expertise du sujet depuis de nombreuses années. Ainsi, les associations de défense des droits humains en ligne, mais également les organisations de lutte contre toutes les formes de discrimination, devraient être consultées aux stades de l'élaboration et du déploiement des

[151] Sénat, « Porno : l'enfer du décor », Rapport d'information n° 900 (2021-2022), 27 septembre 2022.
[152] CNCDH, A-2021-9, *op. cit.,* § 68.

actions éducatives envisagées. En conséquence, ces organisations devraient faire l'objet d'un soutien appuyé de la part des pouvoirs publics.

3. Encourager l'action de la société civile

75. La participation de la société civile est largement considérée comme essentielle à la mise en œuvre effective et efficace de la réglementation numérique européenne[153]. En premier lieu, les associations sont en première ligne de la modération des contenus, en étant désignées comme « signaleurs de confiance » au sens du RSN. Mais leur rôle va plus loin : d'une part, les organisations non-gouvernementales, expertes et experts indépendants, ou membres du milieu académique peuvent apporter leurs connaissances et leur expertise aux processus de contrôle du RSN, tout en alertant l'opinion publique quant aux nouveaux enjeux et aux meilleures solutions pour y répondre. En effet, c'est notamment grâce au travail de recherche d'organisations comme Amnesty International que les effets délétères de l'algorithme de certaines plateformes ont pu être mis en lumière[154]. Le rôle majeur et l'expertise dont bénéficient les associations sont d'ailleurs soulignés par le fait que la Commission européenne elle-même s'appuie sur leurs travaux dans ses enquêtes visant les entreprises de service numérique[155].

76. Plus encore, la CNCDH considère que les organisations de la société civile jouent un rôle de premier plan dans l'autonomisation des utilisateurs et utilisatrices. En effet, par leurs alertes fréquentes quant aux risques émergents, par leurs conseils pratiques adressés au public, ou encore par la mise en évidence des défaillances du cadre légal, elles contribuent directement à alerter l'opinion publique quant aux dangers qui résultent de la conception des services numériques. En particulier, les organisations de la société civile se sont saisies des rapports d'évaluation des risques systémiques publiés par les grandes plateformes (voir *infra*), afin d'en souligner des limites et de proposer des améliorations. Elles dénoncent en particulier le caractère insuffisant de la prise en compte de la conception dans la création de ces risques. C'est enfin par l'action collective et le regroupement au sein d'associations que les individus peuvent trouver les ressources et les moyens d'agir et de se faire entendre. En conséquence, la CNCDH estime qu'il conviendrait de faire de la participation inclusive de la société civile un critère explicite pour évaluer la conformité des très grandes plateformes aux articles 34 et 35 du RSN.

77. D'autre part, la société civile peut contribuer à renforcer globalement la responsabilisation des fournisseurs de services en examinant de manière indépendante les actions et les choix des entreprises de plateformes et des autorités publiques compétentes. Leur action peut être essentielle à la mise en conformité des entreprises de services numériques avec leurs obligations, par le biais d'actions en justice[156]. Une étude concernant l'implication de la société civile dans la mise en œuvre du RSN révèle que ces actions de contentieux stratégique, ainsi que les plaintes déposées auprès des autorités de régulation,

[153] Voir notamment : N. Edler, « *Making Systemic Risk Assessments Work: How the DSA Creates a Virtuous Loop to Address the Societal Harms of Content Moderation* », *German Law Journal*, Vol. 25(7), 2024, pp. 1197-1218.

[154] Amnesty International France, « Entraîné-e-s dans le "*rabbit hole*" », 20 octobre 2025.

[155] Ce fut notamment le cas dans le cadre de son enquête concernant la conception addictive de TikTok.

[156] Voir *supra* l'action intentée par *Bits of Freedom* contre *Meta*.

sont considérées comme les plus efficaces et les plus prometteuses pour agir sur la mise en œuvre du RSN[157].

78. Par ailleurs, l'article 40 du RSN instaure la mise en place d'une procédure d'agrément pour l'accès aux données des très grandes plateformes en ligne et très grands moteurs de recherche, permettant aux chercheurs agréés d'accéder de manière inédite à des données détenues par ces plateformes. Toutefois, cette procédure d'agrément, qui demeure lourde et complexe, est à ce jour inaccessible aux organisations non-gouvernementales. Ainsi, la CNCDH considère qu'il conviendrait d'accorder aux organisations de la société civile, sans se limiter au milieu académique, l'accès aux données des grandes plateformes par le biais de l'article 40 du RSN, dans le cadre d'une procédure d'agrément allégée.

79. Enfin, la CNCDH souligne également le rôle essentiel des lanceurs d'alerte dans la connaissance des risques engendrés par les services numériques. En effet, en raison du secret entourant notamment les algorithmes des plateformes, seuls les employés des entreprises qui les conçoivent sont parfois à même d'identifier et de révéler les abus menant à une conception nocive, addictive et manipulatrice. Ainsi, la CNCDH appuie les conclusions figurant au sein du Rapport bisannuel 2024/2025 du Défenseur des droits portant sur la situation des lanceurs d'alerte visant à renforcer la protection qui leur est apportée[158].

C. Responsabiliser les fournisseurs de services numériques

80. La CNCDH rappelle que l'élaboration d'un espace numérique sain et respectueux des droits humains implique à la fois la responsabilisation et la responsabilité des plateformes[159]. Ainsi, elle estime que la réglementation doit être révisée afin de renforcer les obligations à la charge des entreprises en matière de conception des services numériques (1). Le cadre légal doit également être complété afin de faciliter la mise en cause de la responsabilité des fournisseurs de services numériques du fait de la conception de leurs services (2).

1. Renforcer les obligations pesant sur les fournisseurs de services numériques

81. La CNCDH identifie plusieurs véhicules législatifs envisageables afin de garantir la protection des droits humains par défaut. Une première option consiste à réviser le *Règlement sur les services numériques*, afin notamment d'y inclure explicitement l'interdiction des conceptions dangereuses (a). Il conviendrait par ailleurs de procéder à la réforme du droit de la consommation, au moyen notamment de l'adoption du *Règlement sur l'équité numérique* (REN) (b). En tout état de cause, la CNCDH considère que les pouvoirs publics doivent renoncer à la tentation de la « simplification » à outrance, qui est trop souvent synonyme de standards abaissés de protection des droits humains (c).

a. Réviser le Règlement sur les services numériques et assurer sa mise en œuvre

82. Le RSN contient certaines dispositions permettant d'encadrer la conception des services numériques. En premier lieu, l'article 25 § 1 dispose que les plateformes « *ne conçoivent pas, n'organisent pas et n'exploitent pas leurs interfaces en ligne de manière à*

[157] M. Correia de Carvalho et R. Griffin, « *Who speaks and who is heard? Civil society participation and participatory justice in DSA systemic risk management* », février 2026.
[158] Défenseur des droits, « La protection des lanceurs d'alerte en France : un dispositif à l'épreuve de son appropriation », rapport bisannuel 2024/2025, 28 mai 2026.
[159] Le cadre « STAR » évoqué plus haut implique d'imposer aux entreprises de rendre des comptes (*accountability*) et d'être tenues responsables (*responsibility*) des dommages constatés.

tromper ou manipuler les destinataires ». Au sens de la CNCDH, cette interdiction pourrait être prolongée : plutôt que de se limiter aux interfaces manipulatrices et trompeuses, le RSN pourrait interdire les conceptions toxiques en général, en établissant une « liste noire », ouverte et actualisée tous les six mois, de pratiques prohibées en toutes circonstances parce qu'elles ne respectent pas les droits fondamentaux [160]. Cette liste devrait notamment comprendre celles dont il ressort, à l'issue des travaux de la CNCDH, qu'elles sont désormais établies comme étant particulièrement néfastes :

- la lecture automatique des vidéos ;
- le défilement continu et infini ;
- les récompenses en cas de connexion « en série » ;
- les notifications artificielles ;
- les suggestions fondées sur le profilage, qui reposent sur des informations présumées que les utilisateurs et utilisatrices ne peuvent ni vérifier ni contester ;
- les fonctionnalités et autres méthodes de personnalisation qui exploitent les vulnérabilités cognitives ou émotionnelles, affectent la visibilité des options, le prix ou le consentement des utilisateurs.

83. Par ailleurs, certaines pratiques de conception à haut risque présumées nocives pourraient être réunies au sein d'une « liste grise ». Une telle liste pourrait notamment viser les mécanismes de séries (*streaks*), les incitations à l'engagement formulées selon un registre émotionnel, les fonctionnalités ludifiées, ainsi que les boîtes à butin (*lootboxes*)[161]. Cette liste devrait être conçue comme étant ouverte et non-exhaustive, afin d'être facilement adaptable aux pratiques émergentes. Cette caractéristique est essentielle afin que la législation ne soit pas toujours en retard par rapport aux nouvelles formes de manipulation et d'abus. En cas d'allégation contraire, il reviendrait aux professionnels de démontrer que celles-ci ne portent pas atteinte à l'autonomie des utilisateurs, ni n'induisent d'usage compulsif.

84. Par ailleurs, l'article 28 § 4 du RSN ouvre à la Commission européenne la faculté de publier des lignes directrices aux fins d'aider les plateformes accessibles aux mineurs à se conformer à leur obligation tirée de l'article 28 § 1 de « *garantir un niveau élevé de protection de la vie privée, de sûreté et de sécurité des mineurs sur leur service* ». La Commission européenne a publié de telles lignes directrices le 14 juillet 2025[162]. Celles-ci préconisent notamment de définir les comptes des mineurs en privé par défaut afin que leurs informations personnelles, leurs données et leur contenu sur les médias sociaux soient inaccessibles pour les personnes avec lesquelles ils ne sont pas connectés ; de modifier les systèmes de recommandation des plateformes afin de réduire le risque que les enfants soient confrontés à des contenus préjudiciables ou se retrouvent coincés dans des « terriers de lapin » de contenus spécifiques ; de donner aux enfants les moyens de bloquer tout utilisateur et s'assurer qu'ils ne peuvent pas être ajoutés à des groupes sans leur consentement explicite ; de désactiver par défaut les fonctionnalités qui contribuent à une utilisation excessive, telles

[160] À ce sujet, les propositions du Parlement européen constituent une source d'inspiration particulièrement pertinente. Voir : Parlement européen, Résolution du 12 décembre 2023, *op. cit.*

[161] Voir : EDRi, 2025, *Op. cit*, p. 32.

[162] Commission européenne, « Communication de la Commission. Lignes directrices concernant des mesures visant à garantir un niveau élevé de protection de la vie privée, de sûreté et de sécurité des mineurs en ligne, conformément à l'article 28, paragraphe 4, du règlement (UE) 2022/2065 », (C/2025/5519), 14 juillet 2025, JOUE du 10 octobre 2025.

que les « traces » de communication, le contenu éphémère, les « reçus de lecture », la lecture automatique ou les notifications *push*, ainsi que supprimer les fonctionnalités de conception visant principalement à l'engagement et à mettre en place des garanties concernant l'intégration des *chatbots* fonctionnant par IA dans les plateformes en ligne.

85. Étant dépourvues de valeur contraignante, ces lignes directrices ne peuvent pas, en tant que telles, fonder une décision obligatoire à l'encontre des plateformes. Néanmoins, elles constituent une source importante d'interprétation, à disposition non seulement des plateformes mais également de la société civile ainsi que, le cas échéant, de la Cour de justice de l'Union européenne (CJUE). La CNCDH considère que l'impact de ces lignes directrices pourrait être grandement amélioré si elles étaient étendues à l'ensemble des utilisateurs, et ne concernaient pas uniquement la protection des personnes mineures. Par ailleurs, elle constate que certains paramètres préconisés sont insuffisants à assurer pleinement la protection des droits humains. En particulier, il conviendrait de faire en sorte que, par défaut, les algorithmes ne reposent pas sur la collecte permanente et opaque de données comportementales, mais sur les choix explicitement exprimés par les utilisateurs et utilisatrices. À cette condition, la CNCDH considère que le contenu de ces lignes directrices pourrait constituer le socle d'une obligation générale de conception équitable, au bénéfice de tous les utilisateurs, indépendamment de leur âge.

86. De même, les articles 34 et 35 imposent aux très grandes plateformes en ligne et aux très grands moteurs de recherche de recenser, d'analyser et d'évaluer de manière diligente tout risque systémique découlant de la conception ou du fonctionnement de leurs services. Les entreprises concernées doivent publier un rapport annuel, « *spécifique à leurs services et proportionnée aux risques systémiques, de la gravité et de la probabilité desquels elle tient compte* ». Cette évaluation doit porter sur les risques suivants : la diffusion de contenus illicites par l'intermédiaire de leurs services ; tout effet négatif réel ou prévisible pour l'exercice des droits fondamentaux ; tout effet négatif réel ou prévisible sur le discours civique, les processus électoraux[163] et la sécurité publique ; tout effet négatif réel ou prévisible lié aux violences sexistes et à la protection de la santé publique et des mineurs et les conséquences négatives graves sur le bien-être physique et mental des personnes. La CNCDH constate que les premières évaluations publiées constituent une avancée importante vers la transparence et la responsabilité des plateformes numériques. Toutefois, elle considère que ces évaluations ne prennent pas suffisamment en compte le rôle de la *conception* des services numériques dans la survenance des risques et dans la réduction de ces derniers. Ainsi que l'a montré une réponse collective d'organisations de la société civile après la première série de rapports, ces derniers se sont essentiellement fondés sur des éléments « recyclés », c'est-à-dire déjà connus du public, et se sont concentrés sur la modération des contenus[164]. En particulier, si les principaux fournisseurs de réseaux sociaux considèrent les risques pour la santé et le bien-

[163] Voir les lignes directrices publiées par la Commission en la matière : Commission européenne, « Lignes directrices à l'intention des fournisseurs de VLOP et de VLOSE sur l'atténuation des risques systémiques pour les processus électoraux », 26 avril 2024.

[164] Voir notamment : European Center for Non-Profit Law, « *Five critical lessons from three years of DSA risk assessments* », 3 mars 2026 ; DSA Civil Society Coordination Group, « *Initial Analysis on the First Round of Risk Assessments Reports under the EU Digital Services Act* », mars 2025.

être comme des risques systémiques, ils ne mesurent pas l'ampleur avec laquelle leurs services contribuent à la survenance de ces risques[165].

87. Par ailleurs, en l'absence de lignes directrices précises émanant de la Commission quant à la réalisation de ces rapports, la méthodologie employée, les définitions adoptées et les données fournies ne sont pas harmonisées d'un rapport à l'autre. La CNCDH en conclut qu'il conviendrait que la Commission publie des lignes directrices promouvant une méthodologie harmonisée et imposant aux plateformes des obligations précises aux fins de mesurer l'effet de la conception de leurs services, au-delà de la seule conception de leurs systèmes de recommandation, sur les risques systémiques entraînés[166]. À cette fin, l'outil développé par le Conseil de l'Europe pour évaluer l'impact des systèmes d'IA sur les droits humains pourrait constituer une source d'inspiration particulièrement utile[167]. Dans un premier temps, ces lignes directrices pourraient porter sur l'évaluation des risques en termes de santé mentale. Plus encore, la CNCDH estime qu'il conviendrait pour les plateformes d'attester de l'efficience des mécanismes de réduction des risques mis en place avant la manifestation d'un dommage en résultant. À ce sujet, la CNCDH considère que les fournisseurs devraient collaborer avec la société civile, et notamment avec le milieu académique, pour apprécier les risques potentiels et réaliser ces évaluations.

88. Enfin, la CNCDH invite la Commission européenne à poursuivre les actions entreprises depuis l'entrée en vigueur du RSN. Elle considère également qu'en complément des institutions européennes, les régulateurs nationaux jouent un rôle fondamental dans la détection, la documentation et la sanction des pratiques numériques dangereuses. Elle encourage ainsi à renforcer les moyens humains, matériels et financiers à disposition de l'Arcom, organe essentiel à la mise en conformité des fournisseurs de service numérique à la législation.

b. Renforcer les dispositions du projet de Règlement sur l'équité numérique

89. La CNCDH partage l'avis du Parlement européen selon lequel la règlementation à venir sur l'équité numérique constitue « *une occasion unique d'ouvrir la voie vers une nouvelle génération de dispositions législatives sur la protection des consommateurs, qui inversera les tendances négatives qui ont affaibli la position des consommateurs et réduit leurs droits* »[168]. A l'heure actuelle, le projet concerne « *les pratiques préjudiciables auxquelles les consommateurs sont confrontés en ligne, telles que la conception d'interfaces trompeuses ou manipulatrices, le marketing trompeur par des influenceurs des médias sociaux, la conception addictive de produits numériques et les pratiques de personnalisation déloyales, en particulier lorsque les vulnérabilités des consommateurs sont exploitées à des fins commerciales* »[169].

---

[165] Voir : Knight Georgetown Institute, « *Systemic Risk Assessment under the Digital Services Act* », 1er mai 2025.

[166] Bien que la Commission européenne ait adopté un Règlement délégué, celui-ci ne détaille pas suffisamment la façon dont les risques systémiques doivent être pris en compte. Voir : Commission européene, Règlement délégué *sur les audits indépendants au titre de la législation sur les services numériques*, 20 octobre 2023.

[167] Voir : Comité sur l'intelligence artificielle du Conseil de l'Europe (Cia), « Méthodologie pour l'évaluation des risques et des impacts des systèmes d'intelligence artificielle du point de vue des droits humains, de la démocratie et de l'Etat de droit (Méthodologie Huderia) », 28 novembre 2024.

[168] Parlement européen, Résolution du 12 décembre 2023, *op. cit.* p. 6.

[169] Voir le site de la consultation lancée par la Commission européenne.

90. La CNCDH considère que le REN tel qu'il est envisagé n'est ni suffisamment large ni suffisamment ambitieux pour prévenir les risques issus de la conception nocive des services numériques. Elle se prononce en faveur d'un REN « fondé sur les droits »[170] et partage l'opinion de nombreuses parties prenantes spécialisées, selon lesquelles le REN devrait s'attaquer au cœur du problème, à savoir « *le décalage structurel entre les intérêts commerciaux du fournisseur (à savoir, capter le temps, l'attention et l'argent des utilisateurs) et les intérêts des utilisateurs et utilisatrices (à savoir, avoir les moyens de développer et d'exercer leur autonomie en ligne)* »[171]. En outre, la CNCDH considère que le REN ne devrait pas non plus se limiter aux seules places de marché. Au contraire, le REN devrait être pensé pour s'appliquer également aux plateformes de réseaux sociaux ainsi qu'aux autres services numériques tels que les agents conversationnels, aussi bien généralistes que les compagnons IA.

91. Afin de combler ces lacunes, la CNCDH considère que le droit européen de la consommation pourrait constituer une voie privilégiée. En particulier, la définition des pratiques commerciales issue de la directive sur les pratiques commerciales déloyales (DPCD)[172] pourrait être élargie afin d'inclure les pratiques commerciales numériques. En outre, la définition du « *consommateur moyen* »[173] pourrait être modifiée afin de refléter le constat selon lequel la vulnérabilité est une notion avant tout situationnelle[174] : au sein de l'environnement numérique en particulier, tous les consommateurs sont potentiellement vulnérables[175]. Enfin, le droit de la consommation pourrait constituer le véhicule idoine pour élaborer la liste de pratiques de conception illicites mentionnée au sein du présent *Avis*.

92. Par ailleurs, la CNCDH tient à alerter les autorités françaises et européennes quant aux difficultés d'articulation des textes applicables à la régulation du numérique. Ainsi, les discussions autour du REN devraient également avoir pour objectif de combler les lacunes de la règlementation, et de résoudre les ambiguïtés résultant de l'application du RSN et de la directive DPCD en s'assurant que l'ensemble des services numériques sont couverts par les obligations prévues, de façon graduelle et en fonction des risques identifiés.

---

[170] Voir EDRi, « *Breaking the Extractive Digital Business Model: A Rights-Based Digital Fairness Act* », 3 février 2026.

[171] J. Albert, M. Sax, et N. Helberger, « *Digital Fairness Act: Why we need an ambitious DFA to protect digital consumers from manipulative and addictive design practices* », *DSA Observatory*, mai 2026.

[172] Directive 2005/29/CE du Parlement européen et du Conseil du 11 mai 2005 *relative aux pratiques commerciales déloyales des entreprises vis-à-vis des consommateurs dans le marché intérieur* et modifiant la directive 84/450/CEE du Conseil et les directives 97/7/CE, 98/27/CE et 2002/65/CE du Parlement européen et du Conseil et le règlement (CE) n° 2006/2004 du Parlement européen et du Conseil (« *directive sur les pratiques commerciales déloyales* »).

[173] Le considérant 18 de la directive définit le consommateur moyen comme celui « *qui est normalement informé et raisonnablement attentif et avisé* », par opposition au consommateur vulnérable « *dont les caractéristiques* [le] *rendent particulièrement vulnérable* […] *aux pratiques commerciales déloyales* ».

[174] Constat auquel la Commission elle-même est parvenue, en actant du décalage croissant entre la définition du consommateur type et les réalités du comportement de consommation dans le monde numérique : Commission européenne, « *Commission Staff Working Document Fitness Check on EU consumer law on digital fairness* », p. 46.

[175] Voir le I) A) du présent *Avis*, qui détaille la façon dont la conception des services numériques exploite les vulnérabilités humaines. Voir également : Bureau européen des unions de consommateurs (BEUC), « *Towards the Digital Fairness Act* », décembre 2025.

c. Assurer l'effectivité du cadre législatif et refuser la logique de dérégulation

93. La CNCDH prend acte de la volonté de la Commission européenne, annoncée dans sa communication intitulée « *Une Europe plus simple et plus rapide* »[176], d'alléger la « *charge réglementaire* » pesant sur les entreprises, notamment dans le secteur du numérique. Depuis, la Commission européenne a présenté un paquet de réformes législatives, dit « *omnibus numérique* », qui vise à simplifier le cadre normatif applicable, dans l'objectif annoncé de favoriser l'innovation[177].

94. La CNCDH fait preuve de la plus grande réserve face aux réformes envisagées. En effet, elle s'oppose à la logique de « simplification » qui se confond généralement avec un allègement des obligations pesant sur les entreprises en matière de respect des droits humains. En conséquence, si elle se réjouit de ce que le RSN soit à ce stade exclu des réformes envisagées, elle exprime d'importantes réserves quant aux propositions concernant le Règlement sur l'intelligence artificielle (RIA) et le Règlement général sur la protection des données (RGPD). Elle s'inquiète notamment de ce que les discussions aient été menées sur une période très courte, sans qu'aucune consultation publique ni étude d'impact n'ait été menée. La CNCDH appelle à défendre une approche de l'omnibus sans réduction du niveau de protection des droits fondamentaux.

95. En matière d'intelligence artificielle, la proposition de règlement suggère de procéder à la refonte du RIA, alors que celui-ci n'est même pas encore complètement entré en vigueur. En particulier, l'article 1(14) du projet d'omnibus numérique sur l'IA propose de supprimer l'obligation d'enregistrer les systèmes d'IA que les fournisseurs estiment exemptés de la qualification de système « à haut risque ». Comme le rappellent les réseaux ENNHRI et Equinet, si certains systèmes peuvent être traités comme n'étant pas à haut risque lorsqu'ils remplissent des conditions strictes, cette évaluation doit être documentée, et cette information doit être rendue visible grâce à son enregistrement dans la base de données de l'UE pour les IA à haut risque[178]. La CNCDH est alignée avec l'avis des institutions nationales des droits humains européennes, selon lequel le fait de supprimer les exigences d'enregistrement « *affaiblirait la transparence, entraverait l'accès aux recours et accroîtrait le risque d'usage abusif de l'exception prévue à l'article 6(3), tout en n'apportant qu'un allégement administratif limité compte tenu du caractère relativement léger de l'obligation de documentation* »[179]. En outre, le maintien de l'obligation d'enregistrement des SIA est essentiel pour garantir le bon fonctionnement des analyses d'impact relatives aux droits fondamentaux (AIDF), prévues à l'article 27 du RIA, et pour préserver le droit au recours des personnes concernées. Au vu des risques particuliers présentés par la conception des services numériques, y compris des SIA, ces analyses, qui doivent être réalisées avant la mise sur le marché du produit, apparaissent essentielles à la préservation des droits humains.

---

[176] Commission européenne, « Une Europe plus simple et plus rapide : Communication sur la mise en œuvre et la simplification », COM(2025) 47 final, 11 février 2025.

[177] Commission européenne, Proposition de Règlement du Parlement européen et du Conseil modifiant les règlements (UE) 2024/1689 et (UE) 2018/1139 en ce qui concerne la simplification de la mise en œuvre des règles harmonisées concernant l'intelligence artificielle (train de mesures omnibus numérique sur l'IA), COM(2025) 836 final, 19 novembre 2025.

[178] Equinet-Ennhri, « *Statement on the Digital Omnibus Regulation Proposals on AI and on Data* », 12 mars 2026, p. 5.

[179] *Ibid*., pp. 5-6.

96. Par ailleurs, le 13 mai 2026, le Parlement européen et le Conseil sont parvenus à un accord provisoire concernant l'omnibus IA[180]. Celui-ci envisage notamment d'inclure, parmi les pratiques d'IA interdites par l'article 5 du Règlement, les systèmes capables de générer du matériel pédocriminel (CSAM pour *child sexual abuse material*) ou des images intimes non-consenties (IINC), c'est-à-dire qui représentent les parties intimes d'une personne identifiable, ou encore qui la représentent se livrant à des activités sexuellement explicites, lorsque ces images sont générées sans le consentement de cette personne[181]. La CNCDH accueille favorablement cette proposition, qui renforce la protection des droits humains en ligne, et notamment ceux des femmes, des filles et des enfants[182].

97. En outre, la CNCDH constate qu'en l'état, l'article 5 b) du RIA n'interdit pas les compagnons IA qui reposent sur la *sycophancy*, ni ceux qui reposent sur des biais anthropomorphiques. Or, elle considère qu'au regard des risques qu'ils entraînent, notamment en termes de santé publique, ces systèmes devraient être interdits. En complément, parmi les SIA qui ne sont pas prohibés au titre de l'article 5 du RIA, les services devraient se voir imposer une obligation renforcée de transparence. Celle-ci devrait s'appliquer tant au sujet de l'information fournie à l'utilisateur sur la nature réelle de l'interaction (en rappelant régulièrement qu'il s'agit d'une IA, et en mettant en garde contre les risques d'anthropomorphisation) qu'au niveau du design (sans proposer d'interface conçue sous des traits humanisés ou susceptibles de susciter l'attachement ou encore d'entretenir une confusion sur la nature réelle de l'interaction)[183].

98. Enfin, et pour les mêmes raisons tenant au risque que cela soulève pour la santé et la protection de l'intimité, la CNCDH affirme l'importance d'interdire la publicité fondée sur la collecte de données comportementales. Appliquée aux SIA dès maintenant, cette mesure permettrait d'instaurer un cadre préventivement protecteur, ce qui a manqué en matière de régulation des réseaux sociaux.

---

[180] Parlement européen et Conseil de l'Union européenne, Accord provisoire résultant des négociations interinstitutionnelles relatives à la proposition de règlement du Parlement européen et du Conseil *modifiant les règlements (UE) 2024/1689 et (UE) 2018/1139 en ce qui concerne la simplification de la mise en œuvre des règles harmonisées concernant l'intelligence artificielle (train de mesures omnibus numérique sur l'IA)*.

[181] Parlement européen, « *AI Act: deal on simplification measures, ban on "nudifier" apps* », communiqué de presse, 7 mai 2026. Leur proposition consiste à ajouter à l'article 5 du RIA « *la mise sur le marché, la mise en service ou l'utilisation d'un système d'IA qui génère ou manipule des images, vidéos, enregistrements audio ou autres contenus similaires représentant des parties intimes d'une personne physique identifiable, ou une personne physique identifiable engagée dans des activités sexuellement explicites, sans le consentement libre, spécifique, éclairé, univoque et explicite de cette personne pour cette génération ou cette manipulation* » (traduction libre).

[182] Celle-ci est d'ailleurs alignée avec les considérations développées au sein des précédents travaux de la CNCDH. Voir notamment : CNCDH, *Avis sur la protection de l'intimité des jeunes en ligne*, A-2025-1, 23 janvier 2025 ; *Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux*, A-2022-6, 7 avril 2022.

[183] Voir à ce sujet : Commission européenne, « *Guidelines on the implementation of the transparency obligations for certain AI systems under Article 50 of the AI Act* », 8 mai 2026.

2. Compléter le cadre légal par la reconnaissance d'une responsabilité du fait de la conception

99. Pendant de nombreuses années, le débat autour de la responsabilité des plateformes s'est concentré sur la modération des contenus. Or, l'évolution du rôle et des usages des services numériques, ainsi que les risques que ces derniers font peser sur leurs utilisateurs et utilisatrices, détaillés en première partie de cet avis, amènent à dépasser les réflexions portant sur le contenu, pour les faire porter sur la conception même des plateformes. On observe actuellement un mouvement de reconnaissance de la responsabilité des fournisseurs de plateformes du fait de la conception de leurs services numériques. Ainsi, aux États-Unis, dans le cadre d'une affaire introduite par la mère d'un adolescent s'étant suicidé après avoir entretenu de longues « conversations » avec un compagnon IA, l'exception tirée de la liberté d'expression, qui permettait jusqu'ici aux entreprises de contester leur responsabilité, a été écartée au profit de la responsabilité du fait des produits[184]. De même, les deux décisions étatsuniennes évoquées *supra*[185] dépassent le *statu quo* tiré de la section 230 qui avait, de fait, instauré une immunité des fournisseurs pour les contenus qui figurent sur leurs plateformes. Au Nouveau-Mexique, le procureur a ordonné la création de comptes pour les besoins de l'enquête, afin de démontrer que la manière dont l'algorithme fonctionnait avait pour résultat de faciliter la connexion entre des utilisateurs et utilisatrices mineurs et des prédateurs sexuels. Meta a ainsi été poursuivi sur le fondement du droit de la consommation et des pratiques commerciales déloyales, car l'entreprise n'avait pas suffisamment informé les enfants et leurs parents des risques auxquels les soumettaient Facebook et Instagram. En Californie, Snapchat, YouTube, Meta et TikTok étaient accusées d'avoir sciemment conçu des services particulièrement addictifs pour les personnes mineures. Meta et YouTube ont ainsi été reconnues coupables de négligence intentionnelle.

100. En France, les agissements évoqués dans le cadre de cet avis sont susceptibles d'entraîner l'application de la responsabilité civile délictuelle. Toutefois, la difficile application de ce régime, prévu aux articles 1240 et 1242 du Code civil[186], résulte de la nécessité d'établir un lien de causalité entre le fait générateur (ici, la conception même des services), et le dommage qui est résulté de leur utilisation (qu'il s'agisse de troubles anxieux, troubles du comportement alimentaire, dysmorphie corporelle, troubles dépressifs, automutilation, comportement de dépendance, choc post-traumatique…). C'est là tout l'enjeu des actions en justice poursuivies, en France, par les membres du collectif *Algos Victima*, qui tentent d'obtenir réparation pour les dommages subis par des adolescents et adolescentes à la suite de leur utilisation de réseaux sociaux[187].

101. Une autre solution réside dans le régime spécial de responsabilité du fait des produits défectueux prévu par les articles 1245 à 1245-17 du Code civil. La réforme du droit européen

[184] Les parties ont néanmoins conclu une transaction judiciaire. Voir : *Libération* avec AFP, « Suicides d'adolescents : Google et Character.AI concluent un accord pour éviter les poursuites aux États-Unis », *Libération*, 9 janvier 2026.

[185] *State of New Mexico v. Meta Platforms*, 24 mars 2026, et *K.G.M. v. Meta et al*., 25 mars 2026.

[186] L'article prévoit que l'on est « *responsable non seulement du dommage que l'on cause par son propre fait, mais encore de celui qui est causé par le fait […] des choses que l'on a sous sa garde* ».

[187] *Le Monde*, « Sept familles françaises annoncent assigner TikTok en justice après des suicides d'adolescentes », 4 novembre 2024.

et l'entrée en vigueur de la directive sur la responsabilité du fait des produits défectueux[188] a permis d'inclure les logiciels et les outils d'IA dans la notion de « produit ». Le régime instauré entraîne la responsabilité objective (c'est-à-dire sans faute) du producteur en raison du « *défaut de son produit* », la défectuosité étant caractérisée par le fait qu'un produit « *n'offre pas la sécurité à laquelle une personne peut légitimement s'attendre* »[189]. Toutefois, le champ d'application de la directive apparaît particulièrement restreint. En effet, la directive limite les dommages couverts aux pertes matérielles et immatérielles résultant de la mort, des lésions corporelles et des atteintes aux biens[190]. Si l'atteinte à la santé psychologique est bien incluse, la réparation des atteintes à la vie privée ou à la discrimination est, elle, explicitement exclue[191]. Or, la CNCDH considère que l'ensemble des dommages devraient être inclus. Enfin, le projet de directive sur la responsabilité extracontractuelle de l'IA ayant été abandonné[192], aucun texte ne permet de couvrir l'ensemble des dommages générés par les services numériques de façon objective.

102. Certaines solutions étrangères pourraient inspirer le droit applicable en France et au sein de l'UE. Le régime italien de responsabilité objective pour l'exercice d'activités dangereuses entraîne une inversion de la charge de la preuve, de sorte qu'elle pèse sur le producteur, qui doit démontrer avoir adopté toutes les mesures adéquates pour éviter le dommage[193]. De même, le risque du développement lui incombe également : si, au moment d'entreprendre l'activité dangereuse, la technique n'offre pas de mesures appropriées pour prévenir les dommages, il sera responsable des dommages causés. Toutefois, ce régime repose sur la qualification de l'IA comme une « *activité dangereuse* ». Or, considérer l'IA comme telle par défaut entraîne d'importantes limites, dans la mesure où nombre de ses applications diverses offrent des bénéfices non négligeables à leurs utilisateurs. Une solution pourrait résider dans la distinction entre les différents usages de l'IA, en s'inspirant des distinctions effectuées par le RIA[194]. Ainsi, il serait possible d'engager la responsabilité du programmateur ou de l'opérateur du seul fait qu'il se livre à une telle activité à risque et qu'un dommage en est résulté, sans nécessité de prouver le lien de causalité. Celles-ci sont particulièrement répandues en Amérique du Nord, où des associations s'inscrivent dans un contentieux stratégique dans le cadre de litiges multidistricts. Ainsi, aux États-Unis, l'affaire *K.G.M.* mentionnée *supra* pourrait inspirer d'autres tribunaux, et entraîner des condamnations en chaîne des plateformes au vu de leur conception addictive et des dommages qu'elle entraîne.

103. Dans ce contexte, la CNCDH estime que la régulation et la responsabilité des fournisseurs de services numériques devrait être mieux articulée, de sorte que l'allégation d'un

[188] Directive (UE) 2024/2853 du op. cit. et du Conseil du 23 octobre 2024 *relative à la responsabilité du fait des produits défectueux et abrogeant la directive 85/374/CEE du Conseil*.
[189] *Ibid.*, Article 7.
[190] *Ibid.*, Article 6.
[191] *Ibid.*, Considérant 14.
[192] Euractiv, « Après les critiques de JD Vance, la Commission retire la directive sur la responsabilité en matière d'IA », 12 février 2025.
[193] Voir à ce sujet M. Scalini, « Analyse de la directive (UE) 2024/2853 sur la responsabilité des produits défectueux à l'ère de l'IA à l'aide d'une étude comparative avec les droits français et italien », *MBDE/Droits internationaux*, 11 mars 2026.
[194] Voir les limites soulevées par la CNCDH dans son *Avis relatif à l'impact de l'intelligence artificielle sur les droits humains*, A-2022-6, *op. cit.*

préjudice découlant de la conception défaillante d'un service numérique devrait permettre d'engager la responsabilité du concepteur selon un régime probatoire allégé en faveur de la victime. En premier lieu, lorsqu'une pratique figurant sur la liste noire est constatée, celle-ci entraînerait de plein droit la responsabilité du fournisseur de services à l'égard des victimes du dommage en résultant. En second lieu, et par analogie avec le régime applicable en droit de la non-discrimination, la victime d'une conception nocive serait ainsi tenue d'apporter un commencement de preuve permettant d'identifier ce caractère nocif. Ainsi, l'article L. 1134-1 du Code du travail prévoit : « *Lorsque survient un litige en raison d'une méconnaissance [de certaines dispositions], le candidat à un emploi, à un stage ou à une période de formation en entreprise ou le salarié présente des éléments de fait laissant supposer l'existence d'une discrimination directe ou indirecte […]. Au vu de ces éléments, il incombe à la partie défenderesse de prouver que sa décision est justifiée par des éléments objectifs étrangers à toute discrimination* ». De même, pour toute pratique figurant sur la « liste grise » des pratiques présumées toxiques, il reviendrait, tout d'abord, à la victime de présenter un commencement de preuve laissant supposer du dommage et, ensuite, aux fournisseurs de services de démontrer que leurs systèmes respectent bel et bien l'autonomie des utilisateurs et utilisatrices, n'exploitent pas leurs vulnérabilités, et sont conformes aux différentes obligations auxquelles ils sont soumis. Afin de faciliter l'harmonisation des solutions adoptées sur le continent européen, une présomption de nocivité pourrait être tirée du fait qu'une autorité de régulation ou une juridiction dans un autre État membre a définitivement considéré la pratique en question comme étant nocive[195].

[195] Voir BEUC, 2025, o*p. cit.*, p. 11.

III. Les recommandations de la CNCDH pour une conception des services numériques protectrice des droits humains

Recommandation n° 1. La CNCDH recommande aux autorités françaises de soutenir l'adoption, au niveau européen, d'une « liste noire » des modalités de conception des services numériques présumées de façon irréfragable comme toxiques et interdites à ce titre, et d'une « liste grise » de modalités de conception faisant l'objet d'une présomption simple de toxicité.

Recommandation n° 2. La CNCDH recommande l'instauration d'un régime de responsabilité du fait de la conception des services numériques. Ainsi, d'une part, le constat d'une pratique figurant sur la liste noire entraînera la responsabilité de plein droit du fournisseur de services dès lors qu'un dommage est allégué par la victime. D'autre part, le constat d'une pratique figurant sur la liste grise entraînera un allègement de la charge probatoire en faveur de la victime, au moyen d'un mécanisme analogue à celui prévu en matière de droit de la non-discrimination.

Recommandation n° 3. La CNCDH recommande la reconnaissance d'un droit de ne pas être dérangé, afin de donner aux utilisateurs le pouvoir de désactiver toutes les fonctions qui attirent l'attention et de leur permettre de choisir d'activer ces fonctions par des moyens simples et facilement accessibles, accompagnés d'un avertissement obligatoire sur les dangers potentiels de l'activation de ces fonctions, offrant ainsi aux consommateurs un véritable choix et une autonomie sans les accabler d'une surcharge d'informations.

Recommandation n° 4. La CNCDH recommande de consacrer une obligation générale de conception équitable par défaut, qui imposerait aux services numériques de concevoir des environnements numériques qui respectent l'autonomie, le pouvoir d'agir, et la capacité à prendre des décisions libres et informées, en référence au contenu des lignes directrices de l'article 28 du Règlement sur les services numériques. Ce principe aurait vocation à s'appliquer *ex ante* à l'architecture des services numériques, plutôt que de reposer *ex post* sur une analyse des interactions en ligne. En particulier, la CNCDH recommande de soumettre les plateformes à une obligation de proposer par défaut un algorithme qui ne repose pas sur les données comportementales et de consacrer un droit au paramétrage au bénéfice des utilisateurs.

Recommandation n° 5. La CNCDH recommande aux autorités françaises d'encourager la Commission européenne à publier des lignes directrices relatives aux bonnes pratiques à mettre en œuvre afin de garantir que la conception des services numériques soit respectueuse des droits humains.

Recommandation n° 5. La CNCDH recommande aux pouvoirs publics d'encourager l'utilisation et de soutenir les applications européennes de réseaux sociaux éthiques, fondées sur des logiciels libres.

Recommandation n° 6. La CNCDH recommande de consacrer, au niveau européen, une obligation de pluralisme algorithmique afin que les utilisateurs puissent accéder à des applications tierces depuis la plateforme ou ajouter des fonctionnalités externes aux interfaces originales et s'éloigner ainsi du modèle sur lequel repose le service, de renforcer le droit à la

portabilité des données et de contraindre les plateformes à permettre de paramétrer les systèmes de recommandation.

Recommandation n° 7. La CNCDH recommande d'adopter des lignes directrices relatives à chacun des risques systémiques figurant aux articles 34 et 35 du Règlement sur les services numériques. Ces lignes directrices doivent préciser la mesure dans laquelle la conception des services numériques, au-delà de la seule conception de leurs systèmes de recommandation, contribuent à la survenance de risques systémiques. En priorité, elles doivent concerner les risques en matière de santé physique et mentale.

Recommandation n° 8. La CNCDH recommande de définir une méthodologie spécifique d'évaluation des risques prévus aux articles 34 et 35 du Règlement sur les services numériques, afin d'harmoniser les rapports produits par les plateformes et moteurs de recherche et d'identifier des mesures de remédiation des risques efficaces et efficientes. Cette méthodologie doit imposer aux plateformes concernées de consulter activement la société civile et le milieu académique au stade de l'évaluation des risques systémiques et de l'identification des mesures de limitation de ces risques.

Recommandation n° 9. La CNCDH recommande de modifier l'article 40 du Règlement sur les services numériques relatif à la procédure d'agrément pour l'accès aux données des plateformes, afin, d'une part, d'alléger cette procédure et, d'autre part, de permettre aux organisations non-gouvernementales de procéder à une demande d'agrément.

Recommandation n° 10. La CNCDH recommande, dans le prolongement des discussions relatives au Règlement sur l'intelligence artificielle, d'interdire les systèmes d'intelligence artificielle qui permettent de générer du matériel pédocriminel et/ou des contenus intimes non-consentis (notamment les applications de nudification).

Recommandation n° 11. La CNCDH recommande d'interdire la conception « sycophantique » des systèmes d'intelligence artificielle conversationnelle consistant à flatter l'utilisateur ou l'utilisatrice, à valider ses opinions ou à lui donner systématiquement raison, ainsi que les fonctionnalités qui comportent, par essence, des risques d'anthropomorphisation. En ce qui concerne les autres systèmes d'intelligence artificielle conversationnelle, il conviendrait de les soumettre à des obligations de transparence renforcées. Cette transparence devrait s'appliquer tant au sujet de l'information fournie à l'utilisateur sur la nature réelle de l'interaction, notamment en rappelant régulièrement qu'il s'agit d'une intelligence artificielle, et en mettant en garde contre les risques d'anthropomorphisation.

Recommandation n° 12. La CNCDH recommande d'imposer aux fournisseurs de systèmes d'intelligence artificielle une obligation de mettre en œuvre toutes mesures de sécurité adéquate (*AI safety*) dont des mesures barrières (*guardrails*) visant à bloquer toute interaction toxique. Leur robustesse doit être assurée par le recours à un audit tiers indépendant, ainsi que l'obligation de mettre en œuvre un mécanisme de signalement efficient afin d'identifier tout contournement éventuel (*jailbreak*). Il conviendrait de rendre compte de l'ensemble de ces garde-fous et de l'évaluation de leur efficacité dans le cadre de la publication d'un rapport public périodique.

Recommandation n° 13. La CNCDH recommande d'imposer la possibilité pour les utilisateurs de refuser d'utiliser les agents conversationnels intégrés au sein des plateformes, de ne pas être incités à les utiliser, d'interdire leur installation par défaut et d'offrir la possibilité de les désinstaller aisément.

Recommandation n° 14. La CNCDH recommande que les dispositions du droit de la consommation soient pleinement applicables aux services numériques tels que les réseaux sociaux et tout autre service intégrant des agents conversationnels.

Recommandation n° 15. La CNCDH recommande, afin de mieux appréhender les risques existants et émergents posés par la conception des services numériques en matière de droits humains, de bâtir un système de financement des autorités de régulation du numérique, du secteur de la santé, de la recherche académique et du fonctionnement des organisations de la société civile et en particulier des signaleurs de confiance. Ce financement reposerait sur la contribution financière des fournisseurs de services numériques, qui, en vertu d'un principe de type « pollueur/payeur », serait proportionnelle au nombre de leurs utilisateurs mensuels.

Recommandation n° 16. La CNCDH recommande aux autorités françaises de soutenir la recherche permettant d'améliorer l'état des connaissances en termes d'effets des services numériques et des systèmes d'intelligence artificielle conversationnelle sur la santé physique et mentale des utilisateurs et utilisatrices, en particulier en fonction de leur âge, de leur genre et d'autres facteurs de vulnérabilité. Ces connaissances doivent être mises à jour régulièrement au regard de l'évolution constante de ces technologies.

Recommandation n° 17. La CNCDH recommande de mettre en place, auprès de l'Arcom, un dispositif de veille agile, alimenté par les organisations de la société civile, afin d'identifier et de répondre au plus vite aux risques résultant des pratiques émergentes sur les réseaux sociaux. Elle recommande également d'augmenter les moyens humains, matériels et financiers de l'Arcom afin de garantir un accompagnement de la mise en œuvre du Règlement sur les services numériques et d'assurer la collaboration des différentes parties prenantes.

Recommandation n° 18. La CNCDH recommande de renforcer la mission de l'Agence de protection des droits fondamentaux de l'Union européenne concernant la protection des droits humains dans l'espace numérique, et d'augmenter ses moyens en conséquence.

Recommandation n° 19. La CNCDH recommande d'identifier une autorité administrative en charge de la surveillance des risques et du contrôle des intelligences artificielles conversationnelles quant au respect des droits humains, et de lui conférer des moyens à la hauteur de ses missions.

Recommandation n° 20. La CNCDH recommande l'adoption d'un plan d'action national sur la formation aux usages du numérique dans le cadre scolaire permettant aux élèves de développer un usage critique et autonome des services numériques, en développant une éducation au numérique transversale et intégrée aux programmes scolaires, et en garantissant la sensibilisation des parents ainsi que la formation adéquate du personnel enseignant aux usages des nouvelles technologies, notamment par l'intervention d'acteurs associatifs.

Recommandation n° 21. La CNCDH recommande de sensibiliser la population générale aux risques induits par les services numériques, et en particulier des outils d'intelligence artificielle, en matière de troubles psychiques et psychiatriques, au moyen notamment de campagnes de prévention adaptées à chaque public, et de l'affichage et de la diffusion d'outils dans les structures pertinentes (écoles, transports en commun, hôpitaux, pharmacies…).

Recommandation n° 22. La CNCDH recommande de renforcer la formation des personnels médicaux et paramédicaux, ainsi que la prévention des effets de santé physique et mentale connus et émergents tirés de l'utilisation des services numériques, en premier lieu après des publics particulièrement vulnérables identifiés. Elle recommande d'inscrire ces actions dans

un cadre général d'amélioration de l'accès aux soins en santé mentale, en particulier auprès des jeunes.

Recommandation n° 23. La CNCDH recommande la création d'un Observatoire national du masculinisme et des radicalisations sexistes, confié au Haut-Conseil à l'égalité entre les femmes et les hommes, et de lui conférer la mission d'engager un dialogue structuré avec les plateformes afin de prévenir les risques de dérives radicales.

Recommandation n° 24. La CNCDH recommande d'introduire l'obligation d'inclure des éléments pédagogiques au sein des conditions générales d'utilisation des services numériques, qui porteraient par exemple sur la définition des contenus toxiques ou dangereux ou encore sur les effets de spirale.

Recommandation n° 25. La CNCDH recommande de garantir que les enfants soient acteurs de la protection de leurs droits, en créant, auprès de toute autorité compétente, un comité des jeunes utilisateurs et utilisatrices des réseaux sociaux et outils d'intelligence artificielle conversationnelle. Celui-ci aurait notamment vocation à assurer l'implication des jeunes lors des consultations aux niveaux national et européen.

Annexes

Annexe 1. Remerciements

La CNCDH tient à remercier les étudiantes et étudiants du Master II Droit de la création et du numérique de Paris 1 Panthéon-Sorbonne, promotion 2025/2026, pour leur assistance dans l'organisation et la retranscription des auditions réalisées, ainsi que Margaux Debosque-Trubert, doctorante à l'Université Paris 1 Panthéon-Sorbonne, pour son aide précieuse dans la préparation de cet avis.

La CNCDH adresse également ses remerciements aux associations Square et Après l'école, ainsi qu'aux élèves des collèges de Mantes-la-Jolie, des Mureaux et de Bobigny, dont la rencontre a nourri les réflexions du groupe de travail.

Annexe 2. Liste des personnes auditionnées

Prabhat AGARWAL – Chef d'unité à la Direction générale des réseaux de communication, du contenu et des technologies de la Commission européenne, 24 novembre 2025.

Alice APOSTOLY – Chargée de plaidoyer au sein de l'Association Féministes contre le cyberharcèlement, 8 janvier 2026.

Nacéra BEKHAT – Cheffe de service de l'économie des services à la Commission nationale de l'informatique et des libertés (CNIL), 8 décembre 2025.

Véronique BÉCHU – Directrice de l'Observatoire des violences numériques faites aux mineurs au sein de l'association e-Enfance, 3 novembre 2025.

Elisa BORRY-ESTRADE – Responsable des affaires publiques, Meta, 28 janvier 2026.

Sarah BOUCHAHOUA – Responsable des affaires publiques, Snapchat France, 28 janvier 2026.

Laure BOUTRON-MARMION – Avocate, fondatrice du collectif *Algos Victima*, 17 décembre 2025.

Gaultier BRAND-GAZEAU – Directeur des affaires publiques et gouvernementales, TikTok France, 28 janvier 2026.

Johanna BROUSSE – Vice-procureure, cheffe de la section « J3 » de la JUNALCO au Parquet de Paris, en charge de la lutte contre la cybercriminalité, 12 janvier 2025.

Jean CATTAN – Ancien Secrétaire général du Conseil national du numérique et ancien conseiller du président de l'Arcep, 16 octobre 2025.

David CHAVALARIAS – Mathématicien, directeur de recherche au CNRS et au Centre d'analyse et de mathématique sociales (CAMS) de l'EHESS.

Samuel COMBLEZ – Psychologue, Directeur adjoint de l'association e-Enfance, 3 novembre 2025.

Alice DARMON – Juriste au service de l'économie des services, en charge du marketing digital et de la réglementation des plateformes à la Commission nationale de l'information et des libertés (CNIL), 8 décembre 2025.

Simona DE HEER – Assistante de la députée européenne Kim van Sparrentak, 12 janvier 2025.

Arthur DELAPORTE – Député du Calvados (2e circonscription), président de la Commission d'enquête parlementaire « sur les effets psychologiques de TikTok sur les mineurs », 30 octobre 2025.

Marc FADDOUL – Directeur et cofondateur de l'association *AI Forensics*, 18 novembre 2025.

Alexei GRINBAUM – Président du comité opérationnel d'éthique du numérique du Commissariat à l'énergie atomique et aux énergies alternatives (CEA) et directeur de recherche au CEA-Saclay, membre du Comité national pilote d'éthique du numérique (CNPEN), 18 novembre 2025.

Thibault GUIROY – Directeur des relations institutionnelles de YouTube France, 28 janvier 2026.

Anastasia ILIOPOULOU-PÉNOT – Professeure de droit public à l'Université Paris Panthéon-Assas, 8 janvier 2025.

Bastien LE QUERREC – Chargé des réseaux sociaux et de la régulation des plateformes en ligne au sein de l'association La Quadrature du Net, 8 décembre 2025.

Yann LESCOP – Responsable de projets et études au sein de l'association Point de Contact, 3 novembre 2025.

Margaux LIQUARD – Directrice *Trust & Safety*, Yubo, 17 décembre 2025.

Benoît LOUTREL – Membre du collège de l'Arcom, président du groupe de travail « Plateformes en ligne », 12 janvier 2026.

Florian MARTIN-BARITEAU – Professeur de droit à l'Université d'Ottawa (Canada), 13 mai 2026.

Catherine MORIN-DESAILLY – Sénatrice de la Seine-Maritime, 3 février 2026.

Béatrice OEUVRARD – *AI and Privacy Public Policy Manager*, Meta, 28 janvier 2026.

Laurence PÉCAUT-RIVOLIER – Membre du collège de l'Autorité de régulation de la communication audiovisuelle et numérique (Arcom), présidente du groupe de travail « Protection des publics et diversité de la société française », 13 octobre 2025.

Isabelle PÉRIGNON – Directrice « Consommateurs » au sein de la Direction Générale Justice et Consommateur de la Commission Européenne, 12 janvier 2026.

Katia ROUX – Chargée de plaidoyer « Technologies et droits humains », Amnesty International France, 3 novembre 2025.

Laure SALMONA – Présidente et co-fondatrice de l'association Féministes contre le cyberharcèlement, 8 janvier 2026.

Olivier SYLVAIN – Professeur de droit à *Fordham University*, 18 novembre 2025.

Jonah THOMPSON – EU Policy Manager, *Center for Countering Digital Hate*, 18 novembre 2025.

Capucine TUFFIER – Responsable des affaires publiques en charge de la protection de l'enfance, Meta, 28 janvier 2026.

Arnaud VERGNES – *Government Affairs & Public Policy Manager*, Google, 28 janvier 2026.

Serena VILLATA – Directrice de recherche CNRS au laboratoire d'Informatique, signaux et systèmes de Sophia Antipolis et responsable de l'équipe MARIANE, 8 janvier 2026.

Elodie WEIL – Juriste au sein de la direction de l'accompagnement juridique de la CNIL, 8 décembre 2025.

CNCDH
COMMISSION NATIONALE CONSULTATIVE DES DROITS DE L'HOMME
RÉPUBLIQUE FRANÇAISE

Créée en 1947 sous l'impulsion de René Cassin, la Commission nationale consultative des droits de l'homme (CNCDH) est l'Institution nationale française de promotion et de protection des droits de l'homme, accréditée auprès des Nations unies.

L'action de la CNCDH s'inscrit dans une triple mission :

- Conseiller les pouvoirs publics en matière de droits de l'Homme et de droit international humanitaire ;
- Contrôler l'effectivité des engagements de la France en la matière ;
- Sensibiliser et éduquer aux droits humains.

L'indépendance de la CNCDH est consacrée par la loi. Son fonctionnement s'appuie sur le principe du pluralisme des idées.

Ainsi, seule institution assurant un dialogue continue entre la société civile et les experts français en matière de droits de l'Homme et de droit international humanitaire, elle est composée de 64 personnalités qualifiées et représentants d'organisations non gouvernementales issues de la société civile.

La CNCDH est le rapporteur national indépendant sur la lutte contre toutes les formes de racisme depuis 1990, sur la lutte contre la traite et l'exploitation des êtres humains depuis 2014, sur la mise en œuvre des Principes directeurs des Nations unies sur les entreprises et les droits de l'Homme depuis 2017, sur la lutte contre la haine et les discriminations anti-LGBTI depuis avril 2018 et sur l'effectivité des droits des personnes handicapées depuis 2020.

La CNCDH est en outre la Commission française de mise en œuvre du droit international humanitaire au sens du Comité international de la Croix-Rouge (CICR).