Mykaïa (France) - Cartooning for Peace

Rapport du Groupe de travail visant à

Réaffirmer la liberté d'expression

dans l'espace numérique

HIVER 2026

la villa numeris

Sommaire

Notre groupe de travail

Les travaux de notre groupe de travail sont menés sous la direction de **Thaima Samman**, avocate, fondatrice du Cabinet Samman, avec la contribution de **Marion Boige**, **François Lhémery** et **Benjamin de Vanssay.**

avec la participation et la contribution de nombreuses personnalités que nous remercions chaleureusement :

- **Justine Atlan**, directrice générale Association e-Enfance/3018
- **Pascal Beauvais** agrégé des facultés de droit, professeur de droit privé et sciences criminelles à l'Ecole de droit de la Sorbonne - Université Paris 1, avocat à la Cour
- **Valérie-Laure Benabou**, professeure d'université droit de la propriété intellectuelle, droit européen et diverses branches du droit du numérique à Paris-Saclay
- **Mathias Blandin**, fondateur et PDG de Semiologic
- **Bruno Breton**, fondateur et PDG de Bloom
- **Agathe Cagé**, associée-cofondatrice de Compass Label
- **Monsieur Kak**, dessinateur, président de Cartooning for Peace
- **David Lacombled**, président de La villa numeris
- **Nathalie Laneret**, vice-présidente des Affaires Publiques et Gouvernementales de Criteo
- **Giuseppe de Martino**, co-fondateur et président de Loopsider
- **Barbara Moyersoen**, déléguée générale de Cartooning for peace
- **Rachel-Flore Pardo**, avocate au Cabinet Oplus
- **Arnaud Robert**, secrétaire Général du Groupe Hachette Livre
- **Farah Safi**, professeur agrégée de droit privé et de sciences criminelles
- **Maxime Seno**, avocat, associé en charge de la pratique Droit public économique du Cabinet Veil Jourde
- **Dominique Sopo**, président de SOS Racisme
- **Père Laurent Stalla-Bourdillon**, directeur du Service pour les Professionnels de l'Information (S.P.I) du diocèse de Paris
- **Benoît Tabaka**, secrétaire général de Google France
- **Corinne Thiérache**, avocate associée au Cabinet Alerion Avocats

Merci à **Arthur Brodmann** pour sa relecture attentive

Bâtir un espace d'expression libre, sûr et exigeant

par Thaima Samman, avocate, président du groupe de travail visant à «Réaffirmer la liberté d'expression» de La villa numeris

Pierre angulaire de la démocratie, la liberté d'expression, est une liberté fondamentale de premier rang. Comme toute liberté fondamentale elle n'est pas sans contraintes, strictement encadrées par les principes de légalité, de nécessité et de proportionnalité.

Elle est aujourd'hui mise à l'épreuve des mutations technologiques et de la massification des échanges. Le numérique a libéré des voix, mais il a aussi amplifié ses fragilités : augmentation des discours haineux, désinformation, harcèlement coordonné, montée en puissance d'acteurs privés et présence croissante de systèmes non humains dans l'espace public.

Comment préserver le principe de l'expression libre, exigeant et pluraliste dans un environnement où l'abondance peut parfois étouffer le discernement ?

Cette nouvelle donne pose une exigence : il ne peut y avoir **de droit à la paresse** dans la recherche du modus vivendi d'un monde désirable qui garantisse la sacralité de la liberté d'expression.

A l'heure où l'Europe redéfinit ses normes, où les plateformes recomposent les frontières du dicible, où l'intelligence artificielle redistribue

les capacités d'expression comme celles de manipulation, nous devons regarder ces enjeux avec lucidité et ambition.

C'est dans cet esprit que nos travaux menés ces deux dernières années ont conduit à l'élaboration de **sept recommandations.** Leur ambition est simple : faire en sorte que la liberté demeure la règle et la restriction l'exception.

Déjà, rappelons un principe fondamental : **la liberté d'expression n'existe pas pour les robots**, ce droit est un droit humain et les systèmes automatisés ne peuvent en être les sujets. Par conséquent le principe de liberté qui veut que toute exception engage des

procédures qui peuvent être longues et complexes ne s'applique pas.

Nous invitons ensuite à un constat de lucidité: il est temps de faire une pause législative, **nous disposons déjà d'un cadre juridique robuste**. Le défi n'est pas de créer de nouvelles lois, mais de se donner la peine de maitriser et d'apprendre à faire appliquer le droit existant aux usages contemporains, ce qui nous permettra aussi d'identifier les bonnes mesures complémentaires à prendre plutôt que d'empiler des règles qui ne sont jamais appliquées.

Le **renforcement des moyens de la justice**, l'arbitre ultime des conflits sur la liberté d'expression, nous semble également indispensable.

Mais les tribunaux ne peuvent plus être la seule réponse. La responsabilité et l'action doivent s'élargir, d'abord à **des personnes de confiance** pour aider au signalement de contenus illégaux mais aussi indésirables.

Ensuite, aux côtés du juge et afin de répondre à la masse de contentieux potentiels, nous proposons la création d'une **structure consultative**. Animée par des professionnels du droit à titre bénévole, capable d'éclairer rapidement la décision des plateformes de retrait ou de maintien du contenu, et de les prémunir, autant que faire se peut, à la fois contre le risque de diffusion de contenus illégaux et celui d'autocensure.

Aucune amélioration ne sera bien sur possible sans le **renforcement de l'éducation au numérique et à l'esprit critique**, indispensable pour distinguer l'information, la manipulation et l'artefact.

Enfin, **les technologies ne sont pas des ennemis mais doivent, au contraire, être mises au service de la liberté d'expression**. Les outils d'observation, de modération et de marquage, lorsqu'ils sont transparents et encadrés, contribuent à protéger l'espace public et à restaurer la confiance.

Ces recommandations dessinent une ambition : garantir la liberté d'expression sans naïveté, fondée sur la justice, la connaissance, la maîtrise des technologies et la coopération entre institutions, plateformes et société civile.

Dans un monde où la parole circule plus vite que la loi, il nous revient collectivement de bâtir un espace d'expression libre, sûr et exigeant. La tâche est immense, mais elle est à notre portée — à condition de conjuguer lucidité, ambition et innovation. ❐

Avant-propos

Pourquoi un nouveau rapport sur la liberté d'expression à l'ère numérique ?

La révolution numérique ne cesse de transformer nos sociétés en profondeur, en changeant la manière dont les citoyens se forgent des opinions et les expriment dans la sphère publique. Elle a offert à chacun la possibilité de s'exprimer librement sur les réseaux sociaux sans intermédiation. Ces nouveaux espaces de liberté ont vu en parallèle émerger des effets collatéraux non souhaitables de désinformation, discours haineux et manipulation de l'opinion. Face au rythme soutenu de l'innovation et aux transformations géopolitiques du monde, **il est fondamental de continuellement s'interroger sur la manière de maintenir un cadre solide et protecteur où chaque individu a le droit de s'exprimer librement**. Ce cadre doit cependant prendre en compte d'une part **les limites imposées par la loi**, d'autre part le changement de paradigme d'une expression auparavant limitée par des moyens restreints et contrôlés au niveau national à **une expression débridée portée par des outils qui ne connaissent pas de frontières**.

Loin d'évacuer le débat sur le cadre législatif de la liberté d'expression, l'adoption de **l'Acte sur les services numériques** européen (DSA) **renouvelle en profondeur le débat** sur des sujets aussi cruciaux que le rôle de l'ordre judiciaire dans la protection de cette liberté, la place des plateformes dans la modération des contenus ou encore les marges de manœuvre des États membres dans la définition des limites de la liberté d'expression. Le DSA est un texte européen déterminant les règles de responsabilité pour les intermédiaires en ligne et encadrant la gestion des contenus par ces intermédiaires.

Par sa **démarche originale et inédite**, le présent rapport vise à **rappeler les fondamentaux de la liberté d'expression et formuler des propositions concrètes** afin de répondre aux nouveaux enjeux de protection de cette liberté fondamentale, tels que le traitement des contentieux de masse ou les problématiques liées à l'intelligence artificielle (IA).

Pourquoi est-il nécessaire de « réaffirmer la liberté d'expression ? »

La révolution numérique a un effet paradoxal sur la liberté d'expression. Elle a d'abord largement favorisé **l'épanouissement de l'expression** en démocratisant l'accès aux technologies de l'information et de la communication. Cela a permis à de

nombreuses personnes ou groupes sociaux jusque-là exclus des canaux de communication traditionnels de participer au débat public. L'expression publique n'est en effet plus le monopole d'un nombre limité de médias et de professionnels. Tout individu peut, dans nos sociétés démocratiques, facilement, créer et diffuser du contenu et ainsi contribuer à alimenter des sources diversifiées d'information de « pair à pair ».

Ce mouvement s'est néanmoins accompagné d'une **remise en cause profonde du principe même de liberté d'expression**. La démocratisation des technologies de l'information et de la communication permet en effet la diffusion virale de contenus haineux (provocation à la violence, injure, diffamation), visant des individus ou des communautés humaines, de mener des campagnes visant à censurer de fait l'expression d'idées et d'opinions pourtant légales, propager de la désinformation, diffuser des contenus illégaux à grande échelle ou basculer dans le harcèlement.

Alors que nos sociétés démocratiques et nos gouvernements semblent toujours plus déstabilisés par ces phénomènes, le groupe de travail partage l'intime conviction qu'il appartient de traiter les problématiques actuelles en **réaffirmant certains principes fondamentaux** et en les prenant comme points de départ pour apporter des solutions innovantes.

Quelle est la démarche et l'originalité du groupe de travail ?

Le **groupe de travail « Réaffirmer la liberté d'expression à l'ère du numérique »** a pour objectif de repenser la mise en œuvre du principe de liberté d'expression et de formuler une série de recommandations visant à garantir l'épanouissement souhaitable de cette liberté, tout en corrigeant les effets et impacts, parfois dévastateurs, pouvant remettre en cause d'autres droits fondamentaux ou l'ordre public des sociétés démocratiques.

Il est composé de vingt-deux membres, personnalités reconnues du numérique et/ou de la liberté d'expression, universitaires, chefs d'entreprises innovantes, avocats, ingénieurs, dirigeants d'associations de défense des libertés et/ou de populations ou communautés cibles de violences et vise à protéger, plateformes en ligne, éditeurs, artistes, représentants de religion.

Une trentaine d'experts de premier plan (représentants des pouvoirs publics, plateformes technologiques, prestataires de contenu, entreprises de la tech, universitaires, avocats…) ont été auditionnés afin de participer aux réflexions et d'apporter un éclairage différent sur certains problématiques spécifiques. Le présent rapport a été précédé d'un rapport intermédiaire, publié par La villa numeris en octobre 2024. ❐

Essentiel

Nos recommandations en un coup d'oeil

Recommandation #1

Nul besoin de légiférer davantage à ce stade : Tous les grands principes soutenant l'exercice de la liberté d'expression restent d'actualité. Il importe de maintenir un cadre juridique principiel et évolutif dans lequel la liberté demeure la norme et les restrictions strictement encadrées par la loi.

Les problématiques liées au numérique n'appellent pas nécessairement de nouvelles réformes législatives, mais à repenser les moyens de mettre en œuvre le cadre existant, en particulier lorsqu'il s'agit de traiter un contentieux de masse.

Recommandation #2 :

Permettre à la justice de jouer son rôle : Arbitre de la limite de la liberté d'expression, la justice judiciaire et administrative doit être dotée des moyens de ces enjeux, d'ordre financier, de recrutement et de formation au fonctionnement du numérique. On ne pourra pas faire l'économie d'une compréhension du rôle des différents acteurs et d'une maîtrise du fonctionnement des outils numériques par les magistrats, afin qu'ils puissent traiter ces dossiers sereinement et dans un délai qui fait sens au regard de leurs impacts.

Il est indispensable que la justice pénale joue son rôle lorsque des atteintes, notamment aux personnes, entrent dans son champ de compétence. Les circulaires à destination des parquets doivent donner priorité aux poursuites des auteurs d'infractions sur Internet notamment lorsque la masse des publications identiques ou similaires constitue clairement un harcèlement, ou met en danger des mineurs par du contenu pornographique trop facilement accessible, ou des communautés par des propos racistes, antisémites ou sexistes. La société doit aussi réaffirmer ses valeurs par son bras répressif.

Recommandation #3

Renforcer le dialogue entre des acteurs de la société civile à l'expertise reconnue sur certains sujets (ou signaleurs de confiance), l'État et les plateformes pour accélérer le retrait des contenus problématiques. Un cadre réglementaire clair, fondé sur l'Acte pour les services numériques (DSA), devrait être instauré afin de baliser l'action de ces signaleurs de confiance et s'assurer de leur indépendance.

Recommandation #4

Constituer une structure consultative, composée de juristes bénévoles, permettant aux plateformes d'obtenir rapidement un avis circonstancié sur la légalité d'un contenu.

Recommandation #5

Redoubler d'efforts sur l'éducation aux technologies, aux médias, à l'argumentation rationnelle et à l'esprit critique.

Recommandation #6

Encourager et faciliter le recours aux nouvelles technologies pour contrer certains effets négatifs, tels que les outils d'authentification et de traçabilité des contenus.

Recommandation #7

Accroître le déploiement d'outils de détection des bots et encourager les études visant à leur perfectionnement.

En résumé

Les travaux du groupe de travail ont fait émerger un constat de départ clair : le cadre juridique actuel, établi en droit national par la loi de 1881 sur la liberté de la presse et complété par une multitude de textes ainsi qu'un corpus extensif de jurisprudence, est amplement suffisant, notamment pour circonscrire les limites à la liberté d'expression.

Le cadre juridique européen, garanti la Convention européenne des droits de l'homme et l'Acte sur les services numériques européen (DSA) apparaît également équilibré. Aussi, Le principe selon lequel la liberté d'expression est la règle et les exceptions sont établies par la loi et interprétées strictement doit demeurer le fondement de notre société démocratique.

Dans la poursuite de ses travaux, le groupe de travail s'est penché sur de nombreux sujets, du traitement du blasphème à la propagation de désinformation, et s'est attaché à distinguer les enjeux propres aux nouvelles technologies de ceux liés à l'évolution de la société dans son ensemble afin de formuler des propositions de remèdes adéquats.

Le principal défi posé par le numérique à la liberté d'expression réside dans l'ampleur de l'expression et de la production de contenus, rendue possible par la facilité et l'omniprésence des moyens d'expression. Aussi le groupe de travail a-t-il acquis la conviction que le numérique n'appelle pas forcément à de nouvelles règles, mais à **repenser en profondeur les moyens de les mettre en œuvre au regard du phénomène de masse**.

Autrefois principale gardienne des libertés fondamentales, l'autorité judiciaire ne peut raisonnablement plus assumer ce rôle à elle seule, même renforcée dans ses moyens. Ce rôle ne peut pas non plus être délégué aux seules plateformes, pour des raisons évidentes, ces dernières n'étant pas garantes des libertés individuelles en général et de la liberté d'expression en particulier.

Le groupe de travail considère dès lors que la mise en réseau permise par les technologies numériques appelle à des structures complémentaires de **régulation horizontale** où les interactions entre utilisateurs, plateformes et corps intermédiaires interagissent pour traiter rapidement les premières étapes des conflits d'expression. Ce schéma permettrait de soulager les tribunaux et de ne pas donner un pouvoir de censure ou de surveillance aux plateformes qui ne leur appartient pas.

Le groupe de travail constate que la législation européenne conduit à une forme de **déjudiciarisation** du contentieux de la liberté d'expression et ce faisant **réforme en profondeur les obligations de modération des contenus et la gestion des conflits**.

Face à la massification des litiges sur les contenus numériques et à l'impossibilité matérielle et technologique, pour la justice, d'y répondre, le groupe de travail propose

d'enrichir et perfectionner ce mouvement de déjudiciarisation en renforçant l'intervention de **corps intermédiaires** en lesquels la société peut fonder sa confiance pour traiter des sujets qui la concerne, tels que les signaleurs de confiance. Par leur expertise et leur expérience du terrain, ces tiers pourraient permettre d'aider à la prise de décision dans la modération des contenus en ligne et d'en renforcer la légitimité. Dans la même perspective, le groupe de travail propose enfin **la mise en place d'une structure consultative, composée de juristes bénévoles et de membres de la société civile qualifiés**, permettant aux plateformes d'obtenir rapidement un avis circonstancié sur la légalité d'un contenu**.**

Enfin, à l'ère du numérique, si les technologies sont à l'origine de nombreuses problématiques au regard de la liberté d'expression, elles font tout autant partie de la solution, pour peu que leur utilisation soit transparente, encadrée dans un espace démocratique et que l'humain en garde la maîtrise. Certains outils technologiques de modération ou de veille des réseaux peuvent ainsi permettre de lutter efficacement et à l'échelle appropriée contre des campagnes de haine et de désinformation ou pour détecter des contenus générés ou manipulés par l'intelligence artificielle. ❐

Partie 1

Nul besoin de nouvelles lois : un cadre législatif national et européen complet et évolutif

Recommandation #1

Nul besoin de légiférer davantage à ce stade : Tous les grands principes soutenant l'exercice de la liberté d'expression restent d'actualité. Il importe de maintenir un cadre juridique principiel et évolutif dans lequel la liberté demeure la norme et les restrictions strictement encadrées par la loi.

Les problématiques liées au numérique n'appellent pas nécessairement de nouvelles réformes législatives, mais à repenser les moyens de mettre en œuvre le cadre existant, en particulier lorsqu'il s'agit de traiter un contentieux de masse.

Le maintien d'un cadre juridique national qui a fait ses preuves

La liberté d'expression est un élément central de notre démocratie, garantissant que toute opinion, même celle qui pourrait choquer, puisse être exprimée librement. Cette liberté s'applique par tous les moyens d'expression et canaux de communication, qu'il s'agisse de la presse traditionnelle ou des plateformes en ligne.

Les limites à la liberté d'expression : Maintenir le cadre établi par la loi de 1881 sur la liberté de la presse

Le droit européen et le droit français ont, de longue date, construit un régime juridique hautement protecteur de la liberté d'expression. Il ne s'agit donc pas de remettre en cause cette construction juridique libérale mais de l'adapter et la compléter à l'ère de la communication numérique généralisée et surtout de se donner les moyens de la faire respecter.

Comme pour toute liberté fondamentale, les **limites à la liberté d'expression ne sont admissibles que dans la mesure où elles sont prévues et strictement encadrées par la loi**. Les restrictions à la liberté d'expression se fondent, le plus souvent, sur l'impact potentiel ou réel des contenus et opinions exprimés sur les personnes concernées, plutôt que sur l'acceptabilité du contenu lui-même, dans une approche conséquentialiste basée sur le principe de « non-nuisance » à autrui.

Les limites prévues par la loi peuvent être classées en trois catégories : Celles relevant de la protection de **l'intégrité physique ou morale des personnes** (diffamation, injure publique, incitation à la haine et à la violence, etc.), de la **protection de la vérité** (désinformation) ou de la **protection des droits d'autrui** (exposition de la vie privée, violation de la propriété intellectuelle, violation du secret de la correspondance, etc.).
Outre le fait qu'elles doivent être explicitement prévues par la loi, les restrictions à la liberté d'expression doivent répondre à deux conditions supplémentaires pour être valides. Elles doivent tout d'abord être **justifiées** par un but légitime tel que la défense de l'ordre public, de la sécurité nationale, la protection des droits d'autrui, ou encore la protection de la morale. Elles doivent enfin être « **nécessaires** dans une société démocratique », c'est-à-dire **proportionnées**, entre « les restrictions imposées à la liberté d'expression (…) et le but légitime poursuivi »[1].

En France, la **loi du 29 juillet 1881**[2] organise la liberté de la presse et énonce précisément ses contraintes. Elle prévoit ainsi plusieurs infractions d'expression, telles que l'injure, la diffamation, l'apologie des crimes, le négationnisme, les cris et chants séditieux ou encore la provocation à la discrimination. Le **code pénal** sanctionne également le délit d'outrage, l'apologie du terrorisme ou encore le harcèlement. Ces dispositions visent à protéger la société contre les discours qui incitent à la violence, à la haine, ou qui portent atteinte à la dignité des individus.

Le groupe de travail est **particulièrement**

attaché au principe de la liberté d'expression et à l'interprétation stricte des restrictions prévues par la loi. Malgré son âge avancé, le cadre instauré par la loi du 29 juillet 1881, qui intègre aujourd'hui les contenus en ligne, fait l'objet d'un consensus démocratique et social précieux au regard de la complexité des enjeux et des défis actuels de la liberté d'expression. La force de ce cadre juridique réside dans la constitution au fil du temps d'un corpus jurisprudentiel à la hauteur de l'enjeu et qui continue de faire ses preuves à l'ère des médias numériques.

À ce titre, le groupe de travail s'inquiète d'une **remise en cause profonde du principe de liberté d'expression, constatée à plusieurs reprises lors des travaux et auditions**. Si les contenus haineux (provocation à la violence, injure diffamation) ou attentatoires aux droits fondamentaux d'autrui doivent être modérés ou sanctionnés, le groupe de travail alerte également contre une tendance selon laquelle tout contenu qui heurterait la sensibilité de certaines personnes, par exemple en raison de leurs convictions politiques ou religieuse, ne saurait relever de l'exercice souhaitable de la liberté d'expression. La polarisation et la radicalisation des débats publics sur les réseaux sociaux produit des injonctions d'exclusion - la mise à l'index de personnes privées par d'autres personnes privées - marginalise les contenus argumentés, nuancés et respectueux et, ce faisant, engendre des effets dissuasifs sur l'exercice de la liberté d'expression. Cette tendance se traduit notamment par une remise en cause du droit à la critique et à l'humour, qui comprend la parodie et la satire, comme formes de liberté d'expression.

Ces coups de boutoirs contre le principe de liberté se traduisent le plus souvent par la **proposition de textes visant à modifier, compléter ou contourner la loi de 1881 en réponse à des faits d'actualité**[3]. Au-delà de leur objet immédiat, ces initiatives posent surtout problème en raison de la multiplication de dispositifs dérogatoires ou sectoriels, sans étude d'impact préalable ou d'examen des carences sur la mise en œuvre du cadre existant.

Le groupe de travail insiste sur le fait que **la liberté d'expression vaut également pour ce qui choque, dérange, heurte les sensibilités, ou ce qui est inexact dans la mesure où les propos ne sont pas contraires à la loi**. La démocratie ne peut fonctionner qu'au travers d'un débat contradictoire, dans lequel chacun doit rester libre d'exprimer ses opinions, même si elles ne font pas consensus.

Repenser le contrôle des limites à la liberté d'expression à l'ère numérique

Le groupe de travail considère qu'il est essentiel de maintenir le rôle central du juge comme premier garant de la liberté d'expression. Même renforcé dans ses moyens, l'ordre judiciaire ne peut traiter à lui seul l'ampleur du contentieux lié à la liberté d'expression. Cela appelle donc à repenser la mise en œuvre du droit dans les marges de manœuvre laissées par les législations nationales et européennes.

Le **juge judiciaire est en effet le premier garant de la liberté d'expression, avec le juge administratif, qui a acquis des compétences au cours de ces dernières années**. C'est donc au juge que revient d'appliquer et d'interpréter strictement les limites, fixées par la loi, à la liberté d'expression sur les fondements du code pénal et de la loi de 1881. À travers une

appréciation au cas par cas, il veille également à l'équilibre des différents intérêts en présence en appliquant les critères de nécessité et de proportionnalité au regard des circonstances propres à chaque cas.

Au-delà de l'échelon national, les **normes européennes** et leur mise en œuvre par différentes juridictions, la Cour de Justice de l'Union européenne (CJUE) et la Cour européenne des droits de l'homme (CEDH) ont enrichi et précisé le corpus national.

Le groupe de travail estime que, dans le cadre de la loi, **le juge doit conserver son rôle d'arbitre ultime**. L'application de textes à la portée générale et parfois imprécise laisse inévitablement une certaine marge d'interprétation. Si cette latitude peut conduire à des divergences jurisprudentielles importantes, notamment autour du critère central de l'intérêt général, elle constitue aussi le moteur qui permet au droit et à la jurisprudence d'évoluer en phase avec les transformations de la société.

À ce titre, le groupe de travail observe que **la jurisprudence se fait de plus en plus protectrice de l'expression sur certains sujets de société**. À titre d'exemple, la jurisprudence de la Cour européenne des droits de l'homme a intégré de nombreux sujets dans le critère transversal du « débat d'intérêt général » tels que l'écologie, la polémique syndicale, la religion, la politique, la dénonciation de délits ou de crimes, favorisant une expression plus libre sur des sujets considérés comme importants. La parole des femmes est également beaucoup plus protégée depuis le mouvement #MeToo et les arrêts de la Cour de cassation rendus notamment dans l'affaire «#Balance ton Porc».

Le groupe de travail **s'inquiète toutefois d'un mouvement jurisprudentiel d'extension des limitations à la liberté d'expression**, justifié par la protection d'autres droits. Ainsi, la jurisprudence consacre une vision de plus en plus étendue de l'intégrité morale des personnes et tend également à privilégier la protection de la vie privée, des données personnelles et du secret professionnel à la liberté d'expression. Si le renforcement de la lutte contre les fausses informations est un objectif légitime, il ne doit pas être le fondement à des atteintes substantielle ou disproportionnée à la liberté d'expression.

Plus largement, le groupe de travail s'inquiète du **manque de moyens accordés à la justice**, lequel **nuit gravement à la mise en œuvre de certaines lois clés à l'ère numérique** et **ne permet clairement pas d'appréhender l'ampleur du contentieux** né des contestations sur les plateformes numériques. L'arsenal juridique visant à obliger les plateformes diffusant du contenu pornographique à vérifier l'âge des utilisateurs reste par exemple lettre morte en raison de l'incapacité des tribunaux à agir rapidement et faire face à une multitude de recours dilatoires.

Plus accessoirement, le groupe de travail est préoccupé par la multiplication de **procédures judiciaires dites « bâillon »**, qui consistent à attaquer des cibles sur le fondement de la loi de 1881 ou du code pénal afin de les décourager d'exprimer certaines opinions. Le groupe de travail se félicite de l'adoption, sur le sujet, de la directive européenne du 11 avril 2024[4] qui instaure un cadre protecteur contre ces « procédures abusives ».

[1] Selon les termes de la jurisprudence constante de la Cour européenne des droits de l'homme, *voy. not*.n° 51279/99, 25 juin 2002, Colombani et autres c. France

[2] Loi du 29 juillet 1881 sur la liberté de la presse.

[3] Par exemple, Article 2 bis de la Proposition de loi renforçant la sécurité des élus locaux et des maires, adoptée par le Sénat en première lecture le 10 octobre 2023. Cet article fut abandonné lors de l'examen du texte à l'Assemblée nationale et n'apparaît donc pas dans la version finale de la loi du 21 mars 2024 , décision n° 2021-817 DC du 20 mai 2021 censurant une disposition prévoyant de pénaliser, au sein de la loi du 29 juillet 1881, la diffusion d'images permettant d'identifier des policiers ou gendarmes en intervention, y compris par des journalistes ou la décision n° 2020-801 DC du 18 juin 2020 du Conseil constitutionnel censurant de multiples dispositions de la loi visant à lutter contre les contenus haineux sur Internet (loi dite « Avia »).

[4] Directive 2024/1069 du 11 avril 2024 sur la protection des personnes qui participent au débat public contre les demandes en justice manifestement infondées ou les procédures judiciaires abusives (« poursuites stratégiques altérant le débat public »)

Recommandation #2 :

Permettre à la justice de jouer son rôle : Arbitre de la limite de la liberté d'expression, la justice judiciaire et administrative doit être dotée des moyens de ces enjeux, d'ordre financier, de recrutement et de formation au fonctionnement du numérique. On ne pourra pas faire l'économie d'une compréhension du rôle des différents acteurs et d'une maîtrise du fonctionnement des outils numériques par les magistrats, afin qu'ils puissent traiter ces dossiers sereinement et dans un délai qui fait sens au regard de leurs impacts.

Il est indispensable que la justice pénale joue son rôle lorsque des atteintes, notamment aux personnes, entrent dans son champ de compétence. Les circulaires à destination des parquets doivent donner priorité aux poursuites des auteurs d'infractions sur Internet notamment lorsque la masse des publications identiques ou similaires constitue clairement un harcèlement, ou met en danger des mineurs par du contenu pornographique trop facilement accessible, ou des communautés par des propos racistes, antisémites ou sexistes. La société doit aussi réaffirmer ses valeurs par son bras répressif.

Renforcer les moyens de la justice afin de faire face aux défis du numérique est certes nécessaire, mais ne suffira pas à traiter l'ampleur du phénomène.

En premier lieu, les tribunaux spécialisés, déjà débordés, **ne seront jamais en mesure d'absorber la masse du contentieux lié à l'utilisation des réseaux sociaux et plateformes numériques**. En outre, la temporalité judiciaire s'avère être en décalage complet avec la rapidité de la diffusion des contenus numériques en ligne.

Le numérique revêt également une **dimension extraterritoriale** constituant trop souvent un obstacle majeur dans la mise en œuvre du droit. La saga judiciaire pour la mise en œuvre de l'obligation de contrôler l'âge des utilisateurs incombant aux plateformes diffusant des contenus pornographiques en est un exemple emblématique.

Dans un tel contexte, de **nouvelles approches coopératives**, telles que développées ci-dessous, impliquant des acteurs de la société civile ou des juristes

spécialisés suffisamment légitimes pour se prononcer sur la légalité de certains contenus, semblent essentielles pour répondre efficacement aux défis posés par la diffusion massive et instantanée des propos sur Internet. Il s'agit bien de compléter l'action des tribunaux et non de les remplacer. En outre, l'Union européenne apparaît désormais une échelle minimum pour protéger et réguler la liberté d'expression.

La mise en œuvre de l'Acte européen pour les services numériques

Adopté en 2021, **l'Acte sur les services numérique européen** (« Digital Services Act » ou DSA)[1] constitue un **texte majeur pour la protection de la liberté d'expression** et un nouveau **jalon dans la transformation du paradigme de la gestion de la liberté d'expression** dans l'espace numérique.

Le texte maintient en bonne partie le cadre juridique préexistant, établi par la directive e-Commerce. Il prévoit ainsi les règles de responsabilité des intermédiaires numériques, y compris les plateformes telles que les réseaux sociaux, par rapport aux contenus publiés par leurs utilisateurs.

Le DSA maintient le principe selon lequel les plateformes ne deviennent responsables pour le contenu publié par leurs utilisateurs qu'à partir du moment où elles ont eu connaissance de leur caractère illégal et qu'elles n'ont pas promptement agi pour le retirer (**principe dit de « responsabilité limitée »**). Élément important et souvent incompris, le DSA organise le retrait des contenus illégaux mais **ne définit en aucun cas ce qui est illégal**. La qualification de l'illégalité relève en effet des lois nationales et européennes. Pour les contenus qui restent dans les bornes de la légalité, les intermédiaires déterminent librement les conditions d'utilisation de leurs services et fixent chacun leur niveau de tolérance par rapport à certains types de contenus, tels que les contenus violents ou explicites.

La nouveauté du DSA tient d'abord au fait qu'il édicte une série de **règles plus précises sur la gestion des contenus par les intermédiaires en ligne**. Ces règles, de nature essentiellement procédurale, consistent par exemple en une obligation de mise à disposition des utilisateurs d'un mécanisme de notification des contenus illégaux (« notice and action mechanism ») ou à fournir un exposé des motifs pour justifier toute action de modération d'un contenu illicite ou contraire à leurs conditions générales d'utilisation.

Le DSA marque en outre une nouvelle étape dans le processus de **déjudiciarisation** du contentieux de la liberté d'expression. En effet, le texte institue un nouvel écosystème de traitement des contenus numériques dans un souci d'adaptation à l'ampleur du phénomène. Cet écosystème vise à combiner le retrait rapide de contenus illégaux ou dangereux, tout en permettant à leurs auteurs les outils pour pouvoir contester les décisions des plateformes.

Il implique notamment la mise en place d'une **procédure « interne » de traitement des**

réclamations, **une procédure « externe » de règlement extra-judiciaire** en cas de litige et **la création d'organes de règlements extrajudiciaires des litiges** certifiés par l'autorité publique en charge de la mise en œuvre du DSA au niveau national (l'Autorité de régulation de la communication audiovisuelle et numérique – Arcom en France). Ces organes sont compétents pour traiter du contentieux relatif aux décisions – ou au silence – des plateformes sur les contenus litigieux. Ils doivent rendre leurs décisions dans un délai raisonnable et, sauf exception, au plus tard 90 jours après la réception de la plainte. Cette procédure ne fait toutefois pas obstacle à ce que l'utilisateur du service concerné puisse engager, à tout moment, une procédure de contestation devant les tribunaux.

Au regard de l'ampleur du contentieux de la liberté d'expression et de l'impossibilité pour l'ordre judiciaire de l'absorber, la mise en place du schéma de régulation du DSA est bienvenue. Ce changement soulève cependant des questions fondamentales en droit français, qui nécessitent une coordination de l'ensemble des acteurs, privés, publics et notamment de la justice judiciaire et administrative. **Le groupe de travail insiste sur le besoin d'en encadrer précisément le fonctionnement et formule dans le présent rapport, une série de propositions visant à accompagner ce mouvement, détaillées ci-dessous**. ❐

[1] 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 (règlement sur les services numériques).

Partie 2

Réaffirmer la liberté d'expression dans l'espace numérique

Le rôle des intermédiaires en ligne dans la gestion de la liberté d'expression

Tout au long de ses travaux, le groupe de travail s'est attaché à identifier des solutions de toute nature permettant de faire progresser la liberté d'expression en ligne conformément aux acquis et principes exposés ci-dessus. Le groupe de travail s'est plus particulièrement penché sur le rôle des intermédiaires en ligne, dont les plateformes numériques tels que les réseaux sociaux, en raison de leur rôle clé dans l'espace numérique actuel.

L'essentiel des réflexions a porté sur la gestion par les intermédiaires en ligne de la **« zone grise » de la liberté d'expression**, c'est-à-dire les contenus qui posent question au regard du droit mais dont l'illégalité ne peut être établie *prima facie*, que ce soit en raison de leur nature, de leur mode,

fréquence ou contexte de diffusion (par exemple la pornographie, la désinformation ou les campagnes de harcèlement).

Certains acteurs technologiques ont témoigné au groupe de travail toute la **difficulté de gérer cette zone grise**. Les plateformes n'ont pas la légitimité démocratique et ne sont en effet pas des juges professionnels : elles ne peuvent que décider de retirer les contenus qu'elles diffusent, à l'aune du cadre légal, de leurs conditions générales d'utilisation et de leur politique interne (qui dépend des actionnaires et dirigeants de l'entreprise). Bien qu'elles ne remplacent pas le juge, le rôle qui leur est attribué affecte largement l'exercice effectif de la liberté d'expression dans notre société.

Ce rôle pose d'ailleurs la question de la **place des règles, procédures et stratégies internes par rapport à celle de la loi**. Quels sont les critères objectifs leur permettant de décider de la nature d'un propos afin de savoir s'il entre ou non dans le cadre de la liberté d'expression ? Dans quelle mesure leurs décisions peuvent être orientées par leur politique interne ou influencées par leurs conditions d'utilisation ?
La question est d'autant plus pertinente que dans nos sociétés démocratiques, les plateformes sont libres de définir leur politique de modération dans le respect du cadre légal des pays dans lesquels elles opèrent. Elles **peuvent donc se positionner sur le terrain de la polémique et la provocation**, ou au contraire, **choisir le consensus social**. Ce choix fondé sur des intérêts économiques est porteur de risque de censure ou d'autocensure.
La liberté d'expression étant un « bien commun », au cœur du contrat social démocratique, il est nécessaire d'**identifier et de traiter les effets et les risques inhérents à cette délégation partielle de compétence** du régalien aux seuls acteurs économiques.
Si la déjudiciarisation encadrée constitue un début de réponse à la massification des contentieux, elle ne répond cependant pas encore, ou à tout le moins en l'état, de manière appropriée, à l'enjeu majeur de la construction d'un régime protecteur et consensuel de la liberté d'expression à l'ère du numérique. Il s'agit en d'autres termes de trouver la **voie médiane entre justice publique et modération des contenus par les acteurs privés**, en maîtrisant démocratiquement la problématique de la délégation aux acteurs économiques.
Pour répondre à cet enjeu, le groupe de travail esquisse plusieurs pistes de solution : faire émerger de nouveaux garants de la liberté d'expression dans une démarche coopérative, mieux utiliser les technologies pour identifier des contenus illicites, davantage former les acteurs clés et la population et enfin clarifier le statut de la parole artificielle au regard de la liberté d'expression.

Faire émerger de nouveaux garants de la liberté d'expression

La gestion des limites à la liberté d'expression doit faire l'objet d'une démarche coopérative, passant notamment par des **corps intermédiaires** en lesquels la société peut fonder sa confiance pour traiter des sujets qui la concernent. Ainsi, de nombreux

tiers permettraient, par leurs avis dans leur champ de compétence, d'aider à la prise de décision dans la modération des contenus en ligne et d'en renforcer la légitimité. Pour que cela fonctionne, leur émergence et/ou leur existence ne saurait être décrétée : le processus pour leur donner corps ne peut être que le fruit d'une légitimité construite dans le temps.

Quelques-uns de ces acteurs existent d'ailleurs déjà : le monde éducatif lorsqu'il met en forme et transmet le savoir, le monde de la recherche lorsqu'il fait émerger des « vérités scientifiques », les entreprises de presse ou des journalistes lorsqu'ils font des reportages de terrain ou réalisent des enquêtes, les Think Tanks quand ils réunissent expertise scientifique et professionnelle, les ordres et organisations professionnelles lorsqu'ils produisent des connaissances et des standards de comportements…

Renforcer le rôle des signaleurs de confiance

Le groupe de travail s'est notamment penché sur le **statut de signaleur de confiance** formalisé par le DSA[1]. Il s'agit d'un statut juridique auquel peuvent prétendre des entités indépendantes disposant d'une expertise et de compétences particulières aux fins de détecter, d'identifier et de notifier des contenus illicites. Afin de bénéficier du statut de signaleur de confiance, ces entités doivent être certifiées par l'autorité nationale en charge de l'application du DSA.

Dans l'architecture du DSA, les signaleurs de confiance ont un rôle de **partenaire privilégié des plateformes pour la modération des contenus qui relèvent de leur champ de compétence**. En effet, lorsque les signaleurs de confiance soumettent des notifications pour contenus illicites aux plateformes, ces dernières sont tenues de traiter les signalements de façon prioritaire et de rendre des décisions, sur le maintien ou non de ces contenus, dans les meilleurs délais.

Le groupe de travail considère que l'action des signaleurs de confiance peut être bénéfique dans certains domaines, **lorsque l'analyse du caractère illicite du contenu nécessite un haut niveau d'expertise et une action de terrain**.

L'action des signaleurs de confiance comme e-Enfance peut ainsi être précieuse dans le domaine de la protection des mineurs, par exemple pour signaler des cas dans lesquels la diffusion d'un contenu qui n'est pas forcément illicite de prime abord peut, par son ampleur ou ses modalités, franchir les bornes de la légalité. Des collectifs soutenant des femmes victimes de violence, tels que StopFisha, peuvent également permettre de signaler des contenus à caractère pornographiques illégaux en raison de leur caractère violent ou du non-consentement des participants.

Ces organisations peuvent également avoir un rôle important dans **l'identification de tendances émergentes problématiques**, comme la circulation croissante de contenus pédocriminels par les mineurs eux-mêmes, l'utilisation d'applications de « nudification » ou encore des « challenges » ou effets de mode problématiques.

[1] Article 22 du Règlement (UE) 2022/2065 relatif à un marché unique des services numériques.

Recommandation #3

Renforcer le dialogue entre des acteurs de la société civile à l'expertise reconnue sur certains sujets (ou signaleurs de confiance), l'État et les plateformes pour accélérer le retrait des contenus problématiques. Un cadre réglementaire clair, fondé sur l'Acte pour les services numériques (DSA), devrait être instauré afin de baliser l'action de ces signaleurs de confiance et s'assurer de leur indépendance.

Le groupe de travail souligne cependant la **nécessité d'encadrer le statut de signaleur de confiance** pour éviter que le système ne soit détourné de ses fins et desserve la liberté d'expression. Conformément au DSA, il est crucial que ces organisations travaillent de manière **experte, indépendante**, **diligente, précise et surtout objective**. Leur expertise ne devrait servir qu'à identifier des contenus illicites et non des contenus relevant de la liberté d'expression bien qu'ils puissent heurter certaines sensibilités. Les autorités nationales en charge de l'application du DSA devraient donc s'assurer que les organisations puissent démontrer leur compréhension du cadre juridique de la liberté d'expression.

Outre les signaleurs de confiance, il s'agira de laisser une place pour des personnalités ou organisations de confiance (chercheurs, experts, enseignants, médecins, journalistes…) pour faire des recommandations à propos de contenus illégaux ou non-souhaitables.

Structure collaborative pour accompagner la déjudiciarisation

Le constat du groupe de travail sur la modération des contenus en ligne est double: d'une part la justice ne peut à elle seule traiter la masse du contentieux, et d'autre part les plateformes n'ont pas forcément les moyens de vérifier tous les signalements. Lorsqu'elles en allouent à des cas d'espèce, la détermination du caractère illicite peut être extrêmement complexe, même dans des cas couverts par la jurisprudence.

Fruit de ses réflexions et échanges avec des experts, le groupe de travail a développé une proposition inédite dans le débat public : **constituer une structure collaborative et consultative afin de soutenir la modération des contenus en ligne par les plateformes**. Cette structure pourrait s'articuler comme suit :

- **Composition** : La structure prendrait la forme d'un vaste réseau d'anciens magistrats, juristes, universitaires ou avocats agissant à titre bénévole et disposant d'une solide expertise sur la qualification des contenus au regard du droit français et européen.

- **Mission** : La principale mission de ce réseau serait de rendre des avis sur la qualification de contenus jugés problématiques au regard du droit dans des délais très courts. Les affaires soumises à cette structure devraient être suffisamment représentatives de problématiques auxquelles sont confrontées les plateformes et qui requièrent une expertise technique pour en déterminer la légalité.

- **Nature des avis** : Les avis seraient non-contraignants. Cependant, la plateforme

serait tenue de se justifier lorsqu'elle décide de ne pas les suivre dans l'exposé des motifs qu'elles sont par ailleurs tenues de produire au titre du DSA[1].

Les avantages que comporterait une telle structure sont nombreux. Pour les plateformes, recourir à une telle structure dans des cas pourrait être vecteur de sécurité juridique et témoigner de leur bonne foi dans des procédures ultérieures lorsqu'elles suivent les recommandations émises.

Pour les citoyens et la société, cette structure agirait comme garante de la protection de la liberté d'expression et permettrait d'apporter rapidement une première réponse dans certains cas urgents et particulièrement représentatifs. Lors de ces travaux, le groupe de travail s'est par exemple interrogé sur le traitement des contenus relatifs à l'hydroxychloroquine durant la pandémie du Covid, tout particulièrement en l'absence de consensus scientifique sur l'efficacité de la molécule.

À titre subsidiaire, la structure pourrait également engager un dialogue avec les parties prenantes, et notamment les acteurs de la société civile sur le niveau de tolérance à adopter par rapport à certains contenus problématiques, basé sur les règles fondamentales.

Plus largement, la proposition pourrait efficacement atténuer le revers du principe de responsabilité limitée dont bénéficient les plateformes au titre du DSA, à savoir un réflexe de censure afin de se prémunir contre d'éventuelles actions en responsabilité par la suite.

[1] Article 17 du Règlement (UE) 2022/2065 relatif à un marché unique des services numériques.

Recommandation #4

Constituer une structure consultative, composée de juristes bénévoles, permettant aux plateformes d'obtenir rapidement un avis circonstancié sur la légalité d'un contenu.

Les besoins de financement d'une telle structure pourraient être relativement limités et ne pas reposer sur les deniers publics. Ils consisteraient essentiellement en la formation des juristes et le maintien de leurs connaissances sur le corpus juridique relatif à la liberté d'expression.

Enfin, son fonctionnement opérationnel pourrait s'inspirer du système d'aide juridictionnelle belge, pionnier européen en matière d'aide juridique. Ce système permet aux justiciables de bénéficier d'un premier avis juridique sur des problématiques de droit dans le cadre de permanences assurées par des avocats expérimentés par un système de permanence et de rotation. Ainsi, les juristes volontaires indiqueraient leur disponibilité pour assurer les permanences de consultations. À l'heure de l'intelligence artificielle, la coordination et la synchronisation assurant la présence de plusieurs de ces experts en temps réel ne devraient pas être très compliquées à mettre en place.

Au-delà du droit : L'humain et la technique en soutien de la liberté d'expression

Le groupe de travail est convaincu que les dangers qui guettent la liberté d'expression relèvent d'abord de l'humain et pas seulement des technologies.

Les plateformes, leur modèle économique et le fonctionnement de leurs algorithmes ne peuvent en effet pas être considérés à eux seuls comme étant à l'origine de tous les maux dont souffre la liberté d'expression, qu'il s'agisse de la profusion de discours de haine en ligne ou de la désinformation. **Chercher à résoudre ces problématiques uniquement d'un point de vue technique, en occultant leur dimension humaine ou démocratique, ne donnera lieu qu'à des solutions qui renforcent le contrôle et la sécurité, au détriment des libertés fondamentales des individus**.

Selon le groupe de travail, **les menaces pesant sur la liberté d'expression sont avant tout le résultat de mutations profondes de la société**. Dès le début de ces travaux, le groupe de travail a été confronté au manque de tolérance de certaines personnes ou groupes sociaux à des contenus qu'ils considéraient comme dérangeants au regard de leurs convictions ou sensibilité, malgré le fait qu'ils restaient dans les bornes de la légalité. Cette évolution dangereuse pour la liberté d'expression s'explique en partie par une transformation plus large des attentes sociales, où la protection contre les discours perçus comme offensants ou perturbateurs est davantage revendiquée.
La problématique s'est également renforcée par les politiques de modération menées par les grandes plateformes, ainsi que par les algorithmes de recommandation, conçus pour proposer des contenus alignés avec les préférences personnelles jusqu'à générer des comportements addictifs, du complotisme à la pornographie en passant par le shopping. Les algorithmes de hiérarchisation des contenus, en personnalisant l'accès à l'information selon des profils d'attention et d'affinité, peuvent enfermer les utilisateurs dans des bulles cognitives réduisant l'exposition à la contradiction et altérant les conditions d'un débat public pluraliste.

Dans un tel contexte, les **technologies numériques** apparaissent d'abord comme **accélérateur et amplificateur** des transformations sociales avant d'être, en tant que telles, la source des problèmes. Leur impact le plus marquant a finalement été de permettre un « excès d'expression ». Cet excès a lui-même favorisé l'émergence de ce que l'on peut qualifier de « chaos informationnel » où toute parole, peu importe sa valeur, est mise au même niveau dans un niveau de profusion qui ne permet plus une expression individuelle authentique.

Aussi, la réaffirmation de la liberté d'expression passe donc autant par l'humain que par la technologie, appelant des réponses différentes et complémentaires.

Former les acteurs clés et la population en général

Dans une démocratie, l'exercice de la liberté d'expression doit être intimement lié à une **recherche du bien commun**, lequel est corrélé à une quête commune de vérité. Cela implique de pouvoir s'exprimer de manière éclairée – c'est-à-dire être **informé**, **critique**, **responsable** et soucieux de **contribuer positivement** au débat collectif.

Au cours de ses travaux, le groupe de travail a identifié deux dangers majeurs pour la

liberté d'expression.

D'une part, une **confusion entre l'accès à l'information et le raisonnement**. Ce n'est pas parce que l'information est largement disponible qu'il n'est plus nécessaire de la traiter proactivement, de la retenir et de se l'approprier. Se constituer une base solide de connaissance et de culture générale est indispensable afin de développer un esprit critique et de se forger des opinions. Le groupe de travail constate que cette confusion s'aggrave avec le développement de l'IA, en raison de la méprise sur la capacité des modèles génératifs à raisonner. Les modèles, même les plus sophistiqués, ne sont que des modèles probabilistes, qui ne raisonnent pas au sens humain du terme étant donné qu'elles ne perçoivent pas le sens du langage.

D'autre part, la surabondance d'information, dépourvue de toute hiérarchisation, couplée au fonctionnement algorithmique des plateformes, crée un brouillard **informationnel dans lequel nous peinons à distinguer le fait de l'opinion, ou le réel de l'artificiel**, risquant de mener à ce que certains sociologues, comme Gérald Bronner, qualifient de «scepticisme opportuniste», une posture qui consiste à croire ou non en fonction de notre intérêt, n'étant plus en mesure de distinguer le réel de l'artificiel. Or, pour que le débat public soit constructif et permette une action collective, il est impératif de s'accorder sur l'intangibilité du fait.

Il n'existe aucune fatalité dans ce constat. Le groupe de travail considère que les technologies peuvent être mises au service de la liberté d'expression, dans un cadre démocratique, pour autant que notre société soit éduquée et sensibilisée à leur usage.

Recommandation #5
Redoubler d'efforts sur l'éducation aux technologies, aux médias, à l'argumentation rationnelle et à l'esprit critique.

Pour le groupe de travail il est absolument primordial de renforcer **l'éducation civique** et la **valeur du fait incontestable,** fondamentaux de notre société. Plus que jamais, il importe de former les futures générations à savoir se poser des questions, à la lenteur de la recherche, de développer une éducation qui enseigne à douter avec rigueur, à s'engager. Cet effort devrait aller de pair avec l'enseignement des principes et fondements de l'argumentation et du dialogue.

Dans ce cadre, l'enseignement devrait également porter sur le **fonctionnement des technologies**, en particulier l'IA générative. Même superficielle, la compréhension du fonctionnement des grands modèles de langage suffit pour inciter les utilisateurs à s'interroger sur la véracité des contenus générés. Il est également essentiel de former et de sensibiliser la population à d'autres aspects cruciaux pour la liberté d'expression, comme les techniques de manipulation, les bulles informationnelles, la propagation de fausses informations, etc.

Une meilleure compréhension du fonctionnement des algorithmes de modération des contenus permettra

également de garantir l'effectivité de certaines obligations de transparence imposées par le DSA aux plateformes en ligne. Cela pourrait participer à ce que les utilisateurs puissent être davantage en maîtrise des contenus auxquels ils sont exposés et puissent contribuer à les hiérarchiser selon d'autres critères que ceux paramétrés par défaut par les plateformes.

Enfin, le groupe de travail considère nécessaire de **responsabiliser les jeunes générations sur leurs activités en ligne**. L'actualité témoigne quotidiennement du manque de compréhension des jeunes quant à l'incidence potentielle d'un partage de contenu sur eux-mêmes comme sur autrui, par exemple alimenter une campagne de harcèlement ou commettre une infraction en partageant des contenus pédopornographiques.

Utiliser les technologies pour protéger notre espace informationnel

Comme le mot « pharmakôn » qui désigne en grec ancien à la fois le poison et le remède, la drogue salutaire et celle malfaisante, la technologie peut être à la fois la source du problème et sa solution.

Ainsi l'IA a considérablement amplifié une série de menaces pour la liberté d'expression, telles que la diffusion de fausses informations ou d'hyper trucages. Or, le meilleur moyen de lutter contre ces problématiques consiste justement à utiliser l'IA, par exemple pour suivre la propagation de campagnes de désinformation ou détecter des contenus artificiels ou manipulés. Plus largement, les plateformes en ligne recourent déjà massivement à l'IA pour filtrer et bloquer des contenus illégaux.

Parmi les différentes technologies présentées au groupe de travail, les **outils d'authentification et de traçabilité des contenus** sont apparus comme un moyen efficace pour marquer le réel et contribuer à rétablir la confiance dans l'espace numérique. Ces outils permettent d'authentifier des informations audiovisuelles, des images et des articles en y apposant des marqueurs indécelables et inaltérables. Ils permettent ainsi de contrôler la dissémination des contenus et de détecter d'éventuelles manipulations. Promue par le législateur européen, une série d'outils émerge également afin de marquer les contenus artificiels, tels que les « signatures des modèles ».

Les différents **outils de veille des réseaux sociaux** peuvent également s'avérer efficaces pour protéger l'espace informationnel, en détectant rapidement des tentatives de déstabilisation étrangères par la propagation de désinformation ou de manipulation du débat public.

L'utilisation de ces outils doit être encouragée pour autant qu'ils demeurent sous la supervision et le contrôle d'êtres humains attentifs aux risques et s'assurant du respect des droits fondamentaux et de la liberté d'expression. Aussi, ces outils ne doivent pas s'autonomiser mais être encadrés par des règles claires pour être au service du régime légal de la liberté d'expression et en traduire technologiquement les subtilités.
Ainsi, une modération excessive ou mal calibrée peut provoquer la **suppression de contenus légaux et légitimes, et porter une atteinte grave à la liberté d'expression**. Les algorithmes n'étant pas des humains, ils peuvent parfois mal interpréter le contexte et supprimer des

contenus conformes à la loi et aux règles de la plateforme qui les publie.

Ces technologies étant susceptibles de porter atteinte à la vie privée et produire des effets dissuasifs sur la liberté d'expression, il est important que leur usage continue à être encadré au titre, notamment, de la législation sur la protection de la vie privée et de l'intelligence artificielle. Les outils de modération basés sur l'IA peuvent également être **biaisés** en fonction des données avec lesquelles ils ont été entraînés, ce qui conduit à une application discriminatoire de leurs usages, affectant de manière disproportionnée certains groupes ou types de discours. L'utilisation de ces technologies doit donc être accompagnée d'une **validation humaine responsable** pour les cas les plus complexes et toujours permettre des recours devant les tribunaux.

Enfin, les algorithmes devraient promouvoir les contenus issus d'acteurs identifiés, de manière consensuelle et/ou partenariale, pour le sérieux de leurs méthodes de vérification des faits et pour leur intégrité (par exemple le monde académique, agences ou entreprises de presse) et, ce faisant, réduire la visibilité des contenus faux et manifestement manipulé.

Recommandation #6

Encourager et faciliter le recours aux nouvelles technologies pour contrer certains effets négatifs, tels que les **outils d'authentification et de traçabilité des contenus**.

Rendre la liberté d'expression aux humains

Parmi les impacts de l'IA sur la liberté d'expression, le groupe de travail s'est largement penché sur **le rôle des « bots » sur les plateformes numériques**. Petits programmes informatiques automatisés conçus pour publier, aimer, commenter ou partager des contenus, leur utilisation est consubstantielle aux réseaux sociaux, souvent à des fins légitimes telles que la publication d'actualités ou de la météo.

La présence des bots sur les plateformes connaît cependant un **essor fulgurant**, aujourd'hui lié au développement de l'IA générative qui facilite grandement la création et l'automatisation de ces programmes. L'IA permet surtout de repousser les limites dans l'imitation des comportements humains. Les bots sont désormais en mesure d'interagir proactivement avec d'autres utilisateurs et de participer à des échanges en apprenant du contexte et de leurs interlocuteurs.

L'utilisation des bots à des fins malveillantes, pour manipuler l'opinion publique, diffuser de la désinformation, gonfler artificiellement la popularité d'un compte ou harceler d'autres utilisateurs, est de plus en plus fréquente et pose une problématique réelle pour l'exercice de la liberté d'expression en ligne.

Le groupe de travail a constaté le consensus clair parmi la doctrine juridique que **le bénéfice de la liberté d'expression ne pouvait être reconnu à des non-humains**. La Convention européenne des droits de l'homme ne garantit en effet le droit à la liberté d'expression qu'aux « personnes ».

Nonobstant, le droit à la liberté d'expression recouvre également la **communication des informations et des idées** – qui peut impliquer l'usage de bots sur les réseaux sociaux. Sur ce point, la jurisprudence européenne et française distingue le droit de s'exprimer des modalités techniques de la diffusion. Aussi, la liberté d'expression recouvre la communication des idées ou de l'information mais ne consacre nullement de droit absolu à la visibilité, ni à une amplification algorithmique, médiatique ou publicitaire en tant que telle. De nombreux textes encadrent ou restreignent l'accès à des moyens de communication, comme la réglementation relative à la publicité politique ou les lois sur la communication audiovisuelle, afin de garantir le pluralisme et l'égalité d'accès au débat public.

En pratique, la plupart des grandes plateformes en ligne mettent en œuvre des politiques dites « anti-spams » qui permettent d'identifier et de retirer rapidement les contenus qui ne sont pas véhiculés par des êtres humains. Cette politique de modération repose essentiellement sur des outils de détection et de filtrage automatisés qui permettent d'identifier les bots sur la base de différents critères (tels que des schémas de comportement suspects) et d'éliminer les contenus indésirables, tels que des faits de violence, des paroles haineuses, etc. ❐

Recommandation #7
Accroître le déploiement d'outils de détection des bots et encourager les études visant à leur perfectionnement.

Annexe 1

Panorama du cadre juridique de la liberté d'expression

Le maintien d'un cadre juridique qui a fait ses preuves

La liberté d'expression est un élément central de notre démocratie, garantissant que toute opinion, même celle qui pourrait choquer, peut être exprimée librement. Cette liberté s'applique pour tous les moyens d'expression et canaux de communication, qu'il s'agisse des médias traditionnels, des plateformes en ligne, de la création littéraire ou artistique, ou de tout autre support.

1. Les limites à la liberté d'expression

1.1. La notion de limites à la liberté d'expression

La liberté d'expression n'est pas sans limite. Le cadre constitutionnel et légal, interprété par la jurisprudence, prévoit des restrictions à la liberté d'expression, comme pour toutes les libertés, dans des circonstances bien identifiées et encadrées.

L'interprétation jurisprudentielle des restrictions à la liberté d'expression se fonde, le plus souvent, sur l'impact potentiel des contenus et opinions exprimés sur des personnes, au-delà de la qualification du contenu lui-même, consacrant une approche conséquentialiste basée sur le principe de « non-nuisance » à autrui.

Les restrictions peuvent être classées selon trois grands objectifs :

- La protection de l'intégrité physique ou morale des personnes (interdiction des agressions morales dégradantes, provocation à la violence, propos haineux, diffamation, délit d'injure publique, etc.) ;
- La protection de la vérité (lutte contre les fausses informations portant atteinte à la réputation d'une personne ou à l'ordre public, lutte contre la fraude et l'escroquerie, lutte contre la diffusion massive de fausses informations) ;
- La protection des droits d'autrui (protection de la vie privée, des droits de propriété intellectuelle, du secret des affaires, du secret des correspondances, du secret de l'instruction, du secret professionnel etc.)

1.2. Les différentes sources juridiques encadrant les limites à la liberté d'expression

1.2.1. La Convention européenne des droits de l'homme

La Cour européenne des droits de l'homme (CEDH) a une approche très protectrice du droit à la liberté d'expression, consacré à

l'article 10 de la Convention éponyme qui s'impose dans les Etats signataires, dont l'ensemble des Etats de l'Union européenne. La Cour a posé trois critères cumulatifs à sa restriction :

- Elle doit être prévue par une loi qui doit être « suffisamment accessible »[1] ;
- Elle doit être justifiée par un but légitime tel que la défense de l'ordre public, de la sécurité nationale, la protection des droits d'autrui, ou encore la protection de la morale ;
- Elle doit être jugée « nécessaire dans une société démocratique », c'est-à-dire proportionnée, entre « les restrictions imposées à la liberté d'expression (…) et le but légitime poursuivi »[2] et répondre ainsi à l'exigence d'un besoin social impérieux.

La jurisprudence de la CEDH protège également le droit à la critique et à l'humour qui comprend la parodie et la satire, comme formes de liberté d'expression en vertu de l'article 10 de la Convention. La Cour souligne leur importance dans le débat public, notamment lorsqu'elles visent des personnalités publiques ou des institutions. Si ces formes d'expression peuvent choquer, elles sont considérées comme essentielles au pluralisme démocratique.

1.2.2. Les normes constitutionnelles et la jurisprudence du Conseil d'Etat

Le Conseil d'Etat a également précisé les critères de restrictions des libertés fondamentales :

- Un critère de nécessité : dans le cas de mesures nécessaires à la sauvegarde de certains objectifs tels que la sécurité nationale, la défense de l'ordre et la protection des droits et libertés d'autrui;
- Un critère de proportionnalité : l'entrave aux libertés doit impérativement être proportionnelle au but légitime visé.

1.3. La loi sur la liberté de la presse du 29 juillet 1881

En France, la loi du 29 juillet 1881[3] organise la liberté de la presse et énonce précisément ses limites. Elle prévoit ainsi plusieurs infractions d'expression, telles que l'injure, la diffamation, l'apologie des crimes, le négationnisme, les cris et chants séditieux ou encore la provocation à la discrimination.

Comme toutes dispositions pénales, ces articles de la loi de 1881 sont d'interprétation stricte, le juge ne dispose donc pas d'une marge d'interprétation extensive de la disposition.

Malgré son âge avancé, le cadre instauré par la loi du 29 juillet 1881, qui intègre aujourd'hui la publication de contenus en ligne, fait l'objet d'un consensus démocratique et social précieux au regard de la complexité des enjeux et des défis actuels de la liberté d'expression.

Cette loi repose sur un principe d'équilibre qui permet de protéger la liberté d'expression comme règle première et de ne sanctionner que les expressions identifiées comme posant un risque majeur à la société. Sa force tient à la constitution progressive d'un corpus jurisprudentiel solide, à la hauteur de l'enjeu et éprouvé à l'époque des médias analogiques.

Le législateur a modifié la loi de 1881 au fur et à mesure du temps et de l'évolution de la société, en y intégrant de nouvelles

infractions et en supprimant ou déplaçant des dispositions vers le code pénal.

1.4. Le code pénal

Le code pénal sanctionne lui aussi plusieurs infractions telles que le délit d'outrage, l'apologie du terrorisme ou encore le harcèlement. Ces dispositions visent à protéger la société contre les discours qui incitent à la violence, à la haine, ou qui portent atteinte à la dignité des individus. De nouveaux délits ont été intégrés au code pénal dans le contexte de l'amplification d'Internet et du rôle des réseaux sociaux, comme le délit de mise en danger par diffusion d'information, notamment sur Internet, créé à la suite de l'assassinat du professeur Samuel Paty.

1.5. Le code civil

Le code civil prévoit également des exceptions relatives à la vie privée.

2. Une application très contrôlée des limites à la liberté d'expression

2.1. Le rôle clef du juge judiciaire

Le juge judiciaire est le premier garant des libertés fondamentales. C'est donc à lui qu'il revient d'appliquer et d'interpréter strictement les limites, fixées par la loi, à la liberté d'expression sur les fondements du code pénal et de la loi de 1881. A travers une appréciation au cas par cas, il veille également à l'équilibre des différents intérêts en présence en appliquant les critères de nécessité et de proportionnalité et se détermine en fonction du cas d'espèce (contexte, personnalités…)

Ces dernières années ont vu se multiplier, alors que la justice est déjà débordée, des « procédures bâillon » qui consistent à se plaindre d'infractions à la loi de 1881 ou du code pénal, afin d'étouffer la parole des personnes exprimant des opinions considérées comme dérangeantes par des procédures coûteuses en argent et temps. Souvent infructueuses, elles n'en sont pas moins traitées par la justice et participent à l'engorgement des prétoires. La chambre spécialisée en délits de presse du Tribunal de Paris tend à condamner de plus en plus régulièrement ces « plaignants abusifs » à payer des dommages-intérêts pour « procédure abusive ».

2.2. Les exceptions prévues par la loi du 29 juillet 1881

Les dispositions de la loi de 1881 de limites à la liberté d'expression peuvent être neutralisées par des faits justificatifs de nature légale ou jurisprudentielle permettant d'éviter la sanction, comme la bonne foi ou l'exception de vérité.

Le fait justificatif tiré de la bonne foi a ainsi été développé par la CEDH et repris par le juge français, qui a dégagé quatre critères cumulatifs afin de prouver sa bonne foi :

- La légitimité du but poursuivi ;
- La vérification des sources et le sérieux de l'enquête ;
- L'absence d'animosité personnelle;
- La prudence et la mesure dans l'expression.

La CEDH a également adopté le critère du débat d'intérêt général comme autre fait

justificatif. L'exception de vérité permet quant à elle d'échapper à la condamnation pour diffamation pour la personne poursuivie capable de prouver la véracité de ses allégations.

3. L'adaptation de la jurisprudence aux mutations de la société

L'appréciation de la liberté d'expression par les juges français et européens a évolué dans plusieurs directions.

La jurisprudence se fait de plus en plus protectrice de l'expression sur certains sujets de société. A titre d'exemple, la jurisprudence de la CEDH a intégré de nombreux sujets dans le critère transversal du « débat d'intérêt général » tels que l'écologie, la polémique syndicale, la religion, la politique, la dénonciation de délits ou de crimes, favorisant une expression plus libre sur des sujets considérés comme importants.

La parole des femmes est également beaucoup plus protégée depuis le mouvement #MeToo qui a entraîné une évolution de la jurisprudence nationale à travers plusieurs arrêts de la Cour de cassation rendus notamment dans l'affaire « #Balance ton Porc »[4].

Dans le même temps, les limitations à la liberté d'expression se sont également considérablement multipliées pour protéger d'autres droits. La jurisprudence consacre en effet une vision de plus en plus large de l'intégrité physique ou morale des personnes, une volonté de renforcer la lutte contre les fausses informations, une approche de plus en plus étendue du champ de protection des données à caractère personnel et du secret des affaires.

Les difficultés d'application de la loi contre les contenus illégaux à l'ère du numérique

Si le corpus juridique théorique de la liberté d'expression est assez complet, l'application de la loi contre ses abus, en revanche, laisse largement à désirer.

Ce qui est illégal hors ligne est illégal en ligne, mais l'ampleur des contenus accessibles en ligne souvent diffusés par des moyens numériques difficilement atteignables par les juridictions françaises ou européennes, la vitesse de diffusion et le manque de moyens de la justice, rendent le recours aux tribunaux assez théorique.

Les critères légaux reposent, en partie, sur des notions imprécises, comme « l'intérêt général », qui manquent de clarté et de prévisibilité et se traduisent par des divergences jurisprudentielles importantes. Cette ambiguïté laisse une marge d'appréciation aux juges qui accentue une forme d'insécurité juridique.

L'anonymat en ligne ne permet pas facilement l'identification de l'auteur. Il n'est pas simple ni rapide d'obtenir l'identité de celui ou celle qui se cache derrière un contenu de façon générale, d'autant que la loi

ne permet d'obtenir l'adresse IP que pour les délits punis d'au moins un an d'emprisonnement[5].

Certaines lois s'avèrent particulièrement difficiles à appliquer, parfois par manque de volonté, mais surtout en raison du manque de moyens de la justice. Ainsi, le cadre réglementaire français en matière de protection des mineurs contre l'exposition à la pornographie en ligne est complet et détaillé depuis plusieurs années mais demeure en partie inappliqué. L'article 227-24 du code pénal incrimine en effet la diffusion d'images pornographiques accessibles à un mineur et prévoit qu'un simple clic sur la mention « J'ai plus de 18 ans » n'est pas suffisant pour procéder au contrôle de l'âge. Les sanctions encourues pour la violation de ces dispositions pénales sont sévères : jusqu'à 3 ans d'emprisonnement et 75 000 € d'amende. Pourtant, les poursuites et par extension les condamnations sont très rares, que les contenus soient diffusés depuis la France ou depuis des sites étrangers.

D'autres textes récents, comme la loi du 30 juillet 2020 n°2020-936 visant à protéger les victimes de violences conjugales, le décret n° 2021-1306 du 7 octobre 2021 relatif aux modalités de mise en œuvre des mesures visant à protéger les mineurs contre l'accès à des sites diffusant un contenu pornographique et la loi visant à sécuriser et réguler l'espace numérique (SREN) du 21 mai 2024 n°2024-449, sont venus consolider les pouvoirs de l'Autorité de régulation de la communication audiovisuelle et numérique (Arcom) dans le contrôle des contenus. Bien que l'Arcom se soit saisie de la question du contrôle de l'âge sur les sites pornographiques depuis la loi SREN, en imposant via son référentiel la mise en œuvre d'outils techniques par ces sites, les autres dispositions législatives restent peu ou pas appliquées. Les magistrats du Parquet poursuivent très rarement les sites illégaux, et aucune directive pénale ne semble à ce stade avoir été publiée pour les y encourager. Devant les juridictions civiles, les procédures initiées par l'Arcom[6] ou les associations, sont longtemps restées bloquées en raison de recours dilatoires, notamment mais pas seulement, exercés par les sites pornographiques. Actuellement, des décisions judiciaires relatives au blocage de sites pornographiques présents sur le territoire européen ont été mises en suspens, dans l'attente d'une décision de la Cour de Justice de l'Union européenne qui doit trancher une question de droit sur la compétence de la France à ordonner le blocage de sites européens non établis en France.

L'européanisation des normes de gouvernance de la liberté d'expression : l'Acte sur les Services Numériques (DSA)

Le numérique est par nature transfrontalier et la problématique traitée dans ce rapport est partagée dans toutes les démocraties. L'Union européenne s'est donc emparée du sujet, notamment à travers le DSA, pour tenter d'harmoniser les règles de gouvernance de retraits de contenus au sein de l'Union européenne et de rendre leur exécution plus opérationnelle. Le DSA peut être considéré comme un texte de gouvernance et de procédure pour la gestion des contenus en ligne, la définition de ce qui peut être qualifié « d'illégal » étant renvoyée au droit national et européen.

Le DSA a maintenu le principe préexistant du « pays d'origine », en vertu duquel l'intermédiaire en ligne est soumis principalement au droit du pays dans lequel il

dispose de son principal établissement, même si le service est consommé ailleurs. Des exceptions sont admises mais strictement encadrées. Ce principe soulève donc une problématique juridique liée aux divergences que peuvent avoir certains Etats-membres[7] sur le niveau de protection de la liberté d'expression et la qualification de certains contenus.

Le DSA poursuit avant tout un objectif : renforcer et préciser les obligations de modération et de retrait des contenus illicites qui incombent aux fournisseurs de services intermédiaires. Il confie ainsi aux plateformes la responsabilité de décider si un contenu doit être maintenu ou supprimé.
Ce faisant, il marque une nouvelle étape dans l'évolution du paradigme encadrant la liberté d'expression. Le texte consacre en effet une première phase de traitement déjudiciarisée, confiée à des acteurs privés, alors même que la décision de retirer ou non un contenu relève traditionnellement de prérogatives publiques essentielles, exercées par les autorités de régulation ou les tribunaux.

Le DSA instaure également un nouvel écosystème chargé de gérer, à grande échelle, les contenus numériques. Le texte cherche à concilier la nécessité d'un retrait rapide des contenus illégaux ou dangereux avec la garantie, pour leurs auteurs, de disposer de mécanismes efficaces pour contester les décisions prises par les plateformes. Il leur impose notamment :

- La mise en place d'une procédure « interne » de traitement des réclamations ;
- Une procédure « externe » de règlement extra-judiciaire en cas de litige ;
- La création d'organes de règlements extrajudiciaires des litiges certifiés par le « coordinateur pour les services numériques » de l'Etat-membre dans lequel l'utilisateur du service concerné est établi (en France, il s'agit de l'Arcom). Ces organes sont compétents pour traiter du contentieux relatif aux décisions – ou au silence – des plateformes sur les contenus litigieux. Ils doivent rendre leurs décisions dans un délai raisonnable et, sauf exception, au plus tard 90 jours après la réception de la plainte.

Cette procédure ne fait toutefois pas obstacle à ce que l'utilisateur du service concerné puisse engager, à tout moment, une procédure de contestation devant les tribunaux. ❐

[1] Cour européenne des droits de l'homme, 26 avril 1979, Sunday Times c. Royaume-Uni

[2] Cour européenne des droits de l'homme, n° 51279/99, 25 juin 2002, Colombani et autres c. France

[3] Loi du 29 juillet 1881 sur la liberté de la presse

[4]La Cour de cassation a rendu deux arrêts, le 11 mai 2022, sur les affaires des mouvements #Metoo et #Balancetonporc dans lesquelles deux femmes accusaient l'ancien patron de la chaîne Equidia et ancien ministre, de harcèlement et d'agression sexuelle. La Cour a estimé que les propos étaient diffamatoires mais qu'ils s'inscrivaient « *dans un débat d'intérêt général consécutif à la libération de la parole des femmes* ». Elle a accordé aux deux prévenues le bénéfice de la bonne foi et ne les a pas condamnées.

[5] Article 60-1-2 du code pénal, créé par la loi n°2022-299 du 2 mars 2022.

[6] Il parait important de préciser que la loi SREN nouvellement adoptée élargit les compétences de l'Arcom qui pourra poursuivre et sanctionner et dont les décisions pourront être contestées devant le Conseil d'Etat, conduisant de facto à une concurrence entre les jurisprudences sur la liberté d'expression du juge judicaire et du juge administratif.

[7] Exemple classique des divergences sur la légalité des pratiques homosexuelles dans certains Etats-membres par exemple.

Annexe 2

L'entrée en vigueur du Digital Services Act : une occasion de repenser la liberté d'expression en ligne ?

par Pascal Beauvais
Agrégé des facultés de droit, Professeur de droit privé et sciences criminelles Ecole de droit de la Sorbonne - Université Paris 1, Avocat à la Cour

Dans le cadre d'une réflexion prospective, le cabinet d'avocats Samman & Associés a sollicité du soussigné, M. Pascal Beauvais, professeur agrégé de droit privé et sciences criminelles, une analyse juridique des effets du Digital Services Act sur les règles françaises encadrant la liberté d'expression numérique. Cette étude préliminaire a vocation à être complétée et précisée par des analyses complémentaires au fur et à mesure des discussions et propositions qui accompagneront l'intégration du Digital Services Act au droit français.

Le règlement UE 2022/2065 du Parlement européen et du Conseil relatif à un marché intérieur des services numériques (« *Digital Services Act* » ou « *DSA* ») a été définitivement adopté le 19 octobre 2022 et publié au journal officiel de l'Union le 27 octobre 2022. Cette nouvelle législation européenne sur les services numériques, qui modifie la directive n°2000/31/CE du 8 juin 2000 sur le commerce électronique, obéit au principe général selon lequel « ce qui est illégal hors ligne doit être illégal en ligne ».

Le *DSA* a notamment pour objet d'intensifier les obligations de modération et de suppression des contenus illicites à la charge des fournisseurs de services d'hébergement en les soumettant à des procédures beaucoup plus précises. Il vise à mettre en place un système largement extrajudiciaire de traitement du contentieux de masse des contenus numériques litigieux qui repose sur l'action des fournisseurs de services d'hébergement, des organes de règlement extrajudiciaire accrédités et sur un réseau européen de « coordinateurs nationaux pour les services numériques » - qui, en France, devrait *a priori* être l'ARCOM (Autorité de régulation de la communication audiovisuelle et numérique)[1].

Le *DSA* est applicable à partir du 17 février 2024 sauf pour certaines dispositions qui feront l'objet d'une application anticipée en 2023 pour les fournisseurs de très grandes plateformes en ligne et de très grands moteurs de recherche en ligne.

L'intégration en droit français de cette déjudiciarisation systémique du traitement des contenus numériques litigieux soulève des questions fondamentales, en particulier : comment l'articuler et l'adapter au régime de la liberté d'expression et d'information dont les principaux éléments se trouvent dans la jurisprudence judiciaire relative à la loi du 29 juillet 1881 et à la Convention européenne des droits de l'homme ?

Après avoir présenté, de manière synthétique, les principaux apports du *DSA* relatifs à la modération des contenus (**1**), la présente note examinera les questions que soulève son intégration dans le droit français de la liberté d'expression (**2**).

[1] Mais d'autres autorités de régulation, comme la CNIL par exemple, ont également manifesté leur volonté d'être compétentes pour mettre en œuvre le DSA.

1. Rappel des principaux apports du DSA sur la modération des contenus dans les services numériques

Le *DSA* est applicable à tous les fournisseurs de « services intermédiaires » en ligne dans l'Union européenne, ce qui comprend les opérateurs offrant des services d'« hébergement ». Son champ d'application n'est pas celui du pays d'origine, mais celui de destination du service. Le règlement vise donc tous les services d'intermédiation numérique fournis aux utilisateurs ayant leur lieu de résidence ou d'établissement dans l'Union quel que soit celui du prestataire lui-même. Ce dernier doit toutefois avoir un lien « étroit » avec l'Union[1]. Ce « lien étroit » avec l'Union est caractérisé lorsque le nombre de destinataires du service dans un ou plusieurs États membres est « significatif » au regard de leur population ou sur la base d'un ciblage des activités du fournisseur sur un ou plusieurs États membres.

1.1. Grandes lignes du *Digital services Act*

Les grandes lignes du DSA sont les suivantes : responsabilité limitée des fournisseurs de services numériques intermédiaires en matière de contenu illicite (1), instauration d'obligations organisationnelles et processuelles, plutôt que substantielles (2), déjudiciarisation des procédures relatives aux contenus litigieux (3).

1.1.1. Principe de responsabilité limitée des fournisseurs de services intermédiaires

Le *DSA* ne revient pas sur le principe posé par la directive n°2000/31/CE sur le commerce électronique selon lequel les fournisseurs de services intermédiaires ne sont pas tenus à une obligation générale de surveiller les contenus qu'ils transmettent ou stockent ou de rechercher activement des faits ou des circonstances révélant d'activités illégales[2].

L'objectif du *DSA* est néanmoins de responsabiliser davantage les fournisseurs de services numériques intermédiaires sur les contenus qu'ils transmettent ou stockent. Les obligations instaurées sont graduées en fonction de la nature des services concernés et du nombre d'utilisateurs[3]. Le règlement distingue trois régimes principaux de responsabilité : celui du « simple transporteur »[4], celui de l'activité de « mise en cache »[5] et celui de « l'hébergeur »[6]. Dans ce dernier régime, il distingue les règles applicables aux plateformes en ligne. Enfin le DSA ajoute un régime d'obligations renforcées pour les plateformes en ligne et moteurs de recherche de très grande taille.

Le *DSA* maintient donc l'équilibre fondamental issu de la directive de 2000 sur le commerce numérique - et transposée en France dans la loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique - qui oblige les fournisseurs de services d'hébergement à contribuer à la lutte contre les contenus illicites, mais d'une manière limitée en tenant compte du fait qu'ils ne sont pas les auteurs, mais seulement les dépositaires ou les vecteurs de ces contenus. L'hébergeur ne peut donc être responsable des contenus qu'il stocke, mais à la condition 1) qu'il n'ait pas connaissance d'un contenu illicite ou 2) qu'il agisse promptement, dès le moment où il en prend connaissance, pour retirer ou rendre inaccessible le contenu illicite[7]. Il n'a donc pas d'obligation « proactive », mais seulement une obligation « réactive » d'agir promptement pour ôter un tel contenu dès le moment où il en a connaissance.

Cette exemption de responsabilité n'est donc pas applicable lorsque le fournisseur de services intermédiaires joue un rôle éditorial sur les contenus[8]. En outre, selon la récente proposition de règlement UE sur la liberté des médias, les plateformes qui exercent un contrôle éditorial devront être qualifiées de fournisseurs de services de médias[9].

En revanche, le fait de modérer ou de retirer volontairement des contenus illicites n'a pas pour effet d'écarter, par principe, l'exemption de responsabilité. Cette clause dite « du bon samaritain » constitue l'un des principaux apports du *DSA* par rapport au cadre de la directive de 2000 sur le commerce numérique en ce qui concerne le principe de responsabilité limitée. En effet, pour inciter les fournisseurs à détecter, recenser et combattre les contenus illicites, les activités « proactives » de modération et de retrait peuvent bénéficier de l'exemption de responsabilité, mais à la condition qu'elles soient menées de bonne foi et avec diligence[10]. La condition d'agir de bonne foi et avec diligence comprend « le fait d'agir de manière objective, non discriminatoire et proportionnée, en tenant dûment compte des droits et des intérêts légitimes de toutes les parties concernées, ainsi que le fait de fournir les garanties nécessaires contre la suppression injustifiée de contenus licites »[11]. Si les fournisseurs ne perdent donc pas le bénéfice de l'exemption de responsabilité lorsqu'ils mettent en œuvre, de leur propre initiative, des actions impartiales de modération des contenus, en revanche l'exemption n'est pas présumée du simple fait de prendre, de manière « proactive », ces mesures[12].

Enfin, l'exemption de responsabilité n'affecte pas la possibilité de procéder à des injonctions judiciaire ou administrative à l'encontre des fournisseurs de services intermédiaires.

1.1.2. Des obligations organisationnelles et procédurales renforcées

Le *Digital Services Act* ne porte pas sur le fond, c'est-à-dire sur la définition et le champ des contenus illicites[13], mais impose en revanche aux fournisseurs de services intermédiaires, en particulier aux hébergeurs, la mise en place de procédures précises pour les traiter.

Sur le modèle du *RGPD*, le *DSA* privilégie donc des obligations processuelles plutôt que substantielles[14]. Dans une logique de *compliance,* il contraint les hébergeurs à adopter une organisation et des mécanismes de signalement, de modération/retrait, de motivation et de recours qui sont contrôlés par un maillage d'organes extra-judiciaires et d'autorités nationales, lesquels sont supervisés par un nouveau comité européen pour les services digitaux[15].

1.1.3. Un contentieux déjudiciarisé

Le *DSA* impose aux plateformes en ligne la mise en place d'un système *interne* de traitement des réclamations sur les décisions de modération des contenus en ligne[16]. Il prévoit également que les destinataires des services ayant subi une décision sur les contenus (soit de suppression ou de refus de suppression) puissent saisir un organe de règlement extrajudiciaire des litiges de leur choix parmi ceux qui ont été certifiés par le « coordinateur pour les services numériques » national. Cet organe extrajudiciaire est compétent pour statuer sur les contestations des décisions prises par les plateformes, y compris en cas de silence lorsque le mécanisme interne de traitement des réclamations n'a pas été mis en œuvre. Le *DSA* comble ainsi une lacune de la directive de 2000 sur le commerce électronique en intégrant dans les procédures mises en place l'intérêt des utilisateurs, notamment le droit à la liberté d'expression de celui dont les contenus ont fait l'objet d'un retrait.

Aucune précision n'est en revanche apportée dans le texte sur l'autorité des décisions de ces organes et sur les éventuels recours étatiques auxquels elles pourront donner lieu. Le *DSA* précise seulement qu'elles ne sont pas contraignantes pour les parties[17].

1.2. Principales obligations processuelles des plateformes en ligne en matière de contenus illicites

Les obligations procédurales de lutte contre les contenus illicites sont graduées en fonction de la nature des services concernés et du nombre d'utilisateurs. Les plateformes en ligne sont ainsi soumises aux obligations générales applicables aux fournisseurs de services d'hébergement, mais aussi à des règles qui leur sont propres.

1.2.1. Obligation d'intégrer aux conditions générales (« CGU ») les principaux éléments de la politique de modération des contenus

Les fournisseurs de services intermédiaires doivent inclure dans leurs conditions générales (« CGU ») des renseignements relatifs aux éventuelles restrictions qu'ils imposent en ce qui concerne les contenus qu'ils stockent ou font circuler (article 14). Ces renseignements comprennent des

informations sur les politiques, procédures, mesures et outils utilisés à des fins de modération/suppression des contenus. Les conditions générales peuvent proscrire certains contenus licites, mais que le fournisseur considèrera comme préjudiciable. Elles doivent être facilement accessibles et établies dans un langage clair.

Ces règles sur le traitement des contenus ne relèvent pas de la seule liberté contractuelle, mais sont encadrées : les fournisseurs de services intermédiaires doivent en effet tenir « compte des droits et des intérêts légitimes de toutes les parties impliquées, et notamment des droits fondamentaux des destinataires du service, tels que la liberté d'expression, la liberté et le pluralisme des médias et d'autres libertés et droits fondamentaux tels qu'ils sont consacrés dans la Charte »[18].

1.2.2. Obligation de transparence sur les activités de modération des contenus

Tous les intermédiaires sont tenus d'être transparents sur leurs activités de modération/suppression des contenus : ils doivent mettre à la disposition du public, dans un format lisible et d'une manière facilement accessible, au moins une fois par an, des rapports clairs et facilement compréhensibles sur les activités de modération des contenus auxquelles ils se sont livrés au cours de la période concernée. Ces rapports comprennent, en particulier, des informations précises, qualitatives et quantitatives sur les injonctions, notifications et réclamations soumises ainsi que leurs modes de traitement.

1.2.3. Obligation d'exécuter les injonctions sur les contenus des autorités judiciaires ou administratives

Le *DSA* oblige les plateformes à supprimer les contenus qui enfreignent les réglementations nationales et européennes sur le fondement d'une injonction d'agir émise par une autorité judiciaire ou administrative nationale - comme par exemple l'ARCOM ou éventuellement la CNIL en France[19]. À l'intérieur de l'Union, cette injonction peut être transnationale. L'injonction doit notamment comprendre un exposé des motifs expliquant pourquoi il s'agit d'un contenu illicite, des informations claires permettant au fournisseur de services intermédiaires d'identifier et de localiser le contenu illicite concerné, ainsi que des informations sur les mécanismes de recours dont disposent le fournisseur de services intermédiaires et l'utilisateur ayant fourni le contenu illicite[20]. Lorsqu'une suite est donnée à l'injonction, les fournisseurs de services intermédiaires doivent informer l'utilisateur concerné de l'injonction reçue, de la suite qui lui est donnée et des possibilités de recours qui existent[21].

1.2.4. Obligation de mise en place d'un mécanisme de signalement, de notification et de retrait des contenus illicites

Le *DSA* oblige les fournisseurs de services d'hébergement à mettre en place des mécanismes, faciles d'accès et d'utilisation, qui permettent à tout individu ou à toute entité de leur signaler un contenu qu'il considère comme illicite[22]. Pour pouvoir être traité, ce signalement doit comporter «

une explication suffisamment étayée » des raisons pour lesquelles le contenu est désigné comme illicite. Ces notifications permettent de présumer la « connaissance de l'illicéité du contenu » de l'hébergeur, ce qui lève son exemption de responsabilité et permet d'engager celle-ci en cas d'inaction de sa part[23].

Les hébergeurs doivent fournir à tous les destinataires du service affectés par une modération ou un retrait de contenu illicite (ou contraire à leurs conditions générales) un exposé des motifs clair et spécifique expliquant la restriction imposée. Que la décision concerne des contenus considérés comme illicites ou des contenus incompatibles avec les conditions générales de l'hébergeur, l'exposé des motifs doit comporter une référence au fondement juridique sous-jacent et des explications sur les raisons pour lesquelles ces contenus sont ainsi considérés[24]. L'exposé des motifs de la décision doit également communiquer des informations aisément compréhensibles sur les possibilités de recours contre cette décision de modération ou de suspension de l'hébergeur.

1.2.5. Obligation de mise en place de recours « internes » et « externes » relatifs aux décisions sur les contenus.

L'un des apports du *DSA* est d'intégrer au système de traitement des contenus numériques litigieux les intérêts des utilisateurs, notamment ceux dont le contenu a été supprimé et qui entendent exercer leur droit à la liberté d'expression. Les décisions de suppression ou de refus de suppression doivent donc être motivées et pouvoir faire l'objet de recours internes comme externes.

Le *DSA* introduit donc des obligations procédurales nouvelles permettant aux utilisateurs de contester les décisions des plateformes en ligne sur les contenus :

- d'une part, il impose aux plateformes de mettre en place une procédure « interne » de traitement des réclamations ;
- d'autre part, il prévoit une procédure « externe » de règlement extra-judiciaire en cas de litige.

Le système de traitement interne des réclamations concerne aussi bien les décisions relatives au retrait des contenus, que celles portant sur la suspension ou la clôture des comptes. L'accès à ce recours est ouvert pendant une période de 6 mois après que décision litigieuse ait été rendue. Lorsqu'une réclamation contient suffisamment de motifs pour convaincre la plateforme en ligne que sa décision est infondée, celle-ci doit l'infirmer et en informer les plaignants dans les meilleurs délais[25].

Le *DSA* prévoit également la création d'organes de règlement extrajudiciaire des litiges certifiés par le « coordinateur pour les services numériques » de l'État membre dans lequel il est établi. Ces organes sont compétents pour traiter du contentieux relatif aux décisions – ou au silence – des plateformes sur les contenus. Cette procédure ne fait toutefois pas obstacle à ce que l'utilisateur du service concerné puisse engager, à tout moment, une procédure de contestation devant une juridiction étatique. Les organes de règlement extrajudiciaire des litiges rendent leurs décisions

dans un délai raisonnable et, sauf exception, au plus tard 90 jours après la réception de la plainte.

Le *DSA* crée également un statut de « signaleur de confiance » désignant des entités accréditées, disposant d'expertises et de compétences en matière de liberté d'expression, qui bénéficient de prérogatives renforcées pour notifier de manière prioritaire aux plateformes en ligne des contenus qu'elles considèrent illicites.

Au-delà des procédures de notification-suppression des contenus illicites ou incompatibles avec leurs conditions générales, les plateformes en ligne peuvent également suspendre, pendant une période raisonnable et après avoir émis un avertissement préalable, la fourniture de leurs services aux utilisateurs qui fournissent fréquemment des contenus manifestement illicites[26].

1.2.6. Obligations complémentaires propres aux très grandes plateformes et aux très grands moteurs de recherche

Enfin, des obligations complémentaires sont applicables aux très grandes plateformes en ligne et aux très grands moteurs de recherche dont le nombre mensuel d'utilisateurs actifs au sein de l'UE est supérieur ou égal à 45 millions. Le *DSA* entend responsabiliser spécialement ces acteurs « systémiques » en raison de leur importance économique, politique et sociétale. Il leur impose ainsi de conduire, une fois par an, une analyse des risques systémiques créés par les contenus qu'ils stockent et diffusent et de mettre en place des mesures d'atténuation de ces risques.

Ces risques peuvent notamment résulter d'une diffusion à très grande échelle de contenus illicites, qui peuvent porter atteinte aux droits fondamentaux, aux processus électoraux, à la sécurité publique ou à la santé de leurs utilisateurs. Ces très grands opérateurs ont donc l'obligation, de nature processuelle, d'identifier les facteurs qui favorisent ces risques systémiques et d'en tenir compte dans la conception et la mise en œuvre des services qu'ils offrent[27]. Ils doivent mettre en place des mesures d'atténuation raisonnables, proportionnées et efficaces, adaptées aux risques systémiques spécifiques recensés[28]. Les très grandes plateformes en ligne et très grands moteurs de recherche doivent également désigner un « *compliance officer* » et se plier chaque année à un audit indépendant vérifiant le respect de ces obligations organisationnelles.

La Commission européenne est chargée de contrôler l'activité de ces très grands opérateurs et dispose pour cela de pouvoirs d'investigations administratifs. La Commission peut leur infliger des amendes jusqu'à concurrence de 6 % du chiffre d'affaires mondial annuel réalisé au cours de l'exercice précédent lorsqu'elle constate que les fournisseurs de la très grande plateforme en ligne ou du très grand moteur de recherche en ligne, de manière délibérée ou par négligence, enfreignent les dispositions pertinentes du présent règlement.

En cas de crise caractérisée par une menace grave pour la sécurité publique ou la santé publique, le *DSA* confère à la Commission européenne un pouvoir de décision supplémentaire sur les méthodes et pratiques de ces très grands opérateurs qui peut consister, par exemple, à imposer des mesures accélérant la modération ou la suppression de contenus.

La proposition très récente de règlement sur la liberté des médias complète le *DSA* en prévoyant un régime particulier de modération pour le traitement des informations diffusées par les médias dans l'hypothèse où elles seraient contraires aux conditions générales (CGU) d'une très grande plateforme. Ces dernières doivent en effet offrir une fonctionnalité permettant aux médias éditorialement indépendants de se déclarer auprès d'elles. La proposition de règlement sur la liberté des médias prévoit alors une procédure spéciale amiable de dialogue et de règlement en cas de décision d'une plateforme de suspendre la fourniture de ses services à un média dans l'hypothèse où celui-ci publierait des contenus contraires à ses conditions générales.

2. L'articulation des dispositions du DSA sur le traitement des contenus numériques illicites avec le droit français de la liberté d'expression numérique

Le *DSA* est directement applicable et obligatoire dans les droits des États membres. Les États membres doivent donc adapter et compléter leurs systèmes juridiques pour assurer sa pleine application. Ils ne peuvent adopter ou maintenir des exigences nationales contradictoires ou supplémentaires, car cela porterait atteinte à l'application directe et uniforme des règles pleinement harmonisées du DSA.

Le *DSA* a pour objet d'intensifier et de préciser les obligations de modération et de suppression des contenus illicites à la charge des fournisseurs de services intermédiaires. Ce faisant, il met en place un nouvel écosystème largement déjudiciarisé du traitement de masse des contenus numériques litigieux. Cette déjudiciarisation pose des questions fondamentales en droit français, en particulier celle de son articulation avec le droit de la liberté d'expression et d'information qui, jusqu'à présent, est essentiellement fondé sur la jurisprudence judiciaire issue de la loi de 1881 et de Convention européenne des droits de l'homme.

2.1. Le risque de marginalisation des procédures judiciaires prévues par la loi de 1881 au profit des procédures déjudiciarisées du *DSA* pour le traitement du contentieux de masse des contenus litigieux

Si la loi du 29 juillet 1881, qui définit les principaux abus de la liberté d'expression[29] et sanctionne leurs auteurs à raison d'une publication, n'a pas le même objet que le *DSA*, qui prévoit des procédures de modération/suppression des contenus illicites à la charge des fournisseurs de services numériques qui les stockent et les font circuler, elle pourrait toutefois être rapidement marginalisée par le règlement européen pour le traitement du contentieux de masse des contenus en ligne litigieux.

En effet, comme la Commission nationale consultative des droits de l'homme (CNCDH) l'avait souligné dans son avis « Lutte contre la haine sur Internet » du 12 février 2015, la loi du 29 juillet 1881 n'est pas adaptée à la généralisation de l'expression publique consécutive

à la révolution numérique[30]. Si elle trouve à s'appliquer aux communications numériques, elle n'est aujourd'hui pas adaptée au contentieux de masse que les contenus en ligne sont de nature à engendrer. Pour reprendre les mots de la CNCDH, il s'agit d'une loi complexe, au contenu difficilement accessible, faisant l'objet d'une interprétation jurisprudentielle très nuancée, que seuls des juristes spécialisés maîtrisent. Elle est originellement destinée aux professionnels de la communication (presse, éditeurs, médias) pour encadrer leurs activités et donne lieu à un contentieux sophistiqué devant des magistrats très spécialisés (notamment la 17e chambre correctionnelle du TGI de Paris). Elle n'avait pas initialement vocation à s'appliquer à tout utilisateur des services de communication numériques devenu désormais un éditeur public potentiel. Autrement dit, la loi du 29 juillet 1881 n'a pas été conçue pour une expression publique généralisée, qui n'est plus filtrée en amont par des médias professionnels responsabilisés et soumis à un encadrement déontologique.

Afin de protéger la presse et les journalistes, pour qui elle avait été faite à l'origine, la procédure de la loi du 29 juillet 1881 comporte de très nombreuses « chausse-trappes » enserrant à tout moment la conduite de l'affaire dans un strict réseau d'exigences procédurales originales et très exigeantes[31]. En pratique, les actions des victimes des abus de la liberté d'expression sur ce fondement ne peuvent être engagées qu'assistées par des avocats très spécialisés.

Certes, les procédures de signalement et retrait des contenus litigieux à la charge des fournisseurs de services intermédiaires étaient déjà prévues par la directive de 2000 sur le commerce électronique, et en droit français, à l'article 6 de la Loi sur la confiance dans l'économie numérique du 21 juin 2004 (dite « loi LCEN »)[32], qui instaure un régime complexe de responsabilité en cascade, mais l'ampleur, la précision et la systématisation du nouveau régime de traitement des contenus numériques litigieux introduit par le *DSA* devrait bouleverser cet équilibre. D'une part, l'objectif du *DSA* est bel et bien de prendre en charge le contentieux de masse exponentiel des contenus litigieux en ligne. D'autre part, il met en place des procédures entièrement numériques et accessibles, dont l'efficacité repose à la fois sur les moyens technologiques des fournisseurs, mais aussi sur les obligations de rapidité qu'il fait peser sur eux, qui devraient être plus attractives – même si elles n'ont pas le même objet - que les procédures judiciaires lourdes, lentes et difficiles de la loi de 1881.

Les particuliers souhaitant s'attaquer à un contenu qui leur paraît litigieux auront tout intérêt, à l'avenir, à engager une procédure très simple de signalement/retrait auprès des numériques en faisant l'économie de poursuites judiciaires extrêmement complexes et coûteuses. En pratique, l'encadrement concret de la liberté d'expression numérique devrait donc, d'abord et avant tout, passer par les décisions prises dans le cadre de procédures numériques du DSA. Or, cette délégation aux grands opérateurs de l'encadrement de la liberté d'expression, au détriment de la loi et du juge judiciaire, est problématique tant de point de vue des principes fondamentaux du droit que de la sécurité juridique des acteurs.

Afin de préserver la place du juge judiciaire dans l'encadrement de la liberté d'expression, conformément à la tradition constitutionnelle et européenne, et qu'elle ne soit pas

marginalisée dans le traitement du contentieux de masse par les procédures prévues par le *DSA*, il conviendrait donc : soit de réformer en profondeur la procédure de la loi du 29 juillet 1881 ; soit d'ouvrir le vaste chantier de la codification du droit de la communication afin d'instaurer une justice étatique du numérique à la hauteur de ses enjeux.

En tout état de cause, à l'occasion de l'entrée en vigueur du *DSA* et de la mise en œuvre des États généraux de la Justice, le législateur français devrait lancer le chantier de la justice numérique en instaurant des procédures judiciaires dématérialisées et rapides permettant d'appréhender le contentieux de masse sur les contenus litigieux. Ce contentieux numérique ne pourra être appréhendé par l'autorité judiciaire qu'à la condition d'une transformation en profondeur des procédures et de l'organisation judiciaire. Outre la numérisation des procédures (notamment des assignations et significations ; audiences en visioconférence) ainsi que le développement de procédures d'urgence dématérialisée, cette réforme d'ampleur nécessitera un effort considérable d'équipement, de recrutement de techniciens et experts et de formation des magistrats. Dans la lignée du Rapport des États généraux de la Justice, qui préconise une refonte d'ampleur de la stratégie numérique du ministère de la Justice, une telle réforme implique de repenser l'organisation du pilotage numérique de l'appareil judiciaire[33].

2.2. La marginalisation des interprétations judiciaires strictes des abus de la liberté d'expression, prévues par la loi de 1881, au profit d'interprétations relatives et extensives des fournisseurs de services d'hébergement

Sur le fond, le traitement du contentieux de masse des contenus numériques litigieux par les fournisseurs de services d'hébergement et les organes extrajudiciaires devrait également avoir pour effet de relativiser la place de la jurisprudence judiciaire, fondée sur l'interprétation stricte de la loi du 29 juillet 1881, dans la définition des abus de la liberté d'expression.

Dans le système de la loi du 29 juillet 1881, c'est en effet le juge judiciaire qui qualifie juridiquement les contenus litigieux ; dans le système du *DSA*, ce sont les fournisseurs de services intermédiaires, qui sont amenés à qualifier, en première ligne, ces contenus. Certes, en principe, cette opération de qualification, ainsi que les actions de modération et de suppression des contenus illicites qui s'en suivent, doivent reposer sur les prescriptions du droit national, mais la dimension aujourd'hui très casuistique du droit français de la liberté d'expression rendra en pratique sa mise en œuvre très difficile pour les fournisseurs de services numériques intermédiaires.

- ***L'inadaptation d'un droit jurisprudentiel complexe à un contentieux de masse***

Le droit français de la liberté d'expression est largement d'origine jurisprudentielle : même si elles reposent sur la loi du 29 juillet 1881, les définitions des abus de la liberté d'expression ont largement été dessinées, au cas par cas, par le juge judiciaire sous l'influence de la

Cour européenne des droits de l'homme. Des incriminations essentielles comme la diffamation ou la provocation à la discrimination, la haine ou la violence reposent, dans chaque affaire, sur une appréciation judiciaire sophistiquée s'appuyant sur une multitude de paramètres. Tant que cette complexité ne concernait, dans les faits, qu'un nombre limité de médias, elle n'était pas problématique ; elle le devient en revanche avec la massification de ce contentieux.

Comme l'avait fait le Conseil d'État dans un rapport rendu en 2006[34], il conviendrait donc de s'interroger sur l'intérêt et la faisabilité de préciser et de clarifier le droit français de la liberté d'expression et d'information, en particulier celui issu de la jurisprudence relative à la loi du 29 juillet 1881, en examinant notamment l'hypothèse de sa codification. Il nous semble évident que la codification des abus de la liberté d'expression, aux sources essentiellement jurisprudentielles, contribuerait à l'accessibilité du droit, et améliorerait probablement son intelligibilité et donc son application dans les contentieux de masse. À titre d'exemple, la loi du 29 juillet 1881 sur la liberté de la presse ne donnant aucune précision sur l'élément moral du délit de diffamation, la jurisprudence a créé un fait justificatif, la bonne foi, dont elle a précisé les critères : la légitimité du but poursuivi, l'absence d'animosité personnelle, la prudence et la mesure dans l'expression et le sérieux de l'enquête[35]. Grâce à leur stabilité, ces critères s'apparentent à une véritable norme jurisprudentielle qui pourrait aisément être énoncée dans une règle écrite et codifiée.

La création d'un code de la communication, qui serait fondé, comme la loi de 1881, sur le caractère fondamental du droit à la liberté d'expression et d'information, permettrait d'offrir un cadre général cohérent et prévisible régissant à la fois le régime des auteurs et éditeurs de contenus et celui des fournisseurs de services intermédiaires et articulant, au fond, dans un même ensemble, les dispositions de la loi du 29 juillet 1881 à celles du *DSA*. Le fondement juridique de cette codification ne peut être qu'une nouvelle loi, mais pour éviter des débats parlementaires longs, complexes et risqués sur le droit à la liberté d'expression, le projet de loi pourrait, dans un premier temps, se contenter de proposer une codification principalement à droit constant.

Sans clarification et précision du cadre juridique français applicable, les opérateurs privés auront non seulement des difficultés pour opérer ces qualifications complexes, mais ils seront tentés, compte tenu de cette insécurité juridique et pour minimiser le risque d'engager leur responsabilité, à faire une application large de ces jurisprudences au risque de porter une atteinte disproportionnée à la liberté d'expression.

Reste que la codification du droit de la communication et plus précisément des abus de la liberté d'expression ne pourra pas saisir l'ensemble des nuances de la jurisprudence judiciaire en la matière, qui restera donc une source essentielle du droit sur les contenus numériques illicites.

- ***Une approche large des contenus susceptibles d'être supprimés***

Si le *DSA* étend et intensifie les obligations processuelles de traitement des contenus illicites, il ne régit pas, sur le fond, la définition et le champ des contenus illicites, mais opère, pour cela, par renvoi aux législations nationales et européennes. De surcroît, le *DSA* reconnaît le droit des fournisseurs d'hébergement de modérer et de supprimer les contenus incompatibles avec leurs conditions générales (CGU), ce qui élargit potentiellement beaucoup le champ des contenus susceptibles d'être retirés sur le fondement du droit de propriété et de la liberté contractuelle.

Ce faisant, les fournisseurs de services d'hébergement peuvent ainsi inclure dans les contenus incompatibles avec les conditions générales des contenus dits « préjudiciables », qui, même licites, portent atteinte à certains intérêts particuliers ou collectifs ou généraux, comme les discours de haine ou les *fake news*[36]*.* Toutefois, en prévoyant de manière large et floue ces contenus préjudiciables, les fournisseurs d'hébergement prendraient le risque de porter une atteinte disproportionnée à la liberté d'expression et d'information, qui restera protégée par les juridictions nationales et la Cour européenne. Cette intégration des contenus préjudiciables aux contenus susceptibles d'être supprimés est donc un facteur supplémentaire d'insécurité juridique pour les opérateurs.

En raison de cette approche ouverte et peu prévisible des contenus susceptibles d'être supprimés, les opérateurs pourraient légitimement craindre de voir leur responsabilité facilement engagée et, pour l'éviter, adopter une approche extensive de ces contenus. En effet, « le risque d'une responsabilisation excessive est de débrider un contrôle reposant sur une appréciation discrétionnaire et, le cas échéant, arbitraire de la licéité des contenus par les plateformes elles-mêmes, conduisant au blocage ou au retrait de contenus licites » [37]. En perspective, comme l'affirme le professeur G. Loiseau, « c'est la liberté d'expression qui serait menacée »[10]. Pour cette raison également, les acteurs de ce nouvel écosystème de traitement des contenus numériques auraient tout intérêt à disposer d'un cadre juridique plus certain, précis et prévisible.

2.3. Nécessité d'articuler les procédures du *DSA* avec les procédures judiciaires du droit français de la liberté d'expression

Le *DSA* prévoit que le contentieux de masse des contenus numériques litigieux sera réglé par un réseau d'organes extrajudiciaires, certifiés dans chaque État membre par le coordinateur national des services numériques[38]. Ce droit des utilisateurs à exercer un recours extrajudiciaire ne fait évidemment pas obstacle, à tout moment, au droit des utilisateurs de contester une décision prise par une plateforme devant une juridiction étatique. En effet, l'organe extrajudiciaire n'ayant pas le pouvoir d'imposer un règlement contraignant du litige, le *DSA* ne peut exclure ce recours alternatif aux juridictions étatiques. Pour autant, il ne prévoit pas leur place dans la procédure.

Or, en l'absence de liens ou passerelles entre les décisions de ces organes certifiés et les procédures judiciaires classiques, le système de traitement des contenus litigieux instauré par le *DSA* risque de se développer de manière parallèle à la justice étatique sur la liberté

d'expression[39]. Le traitement de ces contenus risque en effet d'être fragmenté « faute de coordination hiérarchique des organes (…), chaque organe pouvant interpréter la loi applicable et les conditions générales des plateformes à sa façon, différente de celle des juges étatiques, mais aussi des autres organes » [40]. L'organisation décentralisée du traitement déjudiciarisé des contenus litigieux produira une disparité de jurisprudences d'un organe à l'autre, d'un fournisseur à l'autre au détriment de la sécurité juridique, mais aussi de la légitimité et de l'autorité des réponses apportées.

Afin d'assurer le développement d'une jurisprudence cohérente, précise et légitime des contenus illicites, le législateur français devrait donc prévoir des modes d'articulation du traitement privé des contenus litigieux, prévu par le *DSA*, avec les procédures judiciaires en matière de liberté d'expression. Pour garantir la sécurité juridique des acteurs privés, mais aussi du régime de la liberté d'expression, et éviter que la prudence incite les opérateurs à un encadrement trop strict des contenus, le législateur français devrait donc renforcer le rôle de la justice étatique en la matière et préciser ses liens (procédure d'appel ? de validation ?) avec le système déjudiciarisé mis en place par le *DSA*. Pour jouer ce rôle dans un contentieux de masse en devenir, le législateur devra mettre en place un réseau conséquent de juridictions spécialisées dans le numérique ainsi que des procédures judiciaires dématérialisées et rapides.

Même au seul stade de l'appel ou de l'homologation, l'appréhension et la maîtrise de ce contentieux de masse des contenus illicites par la justice française nécessiteront une réforme d'ampleur de l'organisation et des procédures qui devront être largement dématérialisées. Cette exigence trouve un écho dans le Rapport des États généraux de la Justice qui préconise l'émergence d'une véritable « justice numérique » au travers notamment du développement des procédures et des saisines par voie digitale permettant d'améliorer la réactivité et la fluidité des réponses judiciaires[41]. Les actes procéduraux devront être dématérialisés : notification, convocation en ligne, audience en visioconférence. Le programme actuel de « procédure pénale numérique » qui souffre d'un financement insuffisant devrait, par exemple, être considérablement amplifié[42]. Pour instaurer une véritable justice numérique, complémentaire à la justice physique, chaque juridiction devra être dotée d'un véritable service informatique avec des informaticiens nombreux.

L'entrée en vigueur du *DSA* qui déjudiciarise le traitement de masse des contenus numériques illicites appelle donc des réformes structurelles du droit français de la communication pour que celui-ci reste, dans ce nouveau paradigme privatisé, le cadre de référence prévisible et sûr, fondé sur des intérêts publics, de la lutte contre les abus de la liberté d'expression.

2.4. Nécessité de repenser la gestion des contenus en ligne à la croisée de l'action judiciaire et du contrôle des plateformes

Au-delà de ces réformes, le DSA constitue une opportunité unique de repenser le modèle français de la liberté d'expression à l'aune des nouveaux usages numériques et de leur rôle

crucial dans notre société démocratique.

Face à des phénomènes de masse ayant un impact majeur sur les individus et la société, le rôle du juge judiciaire, même renforcé dans ses attributions et ses moyens, ne saurait suffire : la nature *ex-post* du contrôle judiciaire et son approche casuistique ne semblent pas en mesure de répondre à la diversité des enjeux posés. En outre, laisser un pouvoir trop important aux plateformes en ligne dans la définition du caractère acceptable ou non des contenus ne paraît pas souhaitable d'un point de vue démocratique compte tenu de leur nouveau rôle central dans l'organisation du débat public.

Au-delà du contrôle judiciaire et de l'action des plateformes, n'existe-t-il pas un champ des possibles démocratique pour une meilleure gestion des contenus en ligne ? L'actualité récente regorge d'initiatives, dont le législateur pourrait s'inspirer, visant à mettre en place des instances visant à refléter au mieux la diversité de la société civile afin de juger de l'acceptabilité ou de la légitimité de certains contenus[43].

Une piste de solution pourrait être la mise en place d'une structure consultative reflétant la diversité de la société civile qui rendrait des avis sur le caractère acceptable ou légitime de certains contenus en ligne. Sans remplacer le juge judiciaire et en assistant à la prise de décision par les plateformes, une telle structure permettrait d'élaborer et de formuler du consensus social sur des contenus à fort impact.

Le DSA ne ferait pas obstacle à de telles initiatives ; bien au contraire, sa mise en œuvre pourrait permettre de développer en France un cadre plus complet pour la gestion des contenus en ligne.

Pascal Beauvais
Janvier 2023

Notes

[1] Considérant n°8 du DSA.
[2] Article 8 du *DSA*.
[3] Considérant n°19.
[4] Article 4.
[5] Article 5.
[6] Article 6.
[7] Article 6.
[8] Considérant n°18.
[9] Proposition de règlement du Parlement européen et du Conseil établissant un cadre commun pour les services de médias dans le marché intérieur (législation européenne sur la liberté des médias) et modifiant la directive 2010/13/UE, COM/2022/457 final, SWD (2022) 286 final, SWD (2022) 287 final.
[10] Considérant n°26.
[11] Considérant n°26.
[12] Article 7.
[13] La définition et les contours précis des contenus illicites continuent de relever principalement des législations nationales, et bien souvent des jurisprudences casuistiques des juridictions étatiques compétentes. En France, la jurisprudence de la Cour de cassation sur le fondement notamment de la loi du 29 juillet 1881 et la Cour européenne des droits de l'homme joue un rôle déterminant dans la détermination des contenus illicites.
[14] A. Aulas, M. Le Masne de Chermont, « Modération des contenus par les plateformes, quelles obligations

pour demain ? », *Revue Lamy, Droit de l'immatériel*, 2021, n°181.

[15] A. Aulas, M. Le Masne de Chermont, « Modération des contenus par les plateformes, quelles obligations pour demain ? », *Revue Lamy, Droit de l'immatériel*, 2021, n°181.

[16] Article 20 du *DSA*.

[17] Article 21.2.

[18] Article 14.4.

[19] Article 9.

[20] Article 9.2.

[21] Article 9.5.

[22] Article 16.

[23] Article 16.3.

[24] Article 17.3.

[25] Article 20.4 et 5

[26] Article 23.

[27] Article 36.

[28] Article 36.

[29] Certains sont toutefois en dehors de la loi de 1881, comme l'apologie du terrorisme qui figure dans le code pénal.

[30] Sur ce sujet voir également, *L'équilibre de la loi du 29 juillet 1881 à l'épreuve d'Internet*, Rapport d'information du Sénat de MM. F. Pillet et T. Mohamed Soilihi fait au nom de la commission des lois n° 767 (2015-2016), 6 juillet 2016 ; Conseil d'État, *Les réseaux sociaux : enjeux et opportunités pour la puissance publique*, Étude annuelle 2022.

[31] E. Raschel, *La procédure pénale en droit de la presse*, Lextenso, coll. « Guide pratique », 2019, n° 4, p. 16.

[32] L'article 6 de la LCEN prévoit un mécanisme complexe de responsabilité en cascade.

[33] Rendre la justice aux citoyens, Rapport du comité des États généraux de la Justice, avril 2022. Dans la continuité des États généraux de la Justice, le 5 janvier 2023, le **garde des Sceaux, Éric Dupond-Moretti a présenté** le Plan d'action pour une justice plus rapide et plus efficace. Ce plan prévoit notamment une hausse historique des moyens humains et financiers, des mesures novatrices en matière civile et une refonte de la procédure pénale. Un comité scientifique de suivi des travaux sera constitué et les parlementaires seront associés pour suivre les travaux et préparer l'examen législatif.

[34] CE, Section du rapport et des études, *Inventaire méthodique et codification du droit de la communication*, La Documentation française, 2006.

[35] À titre d'exemple, Cass. crim., 21 avril 2020, n°19-81172.

[36] A. Aulas, M. Le Masne de Chermont, « Modération des contenus par les plateformes, quelles obligations pour demain ? », *Revue Lamy, Droit de l'immatériel*, 2021, n°181.

[37] G. Loiseau, « Services de partage de contenus en ligne - Contenus illicites : la responsabilité des plateformes certifiée conforme », Communication, Commerce électronique n° 6, juin 2022, comm. 42.

[38] G. Loiseau, « Services de partage de contenus en ligne - Contenus illicites : la responsabilité des plateformes certifiée conforme », *Communication, Commerce électronique* n° 6, juin 2022, comm. 42.

[39] Article 21du *DSA*.

[40] A. Aulas, M. Le Masne de Chermont, « Modération des contenus par les plateformes, quelles obligations pour demain ? », Revue Lamy, Droit de l'immatériel, 2021, n°181.

[41] A. Aulas, M. Le Masne de Chermont, *« Modération des contenus par les plateformes, quelles obligations pour demain ?* », Revue Lamy, Droit de l'immatériel, 2021, n°181.

[42] *Rendre la justice aux citoyens*, Rapport du comité des États généraux de la Justice, avril 2022, p. 36 et 74.

[43] *Le numérique pour la justice*, rapport remis au comité des États généraux de la Justice le 17 mars 2022.

[44] Par exemple l'initiative « Social Media Council » de l'ONG Article 19 ou l'Autorité de Régulation Professionnelle de la Publicité (ARPP).

Annexe 3

Panorama des technologies d'observation, de filtrage et de marquage des contenus

Les technologies ont permis une démocratisation de la liberté d'expression mais avec les dérives et les effets délétères désormais bien connus. Cette note a pour objectif de démontrer qu'elles offrent également des leviers puissants pour y remédier, notamment pour assurer une meilleure effectivité de la loi dans l'espace numérique. À condition toutefois qu'elles soient **correctement identifiées, comprises et mobilisées de manière transparente, concertée et intelligente** par l'ensemble des acteurs publics et privés.

Ces technologies ont vocation à être au cœur du nouvel espace démocratique et numérique. Leur mode de fonctionnement doit être accessible et pouvoir faire l'objet de recherches et d'évaluations externes sur leurs effets et leur impact.
La note est construite sous forme d'un panorama de ces technologies regroupées en trois grandes catégories :

1. Les technologies d'observation, telles que le *social listening*, qui permettent de capter et d'analyser les dynamiques sociales en ligne ;
2. Les technologies de filtrage sur les plateformes numériques ;
3. Les technologies de marquage qui visent à authentifier des contenus.

Comme indiqué dans le rapport dont cette note est l'annexe :

- Une utilisation efficace et responsable de ces technologies nécessite une **approche coordonnée voire interopérable** afin d'apporter des réponses cohérentes à l'ensembles des enjeux identifiés ;
- Ces technologies d'observation, de filtrage et de marquage étant susceptibles de porter atteinte à la vie privée et produire des effets dissuasifs sur la liberté d'expression, il est important que leur usage soit encadré : consentement effectif des utilisateurs, limitation des finalités, transparence, lorsque nécessaire
- Elles ne peuvent pas faire l'économie d'une intervention humaine pour limiter les risques légaux et éthiques (risques de censure, de sur-modération, de biais, manque de transparence) ou techniques (compatibilité des outils avec les services des différentes plateformes, accessibilité, besoin de compétences spécifiques).

Ces technologies doivent donc être envisagées non comme des solutions automatiques, mais comme **des outils au service des choix et des politiques portés par les sociétés humaines**.

1 | Les technologies d'observation

L'observation des réseaux sociaux, ou le *social listening*, consiste à identifier et à analyser en temps réel les conversations publiques sur les réseaux sociaux, les forums ou encore les blogs. Largement utilisée par les grandes marques pour évaluer leur visibilité et leur e-réputation, elle est aussi exploitée par des institutions publiques et les gouvernements afin de détecter rapidement et d'anticiper des tendances sociales ou des crises informationnelles émergentes. Cette approche proactive offre une compréhension de l'opinion publique et des dynamiques sociales en ligne.

Un écosystème d'entreprises spécialisées se sont développées dans le domaine, comme Talkwalker, Bloom, Meltwater ou encore Synthesio.

En France, le Service d'information du gouvernement (SIG) a sélectionné plusieurs de ces entreprises au terme d'un marché public clos début 2025.

:: Talkwalker

Description : Talkwalker est une entreprise luxembourgeoise spécialisée dans le social listening et l'analyse des données conversationnelles en ligne. Sa plateforme exploite des technologies d'intelligence artificielle (IA) et de traitement automatique du langage naturel (NLP) pour surveiller, collecter et analyser en temps réel les mentions d'une marque, d'un produit ou d'un sujet à travers plus de 150 millions de sites web et plus de 30 réseaux sociaux. L'outil permet d'identifier les tendances émergentes, de mesurer la perception du public, et de suivre la réputation d'une marque ou d'une institution. Il intègre également des fonctionnalités d'analyse d'images (*image recognition*) et d'analyse de sentiment (*sentiment analysis*), afin de comprendre le ton et les émotions associées aux conversations en ligne.

Usages : Talkwalker est utilisé par des entreprises, agences et organisations pour leurs besoins en veille stratégique, gestion de la réputation, communication de crise et analyse de performance marketing. Parmi ses cas clients publics, on trouve par exemple Yves Rocher, STEF Group et l'UNICEF MENA, qui utilisent la plateforme pour analyser les tendances, suivre la perception du public ou lutter contre la désinformation.

Limites : Comme tout outil de *social listening*, Talkwalker présente certaines limites techniques et contextuelles. L'analyse automatique du ton, de l'ironie ou des références culturelles peut parfois générer des biais. La dépendance aux données accessibles via les interfaces logicielles « API »[1] des plateformes sociales peut restreindre l'accès à certaines informations, notamment sur les contenus privés ou éphémères. La qualité des analyses dépend fortement de la précision des requêtes et du paramétrage des filtres choisis par l'utilisateur.

:: Bloom

Description : Bloom est une entreprise française spécialisée dans l'analyse de l'espace informationnel à l'aide de l'IA. Cet outil s'appuie sur un algorithme d'inférence sociale capable de

cartographier les écosystèmes de discussions en ligne en se déployant de lien en lien sur les réseaux sociaux, sans passer par les interfaces de programmation (APIs) des plateformes. Cette technologie est capable de détecter des signaux faibles et d'en dégager des dynamiques d'opinion autour d'un sujet donné, même lorsque celui-ci n'est pas explicitement mentionné. Grâce à un système sémantique avancé et à la détection des émotions générées, elle peut identifier des sujets émergents sur les réseaux sociaux issus des interactions entre internautes.

Usages : La technologie de Bloom est utilisée par le gouvernement et par des organisations internationales comme l'OTAN, mais aussi par des entreprises comme L'Oréal et le groupe LVMH et par des associations comme e-Enfance. Cet outil aide ses utilisateurs à améliorer leur compréhension des dynamiques sociales en ligne et permet aux entreprises de renforcer la sécurité de leur marque (*brand safety*).

Limites : L'émergence des *algospeaks*, langage codé utilisant abréviations, lettres modifiées ou inversées, euphémismes ou métaphores destinées à contourner le filtrage automatique des plateformes, constituent un défi croissant pour les dispositifs de *social listening* tels que ceux déployés par Bloom. Conçues pour contourner les mécanismes de modération des plateformes, ces stratégies de contournement rendent plus complexe la détection des tendances conversationnelles en ligne.

:: Meltwater

Description : Meltwater est une entreprise d'origine norvégienne, ayant aujourd'hui son siège social aux Etats-Unis, spécialisée dans la veille médiatique, l'analyse des réseaux sociaux et l'intelligence consommateur. Meltwater propose une plateforme de *Media Intelligence* qui permet aux entreprises de suivre, analyser et comprendre les conversations en ligne à travers une variété de sources, notamment les actualités en ligne, les réseaux sociaux, les forums, les blogs, les médias imprimés, les émissions de télévision et de radio, ainsi que les podcasts. La plateforme intègre des outils d'IA pour fournir des analyses en temps réel, des alertes intelligentes et des rapports personnalisés, facilitant ainsi la prise de décision stratégique pour les départements de communication, de marketing et de relations publiques.

Usages : Meltwater est utilisé par plus de 27 000 clients dans plus de 125 pays, couvrant divers secteurs tels que les médias, le e-commerce, le luxe, la santé, le gaming et les institutions publiques. L'entreprise propose de la veille médiatique, du *social listening*, de l'analyse de l'audience, de la gestion de la réputation et marketing d'influence : identification des influenceurs pertinents et gestion des collaborations.

Limites : Bien que Meltwater offre une large couverture médiatique, la qualité des données peut varier en fonction des sources et des paramètres de recherche définis. L'interprétation des données nécessite une expertise pour en tirer des conclusions stratégiques pertinentes.

:: Ipsos Synthesio

Description : Ipsos Synthesio est une plateforme française spécialisée dans la veille sur les réseaux-sociaux et la e-réputation. La solution collecte et analyse des données conversationnelles provenant de nombreux canaux, tels que les réseaux sociaux, blogs, forums et sites d'actualités, dans plus de 195 pays et en plus de 80 langues. Elle s'appuie sur des technologies d'IA, de traitement automatique du langage naturel (NLP) et de détection d'émotions pour structurer ces données, identifier des tendances émergentes et aider les marques à comprendre les dynamiques de conversations en ligne.

Usages : Ipsos Synthesio est utilisée par des entreprises et organisations pour surveiller la réputation de leur marque et détecter des signaux de crise, mesurer la performance de leurs campagnes sociales et comparer leur part de voix sur les réseaux sociaux, ainsi qu'identifier des insights consommateurs, des besoins émergents ou des tendances de marché à partir de l'analyse des données sociale.

Limites : L'accès aux informations dépend des sources publiques disponibles, la détection de l'ironie, du ton ou du contexte culturel spécifique peut poser problème, et la qualité des analyses dépend fortement de la pertinence des filtres, des mots-clés et du paramétrage choisi par l'utilisateur.

2 | Les technologies de filtrage

La **modération** est un enjeu majeur pour les plateformes en ligne, visant à tracer les limites de l'espace du débat. La modération permet d'encadrer les discussions ou les contenus produits par les utilisateurs d'une plateforme au sein d'un espace d'échange[2], en déplaçant ou supprimant le commentaire d'un utilisateur qui ne respecterait pas les limites de la liberté d'expression dictées par le droit. Elle permet ainsi l'identification, l'analyse et le filtrage des contenus.

De nombreuses solutions technologiques permettent **d'identifier, de filtrer et de faire remonter des contenus potentiellement illégaux ou problématiques en ligne**. Parmi celles-ci figurent Bodyguard, Atchik, Netino ou encore Semiologic, dont les usages couvrent un large éventail d'environnements numériques, allant des réseaux sociaux aux forums en ligne, en passant par les espaces de commentaires des sites d'actualités.

:: Bodyguard

Description : Bodyguard est une solution technologique de filtrage française s'appuyant sur l'IA et sur des règles linguistiques définies par des experts. A l'origine, Bodyguard a été développée pour protéger les personnes physiques du cyberharcèlement. La technologie s'est ensuite recentrée sur les besoins des entreprises. Son système de détection des propos haineux repose sur un moteur de règles (de l'IA symbolique) et également sur les nouvelles technologies d'intelligence artificielle générative et le *machine learning*, rendant l'IA capable d'identifier, d'analyser, de classifier et de retirer le contenu. Les contenus problématiques sont catégorisés selon les sujets abordés

(racisme, misogynie, harcèlement, etc.) en fonction du contexte -quelle plateforme, quel type de communication et pour quel client-, et la configuration du filtrage dans l'outil est ajustée en fonction de la politique définie par le client sur chacun de ces sujets. Par exemple, la solution de modération de Bodyguard aide les marques de luxe à éliminer les contenus nuisibles à leur image, pour garantir une représentation irréprochable de la marque ainsi que des interactions en ligne plus sûres pour ses clients.

Usages : Utilisée dans les secteurs des médias, du luxe, du sport et du jeu vidéo, la technologie de Bodyguard intervient à trois niveaux, à travers :

- La modération des comptes de la marque ou de l'entreprise sur les réseaux sociaux. L'entreprise va donner accès à Bodyguard à ses comptes, permettant à cet outil de faire une gestion automatisée de la modération des contenus selon des règles de modération préalablement définies avec l'entreprise cliente ;
- La protection des salariés en filtrant et en limitant l'exposition à des contenus haineux ;
- L'intégration technologique aux infrastructures internes des entreprises.

En pratique, la plateforme traite environ trois milliards de contenus par mois, en 45 langues différentes. Bodyguard est en mesure de supprimer un contenu inapproprié posté sur le compte d'un utilisateur de plateforme avant même que les visiteurs du compte de cet utilisateur aient eu la possibilité d'y être exposés. L'outil traite aussi bien les images, que les textes, comprenant les contenus vidéo.

Limites : Les risques associés à un outil tel que Bodyguard sont relatifs à la sur-modération, c'est-à-dire au retrait excessif ou injustifié de contenus jugés problématiques.

:: Semiologic

Description : Semiologic est une entreprise française qui propose un service dénommé Graphcomment, une interface conçue pour améliorer la qualité des discussions en ligne. Elle est dotée d'une interface graphique innovante, le *Bubble Flow*, qui créé un fil de discussion dynamique et couplée à un algorithme de la pertinence et le *Bubble Rank*, destiné à structurer les messages des discussions entre un nombre infini de personne. Chaque commentaire est évalué selon un système de votes par pondération (positif, neutre, négatif) et influencé par la réputation des votants. Des critères comme la structure du message, l'orthographe et des émotions de vote (« pertinent », « troll », « j'aime », etc.) affinent encore cette évaluation. L'outil utilise un filtre de toxicité pour détecter et bloquer les messages offensants ou nuisibles (insultes, propos haineux, menaces, etc.) Cet outil utilise l'intelligence artificielle afin d'analyser le langage et attribuer un score de toxicité à un texte. Au-delà de l'analyse des mots, Graphcomment opère une modération contextualisée (il permet de repérer un mot qui a un double sens ou qui est utilisé avec de l'ironie). En cas de doute, le commentaire est renvoyé à la modération humaine.

Usages : Graphcomment est utilisé dans les espaces de commentaires de médias en ligne

comme Les Échos ou Orange. Il est déployé dans 50 pays, en 20 langues et trois types de modération sont disponibles :

- La pré-modération, automatisée par une IA destinée à pré-modérer les propos ;
- La post-modération, qui repose sur le signalement par les utilisateurs ;
- La smart-modération, plus contextuelle.

L'outil vise à favoriser un climat d'échange respectueux et auto-régulé, où les utilisateurs peuvent débattre tout en maintenant un certain équilibre, sans intervention directe de la plateforme.

Limites : Le système repose sur des algorithmes et sur la **notion de réputation**, ce qui peut poser des questions d'équité pour les nouveaux utilisateurs ou ceux qui tiennent des opinions minoritaires mais toutefois légitimes. Par ailleurs, une modération automatisée pourrait être sujette à erreurs d'interprétation ou à des biais.

:: Atchik

Description : Atchik est une entreprise française, spécialisée dans la modération de contenus en ligne, la veille numérique et la gestion de communautés sur les réseaux sociaux. L'entreprise propose des services de modération humaine, appuyés par des outils technologiques internes comme Lokus, une plateforme permettant de centraliser et d'analyser les messages, commentaires et mentions provenant de différents espaces numériques. Atchik se présente comme un acteur de la « conversation responsable », prônant une approche éthique et qualitative de la modération, avec une équipe basée en France et disponible en continu.

Usages : Les solutions d'Atchik sont utilisées par des médias francophones (tels que France Télévisions, Radio France, Ouest-France, 20 Minutes, TV5Monde ou L'Équipe), ainsi que par certaines institutions publiques et entreprises privées.

Limites : Atchik reste dépendante d'un modèle majoritairement humain, ce qui garantit une lecture fine du contexte mais peut limiter la capacité de gérer des volumes massifs de messages. Bien que la société évoque l'usage d'outils internes et de technologies d'aide à la modération, les détails sur la part d'automatisation ou les algorithmes utilisés ne sont pas rendus publics. La modération humaine demeure un métier exigeant pour ceux qui la pratiquent, soulevant des enjeux de bien-être et de conditions de travail déjà relevés dans plusieurs études du secteur.

:: Netino

Description : Netino, filiale du groupe français Webhelp racheté par l'entreprise américaine Concentrix en 2023, est une entreprise spécialisée dans la modération de contenus. Elle propose un ensemble de services destinés à accompagner les marques, les médias et les institutions dans la maîtrise de leur image en ligne. L'outil principal de Netino repose sur une combinaison d'IA et d'intervention humaine : les algorithmes détectent les contenus problématiques (propos haineux,

spams, fake news, etc.), tandis qu'une équipe de modérateurs assure une validation humaine des décisions sensibles. Netino met aussi à disposition une plateforme de supervision permettant aux clients de suivre en temps réel les performances de modération (taux de détection, temps de réponse, répartition par typologie de contenu). Les solutions s'intègrent avec les principaux réseaux sociaux (Facebook, Instagram, YouTube, TikTok, X/Twitter) et sites de médias.

Usages : Les services de Netino sont utilisés par de grands médias, institutions et marques (comme France Télévisions, TF1, Le Monde, ou encore des institutions publiques).

Trois principaux types d'usages sont proposés :

- Modération préventive (pré-modération) : les contenus sont filtrés avant publication à l'aide d'algorithmes d'IA couplés à une validation humaine pour les cas ambigus ;
- Modération réactive (post-modération) : les messages sont analysés après publication, avec signalement automatique ou manuel selon des seuils de gravité ;
- Social Media Care : Netino assure également le suivi des interactions, réponses aux utilisateurs et gestion de crise en ligne.

Limites : Malgré l'efficacité de la combinaison IA-humaine, la détection automatique peut rester imparfaite, notamment sur les contenus ambigus, ironiques ou contextuels. La dépendance à des algorithmes de filtrage peut entraîner des biais culturels ou linguistiques, impactant la neutralité de la modération.

3 | Les technologies de marquage

Les technologies de marquage peuvent prendre diverses formes. Le ***watermarking***, tel que développé par IMATAG, Google, ou Microsoft repose sur l'intégration d'identifiants ou de filigranes invisibles pour les utilisateurs directement dans un contenu numérique ou dans ses pixels (images, vidéos, textes). Complémentaires des outils de filtrage et d'observation, d'autres technologies, comme celle développée par NewsGuard, évaluent la fiabilité des sources d'information via des critères journalistiques, contribuant à renforcer la confiance dans l'écosystème informationnel en ligne, l'application Ask Vera propose également une solution de vérification de faits en temps réel.

Le watermarking peut être très utile pour tracer un contenu, en revanche son utilité est moindre dans le cadre de la modération car la majorité des contenus n'est pas marquée. Le marquage gagne tout son intérêt si la plateforme cherche à publier seulement des contenus certifiés ou pour tracer un contenu.

:: IMATAG

Description : IMATAG est une entreprise spécialisée dans le tatouage numérique (*digital watermarking*), une technologie de marquage invisible et infalsifiable appliquée aux images, qui

intègre des identifiants uniques et imperceptibles directement dans les pixels des images, des vidéos et des PDF. Sa solution, IMATAG Authenticity, a été conçue pour vérifier et protéger l'authenticité des contenus visuels diffusés sur différentes plateformes numériques. Cette technologie peut identifier l'origine d'un contenu (texte ou image) et détecter ses manipulations. Elle s'inscrit également dans des initiatives internationales telles que la *Coalition for Content Provenance and Authenticity*[3] (C2PA) qui vise à associer des métadonnées sécurisées aux contenus numériques.

Usages : La solution d'IMATAG est utilisée par des créateurs de contenu, des plateformes de réseaux sociaux ou encore des agences de presse telles que l'AFP ou Reuters. Son système comprend trois composants clés :

- L'intégration d'un tatouage numérique intégré au contenu au moment de sa création ou de sa publication ;
- La détection et l'analyse du tatouage numérique même en cas de modification importante du contenu original ;
- La vérification de l'authenticité du contenu, permettant d'accéder aux informations de source d'origine.

Limites : Malgré son expertise technologique, la solution d'IMATAG est confrontée à une prolifération des contenus générés par l'IA qui posent un défi majeur pour la traçabilité et la vérification.

:: SynthID de Google DeepMind

Description : SynthID est une technologie développée par Google DeepMind qui permet d'insérer des filigranes (*watermarks*) numériques invisibles dans des contenus générés par IA (textes, images, vidéos, fichiers audio). Pour les images, le filigrane est intégré dans les pixels de l'image sans altérer l'image en elle-même de manière perceptible. Pour le texte, l'outil ajuste les probabilités de sélection des mots lors de leur génération, créant ainsi un motif identifiable sans altérer la qualité ou le sens du contenu. Cette méthode permet de détecter si un contenu a été produit par un modèle d'IA même après des modifications mineures.

Usages : SynthID est utilisé par les clients de Google Cloud qui utilisent, la plateforme Vertex AI de l'entreprise, le générateur d'images Imagen ou Google's AI tools. Google a également ouvert le code source de SynthID pour le texte (SynthID Text) pour mettre à la disposition des développeurs cette technologie qu'ils peuvent intégrer dans leurs propres modèles d'IA. L'objectif est d'étendre l'utilisation de l'outil afin que les acteurs utilisent les mêmes standards.

Limites : L'efficacité de SynthID est réduite sur les textes courts ou très factuels puisque les possibilités d'ajustement sont limitées sans compromettre l'exactitude. Aussi, les scores de confiance du détecteur peuvent être réduits lorsqu'un texte généré par IA est entièrement réécrit ou traduit dans une autre langue. Par ailleurs, l'outil ne peut pas toujours vérifier l'exactitude des

métadonnées.

:: Microsoft PhotoDNA

Description : **PhotoDNA** est une technologie développée par Microsoft qui repose sur la création d'une signature numérique unique, le hachage, à partir d'images, qui est ensuite comparée aux signatures (hachages) d'autres photos afin de trouver des copies de la même image. Lorsque l'outil est associé à une base de données contenant les hachages d'images illégales précédemment identifiées, PhotoDNA est un outil qui permet de détecter et de signaler la diffusion de contenus illégaux comme les images d'exploitation sexuelle d'enfants (CSAM).

Usages : A l'origine, Microsoft utilisait cet outil sur ses propres services, comme Bing et OneDrive. Depuis 2022, PhotoDNA est également utilisé par des fournisseurs de services en ligne comme Gmail (Google), X (anciennement Twitter), Adobe, Facebook ou encore Reddit. Des organisations gouvernementales et des ONG, comme le *National Center for Missing & Exploited Children* et l'*Internet Watch Foundation*, utilisent également cette technologie. PhotoDNA est très utilisé comme système de hachage des contenus pédocriminels afin d'automatiser leur modération. Microsoft propose PhotoDNA sous forme de service *cloud* gratuit via Azure.

Limites : PhotoDNA repose sur une base de données d'images connues et ne peut donc pas identifier de nouveaux contenus illégaux non répertoriés. L'outil est également moins efficace face à certaines modifications comme des rotations à 90 degrés, des inversions horizontales ou des changements de contraste, pouvant altérer la détection.

:: NewsGuard

Description : NewsGuard utilise le journalisme comme levier de lutte contre la désinformation en ligne, en proposant notamment un système d'évaluation de la fiabilité des sources d'information (presse, sites, blogs), basé sur les neuf critères apolitiques suivants qui évaluent la crédibilité et la transparence d'un site qui :

- Ne publie pas de contenus faux ou manifestement trompeurs de manière répétée ;
- Recueille et présente l'information de façon responsable ;
- Dispose de procédures efficaces pour corriger les erreurs ;
- Gère de manière responsable la différence entre informations et opinions ;
- Evite les titres trompeurs ;
- Indique à qui il appartient et comment il est financé ;
- Identifie clairement la publicité ;
- Indique qui est responsable des contenus et tous conflits d'intérêt possibles ;
- Fournit des informations sur les créateurs de contenu.

Chaque site analysé reçoit un score de confiance entre 0 et 100, accompagné d'une « Etiquette Nutritionnelle » informative. Bien que NewsGuard s'appuie sur des outils d'IA pour le traitement

des données, l'analyse est entièrement réalisée par des journalistes. Les évaluations sont accessibles via un tableau de bord, une API[4] ou un flux de données sur le cloud.

Par ailleurs, NewsGuard contribue à l'interopérabilité des outils technologiques en ligne avec sa solution « Empreintes de la Mésinformation ». Ces Empreintes fournissent un catalogue des principales infox qui circulent en ligne dans un format qui peut être utilisé pour nourrir des outils existants d'IA ou de *social listening*, tel que Bloom, pour suivre la trace de fausses informations sur Internet et les réseaux sociaux. Leur base de données peut se révéler très utile pour entrainer d'autres outils.

Usages : Les outils de NewsGuard sont conçus pour servir à la fois les consommateurs d'information et les acteurs de l'écosystème numérique (instituts de recherches, plateformes, agrégateurs de contenu, entreprises publicitaires, fournisseurs d'IA). Les évaluations de fiabilité couvrent plus de 10 000 sites Internet, 35 000 éditeurs dans le monde entier et couvrent les sources d'information qui représentent plus de 95% de l'engagement avec l'actualité en ligne. Pour le grand public, l'entreprise propose aussi une extension de navigateur permettant de naviguer en ligne tout en connaissant la fiabilité des sites visités, disponible gratuitement dans les bibliothèques publiques grâce à un partenariat avec Microsoft.

Limites : L'analyse humaine reste indispensable pour saisir les nuances et détecter les cas ambigus. Le développement de « robots plagiaires », capables de reproduire et détourner des contenus existants, représente un enjeu croissant.

:: Ask Vera

Description : Ask Vera est une technologie d'intelligence artificielle dédiée à la vérification des faits en temps réel, accessible via un numéro de téléphone ou un message WhatsApp. Ask Vera agit comme un assistant de fact-checking, capable de répondre à des questions sur la véracité d'affirmations circulant en ligne ou dans les discussions, en se basant sur une large base de sources fiables (sites de fact-checking certifiés et médias reconnus). L'IA sous-jacente est alimentée par des modèles de langage (dont GPT-4) et se connecte à des centaines de sources pour fournir des réponses sourcées et neutres plutôt que des suppositions générées sans vérification. Vera a été développée par l'ONG française LaReponse.tech.

Usages : la solution Vera est utilisée pour :

- Vérifier la véracité d'informations ou de rumeurs rapidement ;
- Accéder à des faits sourcés provenant de plus de 300 sites de fact-checking et médias fiables (signataires des chartes IFCN, EFCSN, etc.), ce qui permet de contrer la désinformation en fournissant des preuves documentées plutôt que des opinions.
- Renforcer l'éducation aux médias et l'esprit critique pour les utilisateurs.

Limites : Ask Vera ne peut vérifier qu'une information déjà analysée par au moins une des sources qu'elle interroge. Si aucune source fiable n'a publié de vérification, la réponse peut être incomplète ou indiquer que la vérification n'est pas possible. La technologie ne contient pas non plus de détecteur automatique de fausses images ou vidéos, Ask Vera se concentre sur le texte des affirmations et leur véracité, pas sur l'authenticité technique des fichiers multimédias. Par ailleurs, les réponses peuvent parfois être influencées par la formulation de la question ou par la disponibilité limitée de sources sur certains sujets très récents ou spécialisés.

:: Google CSAI Match

Description : CSAI Match (Child Sexual Abuse Imagery Match) est une API proposée par Google. Elle permet de détecter automatiquement les vidéos contenant des contenus d'abus sexuels sur enfants déjà connus. L'API utilise une technologie de *fingerprinting* vidéo qui génère une empreinte numérique unique pour chaque fichier vidéo. Cette empreinte est ensuite comparée à une base de données de contenus identifiés, ce qui permet de repérer des correspondances même si les vidéos ont été modifiées, recadrées ou partiellement altérées. L'approche combine efficacité algorithmique et réduction de l'exposition humaine directe à des contenus sensibles, offrant un moyen automatisé pour les plateformes de gérer ces contenus.

Usages : CSAI Match est utilisé par YouTube, Snapchat, Adobe, Reddit, Yahoo et d'autres plateformes pour détecter et supprimer les contenus d'abus sexuels sur enfants. L'API facilite également le travail des ONG partenaires, comme Safernet Brasil, en réduisant le temps consacré à la révision manuelle des contenus. Les principales fonctionnalités incluent :

- La détection automatisée de contenus déjà identifiés comme illégaux ;
- Le signalement et blocage rapide des vidéos correspondantes ;
- La prise en charge de vidéos partiellement modifiées ou altérées.

L'outil vise à protéger les utilisateurs et les enfants en ligne tout en soutenant la conformité légale des plateformes.

Limites : L'efficacité de CSAI Match dépend de la qualité et de la mise à jour de la base de données de contenus connus. L'API ne permet pas de détecter de nouveaux contenus entièrement inédits. De plus, l'intégration technique nécessite des ressources spécialisées, et l'accès est limité aux partenaires approuvés par Google, ce qui restreint son usage direct aux entreprises et ONG autorisées.

---

[1] Une API (Application Programming Interface, ou Interface de Programmation Applicative en français) est un ensemble de règles et de protocoles qui permet à des logiciels différents de communiquer entre eux et d'échanger des données ou des services.

[2] Romain Badouard, *Les enjeux de la modération des contenus sur le web*. La Revue Européenne des Médias et du Numérique. https://la-rem.eu/2021/11/les-enjeux-de-la-moderation-des-contenus-sur-le-web/

[3] La *Coalition for Content Provenance and Authenticity* (C2PA) est une initiative lancée en février 2021 par des entreprises telles que Microsoft, Intel ou Adobe. Elle propose un standard ouvert et interopérable permettant d'attester de l'origine et de l'intégrité

des contenus numériques (images, vidéos, audio, texte) grâce à l'intégration de métadonnées cryptographiquement sécurisées directement dans les fichiers numériques.

Annexe 4

Lexique

« Mal nommer les choses, c'est ajouter au malheur du monde ! » - Albert Camus

Dans le cadre de nos travaux, nous avons constaté que certains termes - souvent utilisés dans les débats publics et médiatiques - manquent de définition claire ou sont employés avec des sens différents ou des contresens. Ce lexique a pour objectif de rappeler les définitions de ces notions, afin de favoriser une discussion mieux informée et plus rigoureuse.

Glossaire :

1. Complotisme

Le complotisme désigne une manière d'interpréter les événements sociaux, politiques, scientifiques ou historiques comme le résultat d'un complot orchestré par un groupe occulte, puissant et malveillant. Il s'appuie sur l'idée que des forces cachées manipulent la réalité en secret, indépendamment des preuves disponibles.

Le complotisme se caractérise par une méfiance systématique envers les institutions, les médias ou les connaissances établies, ainsi que par une tendance à rejeter les explications fondées sur des faits au profit de récits alternatifs. S'il n'est pas en soi illégal, il peut contribuer à diffuser des

fausses informations ou à alimenter la défiance sociale, voire à encourager des comportements dangereux lorsque ces croyances incitent à la violence ou à la discrimination.

2. Contenus illicites

Le Règlement européen sur les Services Numériques (***Digital Services Act* - DSA)**, adopté en 2022, définit un **contenu illicite** comme tout contenu – texte, image, vidéo, service ou produit – qui est **illicite au regard du droit national ou européen** applicable dans un État membre[1]. Il ne crée donc pas une liste uniforme de contenus interdits, mais **renvoie aux législations en vigueur** (ex. : code pénal, loi du 29 juillet 1881, code de la propriété intellectuelle, Règlement Général sur la Protection des données personnelles (RGPD)…)

Ce contenu doit correspondre de manière générale aux règles en vigueur dans l'environnement hors ligne[2].

Le DSA impose aux plateformes de réagir rapidement lorsqu'un tel contenu est signalé ou constaté, en mettant en place des **procédures de retrait transparentes et accessibles**, sans pour autant remplacer les autorités judiciaires. Il structure également des voies de **recours pour les utilisateurs** concernés par ces décisions.

Il inclut, par exemple les contenus illicites suivants :

- **Apologie d'un crime ou du terrorisme**

L'apologie consiste à présenter favorablement ou à valoriser un crime ou un acte de terrorisme, de manière à susciter l'adhésion ou à légitimer de tels actes. L'infraction est constituée lorsque les propos sont tenus publiquement - *article 24, alinéa 5, de la loi du 29 juillet 1881 (apologie de crimes) et article 421-2-5 du Code pénal (apologie du terrorisme).*
Exemple : déclarer publiquement qu'un attentat est justifié ou courageux constitue une apologie d'acte de terrorisme.

- **Atteinte à la vie privée**

L'atteinte à la vie privée désigne toute intrusion injustifiée dans la sphère personnelle d'une personne, incluant ses informations, son image ou son intimité. Elle sanctionne la divulgation, la collecte ou l'exploitation de données privées sans consentement.

- **Contrefaçon**

La contrefaçon est la reproduction, l'imitation ou l'utilisation non autorisée d'une œuvre (musique, vidéo, dessin, livre…), d'une marque déposée, ou de tout autre élément protégé par un droit de propriété intellectuelle. Elle constitue une infraction civile et pénale lorsqu'elle est réalisée sans l'accord du propriétaire légitime.

- **Diffamation**

La diffamation consiste en l'allégation ou l'imputation d'un fait précis portant atteinte à l'honneur ou à la considération d'une personne physique ou morale. Elle peut être publique (proférée devant un large public) ou non publique (dans un cadre privé). La véracité du fait ou la bonne foi de l'auteur peut constituer un moyen de défense - *art. 29 alinéa 1, de la loi du 29 juillet 1881 sur la liberté de la presse*.

Exemple : accuser publiquement une personne de détourner des fonds sans preuve.

- **Harcèlement**

Le harcèlement est défini par le Code pénal comme le fait de faire subir à une personne des propos ou comportements répétés ayant pour objet ou pour effet une dégradation de ses conditions de vie, portant atteinte à sa dignité ou altérant sa santé physique ou mentale (article 222-33-2-2 du Code pénal). Il peut s'agir de harcèlement moral, sexuel ou scolaire. Lorsqu'il est commis en ligne, notamment via des réseaux sociaux, messageries ou forums, on parle alors de cyberharcèlement, qui constitue une circonstance aggravante.

- **Injure**

L'injure est une expression outrageante, un terme de mépris ou une invective ne contenant l'imputation d'aucun fait précis. Elle peut être publique (proférée devant un large public) ou non publique (dans un cadre privé) – *art. 29, alinéa 2, de la loi du 29 juillet 1881*.

Exemple : traiter une personne de « sale con » en public, sans faire référence à un fait précis.

- **Pédocriminalité**

La pédocriminalité désigne l'ensemble des infractions sexuelles commises à l'encontre des mineurs, incluant les agressions, viols, actes de corruption ou d'exploitation. Sur les réseaux sociaux, elle peut prendre la forme d'approches prédatrices (grooming), de sollicitations sexuelles, de chantage ou de partage d'images pédopornographiques.

- **Provocation à la haine**

La provocation à la haine vise tout discours, écrit ou propos incitant publiquement à la haine, à la violence ou à la discrimination envers une personne ou un groupe en raison de caractéristiques telles que l'origine, la religion, le sexe, l'orientation sexuelle ou le handicap – *art. 24 de la loi du 29 juillet 1881*.

Exemple : tenir publiquement des propos appelant à l'exclusion ou à la violence envers un groupe en raison de son origine ou de sa religion constitue une provocation à la haine.

3. Contenus inappropriés

Un contenu inapproprié est un contenu qui, sans être nécessairement illicite, peut être considéré comme choquant, offensant, perturbant ou contraire aux règles d'usage d'une plateforme. Sa qualification repose non sur la loi, mais sur un jugement de valeur porté soit par la société, soit par les plateformes à travers **leurs conditions générales d'utilisation (CGU)**.

La société peut considérer certains contenus comme inappropriés (par exemple des propos grossiers, des contenus vulgaires ou la pornographie, qui n'est pas interdite en elle-même mais peut être jugée inappropriée lorsqu'elle n'est pas illégale), tandis que les plateformes appliquent leurs propres critères pour réguler les comportements.

Ce type de contenu peut inclure, par exemple, des images violentes, des théories complotistes non illégales, des discours dégradants, ou des représentations à caractère sexuel ne relevant pas du domaine judiciaire. Les plateformes choisissent souvent de modérer ou de limiter ce type de contenu afin de préserver le « bien-être » de la communauté. La notion de contenu inapproprié reste donc subjective et peut susciter des débats sur la liberté d'expression.

4. Digital Services Act (DSA)

ou Règlement sur les services numériques (RSN)

Le Règlement (UE) 2022/2065 du 19 octobre 2022 relatif à un marché unique des services numériques (DSA ou RSN), est un règlement européen directement applicable qui encadre les obligations des services intermédiaires en ligne (hébergeurs, plateformes, réseaux sociaux). Il vise à assurer un espace numérique sûr tout en garantissant la protection des droits fondamentaux.

Le texte impose des règles de transparence dans la modération des plateformes, l'obligation de motiver les décisions de retrait, la possibilité de recours pour les utilisateurs, ainsi que des garanties contre la censure excessive, notamment lorsqu'elle résulte d'algorithmes.

Le DSA constitue le cadre juridique central de la responsabilité des plateformes au sein de l'Union européenne et en France.

5. Désinformation

La désinformation désigne la diffusion intentionnelle de fausses informations dans le but d'induire en erreur, de manipuler l'opinion publique, de nuire à une personne ou à une institution, ou de servir des intérêts politiques, idéologiques ou économiques.

A la différence de la simple fausse information, la désinformation implique une volonté délibérée de tromper. Elle peut être orchestrée par des individus, des organisations ou des acteurs étrangers pour influencer le débat public ou déstabiliser une société.

Le DSA identifie la désinformation comme un risque systémique pour les très grandes plateformes et leur impose d'évaluer et de réduire ces risques, sans toutefois établir une liste précise de propos interdits. La désinformation n'est pas automatiquement illégale, mais elle peut le devenir lorsqu'elle s'articule à des infractions existantes (ex. : escroquerie, manipulation de marché, discours de haine).

6. Fact-checking (vérification des faits)

Le fact-checking, ou vérification des faits, désigne l'activité qui consiste à vérifier l'exactitude d'une information diffusée publiquement, en la confrontant à des données fiables, des sources reconnues ou des documents officiels. Cette pratique vise à distinguer les faits des opinions et à lutter contre la désinformation.

La fausse information n'étant pas par nature illégale, le DSA n'impose pas aux plateformes d'obligations en termes de fact-checking, tant que les informations diffusées ne le sont pas (ex. diffamation publique).

7. Fausse information

Une fausse information (parfois appelée information erronée ou information inexacte) est une information qui ne correspond pas aux faits, soit parce qu'elle repose sur une erreur, une mauvaise interprétation, un manque de vérification ou une source non fiable.

La fausse information n'est pas nécessairement diffusée dans l'intention de tromper : elle peut résulter d'une simple négligence ou d'un partage impulsif.

En droit, elle n'est pas automatiquement illégale. Elle ne devient sanctionnable que lorsqu'elle porte atteinte à des intérêts protégés (ex. : diffamation, tromperie commerciale, atteinte à l'ordre public, manipulation électorale…) Les plateformes peuvent choisir de la modérer au titre de contenus inappropriés ou pour limiter sa propagation, mais le DSA ne leur impose pas d'obligations générales de *fact-checking*.

8. Libertés

- **Liberté d'expression**

La liberté d'expression est le droit pour toute personne de communiquer librement ses idées, opinions et créations, par tout moyen (écrit, oral, numérique, artistique). Elle fait partie des libertés fondamentales. Mais comme toute liberté fondamentale, elle n'est pas absolue et peut être limitée par la loi pour protéger d'autres droits et libertés tels que le droit au respect à la protection des personnes, de l'ordre public publique, de la vie privée et de la propriété intellectuelle.

- **Liberté d'information**

La liberté d'information désigne le droit de rechercher, recevoir et diffuser des informations sans entrave injustifiée, il s'agit du cadre de base des journalistes. Elle garantit l'accès du public aux données d'intérêt général et favorise la transparence démocratique. Comme la liberté d'expression, cette liberté s'exerce dans le respect des droits d'autrui.

- **Liberté d'opinion**

La liberté d'opinion est le droit de détenir des convictions personnelles, politiques, religieuses ou philosophiques sans subir de pression, discrimination ou sanction. Elle protège la pensée intérieure, qui ne peut être limitée par aucune loi. Elle constitue la base du pluralisme et permet à chacun de se forger librement ses idées.

9. Loi du 29 juillet 1881 sur la liberté de la presse

La **loi du 29 juillet 1881** est une loi pénale spéciale qui constitue le **texte fondamental encadrant la liberté d'expression en France**. Historiquement adoptée pour garantir la liberté de la presse, elle s'applique aujourd'hui à tous les supports d'expression, y compris sur Internet.

Cette loi repose sur un équilibre : elle consacre la liberté d'expression comme principe, mais prévoit des exceptions et régime de sanctions (par exemple en cas de diffamation, injure, provocation à la haine, définis dans la loi). Elle se distingue par sa philosophie libérale, son attachement à la procédure contradictoire et son exigence d'une interprétation stricte des infractions.

10. Modération de contenus

Le DSA désigne la modération des contenus comme l'ensemble des actions entreprises par les plateformes pour détecter, évaluer, invisibiliser, retirer ou désactiver l'accès à des contenus mis en ligne par les utilisateurs, lorsqu'ils sont considérés comme illicites ou contraire aux règles internes de la plateforme (conditions générales d'utilisation).[3]

Le DSA impose aux plateformes de rendre ces actions transparentes, motivées et contestables. Toute décision de modération (ex. : suppression d'un contenu, suspension d'un compte) doit être

notifiée à l'utilisateur concerné avec les motifs précis et les moyens de recours disponibles, y compris via une procédure interne ou un mécanisme extrajudiciaire agréé.

La modération peut être humaine ou automatisée, et porter aussi bien sur des contenus illégaux que sur des contenus simplement jugés inappropriés par les plateformes.

11. Shadow banning (bannissement furtif)

Le *shadow banning* (ou bannissement furtif) est une pratique de modération par laquelle une plateforme limite la visibilité des contenus ou du compte d'un utilisateur sans l'en avertir explicitement. Contrairement à une suspension ou une suppression classique, la personne concernée peut continuer à publier normalement. Ses publications ne sont pas supprimées mais elles sont invisibles ou moins visibles pour les autres utilisateurs.

12. Signaleur de confiance

Le signaleur de confiance (en anglais *trusted flagger*) est une personne physique ou morale reconnue par une plateforme numérique comme étant particulièrement fiable dans le signalement de contenus illicites. Ce statut permet aux signalements effectués par ces acteurs d'être traités en priorité.

C'est le DSA qui a créé cette qualification. Le texte encadre ce dispositif de manière harmonisée à l'échelle de l'Union européenne. Son article 22 établit que les États membres peuvent désigner des signaleurs de confiance, sur la base de critères de compétence, d'indépendance et de qualité des signalements. Une fois reconnus, ces signaleurs bénéficient d'un traitement préférentiel de leurs signalements, sans pour autant que leurs contenus soient automatiquement supprimés – la décision finale relevant toujours des plateformes.

13. Tiers de confiance

Le tiers de confiance, dans le cadre de la régulation des contenus numériques, désigne un acteur externe, neutre, qualifié et légitime, qui assiste les plateformes dans leur mission de modération des contenus. Il peut s'agir d'associations, d'experts, d'institutions ou d'acteurs issus de la société civile reconnus pour leur rigueur, leur indépendance et leur connaissance des enjeux liés à la liberté d'expression. Il n'a pas de définition légale contrairement au signaleur de confiance.

Le rapport LibEx souligne qu'il est essentiel de ne pas laisser les plateformes seules responsables de la modération des contenus, une tâche lourde et parfois arbitraire. Les tiers de confiance apportent alors un regard éclairé, pluraliste et plus représentatif de la société démocratique. Leur

rôle est d'accompagner les décisions de modération et de contribuer à une régulation plus juste, fondée sur des critères transparents et partagés.

Ils s'inscrivent dans une logique de co-régulation, au croisement entre l'autorégulation des plateformes et la régulation publique, et peuvent jouer un rôle déterminant dans la mise en œuvre effective du DSA.

la vi//a .
numeris

*unlock the future, make it human**

hello@lavillanumeris.com
+33 7 80 96 11 11
http://www.lavillanumeris.com

**libérez l'avenir, rendez-le plus humain*