AFNOR SPEC 2305

NOVEMBRE 2023

www.afnor.org

Contacter :
AFNOR – Norm'Info
11, rue Francis de Pressensé
93571 La Plaine Saint-Denis Cedex
Tél : 01 41 62 76 44
Fax : 01 49 17 92 02
E-mail : norminfo@afnor.org

AFNOR SPEC

Novembre 2023

PRÉVENTION DES RISQUES ET PROTECTION DES MINEURS SUR LES RÉSEAUX SOCIAUX

© deagreez

www.afnor.org

afnor

Sommaire

Remerciements

REMERCIEMENTS

À tous les experts et professionnels pour leurs contributions et la qualité des échanges lors des travaux.

Aux institutions, administrations et organisations qui ont compris la nécessité impérieuse de ce travail (notamment la CNIL et l'ARCOM).

À l'animatrice du projet Sharone Franco avec le soutien indéfectible d'Alexandre Eurverte, Margaux Liquard, Marc-Antoine Durand et des équipes de Yubo.

Aux équipes de l'AFNOR et particulièrement à Julie Latawiec, Meriem Oudghough et Capucine Malhomme.

Avant-propos

Le présent document constitue le fruit d'un ambitieux travail de concertation et de collaboration d'un groupe de professionnels, experts dans le domaine des réseaux sociaux, de la sécurité en ligne et de la protection de l'enfance.

Cette AFNOR SPEC se veut une réponse aux préoccupations croissantes concernant la sécurité des mineurs sur les réseaux sociaux. L'objectif du document à destination des plateformes et réseaux sociaux est de rassembler des recommandations pratiques et opérationnelles destinées à façonner un environnement en ligne plus sûr et plus adapté aux besoins des jeunes utilisateurs.

Si les grands acteurs de cette industrie bénéficient d'une expertise et d'un savoir-faire en constante progression, il est bien plus difficile pour les nouvelles plateformes de comprendre comment adresser les problématiques de sécurisation des outils qu'ils proposent et particulièrement l'enjeux primordial que constitue la protection des mineurs.

L'intérêt exponentiel porté sur cette question à l'échelle mondiale, et notamment les très nombreuses évolutions juridiques, particulièrement en France et au niveau européen, complexifie encore davantage le travail de compréhension et d'appréhension que doivent effectuer les plateformes. Les enjeux de vie privée, de contrôle parental, et régulation des contenus sont au cœur des débats, incitant les acteurs de l'industrie à revoir leurs pratiques et mais aussi à penser leur rôle et leur impact dans la société.

Cette AFNOR SPEC adresse quatre principaux sujets : la question de la vérification des comptes (notamment de l'âge et de l'identité des utilisateurs), la détection et la modération des contenus inappropriés, ainsi que la prévention et l'éducation.

En outre, le groupe de travail a souhaité consacrer une partie importante du document au cadre juridique applicable. C'est à la lumière de ces obligations légales et réglementaires que les recommandations et réflexions présentées ont été effectuées. Ce cadre étant depuis quelques années et particulièrement en ce moment très évolutif, une mise à jour de l'AFNOR SPEC sur cette partie devra être effectuée d'ici 18 mois.

Nous sommes fiers du succès dont témoigne le consensus obtenu à l'issue de ce travail concerté et espérons que cet AFNOR SPEC constituera un guide de référence, utile aux nouvelles plateformes ou à celles souhaitant protéger davantage leurs jeunes utilisateurs.

La protection des mineurs est une responsabilité partagée, et il est de notre devoir collectif de veiller à ce que les générations futures puissent profiter pleinement des opportunités qui leur sont offertes sans compromettre leur vie privée, leur sécurité et leur bien-être.

Le présent document a été développé par un groupe de travail ouvert et reflète à ce titre l'accord de personnes et organisations ayant participé à son élaboration. AFNOR a mis à disposition des auteurs son savoir-faire en ingénierie normative afin de coordonner les travaux d'élaboration et éditer le document. En conséquence, le contenu de ce document n'engage que ses auteurs et ne saurait être considéré comme constituant le droit applicable. En effet, AFNOR n'étant ni habilitée à délivrer du conseil juridique ni législateur, AFNOR ne saurait être tenue responsable de l'utilisation qui est faite de ce document, notamment concernant la règlementation éventuellement citée dont la bonne application relève exclusivement de la responsabilité de chacun.

L'AFNOR SPEC :

— est un document technique développé et approuvé dans le cadre d'un processus transparent et ouvert ;

— représente l'approbation de ce seul groupe de travail sur le texte final et ne doit pas être présentée comme une norme française ou comme équivalente à une norme française.

Liste des participants

Contributeurs	Organisme
BENELLI Paul	MYM - AIRMEDIA
BORRY-ESTRADE Elisa	META
BOUTARD Matthieu	BODYGUARD
BRIEND Clotilde	META
CHEVOPPE-VERDIER Florian	YOTI
COMBLEZ Samuel	E-ENFANCE
DAWSON Julie	YOTI
DEFOSSEZ Etienne	DAILYMOTION
DESAINT Axelle	TRALALERE
DURAND Marc-Antoine	YUBO
EUVERTE Alexandre	YUBO
FRANCO Sharone	YUBO
KACOU Arthur	BODYGUARD
LATAWIEC Julie	AFNOR NORMALISATION
LEDAN Baptiste	SORARE
LEROUX Sébastien	DAILYMOTION
LESCOP Yann	POINT DE CONTACT
LIQUARD Margaux	YUBO
MALHOMME Capucine	AFNOR NORMALISATION
MAUSS Laura-Blu	RESPECT ZONE
OUDGHOUGH Meriem	AFNOR NORMALISATION
SERAVANT VOLQUIN Virgile	MYM - AIRMEDIA
TUFFIER Capucine	META

Observateurs	Organisme
CALENDRI Laurence	UNIVERSITE TOULOUSE CAPITOLE
JEHEL Sophie	UNIVERSITE PARIS VIII

Logos des organismes contributeurs et observateurs

PARTIE I

Vérification des comptes

Partie I - Vérification des comptes

Les réseaux sociaux sont aujourd'hui massivement utilisés par les mineurs ; selon une enquête Génération Numérique de mars 2022, 58 % des jeunes de 11 et 12 ans ont un compte sur un réseau social et les 11-18 ans y sont chaque année plus nombreux.

Souvent moins conscients des risques et plus vulnérables aux menaces qui pèsent sur eux, les mineurs doivent faire l'objet d'une attention particulière. Afin de garantir un environnement en ligne sûr et fiable, il convient non seulement de mettre en place un système de modération efficace (cf. Partie II. de l'AFNOR SPEC), mais il est aussi impératif d'adopter des mesures proactives de lutte contre les activités malveillantes, tout en respectant les normes éthiques et les droits fondamentaux des utilisateurs.

Cette Partie I examine de manière approfondie les différentes mesures de vérification des comptes, d'une manière graduée et proportionnée au risque, et propose des pistes de réflexion quant à l'optimisation de ces processus dans le respect des droits individuels et de la vie privée.

1 Limiter les risques de comptes malveillants

1.1 Classification des risques liés à la création et à l'utilisation d'un compte sur un réseau social

Dans le cadre de cette section 1.1, nous appellerons « compte malveillant » tout compte créé sur un réseau social dans le but de nuire à autrui ou à la l'intégrité de la plateforme, ou encore dans le but de commettre une fraude. Le compte malveillant agit souvent sous un faux compte, c'est-à-dire un compte usurpant l'identité de quelqu'un, ou tentant dissimulant son identité ou une information sur son identité (telle que l'âge) pour tromper les autres utilisateurs à des fins malveillantes.

Les comptes malveillants peuvent agir de manière isolée, ou adopter un comportement organisé et coordonné. Les plateformes doivent être particulièrement attentives aux comportements trompeurs organisés, visant à créer des infractions à grande échelle (par exemple, des grandes campagnes de phishing), à manipuler des informations ou des opinions d'une façon contraire aux principes démocratiques (par exemple, tenter d'influencer des élections).

1.1.1 Cartographier les risques inhérents aux plateformes

Quels types de risques ?

Les réseaux sociaux constituent aujourd'hui les outils les plus efficaces pour permettre aux individus du monde entier de communiquer entre eux, de s'exprimer, de s'informer et de se divertir, parfois sur une même plateforme. Cependant, cette accessibilité et cette facilité d'utilisation permettent aussi aux comportements malveillants de s'épanouir dans des proportions jamais vues auparavant.

Pour assurer leur pérennité et attirer de nouveaux utilisateurs en toute confiance, les réseaux sociaux doivent lutter contre les activités illicites. Des infractions telles que la désinformation, le cyberharcèlement, la haine en ligne et les abus commis contre les mineurs font partie des infractions couramment répandues sur les plateformes, dont les auteurs tirent parti du relatif anonymat (ou pseudonymat) permis sur Internet.

Pour lutter contre ces phénomènes, les plateformes doivent tout d'abord prendre conscience des enjeux et des risques inhérents à leurs services. La cartographie de ces risques doit prendre en compte les fonctionnalités offertes et le public auquel ces fonctionnalités s'adressent.

Par exemple :

- une fonctionnalité de messagerie instantanée interpersonnelle présente généralement un risque plus élevé d'activités telles que le spam, les escroqueries, les attaques informatiques de type phishing, et de pédopiégeage des mineurs s'il est accessible à ce public ;
- une fonctionnalité de vidéos en direct devra porter une attention toute particulière aux risques de diffusion en direct de contenus explicites ou violents, rendant ainsi primordial de pouvoir réagir immédiatement pour supprimer ces contenus ;
- un service de microblogging ou d'agrégation de fils d'actualité sera plus susceptible de servir à des phénomènes de désinformation et de propagation de fausses nouvelles.

Il convient aussi de tenir compte du public auquel on s'adresse : un service ciblant les adolescents doit appréhender les risques auxquels les mineurs sont davantage exposés, tels que :

- le pédopiégeage, c'est-à-dire le fait pour un adulte de proposer, au moyen des technologies de l'information et de la communication, une rencontre à un mineur qui n'a pas atteint la majorité sexuelle, dans le but de commettre un abus sexuel ;
- la diffusion de contenus à caractère pornographique ou pédopornographique ;
- la diffusion non consensuelle d'images intimes ;
- les phénomènes de cyberharcèlement, auxquels les jeunes sont tout particulièrement confrontés.

Ainsi, chaque fonction et chaque service du réseau social devrait faire l'objet d'une analyse concrète des risques auxquels les utilisateurs peuvent être confrontés, en fonction des usages et des tendances observées en pratique.

Comment recenser les risques ?

Les plateformes pourraient ainsi constituer et tenir à jour un registre des risques qui les concernent. Ce registre pourrait inclure les informations suivantes :

- une description générale du risque et de ses conséquences, en mettant en avant tout risque pour les mineurs ;
- les équipes internes impliquées dans la gestion du risque ;
- la sévérité de l'impact du risque en cas de réalisation de celui-ci ;

- la probabilité que le risque se produise ;
- les mesures correctives à mettre en place ;
- l'existence d'un risque résiduel éventuel après l'application des mesures correctives.

Ce registre des risques devrait particulièrement aborder :

- les comportements contrevenant aux règles applicables sur la plateforme ;
- le phénomène des comptes malveillants ;
- les comportements visant à abuser de la confiance des autres utilisateurs de la communauté.

Il est important que **ce registre interne soit mis à jour régulièrement (par exemple une fois par an), notamment pour intégrer de nouveaux risques liés à l'évolution du produit, réévaluer les mesures correctrices, ou encore appréhender de nouveaux comportements observés au cours de l'année.**

S'entourer d'experts externes afin de mieux identifier et appréhender les risques :

La plupart des plateformes ont engagé des collaborateurs spécialisés dans des domaines tels que le terrorisme, l'incitation à la haine ou la protection de l'enfance pour définir des règles communautaires et s'assurer qu'ils tiennent compte des risques qui évoluent en permanence. Ces équipes sollicitent l'avis d'experts et d'organisations externes afin de mieux comprendre les différents points de vue, enjeux et tendances sur la sécurité et l'information, et d'adapter au mieux leurs conditions générales et règles communautaires en fonction des publics concernés.

Exemple de registre :

Date de création du document: XXXX
Dernière mise à jour : XXXX

#	Description du risque	Département(s) impliqué(s)	Catégorie de risque	Impact	Vraisemblance	Criticité	Mesures correctives	Gestion du risque	Risque résiduel
#001	**Risque lié à la propagation de comptes malveillants.** **Création de faux comptes à des fins malveillantes, en particulier :** * Activités de pédopiégeage * Diffusion de désinformation et de propagande * Usurpation d'identité * Cyberharcèlement	XXXXX	Risque pour la sécurité et le bien-être des mineurs, risque de fraude en ligne, risque de commission de délits de presse	5	2	10	- XXXXX - XXXXX - XXXXX - XXXXX	90% / 100%	Fully Managed

1.1.2 Agir contre les comptes malveillants

Mettre en balance la minimisation des données, les risques identifiés et les obligations légales

Pour renforcer la confiance de leur communauté, attirer de nouveaux utilisateurs et éviter la multiplication de comportements indésirables, les plateformes doivent mettre en place des mesures pour lutter contre les comptes malveillants.

Lors de l'inscription au service, les plateformes doivent déterminer :

- les données personnelles nécessaires à la création d'un compte telles que, selon les cas, le nom d'utilisateur, la photo du profil, l'adresse email, etc. ;
- les informations nécessaires pour bénéficier du service proposé.

Les informations d'identification nécessaires dépendent généralement du type de services et de contenus proposés sur la plateforme, des contraintes réglementaires applicables, et du type d'audience visée (notamment concernant la tranche d'âge).

Le recours à la vérification d'identité systématique, dès l'inscription, de tous les utilisateurs, ne constitue pas une obligation légale et apparaît même disproportionnée au regard du droit à la protection de la vie privée. Cette mesure risquerait de se heurter à des difficultés techniques importantes et d'entraver d'autres droits et libertés, comme la liberté d'expression.

Cependant, il existe des obligations légales de procéder à des vérifications systématiques de l'âge :

- Pour les sites pornographiques : En application de la loi du 30 juillet 2020 visant à protéger les victimes de violences conjugales, les sites pour adultes proposant des contenus à caractère pornographique sont désormais dans l'obligation de contrôler l'âge des utilisateurs souhaitant y accéder. Cette mesure est imposée sous peine de violation de l'article 227-24 du Code pénal, condamnant le fait de donner accès à des contenus pornographiques par des individus de moins de 18 ans, et ce même s'ils ont déclaré être majeurs. Le projet de loi visant à sécuriser et réguler l'espace numérique confie à l'Arcom la responsabilité de veiller au respect de ces obligations par les sites pornographiques, ainsi que le soin d'élaborer un référentiel pour le contrôle de l'âge des internautes, pris après avis de la CNIL (voir Annexe juridique, Section 5, pour plus d'informations).

- L'ensemble des réseaux sociaux doivent bloquer l'accès à leur plateforme aux mineurs de moins de 15 ans en France, à moins que le consentement d'un titulaire de l'autorité parentale n'ait été obtenu. C'est ce qui résulte de la loi du 7 juillet 2023 visant à instaurer une majorité numérique et à lutter contre la haine en ligne. Cette obligation implique de mettre en place une solution technique pour 1) vérifier l'âge de tous les utilisateurs dès l'inscription, et 2) obtenir un consentement parental, le cas échéant, pour les utilisateurs de moins de 15 ans. Pour plus d'informations sur ces sujets, voir les sections 2.1 et 2.2 ci-dessous.

Hormis ces cas précis, une approche équilibrée consisterait à définir des critères objectifs pour procéder à des vérifications ciblées et ponctuelles, notamment en cas de détection de signaux d'alerte.

Quels sont les signaux d'alerte ?

En fonction des usages et du public visé, voici une liste non exhaustive de signaux pouvant caractériser un doute objectif et raisonnable pour lesquels des mesures de contrôle ponctuelles et ciblées pourraient être prises :

- Présence suspectée de mineurs de moins de 13 ans sur une plateforme réservée à un public de plus de 13 ans, ou présence de mineurs de moins de 18 ans sur une plateforme réservée à un public adulte.
- Violation manifeste des règles communautaires de la plateforme par un utilisateur ayant menti sur son identité (par exemple : usage de photos ou d'informations appartenant à une autre personne).

- Envoi de messages promotionnels ou commerciaux massifs depuis un même compte.
- Détournement des mécanismes de signalement des utilisateurs, notamment à des fins malveillantes (par exemple : harcèlement ou discrimination).
- Création de comptes multiples depuis un même appareil avec des données d'identification comportant d'importantes différences (comme un écart d'âge important).
- Proposition d'ajout sur des plateformes tierces, notamment des plateformes réservées aux adultes.
- Promotion de services comportant un risque d'arnaques ou d'activités illicites (notamment dans le domaine des services financiers et le domaine des services à caractère sexuel).

Quels types de mesures mettre en place ?

Lorsque des signaux d'alerte sont remontés auprès des équipes de sécurité et/ou de modération, une réponse proportionnée et graduelle doit être apportée. Ces mesures visent un triple objectif de s'assurer que l'utilisateur :

1. n'a pas menti sur son profil ou fourni des informations erronées afin de tromper ses interlocuteurs ou commettre une fraude ;
2. qu'il n'a pas commis une violation des règles en vigueur sur la plateforme ;
3. qu'il respecte les conditions d'âge prévues par la réglementation (par exemple qu'il n'a pas moins de 13 ans).

Parmi les mesures les plus courantes pour vérifier l'authenticité et la légitimité des comptes, on retrouve notamment :

- la vérification du numéro de téléphone (généralement via l'envoi d'un code par SMS pour s'assurer de l'authenticité dudit numéro) ou de l'adresse e-mail. Il s'agit de la méthode de la double authentification ;
- la vérification de profil par le visage de l'utilisateur : cette méthode nécessite souvent que l'utilisateur prenne un selfie ou une courte vidéo représentant son visage. Le modérateur humain ou l'intelligence artificielle vérifie ensuite que cette photo ou cette vidéo est conforme à celles sous lesquelles l'utilisateur prétend agir via son compte. Si cette vérification est automatisée et implique le traitement de données biométriques, il est obligatoire d'obtenir le consentement préalable de l'utilisateur. Le consentement peut, par exemple, consister à laisser à l'utilisateur une alternative qui n'implique aucune collecte de données biométriques ;
- la vérification par pairs, c'est-à-dire la vérification du profil grâce à des informations fournies par d'autres utilisateurs du service, par exemple, grâce à un système de confirmation ou de vote. L'utilisateur ayant obtenu une majorité de votes favorables de la part d'un certain nombre d'autres utilisateurs du service est considéré comme valablement vérifié ;
- la vérification ou l'estimation de l'âge de l'utilisateur (voir la section2.1 ci-dessous) ;
- la vérification d'identité grâce à un document officiel (voir la section 2.3 ci-dessous).

Aucune de ces mesures ne doit être considérée infaillible. Elles ne sont, d'ailleurs, pas exclusives les unes des autres, et doivent faire l'objet d'une évaluation approfondie par la plateforme pour s'assurer qu'elles sont justifiées et limitées à ce qui est strictement nécessaire.

Ces vérifications ne sont pas de même nature. Par exemple, du fait de la sensibilité de la pièce d'identité, celle-ci doit être demandée par la plateforme en dernier recours. En outre, si la plateforme est seulement soumise à une restriction d'âge, elle n'a pas besoin de connaître l'identité complète de l'utilisateur ; la vérification d'identité devrait être écartée dans ce cas au profit d'un processus de vérification de l'âge. Il convient de toujours privilégier la mesure la moins intrusive possible au regard de l'objectif poursuivi.

Lorsque la plateforme a caractérisé l'existence d'un compte malveillant, elle doit prendre une action pour bannir ce compte ou lui bloquer l'accès au service, que ce soit temporairement ou définitivement (selon la dangerosité du compte).

1.2 La proportionnalité et la conformité des mesures : principes à respecter dans le cadre du *Safety* et *Privacy by Design*

1.2.1 Minimiser l'impact pour la vie privée

Minimisation des données

Pour détecter et bannir les comptes malveillants, les mesures prises impliquent nécessairement de traiter des données personnelles avec des risques d'entraver les droits et libertés des comptes associés. En effet, ces mesures peuvent inclure le retrait des contenus problématiques, la suspension de l'accès au service, ou des mesures restrictives ayant pour effet de limiter leur liberté d'expression et de communication, etc.

Ces procédures requièrent généralement la mobilisation des équipes de modérateurs de la plateforme, et peuvent également s'appuyer sur des technologies de détection spécifiquement conçues pour détecter des signaux susceptibles d'identifier les comptes suspects (comme évoqués en Section 1.1 ci-dessus). Ces outils peuvent s'avérer particulièrement efficaces pour créer des rapports internes, que les spécialistes peuvent ensuite examiner afin de prendre les mesures adéquates. Les plateformes doivent identifier les données qu'elles ont besoin de traiter pour identifier les comptes malveillants.

Il est essentiel, dans la détermination des moyens et ressources à mettre en place, de privilégier les outils et les mesures les moins intrusifs possibles pour la vie privée. Les données traitées par ces outils doivent uniquement servir les impératifs de sécurité de la plateforme, et non pas ses intérêts commerciaux.

Principe de licéité

Ces traitements doivent tout d'abord reposer sur une base juridique valable conformément à l'article 6 du Règlement européen 2016/679 du 27 avril 2016 (« **RGPD** »). Dans le cadre de la lutte contre les comptes malveillants, l'intérêt légitime de la plateforme peut constituer une base légale appropriée, dans la mesure où la lutte contre la fraude est, de façon générale, reconnue comme un intérêt légitime du responsable de traitement[1)]. Le recours à l'intérêt légitime doit toutefois être évalué et soupesé pour s'assurer que le traitement ne porte pas atteinte aux droits et libertés des individus. À cette fin, il convient de créer une fiche documentant cet intérêt légitime, ainsi que son impact sur les droits et libertés.

1) Considérant 47 du RGPD.

Les personnes concernées par le traitement disposent également de droits spécifiques. Notamment en cas d'usage de technologies reposant sur l'intelligence artificielle. Le RGPD accorde aux individus le droit aux individus de ne pas être soumis à une « ***décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire*** » (article 22 du RGPD).

Un système d'intelligence artificielle collectant des informations sur les utilisateurs afin de détecter d'éventuelles infractions sur la plateforme constitue un traitement automatisé, pouvant potentiellement produire des effets juridiques ou affecter significativement un individu. En effet, selon les circonstances, l'utilisateur détecté pourrait être banni de la plateforme, ou même faire l'objet d'un signalement aux autorités compétentes en cas d'infraction grave. Dans ce genre de décision, il est essentiel que des spécialistes humains gardent le contrôle sur la décision finale, en pouvant, *a minima*, examiner et valider la proposition faite par le système d'IA.

En outre, lorsque des actions de modération sont mises en œuvre de façon automatisée pouvant aboutir à une forme de sanction de l'utilisateur, il est essentiel d'en informer ce dernier et de lui offrir un moyen de recours, lui permettant de demander un nouvel examen - manuel et humain - de la situation. Ce moyen de recours devrait être mis en évidence au moment où l'utilisateur est informé qu'une action de modération est prise sur son compte.

Ces obligations proviennent, de façon générale, du RGPD (notamment l'article 22), et plus spécifiquement, du Règlement européen 2022/2065 du 19 octobre 2022 sur les services numériques (notamment les articles 12, 14, 17).

Limitation de la conservation

Minimiser l'impact pour la vie privée implique également de réduire la durée de conservation des données traitées au minimum requis pour atteindre l'objectif recherché. Ainsi, les informations détectées relatives à des activités suspectes, à des signaux d'alerte ou à des actions de modération, et les données de contenus et de trafic associées, incluent presque toujours des données personnelles.

Il convient dès lors de définir des durées de conservation appropriées, permettant aux plateformes de se conformer à leurs obligations légales de conservation des données, de constituer des preuves en cas de réclamation ou de procédures judiciaires, de répondre aux éventuelles demandes des autorités, et de conserver un historique adéquat des actions de modération.

1.2.2 Rédiger la documentation obligatoire

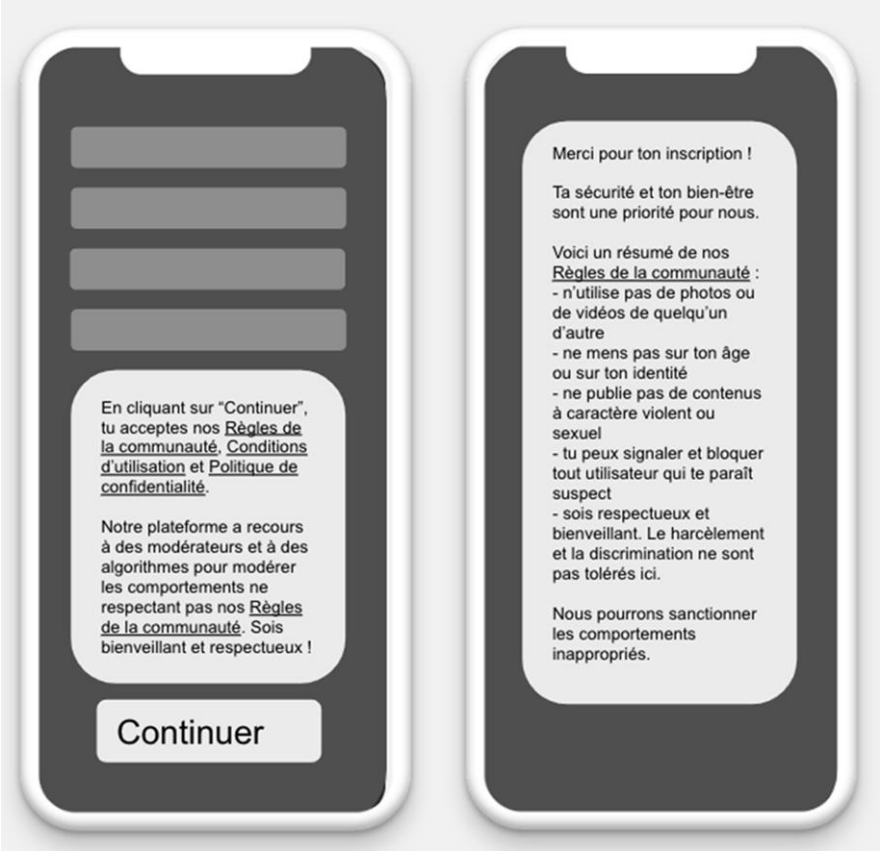
Document d'information des utilisateurs

Les mesures de sécurité mentionnées dans la section 1.1.2 doivent être expliquées aux utilisateurs, par exemple dans la politique de confidentialité ou dans les conditions générales d'utilisation. La transparence due aux utilisateurs résulte tant du RGPD, concernant l'usage des données personnelles, que du DSA, qui impose d'indiquer les raisons pouvant conduire la plateforme à restreindre l'utilisation du service, les mesures prises pour modérer les contenus, y compris celles reposant sur des outils algorithmiques, renforcer l'accessibilité de l'information pour les mineurs, etc.

Lorsqu'un individu navigue sur un réseau social, il n'a pas toujours conscience que ses informations peuvent être traitées à des fins de détection d'activités frauduleuses et de modération des contenus. C'est pourquoi une information claire doit apparaître à ce sujet dans la politique de confidentialité ou dans des mentions d'informations.

En outre, il est toujours bénéfique de renforcer la transparence via des mentions d'information courtes et contextualisées.

Dans les exemples ci-dessous, les règles à suivre sur la plateforme, ainsi que le fait que les contenus peuvent être modérés (y compris de façon automatisée) sont rappelés à l'utilisateur :



Documentation interne

L'ensemble des moyens technologiques et humains mis en place pour lutter contre les activités frauduleuses et malveillantes doivent être documentés dans le registre des activités de traitement de l'entreprise, conformément à l'article 30 du RGPD.

En ce qui concerne la gestion des risques en interne, il est recommandé de réaliser une analyse d'impact sur la protection des données ***(« AIPD »)*** pour documenter les risques pour la vie privée. Cette analyse peut être obligatoire en application de l'article 35 du RGPD, dès lors que le traitement de données personnelles est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes. L'AIPD est obligatoire si au moins deux des neufs critères issus des lignes directrices du G29 sont réunies.

Dans le cas de la lutte contre les comptes malveillants, potentiellement six des neuf critères pourraient être pertinents (à vérifier au cas par cas) :

- Évaluation/scoring.
- Décision automatique avec effet légal ou similaire.
- Collecte de données personnelles à large échelle.
- Personnes vulnérables (dont mineurs).
- Usage innovant.
- Exclusion du bénéfice d'un droit/contrat.

En plus de l'application de ces critères, l'AIPD peut être rendue obligatoire :

- En vertu de textes réglementaires, tels que le Règlement européen 2021/1232 du 14 juillet 2021 qui impose aux fournisseurs de messageries qui mettent en place des mesures de détection de contenus pédopornographiques ou de comportements type « grooming » de réaliser une AIPD, laquelle doit ensuite être soumise à l'autorité de contrôle dans les conditions de l'article 36 du RGPD.

- En vertu de la liste des traitements pour lesquels une AIPD est requise, publiée par la CNIL[2)]. Parmi ceux-ci, l'autorité inclut : tout *« traitement reposant sur une analyse comportementale visant à détecter des comportements « interdits » sur un réseau social »*, ainsi que tout *« dispositif de signalement de mineurs en danger »*.

L'AIPD peut également intégrer l'impact de l'utilisation d'algorithmes pour aider à la détection des comptes malveillants, pour documenter les risques que cela peut entraîner. Pour s'assurer que les effets des algorithmes soient bénéfiques et qu'ils servent les valeurs promues par la plateforme, il est essentiel de les maintenir sous contrôle en permanence. Cela implique notamment d'évaluer régulièrement leurs effets dans le temps afin de détecter d'éventuels biais ou effets discriminatoires, de mesurer leur efficacité, leur taux de précision (y compris les faux positifs), et de limiter leur capacité d'action d'une manière appropriée.

2 L'identification minimale et appropriée des utilisateurs

Le principe de minimisation impose de collecter le moins de données possibles pour proposer un service en ligne. Néanmoins, certaines normes juridiques et certaines circonstances factuelles nécessitent que des mesures soient prises pour obtenir des données d'identification fiables.

Cette partie propose d'examiner trois mesures parmi les plus répandues : le contrôle de l'âge (2.1), le consentement parental (2.2), et la vérification d'identité (2.3).

---

2) https://www.cnil.fr/sites/cnil/files/atoms/files/liste-traitements-aipd-requise.pdf

Les points d'attention

- **La complexité** : l'objectif est de minimiser les « frictions » pour l'utilisateur, car plus les démarches sont complexes, plus cela entraîne un coût élevé pour l'entreprise qui risque de perdre des utilisateurs.

- **La transparence** : l'objectif est de garantir la confiance des utilisateurs. Plus la procédure est transparente sur ses objectifs et les moyens qui sont mis en œuvre, plus l'utilisateur accepte facilement de s'y soumettre. Lorsqu'on lui propose plusieurs méthodes, l'utilisateur peut faire un choix informé.

- **La nécessité** : l'objectif est de s'assurer de l'acceptabilité de la mesure pour les utilisateurs. Il est important que les contrôles soient proportionnés et adaptés aux risques pour ne pas susciter d'hostilité des utilisateurs.

Les avantages

- **La confiance** : les garanties apportées par un système de vérification des comptes permettent aux utilisateurs d'être assurés que leurs interlocuteurs ont été vérifiés.

- **La responsabilité** : pouvoir s'assurer de certaines informations comme l'âge des utilisateurs permet aux plateformes et aux entreprises de développer en toute sérénité des activités (d'accès aux contenus, de vente, de jeux, de rencontres) qui seraient réservées aux majeurs.

- **La souplesse** : certaines solutions permettent de garantir l'anonymat des utilisateurs tout en bénéficiant des avantages d'une information confirmée sur le profil de l'utilisateur.

2.1 Le contrôle de l'âge

2.1.1 Est-ce que l'identifiabilité suffirait ?

La différence entre les approches de vérification et d'estimation de l'âge

Les mesures de contrôle de l'âge permettent à une plateforme d'offrir à ses utilisateurs une expérience et un parcours numérique sécurisé, en leur proposant des contenus et services appropriés. Chaque méthode d'estimation ou de vérification de l'âge présente des avantages et des inconvénients. Selon la méthode retenue, la plateforme obtient un niveau d'assurance plus ou moins élevé quant à l'âge réel de l'utilisateur.

La ***vérification de l'âge*** consiste à effectuer un contrôle de l'âge exact d'un individu ou un contrôle que son âge se situe au-dessus ou au-dessous d'un âge (souvent 18 ans), à partir de sa date de naissance, supposant généralement la communication d'un document officiel tel qu'une pièce d'identité.

L'***estimation de l'âge*** consiste à attribuer à un individu un âge approximatif, avec une faible marge d'erreur tolérée, généralement via un système d'intelligence artificielle capable d'analyser les traits du visage d'un individu à partir d'une photo.

Comme évoqué dans la partie 1, certaines plateformes sont soumises à des obligations légales leur imposant un âge minimum à partir duquel un individu peut avoir accès à leurs services : c'est notamment le cas des sites proposant des contenus pornographiques, dont l'accès est réservé aux plus de 18 ans, ou encore des réseaux sociaux dont l'accès est réservé aux plus de 13 ans, ou aux plus de 15 ans en l'absence d'un consentement parental.

La pratique a connu une évolution significative au cours de la dernière décennie. Cette industrie est aujourd'hui plus mature et structurée y compris au sein d'organisations sectorielles.

Par souci d'inclusion, l'industrie a développé des méthodes comme celles de l'estimation de l'âge par le visage ou de la voix, qui peuvent être utilisées par une plus grande majorité d'individus. En effet, tous les individus ne possèdent pas de documents d'identité (en particulier dans le cas de ressortissants de pays dans lesquels la possession d'un document d'identité n'est pas obligatoire), ou ne souhaitent pas les utiliser. Les expérimentations récentes menées par des plateformes et tiers de confiance qui offrent ces techniques montrent que les individus préfèrent, dans leur grande majorité, les techniques d'estimation de l'âge.

Cas d'usage n°1 : Suite à des tests réussis aux États-Unis, en Inde et au Brésil, Meta a déployé sur ses applications une technologie d'estimation de l'âge dans l'Union européenne, au Royaume-Uni et dans plus de 124 autres pays et territoires.

Sur Instagram, par exemple, certains utilisateurs sont invités à vérifier leur âge et peuvent choisir entre l'estimation de l'âge du visage ou la vérification par pièce d'identité. Au cours de ces essais, il a été observé que 81 % des utilisateurs qui tentent de modifier leur date de naissance choisissent l'estimation de l'âge du visage.

Cas d'usage n°2 : Depuis septembre 2022, Yubo a déployé sur l'ensemble de sa base d'utilisateurs une technologie d'estimation d'âge. Cet outil vient s'ajouter à d'autres technologies et mesures déjà mises en place sur l'application pour limiter les risques que les utilisateurs n'indiquent pas leur âge réel. Ce processus permet d'assigner les utilisateurs à un groupe d'âge, dans lequel ils peuvent interagir avec des utilisateurs appartenant à leur groupe. Garantir une plus grande fiabilité des âges permet notamment de limiter les risques que des personnes de moins de 13 ans (âge minimum requis) s'inscrivent sur l'application, que des utilisateurs de moins de 15 ans contournent la procédure de consentement parental, ou encore que des utilisateurs ayant une importante différence d'âge puissent interagir.

Les méthodes d'estimation de l'âge peuvent impliquer un traitement de données biométriques et incluent généralement une détection du 'vivant'. Il existe une grande variété de catégories telles les estimations d'âge basées sur le visage, le comportement (cela inclut le contenu auquel l'utilisateur accède, sa façon d'écrire, son cercle d'activité et d'amis), la voix et la main. Ces méthodes ne nécessitent pas que les personnes possèdent un document d'identité.

Il convient de préciser que le régime juridique est plus souple lorsque la solution se limite exclusivement à estimer l'âge d'une personne à partir d'une seule caractéristique physique, sans chercher à identifier ou à authentifier cette personne en particulier. En effet, le régime contraignant de l'article 9 du RGPD sur les catégories particulières de données (dont les données biométriques) ne s'applique qu'aux traitements de *« données biométriques aux fins d'identifier une personne physique de manière unique »*. Ainsi, si la solution vise seulement à estimer l'âge d'une personne sans savoir de qui il s'agit, l'article 9 pourrait être écarté. Si la solution vise à identifier une personne de manière unique, il s'agit d'un traitement de données biométriques, par principe interdit (sauf avec le consentement de la personne ou sur le fondement d'une autre exception de l'article 9.2 du RGPD).

Les méthodes d'estimation peuvent, dans certains cas et notamment lorsqu'elles ont recouru à un palier d'âge suffisamment élevé, offrir des niveaux d'assurance tout aussi élevés que ceux des techniques dépendantes d'une vérification de pièces d'identité, qui peuvent par ailleurs être falsifiées.

Contrairement aux approches d'estimation de l'âge, certaines méthodes de vérification de l'âge nécessitent que leurs utilisateurs possèdent un objet, comme un document d'identité délivré par une source de confiance. Ces preuves d'identité peuvent être présentées au cas par cas ou bien à partir d'un portefeuille d'identité numérique. D'autres méthodes sont conditionnelles et dépendent d'une vérification antérieure, comme lors du processus de vérification d'un compte bancaire ou de téléphonie mobile.

En outre, les approches de vérification nécessitent un haut niveau de précision afin d'atteindre un niveau d'assurance supérieur aux approches d'estimation de l'âge. Une vérification de l'âge via une pièce d'identité requiert des opérations de vérification de l'authenticité des documents, de l'identité des visages et de l'aspect « vivant » ou « réel » de l'utilisateur. Une méthode qui inclurait toutes ces étapes est susceptible d'assurer une plus grande fiabilité, mais s'avère plus intrusive ; elle devrait donc être employée si aucune méthode moins intrusive ne peut être envisagée.

De manière générale, la déclaration sur l'honneur devrait être considérée comme une méthode obsolète. Elle n'est d'ailleurs plus considérée comme une méthode appropriée et suffisante pour vérifier l'âge des utilisateurs souhaitant accéder à des contenus réservés aux adultes. Elle est même interdite pour l'accès aux sites pornographiques en vertu de l'article 227-24 du Code pénal, les sites contrevenants pouvant être condamnés pour avoir exposé des mineurs à des images pornographiques.

La mise en balance entre la nécessité de protéger les mineurs et la protection des droits et libertés (dont la vie privée)

L'objectif de protection de l'enfance ne peut et ne doit pas mener les acteurs du numérique (sites, plateformes et tiers de confiance) à une surveillance disproportionnée de leurs utilisateurs, à une collecte d'informations personnelles excessive, ni à une utilisation de ces données pour des objectifs autres que la protection de l'enfance.

Lors de la mise en place d'un système de contrôle de l'âge, le site ou la plateforme doit respecter les obligations légales ou réglementaires et trouver un juste équilibre entre la sécurité et le bien-être des mineurs et la protection de la vie privée.

Souvent, les entreprises n'ont pas besoin de vérifier l'identité complète d'un client ou d'un utilisateur, mais elles ont simplement besoin de vérifier si l'âge requis est atteint, sans connaître leur identité ni leur âge. C'est la notion d'identifiabilité : l'identification complète n'est pas requise, mais un élément de l'identité doit être vérifié. Dans ce cas, une méthode d'estimation de l'âge se montre plus pertinente et respectueuse du principe de minimisation qu'une méthode de vérification. C'est cette démarche progressive qui doit être privilégiée.

La CNIL a émis des recommandations concernant les principales méthodes de contrôle de l'âge existant sur le marché. En synthèse, il en ressort notamment que :

- Il convient de privilégier les solutions tierces et certifiées, qui pourront transmettre une confirmation à la plateforme que tel utilisateur a l'âge requis pour accéder aux services dont l'accès est restreint à une certaine tranche d'âge. L'objectif est de préserver la confidentialité de l'utilisateur par deux éléments :

  1) le tiers indépendant qui connaît l'âge (voire l'identité) de l'utilisateur ne sait pas quel site ou application l'utilisateur souhaite consulter ;

  2) la plateforme consultée ne connaît pas l'identité complète de l'utilisateur.

- Les solutions présentes actuellement sur le marché ne sont pas toutes infaillibles pour préserver les droits et libertés des individus, en plus de pouvoir être contournées dans certains cas.

- Les solutions d'estimation de l'âge devraient être testées pour assurer un haut niveau de précision et de fiabilité. Elles devraient en outre permettre aux utilisateurs de contester la décision résultant de l'estimation, en cas d'erreur, et de se faire vérifier par un autre mode de vérification.

Il revient à chaque acteur de déterminer la solution la plus adéquate et qui répond à ses besoins, en fonction des règles légales qui lui sont applicables et des risques existants pour la sécurité et le bien-être des mineurs.

2.1.2 Appréhender les risques techniques

Les plateformes doivent prendre en considération les risques suivants : détection du vivant et attaques informatiques, interception des données, authenticité et correspondance entre propriétaire du document ou de l'élément d'identifiabilité et l'utilisateur.

Les fraudeurs peuvent utiliser une variété d'attaques, telles que des images imprimées, des masques, des images ou des vidéos sur un écran, des hypertrucages, des attaques de bots, etc. Par conséquent, les plateformes doivent s'assurer que les fournisseurs de solutions de contrôle de l'âge mettent en œuvre des technologies de détection des tentatives d'usurpation d'identité sophistiquées, comme par exemple des technologies répondant au standard NIST de niveau 2.

Cela signifie que les offres des fournisseurs doivent être capables de résister aux attaques plus spécialisées telles que les masques en latex ou les imprimantes 3D. Pour répondre aux critères du standard NIST de niveau 2, ces fournisseurs doivent par exemple être en mesure de détecter 99 % des attaques et de les limiter les faux négatifs à moins de 15 %.

La transmission d'informations au cours de la vérification ou de l'estimation de l'âge présente un risque d'interception des données, et doit donc reposer sur une technologie de cryptographie adaptée. Les plateformes doivent donc s'assurer que leurs fournisseurs de service opèrent également selon les standards de cybersécurité et de stockage des données en vigueur, et que cette conformité aux standards soit vérifiée de manière indépendante, conformément aux recommandations de la CNIL précitées.

Concrètement, lorsque les données ne sont pas stockées localement sur le terminal de l'utilisateur, il convient de recourir à un protocole de sécurité qui permet de chiffrer les communications électroniques afin de garantir la confidentialité et l'intégrité des données échangées entre les utilisateurs et les serveurs. Le protocole TLS (*Transport Layer Security*) est utilisé pour sécuriser un large éventail de protocoles, tels que HTTPS (sécurisation des sites web), les connexions de messagerie électronique (SMTP, IMAP, POP3), les connexions VPN, etc. Il y a plusieurs versions de TLS, allant de TLS 1.0 à TLS 1.3 (la plus récente). Les versions plus récentes offrent de meilleures garanties de sécurité. Si elles sont compatibles avec les systèmes utilisés, celles-ci devraient être privilégiées.

Lorsque la vérification est basée sur un document, la technologie de scan NFC (*Near Field Communication*) doit être capable de détecter toute altération frauduleuse des données. Au cours de la procédure d'estimation ou de vérification, il est crucial de s'assurer que la personne est 'vivante' (c'est-à-dire représentée en temps réel) et que si elle présente un document, celui-ci correspond à la même personne figurant sur le document.

Enfin, les plateformes devraient mettre en place une politique de sécurité intégrant les mesures conformes à l'état de l'art, telles que : l'authentification forte, le contrôle et la gestion des accès, les pare-feu, la sécurité physique, la sensibilisation et la formation du personnel, la gestion des vulnérabilités, etc.

2.1.3 Présentation d'exemples et de cas d'usage

Une plateforme qui propose uniquement ou en partie du contenu restreint en fonction de l'âge par la loi ou une institution (telle que la Commission de classification des œuvres cinématographiques pour les films) pourrait mettre en place une vérification ou une estimation de l'âge pour se conformer aux règles en vigueur. Cette vérification, qui peut s'avérer utile pour les plateformes de streaming proposant des œuvres pour différentes catégories d'âge, peut être effectuée lors de la création de compte ou au moment d'accéder au contenu en question.

La question se pose aussi quant à la nécessité d'authentifier à nouveau à un utilisateur ou de procéder à un contrôle de l'âge à intervalles réguliers afin, par exemple, d'empêcher des adultes de créer un compte pour des mineurs, ou qu'un mineur utilise le compte d'un adulte (avec ou sans autorisation).

Un réseau social peut aussi estimer que certaines de ses fonctionnalités (telles que la messagerie instantanée gratuite ou le partage de contenu généré par ses utilisateurs) peuvent présenter un risque pour un utilisateur mineur s'il lui est permis d'entrer en contact avec un majeur. Le site peut alors mettre en place une vérification ou une estimation de l'âge afin de séparer les mineurs et les majeurs au moment où l'utilisateur crée un compte ou souhaite accéder à ces fonctionnalités.

Le réseau social peut également souhaiter offrir une expérience et des fonctionnalités uniques selon les tranches d'âge de ses utilisateurs, et pour cela, il a besoin de connaître la tranche dans laquelle l'utilisateur se situe. Dans ce cas, un système de contrôle de l'âge serait pertinent pour vérifier ou pour estimer la tranche d'âge de l'utilisateur, et non pas son âge exact.

2.2 Vérification du consentement parental

L'article 8 du RGPD constitue la première pierre à la construction d'un concept de **majorité numérique**. Il prévoit que la fourniture de services de la société de l'information à un mineur de moins de 16 ans doit se faire avec le consentement parental, les États membres étant libres d'abaisser cet âge jusqu'à 13 ans. La France a, depuis lors, fait le choix de ramener l'âge limite à 15 ans, ce qui est désormais prévu à l'article 45 de la loi Informatique et libertés.

Ainsi, en principe, un mineur de moins de 15 ans ne peut pas consentir seul au traitement de données personnelles dans le cadre d'un service numérique. Il doit pour cela avoir une autorisation parentale. En outre, le législateur français a ajouté que les informations à fournir au mineur doivent être rédigées *« en des termes clairs et simples »,* et *« aisément compréhensibles ».*

Néanmoins, cette limite d'âge s'est confrontée à des difficultés d'application depuis plusieurs années. Le législateur français a récemment adopté la **loi n°2023-566 du 7 juillet 2023 visant à instaurer une majorité numérique et à lutter contre la haine en ligne**. Cette loi complète la LCEN et s'applique spécifiquement aux plateformes de réseaux sociaux.

La majorité numérique est donc fixée à 15 ans en France. Cela signifie que les mineurs de moins de 15 ans ne peuvent pas créer de compte sur les plateformes en ligne sans l'autorisation d'au moins un titulaire de l'autorité parentale. Les plateformes ont en conséquence l'obligation de refuser l'accès à leurs services par les mineurs de moins de 15 ans ne justifiant pas d'une autorisation parentale. Aussi, le parent a le droit de demander de suspendre le compte de son enfant de moins de 15 ans, et donc retirer le consentement parental qu'il aurait préalablement donné.

D'un point de vue opérationnel, les plateformes doivent mettre en place une solution technique au sein de leurs services, laquelle devra se conformer à un référentiel élaboré par l'Arcom, pris après consultation de la CNIL.

Les plateformes n'ayant pas mis en place une solution technique pourront recevoir une mise en demeure du président de l'Arcom. Elles auront alors 15 jours pour se conformer au dispositif, à défaut de quoi le président de l'Arcom pourra saisir le président du tribunal judiciaire de Paris pour ordonner la mise en place de la solution technique.

Le non-respect du dispositif instauré pour la majorité numérique pourra être sanctionné par une **amende pouvant aller jusqu'à 1 % du chiffre d'affaires mondial annuel** de l'entreprise.

Les modalités pratiques d'application de ces obligations seront déterminées par décret (lequel n'est pas encore paru à la date de cette AFNOR SPEC). Les plateformes ont un délai d'un an pour mettre en place le système d'autorisation parentale pour les nouveaux utilisateurs et un délai de deux ans pour les utilisateurs ayant déjà un compte sur leurs services.

D'autres obligations complètent le dispositif de la majorité numérique :

- Informer les mineurs de moins de 15 ans sur *« les risques liés aux usages numériques et les moyens de prévention »* et sur les conditions d'utilisation de leurs données personnelles.
- Installer un système de contrôle du temps passé en ligne sur la plateforme.
- Diffuser des messages de prévention contre le cyberharcèlement, en indiquant notamment le numéro 3018, le numéro pour les jeunes victimes de harcèlement et de violences numériques.

Ce nouveau système de contrôle parental devra trouver un juste équilibre entre la sécurité des mineurs et la protection de leur vie privée et de leur intimité. L'autorisation donnée par le parent ne devrait en aucun cas permettre à ce dernier de surveiller en continu l'activité en ligne de son enfant. L'exercice de l'autorité parentale, si elle est nécessaire pour accompagner le mineur dans la découverte d'Internet, ne devrait pas pour autant le priver de son autonomie et de son intimité. Les réseaux sociaux devraient aussi être un espace où les jeunes peuvent exercer leurs libertés individuelles, exprimer leurs opinions, et créer du lien social indépendamment de leurs parents.

Les plateformes devraient donc éviter de proposer des fonctionnalités favorisant la surveillance des mineurs d'une façon disproportionnée par des mesures comme la géolocalisation en continu, ou l'accès au contenu des conversations privées ou à l'historique de navigation. Il est préférable de favoriser l'accompagnement dans la vie numérique via des campagnes de prévention et d'information, la modération des contenus et des interactions indésirables, la mise en place de fonctions de limitation du temps d'écran ou d'horaires que le parent pourrait paramétrer, etc. Ces initiatives sont déjà largement explorées par les grandes plateformes, qui permettent notamment la création de comptes communs familiaux.

Actuellement, les solutions disponibles se limitent à la vérification qu'un adulte donne son consentement. Ceci est souvent réalisé en vérifiant la possession d'un compte bancaire, d'une carte de crédit, ou par une analyse faciale. Si ce type de dispositif ne permet pas de vérifier la réalité d'un lien de parenté, elles constituent néanmoins des mesures pragmatiques, peu intrusives, et relativement simples à mettre en place.

Des vérifications plus approfondies, comme demander un extrait du livret de famille, semblent prématurées et excessives en l'état de la réglementation et dans l'attente de la publication du référentiel technique.

2.3 Vérification de l'identité

2.3.1 Proportionnalité de la mesure

Comme indiqué dans la section 2.1, la notion ***« d'identifiabilité »*** signifie qu'il convient de privilégier une identification raisonnable et minimale des utilisateurs, au regard de l'objectif à atteindre concernant la sécurité et l'intégrité de la plateforme. La vérification d'identité ne peut être une mesure généralisée en l'absence d'une justification particulière.

Elle pourrait être considérée comme proportionnée, par exemple, lorsqu'un ***« risque objectif »*** est caractérisé, ou si la réglementation exige de la plateforme qu'elle procède à un véritable contrôle de l'utilisateur. Par exemple : en cas de lois anti-blanchiment d'argent et de lutte contre le financement du terrorisme, des règles dites de « KYC » (***« Know Your Customer »***) imposent de connaître l'identité de ses clients ou utilisateurs. Cela peut être le cas si la plateforme propose des services financiers ou permet de transférer de l'argent.

Le DSA prévoit notamment que les marketplaces doivent s'assurer de la **traçabilité des vendeurs** professionnels sur leur plateforme, dont la vérification de documents d'identité peut constituer une étape dans l'évaluation de la fiabilité de ces professionnels.

En l'absence de justification légale ou réglementaire, les plateformes devraient évaluer si une méthode alternative plus respectueuse de la vie privée n'est pas disponible pour atteindre le même résultat, ou du moins proposer une méthode alternative aux utilisateurs. Dans ses recommandations sur les systèmes de contrôle de l'âge, la CNIL met en garde sur les risques découlant de cette mesure, notamment en raison du manque de fiabilité et de la nécessité de traiter un document officiel[3)].

Pour éviter le contournement de cette mesure, la CNIL suggère qu'un « ***test de détection du vivant*** » vienne compléter la mesure pour s'assurer que l'utilisateur est bien la personne qu'il déclare être. Sur un réseau social, la prise d'une photo ou d'une vidéo en direct via l'appareil de l'utilisateur est généralement efficace. Le dispositif devrait véritablement permettre de détecter « le vivant », c'est-à-dire qu'il soit en mesure d'établir qu'il s'agit d'une représentation en temps réel de l'utilisateur, et non pas d'une photo de photo par exemple. Pour cela, les technologies généralement employées permettent de détecter le mouvement, les textures et la profondeur dans l'image.

De plus, tous les documents d'identité ne présentent pas les mêmes fonctionnalités et niveaux de protection, notamment en ce qui concerne les puces biométriques. Par conséquent, la vérification de l'identité ne devrait être effectuée qu'à partir de documents offrant un niveau d'assurance suffisant, comme un passeport, une carte d'identité nationale, un permis de conduire ou un titre de séjour. Il est également possible de réaliser cette vérification à partir d'un fournisseur d'identité numérique proposant un service conforme à la régulation européenne **eIDAS**[4)].

En tout état de cause, la vérification d'identité devrait être utilisée de façon ciblée, par exemple en cas de risque de sécurité important pour la communauté d'utilisateurs (par exemple : usurpation d'identité, fraude, risque d'abus sur les mineurs, etc.).

Il existe de nombreux exemples : signalements répétés d'un même utilisateur pour des activités frauduleuses, faux comptes se livrant à des comportements nocifs pour la communauté, suspicions d'usurpation d'identité, risques d'extorsions sexuelles ou risque d'atteinte à l'intégrité physique d'un utilisateur (par exemple sur un site de rencontre).

Des débats existent, par exemple dans l'industrie des sites de rencontre, concernant la proportionnalité de certaines mesures, telles que l'accès aux registres de personnes condamnées pour des faits de violences sexuelles ou d'autres comportements illégaux.

La question de la proportionnalité de la mesure de la vérification de l'identité est essentielle, et les processus doivent être expliqués d'une manière claire et transparente par la plateforme. Cela implique de définir des processus de gouvernance, d'appel des décisions, et de rectification en cas d'erreur.

---

3) https://www.cnil.fr/fr/verification-de-lage-en-ligne-trouver-lequilibre-entre-protection-des-mineurs-et-respect-de-la-vie#:~:text=La%20CNIL%20a%20analys%C3%A9%20plusieurs,vie%20priv%C3%A9e%20des%20individus%20

4) Règlement européen (EU) n°910/2014 du 23 juillet 2014 sur l'identification électronique et les services de confiance pour les transactions électroniques au sein du marché intérieur et abrogeant la directive 1999/93/CE.

2.3.2 Sécurisation des documents d'identité

En tout état de cause, tout réseau social amené à traiter des pièces d'identité devrait se conformer aux recommandations de la CNIL[5)] en la matière. La Commission préconise, en effet, que les pièces d'identité utilisées pour vérifier l'identité d'une personne ne doivent pas être conservées pour une durée allant au-delà du temps nécessaire pour effectuer cette vérification. De cette façon, l'entreprise limite les risques que ces documents fassent l'objet d'une violation de données et tombent entre de mauvaises mains. En pratique, la durée nécessaire pour effectuer une vérification d'identité ne devrait pas excéder quelques jours.

Même si la durée de conservation doit être la plus courte possible, la plateforme doit stocker le document d'identité de manière sécurisée, et en définissant des règles d'accès strict en interne en fonction du besoin des employés d'y accéder. Les employés de l'entreprise doivent être formés sur la sécurité des données afin de pouvoir identifier et signaler les activités suspectes. Ils doivent être soumis à des obligations de confidentialité appropriées. Les éventuels sous-traitants engagés pour héberger ou traiter ces éléments doivent être soumis à des clauses de confidentialité et de protection des données personnelles conformes à l'article 28 du RGPD.

5) Référentiel relatif aux traitements de données à caractère personnel mis en œuvre à des fins de gestion des activités commerciales : https://www.cnil.fr/sites/cnil/files/atoms/files/referentiel_traitements-donnees-caractere-personnel_gestion-activites-commerciales.pdf

PARTIE II

Détection, modération et signalement

Partie II - Détection, modération et signalement

La modération du contenu et la sécurité des mineurs en ligne sont aujourd'hui des enjeux cruciaux dans notre société connectée. Les plateformes en ligne sont confrontées à un défi majeur : assurer la sécurité des utilisateurs tout en préservant la liberté d'expression et en respectant les obligations légales. Dans ce contexte, il est essentiel d'établir des normes et des bonnes pratiques pour la modération de contenu, afin de garantir un environnement en ligne sain et sécurisé.

Ce document fournit des lignes directrices claires et cohérentes pour la gestion des contenus inappropriés et/ou illicites sur les plateformes en ligne. Il aborde différents aspects, de la définition d'une politique de gestion à la mise en œuvre de procédures et d'outils de modération efficaces.

Il convient également, pour renforcer la confiance des utilisateurs, de les impliquer activement dans le processus de modération et de favoriser un environnement en ligne sécurisé et respectueux pour tous. Cette démarche passe par ces 5 piliers :

- **Transparence** : Une communication transparente est essentielle. Les utilisateurs doivent recevoir des notifications de confirmation de signalement, des mises à jour sur l'état de leur plainte et des informations sur les mesures prises. Cette transparence renforcée est par ailleurs requise par le DSA.

- **Réactivité/Célérité** : Il est primordial de répondre rapidement aux utilisateurs lorsqu'ils signalent un contenu inapproprié. Les signalements doivent être traités de manière appropriée et les réponses doivent être fournies dans les meilleurs délais, conformément aux règles légales.

- **Proactivité** : Protéger de manière proactive les utilisateurs en identifiant et en agissant sur les contenus problématiques avant que ceux-ci ne soient visibles sur la plateforme. Cette démarche proactive n'est pas imposée par la réglementation ; les plateformes sont toutefois invitées à mettre en place des mesures proactives pour renforcer la sécurité de leurs services.

- **Clarté** : La communication doit être claire, compréhensible et accessible. Les politiques de modération et les décisions prises doivent être expliquées de manière concise. Les utilisateurs doivent pouvoir comprendre pourquoi un contenu est jugé inapproprié et quelles actions ont été prises.

- **Empathie & Objectivité** : Il est essentiel de construire et développer des politiques de modération fondées sur le respect des droits fondamentaux et la protection des plus vulnérables. Les plateformes ont donc un devoir d'objectivité et de neutralité dans l'application de ces politiques et le traitement des signalements des utilisateurs ; l'objectif étant de créer une relation de confiance, et d'encourager les utilisateurs à continuer à signaler des contenus problématiques.

- **Amélioration continue** : Une démarche d'amélioration continue est nécessaire dans le développement d'une modération proactive efficace et éthique. Les commentaires des utilisateurs doivent être recueillis afin d'identifier les lacunes et d'ajuster les pratiques de modération et de communication en conséquence.

1 La création d'un environnement de confiance

1.1 La lutte contre les contenus *« manifestement illicites »*

Nous détaillerons dans cette section ce qui est attendu des réseaux sociaux lorsque des contenus illicites sont identifiés sur leurs services. Pour un rappel des sources législatives et réglementaires applicables à la régulation des plateformes et des contenus, se reporter à l'Annexe juridique - Cadre Juridique, Sections 2 et 3.

Les réseaux sociaux, qualifiés juridiquement d'hébergeurs de contenus, ont l'obligation de mettre en place des moyens efficaces de retirer de leurs services les contenus « manifestement illicites » qui leurs sont signalés. Cette approche vise à préserver la liberté d'expression sur Internet et la neutralité des plateformes ; celles-ci ne devraient pas pouvoir se substituer à une institution judiciaire pour apprécier la licéité d'un contenu au regard du droit applicable. C'est pourquoi leur responsabilité est limitée aux actions qu'elles entreprennent une fois informées de la présence de contenus manifestement illicites sur leurs services.

Les politiques de modération des plateformes doivent intégrer ce régime de responsabilité.

Notons, par ailleurs, qu'il n'y a pas d'obligation générale, pour les plateformes de détecter proactivement des contenus manifestement illicites. Le Considérant No. 30 du DSA dispose en effet que : ***« Aucune disposition du présent règlement ne devrait être interprétée comme imposant une obligation générale de surveillance ou une obligation de recherche active des faits, ou comme une obligation générale, pour les fournisseurs, de prendre des mesures proactives à l'égard des contenus illicites »***.

Il convient de définir ce qu'est un contenu ***« manifestement illicite »***. Le Conseil constitutionnel a considéré que la responsabilité d'un hébergeur ne saurait être engagée s'il n'a pas retiré une information notifiée comme 'illicite' si elle ne présente pas manifestement un tel caractère ou si son retrait n'a pas été ordonné par un juge (Cons. 10 juin 2004).

La jurisprudence française considère que le caractère manifestement illicite d'un contenu signalé à une plateforme *« ne peut être la conséquence que d'un manquement délibéré soit à la loi pénale, soit à une disposition de droit positif, explicite et dénuée d'ambiguïté »* (TGI Paris, 17e ch., 15 nov. 2004 ; TGI Paris, 30 mai 2017). Ainsi, le caractère illicite de l'information en question doit être flagrant et indiscutable.

Exemples d'infractions ayant pu être jugées comme manifestement illicites (en fonction du cas d'espèce) :

- Contrefaçon (CA Paris, 12 sept. 2017).
- Compilation de propos antisémites et révisionnistes (CA Paris, 24 nov. 2006).
- Services de gestation pour autrui (TGI Versailles, 26 févr. 2019, C. Cass, 23 novembre 2022).
- Atteinte évidente à l'intimité de la vie privée (CA Paris, 6 juin 2007).
- Harcèlement moral aggravé (TGI Paris, 30 mai 2017).

À l'inverse, les infractions complexes dans leur caractérisation devraient être exclues du champ de ce qui est « *manifestement illicite* ». C'est le cas, par exemple, de la diffamation (CA Paris, ordonnance de référé, avril 2013).

Cette jurisprudence française n'est, a priori, pas remise en cause par le DSA, lequel retient dans son Considérant 63 que *« lorsqu'il est évident pour un profane, sans aucune analyse de fond, que le contenu est illicite [...], alors celui-ci est manifestement illicite »*.

Concrètement, on peut tirer de ce régime juridique des critères objectifs pour qualifier le caractère ***« manifestement illicite »*** d'un contenu :

- Il peut être considéré comme illicite par un non-juriste.
- Il doit être analysé dans son rapport au cadre réglementaire national (exemple de la jurisprudence sur la promotion de la gestation pour autrui).
- Il est réprimé ou interdit en vertu d'une disposition dépourvue d'ambiguïté.
- Il ne nécessite aucune analyse de fond, en ce sens qu'il n'a pas besoin d'être interprété.

Les plateformes peuvent en faire une feuille de route pour leurs équipes de modération. Ainsi, lorsque les critères sont réunis, une mesure de retrait du contenu et/ou de blocage du compte peut être prise très rapidement, sans examen approfondi.

En plus des actions de modération, l'article 6-I-7 de la LCEN impose aux plateformes de **signaler aux autorités compétentes** des infractions particulièrement graves dont les suivantes : atteinte à la dignité humaine, harcèlement scolaire, harcèlement sexuel, importation illégale de tabac, incitation à la discrimination (genre, sexualité, handicap), incitation à la haine raciale, jeux d'argent illégaux, négation/banalisation crimes contre l'humanité, pédopornographie, promotion de thérapies de genre, prostitution, proxénétisme, terrorisme, traite d'êtres humains, violence ou menace, violences sexuelles et sexistes.

Les plateformes doivent donc être particulièrement attentives aux signalements effectués par leurs utilisateurs et qui pourraient entrer dans le champ de ces infractions. Leur particulière gravité est susceptible de les qualifier de *« manifestement illicites »*, bien davantage que d'autres faits qui nécessitent un examen ou des investigations approfondis (diffamation, insulte, droit d'auteur, atteinte au droit à l'image, etc.).

À noter qu'il y a des obligations renforcées concernant les contenus à caractère terroriste (cf. Section 3.2 de la Partie Cadre juridique).

Les plateformes de réseaux sociaux peuvent être sanctionnées si elles ne respectent pas les obligations mentionnées ci-dessus, en particulier :

- si elles laissent en ligne des contenus manifestement illicites dont elles avaient connaissance ou qu'elles ne les ont pas retirés *« promptement »* ;
- si elles ne mettent pas en place un dispositif de signalement facilement accessible et visible pour leurs utilisateurs ;
- si elles n'ont pas traité les signalements effectués par leurs utilisateurs ou qu'elles ne les ont pas signalés aux autorités compétentes lorsqu'elles en avaient l'obligation (ou n'ont pas respecté les délais applicables pour les signaler aux autorités).

Le DSA a considérablement renforcé les sanctions prévues en cas de manquement aux obligations prévues ; elles peuvent désormais atteindre jusqu'à **6 % du chiffre d'affaires mondial annuel** de l'entreprise.

Sur le fond, il a précisé et consolidé le régime applicable aux hébergeurs de contenus. Il a notamment instauré de nouveaux droits pour les utilisateurs concernant leurs contenus : droit d'être informé qu'une décision de modération a été prise à propos d'un contenu ou de leur compte, ainsi que le droit de pouvoir contester cette décision de modération.

Les plateformes doivent désormais être plus transparentes dans leurs pratiques de gestion des contenus. Parmi les nouvelles obligations de transparence :

- Expliquer le fonctionnement de leur système de modération de leurs algorithmes de recommandation (c'est-à-dire la façon dont les contenus sont présentés aux utilisateurs) avec un droit pour les utilisateurs de ne pas faire l'objet de recommandations de contenus sur la base de profilage.
- Informer les utilisateurs sur le système de ciblage publicitaire présent au sein de leurs services et sur les outils mis à leur disposition pour modifier ces paramètres de ciblage.
- Produire des rapports annuels contenant des informations sur le nombre de contenus qu'elles ont modérés, les réclamations qu'elles ont reçues et traitées, les outils informatisés qu'elles utilisent pour soutenir leurs activités de modération, ainsi que des informations relatives au niveau d'efficacité de ces outils.

Pour plus d'informations sur le DSA, voir la section 2.4 de l'Annexe sur le cadre juridique.

1.2 Établir des politiques de la plateforme, règles communautaires et guidelines internes

1.2.1 Politiques et règles de la plateforme

Au-delà de constituer une sorte de « règlement intérieur » de la plateforme, les politiques d'une plateforme et les règles communautaires reflètent les valeurs de cette dernière, elles participent à créer l'identité de la plateforme en définissant ses contours.

Il est également indispensable de penser l'articulation de ces règles avec les autres politiques/documents juridiques tels que les conditions générales d'utilisation, ainsi que la politique de confidentialité des données.

L'élaboration des résiliations des plateformes (règles de la communauté) découle à la fois d'une obligation légale d'informer les utilisateurs et d'une nécessité opérationnelle.

Les règles communautaires

Les règles communautaires sont un élément vital pour créer un environnement en ligne plus sûr et positif. Elles peuvent aller au-delà des réglementations pour garantir la sécurité des utilisateurs et pour s'adapter en permanence aux défis changeants du monde en ligne. Des règles claires et transparentes sont essentielles afin de favoriser une expérience en ligne constructive, enrichissante et respectueuse pour tous.

Aussi appelées politiques, cet ensemble de règles vise à régir tout ce qui pourrait potentiellement mettre en danger la sécurité des utilisateurs ou affecter leur bien-être. Cela inclut une variété de sujets qui vont du contenu injurieux ou à caractère sexuel jusqu'à des activités plus graves comme l'automutilation ou le trafic d'êtres humains.

Les règles communautaires ne sont pas fixes. Elles sont conçues pour évoluer en permanence, afin de faire face aux nouvelles menaces et défis qui émergent en ligne. Cette adaptabilité est primordiale et garantit que les règles de la plateforme restent pertinentes à l'épreuve du temps. Cela permet d'assurer la pérennité de la plateforme en ce qu'elle saura protéger efficacement ses utilisateurs contre de nouveaux risques.

Les règles de la plateforme sont élaborées en adéquation avec les guidelines internes de modération. Cela signifie qu'elles sont conçues pour correspondre aux processus de modération relatifs aux thématiques ainsi qu'aux comportements autorisés et interdits sur la plateforme. Cette cohérence assure que les règles sont appliquées de manière équitable et compréhensible.

Il est essentiel que les règles de la plateforme soient rédigées de manière claire, lisible et transparente. Elles doivent être compréhensibles pour tous les utilisateurs, quel que soit leur niveau de familiarité avec la plateforme et quel que soit leur âge. Cela implique d'utiliser un langage simple et accessible, ainsi que de prendre en compte la diversité des formats de contenu utilisés sur la plateforme.

Devoir de communication et d'accessibilité : en pratique, peu de personnes lisent les documents juridiques des plateformes, en raison de leur longueur et de leur technicité juridique. Il est donc recommandé de détailler les règles relatives aux contenus de manière intelligible dans un endroit dédié de la plateforme (type blog, support etc.), et de les faire traduire dans les langues des pays où la plateforme a une communauté importante. Ces règles doivent être renseignées dans les conditions générales de la plateforme.

1.2.2 L'amélioration continue de ces règles et des politiques de la plateforme

L'élaboration et la mise à jour de politiques efficaces pour une plateforme en ligne constituent une démarche indispensable pour instaurer un environnement sécurisé et bienveillant. Afin de mener à bien cette entreprise complexe, plusieurs axes stratégiques peuvent être empruntés et se superposer :

- **Expertise multidisciplinaire** : La constitution d'une équipe diversifiée d'experts en provenance de différentes régions et domaines d'expertise tels que les discours incitant à la haine, la sécurité des mineurs, le terrorisme, ou encore les droits humains offre une perspective globale et holistique dans la conception des politiques. Cette diversité garantit la prise en compte de nombreuses nuances et complexités inhérentes à la sécurité des utilisateurs et à la modération des contenus sensibles.

- **Anticipation des répercussions** : Mettre en place un groupe dédié à l'évaluation des conséquences des modifications apportées aux politiques peut se révéler essentiel. Cela permet non seulement d'anticiper les éventuels effets secondaires des nouvelles mesures, mais également de développer des outils technologiques aptes à appliquer ces directives avec précision et cohérence.

- **Analyse des insuffisances** : Une veille constante sur d'éventuelles lacunes dans les politiques existantes est essentielle. En encourageant la participation des parties prenantes externes et en surveillant les signaux médiatiques, il est possible d'identifier et de combler rapidement les manques dans les politiques en vigueur.

- **Démarche de perfectionnement continu** : Il faut continuellement encourager les équipes dédiées à collecter des données et des réactions d'utilisateurs à utiliser ces informations pour façonner de nouvelles politiques en adéquation avec des besoins en constante évolution. Il est par ailleurs crucial de recueillir les remontées de terrain issues des personnels en mesure de suivre les tendances émergentes. Finalement, l'amélioration perpétuelle des politiques se fonde sur l'analyse continue de leur efficacité. Cette approche proactive implique d'apporter des ajustements au fil du temps pour créer un environnement en ligne qui reflète fidèlement les valeurs de la plateforme.

1.2.3 L'opérationnalisation des politiques internes

Afin de faciliter leur application et futur développement, il est important de catégoriser les politiques internes autour de grands piliers qui regroupent les comportements et contenus spécifiques susceptibles de survenir sur la plateforme.

Exemple de méthode & développement des politiques internes

Catégories	Définition et cadre	Types de violations	Contenus/comportements modérés	Contenus/comportement tolérés
Grandes thématiques ayant pour but de segmenter les types de comportements/contenus susceptibles de survenir sur la plateforme	Permet de donner la définition et le cadre de la catégorie en question pour comprendre quel comportement/contenu est susceptible d'être applicable ou non	Le type de comportements/contenus spécifiques rattachés à la catégorie en question (il peut prendre la forme de sous catégories)	Le type de contenus/comportement modéré et sanctionné	Le type de contenus/comportement qui dans un contexte de liberté d'expression sont tolérés, et ne vont pas à l'encontre des règles de la plateforme
Harcèlement	Comportements abusifs caractérisés par : - une intention de causer un mal psychologique et une humiliation envers une autre personne - comportement qui peut être répété et persistant Ils peuvent inclure : Les moqueries, les insultes (physiques ou morales), la violation de la vie privée, etc. Hors cadre : la discrimination et les contenus haineux	- Les contenus injurieux - Les campagnes de harcèlement - etc.	Exemple d'insultes	Certains contenus dans un contexte amical, partage sous le ton de l'humour ou il n'y pas d'intention d'humilier ou causer du mal

Template d'une politique interne

Politique interne - Template

1. Justification de la politique

- Rappel de la mission de la plateforme
- Justification et explication du besoin de la politique en question (risque et impact)
- Definition generale de la politique en question

2. Constat

- Comment se manifeste le probleme expose aujourd'hui
- Comment est-il traite (Rercherche, benchmark etc)
- Impact sur l'utilisateur et la plateforme

3. Cadre

- Type de contenus/comportements applicables

Limites/hors cadre

- Type de contenus/comportements applicables à une autre politique interne.

Exceptions

- Types de contenus/comportements qui pourraient etre toleres par la plateforme (liberté d'expression, contexte etc)

4. Types de violations - (sous categories)

Sous thematique	
Definition	**Facteurs contextuels**
Definition et explication generale du type de violation en question	**Aggravants:** le contexte va parfois determiner un niveau d'urgence, ou de severite. **Attenuants:** A l'inverse, le contexte peut parfois attenuer le risque en question et donc etre modere de maniere differente.
scenarios	
Les scenarios susceptibles de rentrer dans la thematique en question	

Examples

- Illustrer les scenarios avec des exemples concrets.

5. Questions Frequentes

Ajouter les questions frequentes des professionels pour faciliter les process operationnels.

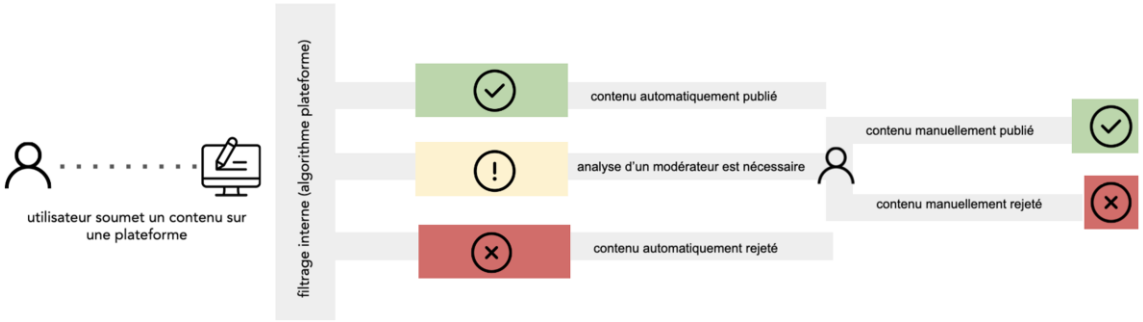
2 Façonner des procédures et des outils de détection des contenus inappropriés et/ou illicites

2.1 Processus et solutions de détection

Afin de garantir une expérience sûre à ses utilisateurs, une plateforme doit adopter une approche équilibrée et complète, ce qui implique de travailler sur deux fronts : la détection proactive et la détection réactive. Chacune de ces méthodes joue un rôle distinct mais complémentaire.

Incluant aussi la pré-modération, la modération proactive consiste à identifier et à agir sur des contenus problématiques avant que ceux-ci ne soient visibles sur la plateforme (et donc sans intervention d'un utilisateur). Cette approche réduit les risques en limitant l'exposition aux contenus inappropriés, notamment grâce à des mécanismes de détection automatique (mots-clefs, intelligence artificielle, bases de données de contenus sous forme de *hash*, etc.). La modération réactive intervient quant à elle en réponse aux signalements des utilisateurs. Elle porte ce nom car elle s'enclenche après que les utilisateurs ont été exposés à des contenus ou comportements inappropriés. Un de ces enjeux inclut l'amélioration de la détection proactive en identifiant les lacunes et en complétant les mesures de prévention essentielles.

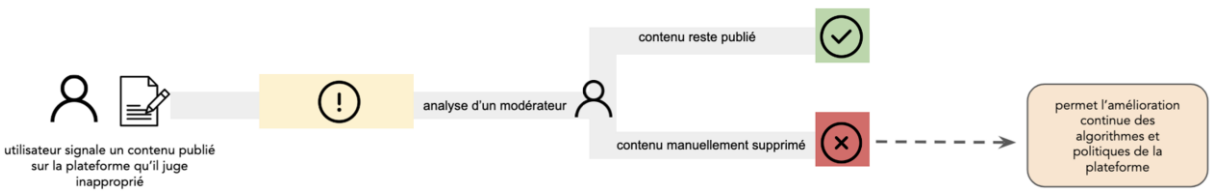
La détection proactive



Tout contenu publié manuellement ou automatiquement peut être reporté par un utilisateur,ce contenu reporté doit être évalué par une équipe dédiée et peut être supprimé à tout moment si il enfreint les politiques de la plateforme, ou si l'auteur/les autorités en font la demande

Les auteurs des contenus supprimés peuvent faire recours à des mécanismes d'appels mis à disposition par la plateforme

L'amélioration des systèmes proactifs



2.1.1 Détection proactive

La modération proactive est le premier type de modération pouvant intervenir dans le traitement des contenus. Elle consiste en l'analyse d'un contenu dès sa soumission, que ce soit par une machine ou par un être humain, afin de décider s'il doit être retiré ou non. Contrairement à la modération réactive qui dépend des signalements des utilisateurs, la modération proactive intervient avant tout signalement.

Pour mettre en œuvre la modération proactive, il est par exemple possible d'utiliser des systèmes d'intelligence artificielle capables de reconnaître les éléments présents dans une image ou d'analyser le texte d'une publication. Le système de modération proactive peut ensuite prendre automatiquement des mesures telles que la suppression du contenu ou la réduction de sa diffusion.

Dans certains cas, un contenu nécessite un examen approfondi. Le système d'analyse automatique le transmet alors à une équipe d'examinateurs pour une analyse détaillée. Dans ce scénario, ce sont les examinateurs humains qui prennent la décision finale. Les contenus analysés par les examinateurs ainsi que la décision prise par ceux-ci peuvent être utilisés pour améliorer la performance des systèmes de détection automatisé, à condition de justifier d'une base légale permettant cette réutilisation des données des utilisateurs. Actuellement, sur les plus grandes plateformes de réseaux sociaux, une part croissante des contenus problématiques est identifiée et retirée avant tout signalement.

Plusieurs facteurs rendent difficilement contournable le recours à la détection automatisée dans le cadre de la modération proactive lorsque celle-ci est mise en place :

- **Volume massif de contenus** : le flux constant de contenus créés et publiés sur les réseaux sociaux est immense. Par exemple, Facebook compte plus de 2 milliards d'utilisateurs actifs par jour.
- **Impossibilité de mobiliser une main d'œuvre humaine suffisante** : la mise en place d'une infrastructure humaine pour modérer proactivement l'ensemble de ces contenus est irréalisable.
- **Exigence d'instantanéité** : les besoins d'une réaction rapide face aux contenus manifestement illicite imposent des contraintes temporelles strictes.

La modération automatique offre des avantages significatifs :

- **Rapidité d'action** : contrairement à une pré modération humaine, une solution automatisée peut analyser rapidement de grandes quantités de données et prendre des mesures de modération en moins de 100 millisecondes, permettant que les contenus soient immédiatement visibles malgré la modération proactive garantissant une protection instantanée.
- **Maîtrise des langues** : les systèmes automatisés peuvent analyser les contenus dans de nombreuses langues, ce qui réduit la charge opérationnelle de devoir constituer et encadrer des équipes de modérateurs compétents dans chaque langue.
- **Uniformité dans la modération** : la modération automatisée peut permettre l'application d'une modération uniforme pour tous les utilisateurs. Cependant, il convient de noter que l'objectivité totale et l'absence de biais n'est pas garantie, car les machines sont les produits de leurs créateurs. Ainsi, l'intervention humaine reste indispensable pour équilibrer et corriger les éventuelles discriminations inhérentes aux systèmes automatisés de modération.

Les technologies de modération automatique sont variées. Le *Natural Language Processing* (NLP) ou traitement automatique de la langue naturelle permet de comprendre la langue et son contexte. ***Le Machine Learning*** et ***deep Learning*** (ML/DL) ou apprentissage automatique et apprentissage profond apprend à la machine à accomplir des tâches cognitives. Les mots-clés détectent des termes spécifiques, mais leur efficacité peut être limitée par le manque de contexte. Cette méthode est particulièrement utile pour la détection proactive effectuée par les utilisateurs, c'est-à-dire pour les cas de figure où les utilisateurs peuvent bloquer des termes indésirables, mais elle peut manquer de précision.

2.1.2 Détection réactive

Un second type de modération est la modération réactive : un utilisateur est exposé à un contenu qu'il juge indésirable. Il remonte alors le contenu à la plateforme pour que celle-ci l'analyse et prenne une décision de modération. Dans ce cas, disposer d'équipes de modération humaine est fondamental, bien que, selon la volumétrie des contenus, certaines tâches puissent être automatisées.

Cette autre modalité de la modération se matérialise par l'intégration d'un mécanisme de signalement direct et accessible sur chaque contenu. Ce procédé résulte d'une obligation légale destinée à garantir une modération réactive efficace. Cette approche permet aux utilisateurs de jouer un rôle actif dans la préservation d'un environnement en ligne sécurisé et conforme aux normes des plateformes :

- Les utilisateurs doivent disposer d'un accès aisé à ce bouton, leur permettant ainsi de signaler promptement les contenus inappropriés ou problématiques qu'ils pourraient rencontrer lors de leur navigation.

- Il est impératif de garantir aux utilisateurs la possibilité de signaler divers types de contenus, qu'il s'agisse de profils, de pages, de publications, de commentaires, de messages, de groupes, d'événements ou de publicités. Dans le but de simplifier cette démarche, il est recommandé de fournir des informations exhaustives sur les différentes catégories de signalement, et éventuellement, orienter les utilisateurs vers une page informative dédiée.

- Il est impératif de veiller à ce que le processus de signalement soit à la fois simple et intuitif, de manière à ne pas décourager les utilisateurs. L'idée d'adopter un canal de signalement unique pourrait faciliter davantage cette démarche. Toutefois, il est primordial de spécifier clairement les contenus manifestement illégaux, en insistant sur leur nature problématique.

- Dès que le signalement est soumis, une notification informe l'utilisateur que la demande est en cours de traitement. La communication de l'issue de la modération, qu'elle aboutisse ou non au retrait du contenu, doit être assurée. En cas de retrait, un rappel aux règles qui régissent les usages de la plateforme vient au soutien de cette décision pour une meilleure compréhension de cette dernière par l'utilisateur.

- Il est à noter que le processus de signalement doit être conçu pour préserver l'anonymat de l'utilisateur. Son identité demeure confidentielle à l'égard des autres utilisateurs, tout en étant connue et dûment conservée à des fins de sécurité et/ou d'enquête par les services de la plateforme.

Pour garantir une gestion efficace des signalements au sein des équipes de modération, il est essentiel de mettre en place une procédure opérationnelle claire. Une approche visant à optimiser et rendre intelligible cette procédure doit guider sa mise en place. Un certain nombre d'éléments clés peuvent être identifiés aux fins de répondre à ces besoins :

- **Priorisation basée sur la gravité** : Débuter par la hiérarchisation des signalements en fonction de la gravité de l'infraction. Cela implique de donner la priorité aux cas les plus graves (pédocriminalité, terrorisme) pour garantir une réponse rapide et adéquate.

- **Priorité aux « signaleurs de confiance », conformément au DSA** : Accorder la priorité opérationnelle aux signalements émis par les signaleurs de confiance, dont le rôle est défini au sein du DSA. L'expertise reconnue de ces entités en font des collaborateurs privilégiés, lesquels soutiennent la démarche dans laquelle s'engage la plateforme.

- **Mécanisme d'appel** : Cette fonctionnalité devient obligatoire avec le DSA. Elle doit permettre qu'un nouvel examen soit réalisé sur une décision de modération. Certaines plateformes cherchent à perfectionner ces mécanismes, en instituant des organes indépendants spécifiquement dédiés à cette mission (par exemple, le Conseil de surveillance de Meta[6)]).

- **Utilisation d'API à des fins de fluidification** : Considérer l'automatisation de certaines étapes du processus grâce à l'utilisation d'API. Cette approche permet d'optimiser l'efficacité opérationnelle, en mettant en relation les systèmes des différents acteurs jouant un rôle dans la modération des contenus.

- **Précision et collégialité dans la prise de décision** : Encourager l'équipe à analyser la nature de l'infraction, sans être limitée par une première impression. Une requalification du contenu peut s'avérer nécessaire dans certains cas et peut intervenir à l'issue de concertations collectives, en sollicitant par exemple l'avis de collègues.

2.2 La modération

2.2.1 Application des politiques internes

Mesures de protection à l'échelle des contenus

Pour assurer le respect des politiques internes, une approche en trois volets est adoptée pour la mise en application des règles relatives au contenu : supprimer, limiter et informer.

- **Supprimer** : Dès que des contenus en violation des politiques sont identifiés, ils sont supprimés immédiatement. Par exemple, si une publication contient des discours de haine ou de la violence explicite, elle est supprimée afin de maintenir un environnement sûr et respectueux pour les utilisateurs.

6) https://www.oversightboard.com/

- **Limiter** : Certains contenus problématiques, même s'ils ne répondent pas entièrement aux critères de suppression, peuvent être restreints dans leur diffusion. Par exemple, si une publication contient du langage vulgaire, elle peut être réduite dans sa visibilité pour minimiser son impact négatif. Ainsi, les utilisateurs sont moins susceptibles d'être exposés à des contenus potentiellement nuisibles.

- **Informer** : Lorsque les contenus sont potentiellement sensibles, comme par exemple quand un utilisateur partage une de ses données privées, des avertissements peuvent être partagés avec l'utilisateur en question, afin de l'informer sur les risques de communiquer une telle information sur la plateforme. L'approche informative peut être utilisée à plusieurs égards : avertissement avant d'accéder à des contenus sensibles mais non proscrits, ou encore alerte lorsqu'un contenu porte un risque de désinformation ou de propos complotistes mais qui reste dans les limites de la liberté d'expression.

Ces mesures de sanctions graduelles permettent d'assurer une gestion proactive des contenus problématiques, en supprimant ceux qui enfreignent clairement les règles, en restreignant la diffusion de ceux qui sont potentiellement nuisibles et en informant les utilisateurs sur les contenus douteux. Ainsi, l'objectif est de maintenir des plateformes sécurisées, respectueuses et fiables pour les utilisateurs.

Solutions de dissuasion à l'échelle de l'utilisateur et incitations positives

Une autre composante essentielle de la gestion de la modération des contenus est la mise en place de solutions de dissuasion à l'échelle des utilisateurs, ainsi que des mécanismes d'incitation positive. Ces mesures visent à décourager les comportements indésirables et à promouvoir un environnement en ligne respectueux et constructif. Voici deux exemples concrets de telles mesures :

- **Dissuasion** : Politiques de lutte contre la récidive : Les plateformes de réseau social mettent en place des politiques de lutte contre les récidivistes (***« Repeated Infringers »***). Par exemple, si un utilisateur enfreint de manière répétée les règles de la plateforme, des sanctions peuvent être prises. Cela peut aller de la restriction de la visibilité du contenu à la suspension temporaire voire à la suppression définitive du compte. Ces mesures dissuasives visent à responsabiliser les utilisateurs et à réduire les comportements nuisibles répétitifs.

- **Incitation positive** : Récompenses et reconnaissances : Certaines plateformes mettent en place des mécanismes d'incitation positive pour promouvoir un comportement vertueux et positif. Par exemple, X (ex-Twitter) a introduit les *« Notes de la Communauté »* (*« Community Notes »*), qui sont des badges spéciaux attribués aux utilisateurs qui se distinguent par leur contribution positive à la communauté en ligne. Ces badges peuvent être accordés aux utilisateurs qui interagissent de manière respectueuse, partagent des informations utiles ou contribuent de manière constructive aux conversations, en particulier en apportant des précisions ou du contexte à des publications potentiellement trompeuses. Les Notes de la Communauté servent à mettre en avant et à reconnaître les comportements positifs des utilisateurs, créant ainsi une dynamique d'incitation à adopter des attitudes bienveillantes et constructives.

En mettant en œuvre ces solutions de dissuasion et d'incitation positive, les plateformes cherchent à encourager un comportement responsable, respectueux et constructif de la part des utilisateurs.

Mise en place des moyens techniques pour permettre la prise de décision

La mise en place de moyens techniques pour permettre la prise de décision est essentielle, en considérant bien l'ensemble des actions nécessaires :

- des interfaces d'administration, pour gérer les contenus et les utilisateurs de la plateforme, indispensables pour s'assurer de pouvoir intervenir en toute circonstance ;

- des consoles de modération, lorsque les volumes de contenus à vérifier manuellement deviennent importants. Une interface adaptée à l'activité de la plateforme est importante pour l'efficacité de l'activité de modération (l'affichage uniquement des informations clés permet une prise de décision à la fois rapide et précise).

Des outils automatisés peuvent également se montrer utiles en fonction des volumes et/ou des complexités propres aux besoins de la plateforme. Ces outils peuvent avoir plusieurs fonctions :

- détection d'un risque d'infraction ;

- aide à la décision, en mettant à disposition des modérateurs des indicateurs qui améliorent l'efficacité de la prise de décision (attention toutefois à ne pas créer de biais ou de dépendance dans la prise de décision) ;

- la qualification, en prenant des décisions automatiques.

Il est essentiel de mettre en place des processus de contrôle qualité des activités de modération en établissant des KPI (*Key Performance Indicators*) ou indicateurs à superviser pour mesurer l'exactitude des actions prises.

En mettant en œuvre ces approches, la gestion de la modération des contenus se trouve renforcée, garantissant une application cohérente des politiques internes tout en facilitant la prise de décision efficace.

Il est possible de mettre en place par ses propres moyens ces 3 approches, mais il est également possible de s'appuyer sur l'expertise de solutions tierces.

L'intégration des politiques internes dans les outils de modération repose sur deux aspects importants :

- **Processus d'amélioration continue** : Un processus d'amélioration continue est mis en place pour améliorer la détection proactive des contenus problématiques et ajuster les politiques en fonction des évolutions.

- **Développement d'expertises des professionnels** : Des professionnels formés et expérimentés sont chargés de faciliter l'application cohérente des politiques internes.

2.2.2 Les professionnels de la Trust & Safety

Au sein du secteur de la sécurité en ligne (communément appelé *« Trust & Safety »*), l'épanouissement et le développement des compétences professionnelles revêtent une importance capitale pour une modération efficiente. Ces préoccupations se révèlent d'autant plus importantes que le secteur se trouve au croisement de multiples domaines d'activités, de telle sorte que les formations et diplômes spécialisés peinent encore à émerger. Un recrutement éclairé et des formations adaptées viennent ainsi soutenir ce processus :

- **Recrutement ciblé et avisé** : La création d'une équipe *Trust & Safety* solide requiert un recrutement avisé. La présence de juristes au sein de ces équipes, ou une collaboration étroite avec le service juridique, garantit une compréhension pointue des normes et des réglementations. Une transparence quant aux aspects du métier, notamment l'exposition à des contenus violents, est essentielle, particulièrement lors du processus de recrutement. Prévoir des clauses de confidentialité et de protection des données personnelles, tout en évaluant la résilience psychologique des futurs employés, instaure un environnement favorable.

- **Développement des compétences** : Pour garantir une modération de qualité, une formation complète et continue est indispensable. Les biais humains sont inévitables, influencés par nos contextes sociaux et culturels. Pour une détection précise et une modération efficace, les règles de la plateforme doivent être éprouvées au principe d'égalité et de neutralité. Offrir un accès régulier à des formations permet aux modérateurs de maintenir les compétences requises. Ces formations devraient couvrir une formation théorique (juridique et règles de la plateforme) ainsi qu'une formation technique (interface de travail). L'intervention d'experts et le partage d'expériences enrichissent cette démarche.

- **Adaptation individuelle** : Afin de préserver la santé mentale et émotionnelle des professionnels, offrir un environnement souple est crucial. Chaque personnel devrait pouvoir exprimer ses besoins et idéalement ajuster son niveau d'exposition aux contenus, afin de tenir compte de son équilibre psychologique, en particulier lorsqu'il est exposé aux contenus les plus choquants. Un soutien psychologique devrait leur être proposé.

De manière significative, les efforts pour renforcer ces bonnes pratiques dans la gestion des personnels sont également portés par des structures associatives telles que Point de Contact[7)]. Ces initiatives jouent un rôle central dans la conceptualisation et la promotion de recommandations émises dans le but de mieux protéger les professionnels de la *Trust & Safety*.

7) « Modération des contenus illicites en ligne. Opérations et protection des professionnels » Livre blanc portant sur le signalement de contenus illicites et la prise en charge des professionnels exposés à des contenus choquants. Point de Contact, Paris, 2023 : https://www.pointdecontact.net/livre-blanc-pedopornographie-terrorisme/

3 Signalement aux autorités de police et coopération

Deux sujets principaux vont se poser pour une plateforme en matière de coopération avec les autorités de police :

- Le signalement de contenus illicites et de situations d'urgence impliquant un risque d'atteinte à la vie humaine.
- La réception et le traitement des réquisitions judiciaires.

3.1 Coopération proactive avec les autorités ou avec les organisations dédiées

Comme rappelé plus haut et dans la section 3 de l'Annexe sur le Cadre juridique, les plateformes sont assujetties à des obligations de signalements de certains contenus aux autorités de police, au premier rang desquels les contenus à caractère terroristes ou pédopornographiques.

En plus de l'obligation de retrait, ces contenus doivent aussi être signalés le plus rapidement possible, quelques heures pour les plus dangereux, afin que les autorités puissent intervenir, le cas échéant, avant qu'un dommage réel ne soit infligé.

Une fois ces contenus identifiés conformément aux critères définis juridiquement, leur qualification précise revient aux autorités de police à qui ces contenus ont été signalés.

L'une des difficultés majeures qui peut se poser pour une plateforme est d'identifier l'autorité compétente pour effectuer ce signalement et ce notamment au regard des considérations géographiques et du type de comportement dont il est question.

Une plateforme de droit français dont les utilisateurs peuvent se situer dans d'autres pays du monde se posera donc nécessairement la question de savoir à quelle autorité et par quel moyen effectuer le signalement en question :

- Moyens techniques existants pour procéder à ces signalements en intégrant la dimension internationale (e.g. Pharos, NCMEC...).
- Mise en place de procédures internes permettant la bonne identification des contenus à signaler et leur suivi.
- Mise en place d'une politique de conservation spécifiques de ces données.
- Conservation de données agrégées, statistiques (boucle de rétroaction avec les équipes de sécurité en ligne, constitution du Transparency Report...).

Les plateformes travaillent de concert avec les autorités pour assurer un environnement sûr aux utilisateurs. Cela implique parfois de transmettre des informations aux autorités en cas d'urgence. Dans les cas de risque imminent pour un enfant ou de risque de mort ou d'atteinte sérieuse à l'intégrité physique d'une personne, et lorsque le cas en question oblige à la divulgation d'informations sans délai, un représentant des autorités de police peut soumettre une demande par le biais d'un système de demande en ligne destiné aux forces de l'ordre[8)]. Tout utilisateur ayant connaissance d'une situation d'urgence doit contacter immédiatement et directement les autorités des forces de l'ordre locales compétentes.

Pour ce qui concerne la protection des mineurs en particulier, des dispositifs de signalements spécifiques aux autorités existent. Sur le territoire des États-Unis et du Canada par exemple, les plateformes signalent généralement tous les cas d'exploitation des enfants au ***« National Center for Missing and Exploited Children » (NCMEC)***, y compris les contenus portés à leur attention par les autorités. Le NCMEC coopère avec les autorités de police partout dans le monde.

3.2 Répondre aux réquisitions judiciaires

Il est fréquent que les autorités de police demandent aux plateformes de coopérer dans le cadre d'enquêtes judiciaires, notamment pénales. Ces demandes portent le plus souvent sur la communication de données relatives à des utilisateurs : localisation, adresse IP, adresse email, contenus, etc. Ces éléments permettent non seulement d'identifier précisément un utilisateur ou de le localiser, mais aussi de constituer des preuves ou de rechercher des indices dans le cadre d'enquêtes. C'est pourquoi la politique de conservation des données des plateformes doit prendre en compte cette nécessité de coopérer avec les autorités.

L'autorité peut aussi demander de préserver les données d'un utilisateur pendant un certain temps (en général quelques mois). La préservation des données est une mesure préventive ciblée pouvant, le cas échéant, être suivie d'une demande de communication de données.

Les plateformes doivent mettre en place des processus opérationnels permettant de faciliter cette coopération, tout en s'assurant du respect des obligations légales applicables et de la protection de la vie privée :

- Mise en place d'un processus interne destiné à traiter les réquisitions judiciaires et demandes de gel de données etc. (Équipes dédiées, suivi des réquisitions). Il est impératif que le service chargé de répondre aux autorités soit encadré par le département juridique, lequel fournit les instructions nécessaires pour vérifier la légitimité et l'authenticité des requêtes, ainsi que leur compatibilité avec la règlementation (RGPD, DSA, etc.).

8) Par exemple, celui-ci pour Meta : https://fr-fr.facebook.com/help/494561080557017

- Pour les réquisitions judiciaires provenant de l'étranger : il convient de mettre en place des moyens de traitement si elles sont adressées directement à la plateforme. Pour une plateforme française, il est recommandé de faire approuver les requêtes étrangères par la Direction Nationale de la Police Judiciaire (sous-direction de la lutte contre la cybercriminalité).

- Conservation de données agrégées, et de statistiques (boucle de rétroaction avec les équipes de sécurité et de modération, constitution de rapports de transparence...). Des durées de conservation spécifiques doivent être définies ; elles peuvent tenir compte des délais de prescription légale, des obligations issues de la LCEN et de ses décrets d'application, ainsi que des délais d'intervention des autorités.

Un contrôle de conformité est nécessaire afin de garantir un équilibre entre la sécurité des utilisateurs et protéger leurs données personnelles. Il est essentiel de ne communiquer que les données des comptes en accord avec les conditions de service et les lois applicables sur la protection des données, comme le RGPD. Une demande dans le cadre d'un traité d'assistance judiciaire mutuelle ou une lettre rogatoire peut être nécessaire avant toute communication des contenus d'un compte.

Les plateformes recherchent et communiquent les données spécifiées dans la demande officielle, dans la limite d'efforts raisonnables de recherche et de récupération. Elles ne peuvent conserver les données, à des fins policières, si elles n'ont pas reçu une demande de préservation valable avant la suppression de ces données par l'utilisateur.

Certaines plateformes rendent compte du nombre de demandes reçues et traitées par pays[9)].

Beaucoup de réseaux sociaux ont aujourd'hui développé des canaux de signalement dédiés aux autorités afin que celles-ci puissent formuler des demandes d'accès à des données dans le cadre de procédures judiciaires.

Voici quelques recommandations et bonnes pratiques :

- Publier des instructions claires à l'attention des autorités et des forces de police, en mettant notamment à leur disposition un canal de communication efficace.

- Les plateformes peuvent être amenées à partager des informations avec des autorités en réponse aux demandes légales telles que les mandats de perquisition, les ordonnances du tribunal, les ordonnances de communication ou les citations à comparaître. Ces demandes émanent de tiers, notamment de parties à un procès civil, de la police et d'autres autorités gouvernementales.

- Dans les cas de risque imminent pour un enfant ou de risque de mort ou d'atteinte sérieuse à l'intégrité physique d'une personne, et lorsque le cas en question oblige à la divulgation d'informations sans délai, un représentant des autorités de police peut soumettre une demande par le biais du système de demande en ligne destiné aux forces de l'ordre. Cependant, il convient de toujours vérifier l'authenticité et la légitimité des demandes et de s'abstenir de répondre aux demandes envoyées par des personnes autres que celles représentant les autorités compétentes dans les juridictions d'où proviennent ces demandes.

9) Voir par exemple, pour Meta, : https://transparency.fb.com/data/government-data-requests/

- Mettre à la disposition des utilisateurs des numéros d'urgence locaux et les inciter à signaler aux autorités toute situation d'urgence dont ils pourraient être témoins

- Ne répondre qu'aux réquisitions judiciaires ou administratives, pas aux banques ou à d'autres interlocuteurs car les plateformes sont tenues au secret professionnel et au respect de la confidentialité des données personnelles de leurs utilisateurs.

- Toujours se limiter au périmètre et aux données expressément mentionnées dans la requête.

Créer un système d'amélioration continue et de qualité, via notamment un système de *reporting* interne, permettant de garder une trace des demandes reçues et traitées, et des délais moyens de réponse.

PARTIE III

Transparence et sensibilisation

Partie III - Transparence et sensibilisation

Évoquer la protection des mineurs sur les réseaux sociaux implique nécessairement de mentionner la conception de l'enfant qui prévaut depuis l'adoption, en 1989, de la Convention internationale des droits de l'Enfant (CIDE). Celle-ci indique notamment :

- L'enfant n'est pas la propriété de ses parents, ni le bénéficiaire passif d'un régime d'aides et de protections juridiques qui s'imposeraient à lui comme aux adultes.
- Comme tout individu, il est doté de droits (et aussi de responsabilités), qui sont juridiquement effectifs.
- Certains de ces droits peuvent lui être spécifiques, pour tenir compte de ses besoins propres.
- Il peut les exercer directement, dès que son âge et/ou son degré de maturité le permet.

L'enfant n'est plus cantonné dans un rôle où il n'aurait qu'à recevoir passivement des aides pour être détenteur de droits. Le RGPD s'inscrit pleinement dans cette conception lorsqu'il définit les droits numériques des enfants.

Ce principe d'autonomisation progressive des mineurs justifie à lui seul la mise en place par chaque plateforme d'une politique de transparence et de sensibilisation tout particulièrement destinée aux différentes catégories de mineurs réellement présents sur ses réseaux sociaux.

Cela ne signifie pas que les plateformes ne devront pas également chercher à s'adresser aux parents/responsables légaux de ces enfants, afin de faciliter leur intervention en cas de besoin. L'association des adultes référents reste indispensable pour prémunir les enfants des risques inhérents à l'utilisation d'Internet.

Cet équilibre nécessite la création d'un environnement de confiance (1), tant pour les moins de dix-huit ans que pour leurs parents et adultes référents, et la mise en place d'actions et de relais de sensibilisation (2).

1 La création d'un environnement de confiance

L'environnement proposé sur une plateforme doit tenir compte des obligations légales à respecter (les règles de la protection des données personnelles, de la lutte contre la pédocriminalité, le cyberharcèlement, etc.). La confiance que le réseau social doit susciter passe par la transparence (1.1) et les moyens offerts aux utilisateurs (1.2), notamment pour exercer leurs droits.

1.1 Assurer la transparence et l'information des utilisateurs par des outils et des paramètres communs

1.1.1 Règles de la communauté et de fonctionnement de la plateforme

Afin de pouvoir agir sur les comportements inacceptables sur une plateforme, celle-ci doit mettre en place un règlement, que nous appelons les « **règles de la communauté** », visant à expliquer les règles de conduites à adopter ainsi que les comportements interdits, afin de promouvoir les attitudes vertueuses et conformes aux lois en vigueur.

Ces règles doivent être explicites et adaptées à l'âge de l'utilisateur. C'est pourquoi les informations les plus importantes doivent être mises en exergue auprès l'utilisateur.

Il est particulièrement opportun de communiquer sur ces règles et le fonctionnement de la plateforme dès l'inscription de l'utilisateur, afin de s'assurer qu'il comprenne les enjeux pour son intégrité et son bien-être. La plateforme met alors en valeur auprès de l'utilisateur, et tout particulièrement des mineurs, ces informations, qui doivent être facilement accessibles et compréhensibles.

Il convient de faire accepter les règles de la communauté par les utilisateurs, par exemple via une case à cocher. En tout état de cause, ils doivent pouvoir les consulter avant de s'inscrire sur la plateforme.

Il est important de distinguer :

- L'information poussée à l'utilisateur dans des situations particulières (exemple : inscription, première ouverture d'une fonctionnalité, utilisation d'un dispositif de signalement). Cette information est généralement contextualisée et présentée d'une façon très synthétique et visuelle.
- L'information toujours accessible à l'utilisateur. Il s'agit de permettre aux utilisateurs de pouvoir consulter l'ensemble des documents juridiques depuis la plateforme (par exemple dans les paramètres). Une rubrique dédiée devrait permettre de pouvoir accéder à la politique de données personnelles, aux règles de la communauté, et aux conditions d'utilisation.

Lorsque l'audience de la plateforme comprend des mineurs, il convient de leur éviter la charge de lire des documents juridiques longs et techniques. Il est préférable de doubler les documents officiels de « résumés », présentés dans des formats ludiques : langage adapté, vidéo, infographies, schémas, dessins, bandes dessinées, etc.

Bien entendu, les règles de la communauté ne sont pas les mêmes pour toutes les plateformes : elles doivent être adaptées en fonction de l'âge du public ainsi que des mœurs locales et des lois applicables dans les pays ciblés. Elles doivent, en tout état de cause, être claires et intelligibles. Il s'agit d'un document opérationnel, qui énonce des règles concrètes ; il convient d'éviter d'employer des termes abstraits ou trop juridiques et de privilégier la simplicité rédactionnelle, la fourniture d'exemples et de cas d'usage.

Pour un meilleur accès à ces informations, multiplier les canaux de communication constitue une bonne pratique : acceptation lors de l'inscription, rappel des règles dans des mentions d'informations dédiées, disponibilité dans les paramètres de la plateforme, etc.

Toutefois, même en faisant usage d'un langage clair, la bonne compréhension des règles applicables sur un réseau social peut s'avérer difficile pour un jeune. C'est pourquoi il serait bénéfique d'inciter les jeunes à demander l'appui d'un parent ou d'un adulte de confiance, pour bien comprendre ce qui est attendu de lui sur la plateforme, mais aussi pour l'aider à faire des choix en ligne (par exemple pour le paramétrage des options de confidentialité disponibles).

En outre, les plateformes devraient prévoir des ressources et une information dédiée et adaptée aux parents, éducateurs, et/ou responsables légaux, par exemple via la création de pages dédiées aux outils et ressources mises à la disposition des utilisateurs. Ces pages, pouvant être regroupées dans une rubrique intitulée « *Centre de sécurité* », sont des initiatives qui peuvent aider à promouvoir un environnement en ligne plus sûr et transparent pour leurs utilisateurs. Ces centres sont conçus pour favoriser un usage sain du réseau social, offrir un large éventail de ressources, d'outils et d'assistance afin de prévenir les abus (haine en ligne, LGBTphobie, sexisme..., etc.), et les dangers propres aux plateformes, ou bien comment gérer ces abus lorsqu'ils sont commis.

Ces ressources peuvent proposer différents formats de contenus : guides pratiques, conseils, articles de blog, fiches thématiques… Elles devraient fournir des lignes directrices claires concernant l'utilisation responsable de la plateforme : par exemple, sur la consommation de drogues, en expliquant que la promotion de drogues illicites ainsi que la consommation de drogues illicites sont proscrites sur la plateforme, mais que partager son expérience ou mentionner des drogues à des fins de prévention sont autorisés et peut même être bénéfique pour la communauté. Afin de rendre l'ensemble de ces informations importantes accessibles, elles doivent être traitées en fonction du sujet (LGBTQ+, femmes, parentalité numérique, haine en ligne, prévention du cyber harcèlement, troubles alimentaires, bien-être animal, etc.), mais aussi du type d'utilisateurs à qui elles sont destinées (femmes, hommes, parents, mineurs...).

Le centre de sécurité constitue un espace privilégié pour sensibiliser les utilisateurs aux risques potentiels liés à l'activité en ligne et leur transmettre des conseils pour maintenir une communauté respectueuse et tolérante. Ces pages ont encore plus d'impact quand elles sont réalisées avec des experts en sécurité en ligne, mais aussi avec des associations spécialisées.

Afin de rester pertinents, ces centres doivent non seulement être régulièrement mis à jour mais surtout rendus accessibles et visibles à tous les utilisateurs, de manière intuitive, par exemple en apparaissant dès les pages de connexion ou en étant mis en avant dans le cadre de campagnes de prévention.

Enfin, ces espaces incitent les jeunes et les parents à se rapprocher d'autres ressources plus spécialisées, comme celles du gouvernement[10)] et d'associations.

1.1.2 Protection des données personnelles

Le RGPD a créé une obligation générale de transparence qui constitue une exigence juridique et opérationnelle forte.

Il en résulte que l'obligation d'information sur :

- les caractéristiques des traitements mis en œuvre : identité du responsable de traitement, finalités et base légale des traitements envisagés, destinataires des données traitées, transferts de données vers l'étranger, règles de conservation des données, systèmes de prise de décision de décision automatisée, coordonnées du délégué à la protection des données, droit de porter plainte auprès de la CNIL ;

- les modalités d'exercice des différents droits individuels - pour faciliter leur exercice - lesquelles doivent être suffisamment claires, intelligibles et précises.

La même obligation générale de transparence s'applique aux réponses du responsable de traitement en cas d'exercice individuel d'un des droits ouverts aux personnes au titre du traitement de leurs données, comme le droit d'accès à leurs données personnelles.

10) Par exemple : www.jeprotegemonenfant.gouv.fr

Le niveau d'exigence attendu des professionnels implique la fourniture, gratuite, d'informations claires, précises, simples, facilement accessibles, compréhensibles, ce qui peut inclure la présence d'icônes faciles à comprendre, notamment, par les jeunes utilisateurs.

Les obligations qui se rapportent à l'information des personnes constituent la clé de voûte de la protection des données : elles sont la condition *sine qua non* du recueil de leur consentement - lorsque le consentement est nécessaire - et de l'exercice de leurs droits individuels.

Le RGPD tient compte de la particularité de la situation des mineurs, très présents sur Internet, notamment sur les réseaux sociaux, qui justifie une attention particulière et une protection spécifique. Ils risquent, en effet, de ne pas être pleinement conscients des risques inhérents au traitement de leurs données, des conséquences et de leurs droits. Cette protection s'applique notamment aux traitements effectués à des fins de marketing, à la création de profils de personnalité ou d'utilisateur et à la collecte de données lors de l'utilisation de services proposées directement à un mineur.

Pour toute collecte de données concernant un mineur de moins de quinze ans (en France) qui est liée à la fourniture, à distance et sur demande individuelle, d'un service par voie électronique, le consentement d'un titulaire de l'autorité parentale est nécessaire en plus de celui de l'enfant concerné. Cette protection s'applique notamment aux traitements effectués à des fins de marketing et à la création de profils de personnalité ou d'utilisateur.

Le responsable du traitement doit prendre toute mesure appropriée pour fournir des informations destinées aux enfants qui soient dans des termes clairs et simples, concises, adaptées à leur niveau de compréhension et facilement accessibles.

Lorsque des données ont été collectées sur la base du consentement de la personne concernée alors qu'elle était mineure dans le cadre de l'offre de services de la société de l'information, celle-ci peut exercer son droit à l'effacement sans avoir à en donner les raisons. Le responsable du traitement est tenu d'effacer les données dans les meilleurs délais. Cela vaut notamment pour les images.

Seules ne seront pas effacées les informations nécessaires pour l'un des motifs suivants :

- L'exercice du droit à la liberté d'expression et d'information.
- Le respect d'une obligation légale.
- L'exécution d'une mission d'intérêt public.
- La constitution d'archives dans l'intérêt du public.
- La recherche scientifique ou historique.
- L'exercice d'un droit en justice.

Il convient néanmoins de souligner que certaines de ces exceptions n'impliquent pas que les données continuent à être communiquées au public si la personne concernée s'y oppose.

Pour l'utilisateur, la transparence va au-delà de la mise à disposition d'une simple politique de confidentialité. Il convient à ce titre :

- Qu'un document définissant la politique de confidentialité du réseau social soit facilement accessible et proposé sous une forme compréhensible pour un mineur.

- Que des mentions d'informations courtes soient montrées à des moments clé du parcours utilisateur (inscription au service, premier accès à une fonctionnalité, décision de modération prise à un instant donné, changement de paramétrage), notamment pour expliquer la nature des traitements susceptibles d'être effectués pour détecter des contenus illicites à des fins de signalement ou de modération.

- De proposer des paramètres facilement identifiables et accessibles pour modifier les choix en matière de suivi publicitaire, de cookies, et de partage de données. Ces paramètres doivent s'accompagner d'intitulés et d'explications claires pour que les utilisateurs comprennent les conséquences de leurs choix et de leur impact sur leur vie privée.

- D'offrir un paramétrage simple et intuitif concernant la géolocalisation. Il faut notamment que la fonction de géolocalisation ne soit pas cochée par défaut et, s'agissant de mineurs de moins de 15 ans, dans le système de contrôle parental proposé par la plateforme.

1.2 Interactions spécifiques/individuelles (dans les interactions avec les utilisateurs) : alerte, signalement, exercice des droits

1.2.1 Exercice des droits RGPD

Description générale des droits issus du RGPD

Le RGPD vise à renforcer la protection de la vie privée et des données personnelles des citoyens de l'Union européenne. Il accorde plusieurs droits fondamentaux aux personnes dont les données personnelles font l'objet d'un traitement, dont voici les principaux :

Droit à l'information : Il découle du principe de transparence, qui est essentiel pour que les personnes puissent exercer un contrôle sur leurs données. Les individus ont le droit d'être informés de manière claire et transparente sur la collecte et l'utilisation de leurs données personnelles.

Concrètement, pour un réseau social, ce droit se matérialise le plus souvent dans la Politique de confidentialité qui est mise à la disposition des utilisateurs. Néanmoins, la politique de confidentialité est souvent matérialisée dans un long document, dont les termes technico-juridiques pourraient rebuter l'utilisateur d'attention moyenne, et en particulier les plus jeunes. Les bonnes pratiques consistent ainsi à multiplier les supports d'information tout au long du parcours utilisateur : lors de son inscription, lors d'étapes clé durant lesquelles il est amené à partager des informations personnelles, et lors de la fermeture de son compte. Il est ainsi fortement recommandé de multiplier les supports et donner des indications précises en fonction du contexte et de l'activité de l'utilisateur, de sorte à lui offrir un environnement transparent.

À l'inverse, dans un environnement en ligne de plus en plus complexe, se contenter d'une politique de confidentialité acceptée au moment de l'inscription risque d'être considéré comme insuffisant.

À l'égard des mineurs, il est en outre recommandé de rédiger un résumé de la politique de confidentialité, en le composant par exemple d'éléments graphiques ou visuels (par exemple, une vidéo) qui faciliterait leur compréhension par un mineur. Aussi, des contenus spécifiques mis à la disposition des parents devraient être prévus, pour les sensibiliser à la protection de la vie privée de leur enfant et aux spécificités légales propres aux mineurs.

Droit d'accès *:* Les individus ont le droit d'accéder aux données personnelles les concernant détenues par un organisme. Ils peuvent demander une copie de leurs données et obtenir des informations sur la manière dont ces données sont traitées.

Une bonne pratique pour les réseaux sociaux consiste à permettre aux utilisateurs d'exercer ce droit de manière simplifiée, grâce à une procédure automatisée. Il peut s'agir d'une rubrique disponible dans les paramètres de la plateforme permettant de demander, en quelques clics, de recevoir un fichier comportant l'ensemble des données personnelles que la plateforme détient sur l'utilisateur concerné. Ce fichier devra donc comporter des informations comme les données du profil (coordonnées, photo de profil, etc.), les contenus publiés, les données techniques de connexion (adresse IP, données de localisation, données relatives au terminal de l'utilisateur, etc.).

Le processus d'accès aux données personnelles doit, bien entendu, être sécurisé de bout en bout. Voici quelques exemples de mesures à prévoir pour garantir la confidentialité du fichier qui sera envoyé :

- Mécanisme de double authentification pour s'assurer que le demandeur soit bien l'utilisateur en question.
- Protection du fichier par un mot de passe.
- Limitation de la durée de validité du fichier.
- Vérification de l'identité du demandeur si son identité n'est pas connue (par exemple, lorsque le demandeur n'est pas authentifié en ligne).

Droit de rectification *:* Si les données personnelles sont inexactes ou incomplètes, les individus ont le droit de demander leur rectification ou leur mise à jour.

En pratique, pour une plateforme permettant à des individus de s'exprimer, il est important de permettre aux utilisateurs de rectifier eux-mêmes la plupart des informations qui les concernent :

- d'une part, s'agissant de leurs données d'identification : ils doivent pouvoir facilement mettre à jour eux-mêmes leurs coordonnées, comme le nom, adresse email, numéro de téléphone, etc. ;
- d'autre part, s'agissant des contenus qu'ils ont publiés et qui restent visibles en ligne : il est important de donner aux utilisateurs un contrôle sur leurs contenus. Cela relève tant de la protection de leur vie privée que de leur liberté d'expression et de communication. Ainsi, ils doivent pouvoir retirer ou modifier les contenus qu'ils ont publiés et avec lesquels ils ne sont plus en accord, ou qu'ils ne souhaitent plus voir apparaître en ligne, etc. Les plateformes conservent toutefois la possibilité de conserver en interne les contenus illicites ou qui seraient en violation de leurs politiques internes, à des fins de modération et/ou de signalement aux autorités compétentes.

La mise à jour de certaines informations peut parfois constituer un risque de sécurité, tels que le changement de date de naissance ou de genre. En pratique, ces changements peuvent être à l'origine de comptes malveillants (fraude, arnaque, pédopiégeage). Il appartient à la plateforme de s'assurer de la légitimité de ce type de requête et de prendre les mesures qui s'imposent en cas d'activité suspecte. La plateforme peut, par exemple, limiter les possibilités de mettre à jour certaines informations et, si ces changements interviennent à plusieurs reprises depuis un même compte, demander un justificatif ou une pièce d'identité. Par exemple, la plateforme peut considérer qu'il existe un risque objectif et important lorsqu'un utilisateur enregistré comme adulte change sa date de naissance pour être identifié comme un mineur.

Droit à l'effacement (ou « droit à l'oubli ») : Ce droit permet aux individus de demander la suppression de leurs données personnelles dans certaines circonstances, comme lorsque les données ne sont plus nécessaires aux finalités pour lesquelles elles ont été collectées.

En pratique, les réseaux sociaux devraient mettre à la disposition de leurs utilisateurs un outil automatisé et simple d'utilisation leur permettant de supprimer leur compte et toutes les données s'y trouvant. Toutes les données ne seront pas nécessairement immédiatement effacées. Certaines peuvent être conservées par la plateforme, en cas d'obligation légale ou d'intérêt légitime, pour une durée limitée et conforme à la politique interne de conservation des données personnelles. En tout état de cause, il convient que l'ensemble des contenus et informations relatives à l'utilisateur ayant demandé la suppression de ses données ne soient plus accessibles et visibles par les autres utilisateurs.

Les plateformes devraient aussi permettre d'organiser le droit à l'effacement d'informations qui n'auraient jamais dû être collectées ou partagées (par exemple : une information piratée, une image obtenue sous la contrainte), ou qui sont gênantes pour eux (par exemple : une photo intime). Ce droit devrait être renforcé pour les mineurs, ce qui signifie que les personnes en charge de répondre à ces demandes doivent être spécialement formées pour communiquer avec des jeunes et les accompagner dans cette démarche. Les délais de réponse devraient être le plus court possible, pour ne pas laisser les utilisateurs démunis face à des situations inconfortables pour eux, et limiter autant que possible un potentiel préjudice moral.

Les moteurs de recherche devraient informer sur le droit au déréférencement, grâce auquel les utilisateurs peuvent demander de désindexer certains résultats de recherche associés à ses nom et prénom.

Conformément aux lignes directrices de la CNIL sur la gestion des relations commerciales, les réseaux sociaux doivent aussi mettre en place une politique de gestion des comptes inactifs. L'autorité française recommande ainsi que les comptes inactifs depuis plus de deux ans devraient être supprimés. Fixer une durée maximale de validité d'un compte sur un réseau social permet d'éviter de conserver les données personnelles des utilisateurs de façon illimitée, ce qui risque d'être non conforme au RGPD.

Droit de ne pas faire l'objet d'une décision automatisée : Les personnes ont le droit de ne pas faire l'objet d'une décision basée exclusivement sur un traitement automatisé, y compris le profilage, si cette décision peut produire des effets juridiques significatifs à leur égard.

En pratique, les utilisateurs d'un réseau social ont le droit de ne pas faire l'objet d'une décision qui reposerait uniquement sur un algorithme, que ce soit dans le domaine de la modération, de la publicité ciblée ou d'un système de recommandations personnalisées basé sur le profilage.

Ainsi, il convient d'évaluer l'impact sur les utilisateurs des services et outils automatisés mis en place, pour déterminer si les utilisateurs sont susceptibles de subir un « effet significatif ». Si tel est le cas, des garde-fous doivent être alors prévus pour leur permettre de voir leur situation examinée à nouveau, cette fois par un humain.

Ce principe trouve particulièrement à s'appliquer dans le cas du bannissement d'un utilisateur en cas de violation des règles de la plateforme. Un utilisateur banni en raison d'un comportement détecté grâce à un algorithme doit pouvoir demander à faire réexaminer la situation par un humain (un modérateur).

Le DSA est venu renforcer ces obligations prévues par le RGPD, en indiquant que les utilisateurs doivent pouvoir contester facilement les décisions prises à l'égard des contenus. Le système mis en place doit garantir un réexamen par un être humain lorsque des moyens automatisés sont utilisés.

Droit d'opposition : Les personnes concernées ont le droit de s'opposer à l'utilisation de leurs données personnelles à des fins de prospection commerciale, y compris au profilage qui serait lié à cette prospection. Dans les autres cas, le droit d'opposition peut s'exercer en justifiant d'une raison particulière. Il est précisé que, dans un environnement numérisé, le droit d'opposition doit pouvoir s'effectuer de manière automatisée. Si le droit d'opposition s'exerce dans certaines limites prévues par le RGPD, il convient de ne pas entraver ce droit pour les mineurs, et de pouvoir examiner leur demande en tenant compte de leur meilleur intérêt.

Droit à la portabilité : Ce droit permet aux utilisateurs de demander à un responsable de traitement de leur fournir une copie de leurs données personnelles dans un format structuré, couramment utilisé et lisible par machine. Dans l'univers des réseaux sociaux, ce droit est généralement relié au droit d'accès aux données personnelles, étudié plus haut. Ainsi, la meilleure pratique consiste à permettre aux utilisateurs de télécharger un dossier comprenant l'ensemble de leurs informations et contenus postés, facilement et en quelques clics dans un environnement « logué ». Le format du fichier téléchargé doit être *« couramment utilisé et lisible par machine »*. Il est donc recommandé de proposer aux utilisateurs au moins un format grand public, tels que PDF ou HTML.

Pour faciliter l'exercice de ces droits, il convient d'indiquer clairement comment contacter l'équipe de protection de la vie privée de la plateforme. Ainsi, les utilisateurs pourraient avoir plusieurs moyens d'exercer leurs droits : (i) au travers de fonctionnalités techniques disponibles sur le site ou l'application, et (ii) en contactant une équipe support chargée de la protection de la vie privée.

Dans tous les cas, les utilisateurs mineurs devraient être accompagnés de façon adéquate pour qu'ils puissent exercer leurs droits, ce qui signifie que la plateforme devrait, par exemple, mesurer la satisfaction et prendre en compte les retours de leurs utilisateurs concernant ces démarches, ou encore encourager le jeune à se faire assister de l'un de ses parents ou d'un adulte de confiance lorsque c'est pertinent. Améliorer l'efficacité des procédures et faciliter leur utilisation par les mineurs est important pour que l'exercice de ces droits ne soit pas seulement théorique.

L'exercice opérationnel des droits

S'il est important d'accorder aux utilisateurs une certaine autonomie pour exercer eux-mêmes les droits qu'ils tiennent de la réglementation de façon rapide et effective, les plateformes doivent également disposer de ressources humaines calibrées de façon appropriée et formées de façon adéquates pour répondre aux éventuelles demandes, réclamations ou difficultés qu'éprouvent les utilisateurs pour faire valoir leurs droits.

Ainsi, la plateforme doit mettre à la disposition de chacun une adresse de contact ou un espace permettant à tout utilisateur de contacter un personnel qualifié pour répondre à sa demande en lien avec ses données personnelles. Le canal de communication choisi ne doit pas être excessivement dissimulé sur le site de sorte à décourager les utilisateurs de l'employer.

Les plateformes devraient recruter des spécialistes formés aux questions de protection des données, ces équipes pouvant être sensibilisées, encadrées et supervisées par le service du délégué à la protection des données (DPO) de l'entreprise. Les procédures de demandes d'exercice des droits, la prise en charge des demandes, et les types de réponses apportées aux utilisateurs, devraient être approuvées par le service du DPO, lequel est aussi responsable de la formation des spécialistes en charge de répondre aux demandes d'exercice des droits. Une attention toute particulière doit être portée à ce type de prises de contact avec le support, compte tenu du droit fondamental à la vie privée et aux enjeux réglementaires applicables.

La formulation des réponses apportées aux utilisateurs devrait tenir compte de l'âge de l'utilisateur concerné. Lorsqu'un échange est engagé avec un mineur, il convient de lui écrire dans un langage clair, simple et direct, et de lui apporter toutes les précisions nécessaires pour qu'il comprenne ses droits. Au besoin, il est recommandé de lui indiquer qu'il devrait consulter ses parents ou un adulte de confiance pour saisir pleinement l'étendue de ses droits. Si besoin, le spécialiste devrait le renvoyer vers des ressources adaptées.

Compte-tenu de la sensibilité et du volume de données personnelles qu'elles peuvent traiter, les plateformes doivent protéger la confidentialité des données de leurs utilisateurs. Elles doivent notamment, en cas de doute raisonnable, vérifier l'identité d'un demandeur qui souhaite accéder à ses données personnelles. Cette formalité n'est pas nécessaire dans les cas où l'utilisateur est déjà authentifié (c'est-à-dire déjà connecté à la plateforme). En tout état de cause, les plateformes doivent vérifier l'identité d'une personne souhaitant avoir accès à des informations personnelles d'un utilisateur, et en toute hypothèse s'assurer de ne pas envoyer de données personnelles à un destinataire inconnu ou à une personne sur laquelle il existe un doute raisonnable quant à son identité.

Les réponses apportées doivent respecter les délais maximaux prévus par la réglementation, à savoir un mois pour une demande ne nécessitant pas de difficulté particulière ou deux mois pour les demandes complexes ou nombreuses (article 12 du RGPD). En cas de difficulté particulière, l'équipe dédiée au support devrait pouvoir consulter le service du DPO de la plateforme.

Par ailleurs, les plateformes peuvent être amenées à recevoir des demandes d'exercice de droits de la part de parents à l'égard de leurs enfants mineurs. Les plateformes devraient anticiper ce type de demandes et créer une procédure spécifique, afin de pouvoir répondre correctement et dans les délais requis. Dans sa recommandation 2 *« encourager les mineurs à exercer leurs droits »*, en dépit du flou juridique existant, la CNIL milite pour que les mineurs puissent exercer eux-mêmes leurs droits directement, et ce en particulier sur les réseaux sociaux où ils partagent d'eux-mêmes un grand nombre d'informations. Selon l'autorité, il est nécessaire de donner aux mineurs les moyens de leur autonomie dans leur vie numérique, et notamment pour se défendre face aux risques de cyberharcèlement.

Lors de l'évaluation d'une demande faite par un parent, plusieurs points de vigilance doivent être considérés :

- La demande du parent va-t-elle dans le sens du meilleur intérêt de l'enfant ? Ou au contraire, va-t-elle à l'encontre de cet intérêt, au regard de la situation d'espèce ? Pour ce faire, la plateforme pourrait rechercher si d'autres droits fondamentaux que le respect de la vie privée devraient être évalués, par exemple la sécurité et l'intégrité physique du mineur, le respect de sa liberté d'expression, son droit au développement, son droit à la protection contre toutes les formes de violence et d'exploitation, etc.

- Le mineur a-t-il l'âge requis pour consentir au traitement de ses données personnelles sur un réseau social, à savoir 15 ans en France ? S'il a plus de 15 ans, il est généralement apte à exercer ses droits sans l'intervention de ses parents, sauf en cas de circonstances particulières.

En tout état de cause, l'exercice d'un droit RGPD par un parent ne devrait pas se faire à l'insu de l'enfant mineur, ni sans son accord. En outre, il convient de veiller à ce que l'intervention des parents ne conduise pas à créer une surveillance généralisée ou disproportionnée de l'activité en ligne du mineur. Les plateformes devraient, de manière générale et sauf circonstances exceptionnelles[11)], éviter de permettre aux parents d'accéder aux informations intimes d'un mineur, en particulier le contenu de leurs conversations privées, ou encore leur localisation précise. La nécessité de protéger la sécurité d'un enfant ne doit pas être interprété comme un blanc-seing donné pour sacrifier son intimité et sa vie privée.

1.2.2 Continuité de la construction de l'environnement de confiance dans les interactions individuelles

Pour mieux protéger les mineurs, il est essentiel que les plateformes soient transparentes sur les paramètres de confidentialité et de sécurité qu'elles offrent. Ces paramètres doivent être mis en avant d'une manière appropriée pour permettre aux utilisateurs de contrôler la façon dont leurs données sont utilisées, mais aussi pour contrôler leurs interactions avec les autres utilisateurs et les aider à se protéger.

La plateforme doit donc communiquer autour de ces paramètres, afin que leurs communautés sachent qu'ils existent, et qu'ils peuvent les utiliser facilement. Les systèmes complexes visant à décourager l'usage de paramètre de confidentialité ou de sécurité doivent être proscrits, dès lors qu'ils auraient pour effet de limiter ou d'entraver les droits des utilisateurs.

La conception des interfaces et des fonctionnalités doit rendre particulièrement visible ces paramètres ; ces derniers pourraient même être promus au sein du système de recommandation de la plateforme (par exemple, dans un fil d'actualité), en ciblant tout particulièrement les mineurs pour les encourager à les utiliser.

Pour s'assurer que les paramètres sont opérationnels et utiles pour les utilisateurs, il est recommandé que les plateformes tiennent régulièrement des statistiques de leur fréquentation, ainsi que sur leurs effets pratiques.

Cas pratique

La consultation d'un panel de jeunes composé d'adolescents âgés de 13 à 17 ans a permis d'identifier leurs pratiques de création et paramétrages de comptes et collecter leurs souhaits d'optimisation.

Connaissance des fonctionnalités de paramétrages et des données collectées

On constate que les jeunes consultés ne sont pas accompagnés par des adultes dans leur phase d'inscription et ont peu de connaissances des paramétrages existants sur leurs réseaux sociaux et de leur utilité pour leur protection. Ils reçoivent parfois de mauvaises recommandations de leurs parents (comme s'inscrire en tant que majeur) pensant les protéger d'avantage des prédateurs sexuels notamment. Ils connaissent mal le droit encadrant leurs pratiques des réseaux sociaux (droits à l'image, partage de données, liberté d'expression...).

11) Ces « circonstances exceptionnelles » doivent être déterminées au cas par cas, mais pourraient notamment prendre en compte les situations d'urgence comportant des risques importants pour le mineur.

Les problèmes identifiés

- Les CGU ne sont pas compréhensibles par des mineurs et conçus pour donner envie de les lire.
- Les jeunes se sentent écoutés par les plateformes et manipulés par des algorithmes sur lesquels ils n'ont pas la main.
- Les outils de signalement sont dissuasifs et ne considèrent pas toujours l'expérience vécue comme légitime.
- Trop d'arnaques et de fausses informations sont accessibles dans leur fil.

Les souhaits formulés

- Être informés dès l'inscription des données collectées et de leur utilisation.
- Intégrer ses informations à l'expérience utilisateur avec des messages adaptés à leur âge et leur attention (des messages clairs, courts et des pictogrammes).
- Être guidés dès l'inscription vers les fonctionnalités de paramétrage avec une explication claire de leur rôle dans la protection des mineurs.
- Comprendre et maîtriser les algorithmes des plateformes.
- Être informés du droit encadrant leurs pratiques au cours de leur parcours utilisateur (quand ils publient, partagent un contenu ou rédigent un commentaire...).
- Être mieux informés sur le droit à l'oubli et les procédures à leur disposition.
- Avoir des outils et des bonnes pratiques pour repérer les fausses informations.
- Simplifier les procédures de signalement, apporter une réponse quelle que soit la situation et donner des informations de suivi du signalement.

Afin de répondre à ces attentes, plusieurs actions peuvent être mises en place par les plateformes :

- **Inciter les utilisateurs à adopter un comportement responsable et approprié en ligne :** Objectif de prévenir les violations et infractions. Par exemple : envoyer des alertes pour faire cesser une conduite inappropriée avant de modérer, communiquer clairement autour des règles de conduite à suivre sur la plateforme.
- **Inciter les utilisateurs à limiter les excès d'usage et les effets potentiellement addictifs** : Mise en place de fonction de gestion du temps d'écran, alerte pour alerter en cas d'usage continu de plus de x heures etc. ou en cas de paiement excessif, désactivation des notifications.
- **Permettre aux utilisateurs de bannir les interactions nocives pour eux** : Mots cachés, bloquer des utilisateurs sans risque de représailles, bloquer des contenus réservés à des adultes.
- **Envoyer une alerte automatisée lorsque le mineur prend un risque.**
- **Risque d'enfreindre les règles de conduite.**

- **Risque de commettre une infraction en ligne.**
- **Risque de faire un choix compromettant pour le respect de sa vie privée.**
- **Faciliter le contact avec un humain lors de l'expérience utilisateur** : Poser des questions, exprimer un avis, faire une contestation auprès d'une équipe support dédiée, avec une intervention humaine.
- **Prévoir un lieu ressources comme les centres de sécurité**, pour que l'utilisateur ait un interlocuteur extérieur à la plateforme si celle-ci ne répond pas ou pas ce qui est attendu : Service 3018 (le numéro pour les jeunes victimes de harcèlement et de violences numériques), point de contact, Pharos, Stop Fisha, Respect Zone et tout autre signaleur de confiance et service tiers.
- **Informer les personnes modérées** : Expliquer la mesure dont un utilisateur fait l'objet, pourquoi la règle qui a été enfreinte, la durée de la sanction.
- **Informer les personnes faisant un signalement** : Informer des suites données à un signalement de façon proportionnée (pas nécessairement la sanction prise mais le fait que le signalement était fondé et qu'une mesure a été prise) et les informer de leur droit si besoin. Préserver la confidentialité des personnes qui font des signalements pour éviter les risques de représailles.
- **Fournir des recommandations pour accompagner les témoins et les renforcer dans leur capacité de signalement** : Informations à transmettre, signalement ou non à la personne ciblée, point spécifique sur les personnes qui font des signalements qui ne les concernent pas directement etc.

2 Actions et relais de sensibilisation

2.1 Campagnes d'éducation et/ou de sensibilisation, partenariats associatifs et pouvoirs publics (enjeux de compréhension et de prévention)

Considérant leur positionnement, le trafic et l'engouement qu'elles suscitent, les plateformes ont un rôle primordial à jouer dans la diffusion, l'accès aux informations et à la construction d'un cyberespace respectueux des lois. Toutefois, cette ubiquité numérique s'accompagne de responsabilités cruciales en matière de compréhension et de prévention des enjeux liés à la sécurité en ligne, au bien-être des utilisateurs, et à la diffusion d'informations pertinentes.

À cette fin, les plateformes travaillent à la mise en place de règles et outils clairs et accessibles à tous, de la création d'équipes dédiées et des investissements en intelligence artificielle pour garantir l'application de ces règles.

Cependant, les partenariats entre les plateformes de réseaux sociaux, les associations, et les pouvoirs publics revêtent une importance capitale. Cette collaboration permet non seulement de sensibiliser davantage le public aux risques en ligne, mais aussi de mettre en place des mesures préventives efficaces, afin d'assurer un environnement numérique sûr, inclusif, et informatif pour tous.

La plupart des plateformes ont développé des partenariats forts avec des associations en charge de sujets liés à la protection de l'enfance, notamment : Association e-Enfance/3018, Internet Sans Crainte Génération Numérique, Respect Zone, Point de Contact, Stop Fisha, UNAF, Cemea et bien d'autres. Ces partenariats comportent de nombreux aspects et se matérialisent de différentes manières :

Sur les outils et ressources des plateformes elles-mêmes

Comme vu précédemment (partie 1.1.1), les plateformes s'investissent dans le développement d'outils et de ressources de sensibilisation qui leur sont propres.

Dans leur développement, elles peuvent faire appel à des associations, experts et pouvoirs publics pour les mettre en place. Pour ce, elles organisent des ateliers internes afin de récolter des retours « terrain » et optimiser des produits et des ressources développés par les plateformes (centre de sécurité, fiches d'aide ou encore action de prévention...).

La mise en ligne de ces outils et ressources, s'accompagne de campagnes internes et régulières pour mettre en avant leurs ressources et faire de la sensibilisation sur des sujets en lien avec la protection des mineurs.

Ces campagnes permettent également de promouvoir les outils et produits spécifiques dédiés sur leurs plateformes afin de renforcer la notoriété de ces derniers auprès des publics visés notamment les jeunes et les parents.

Sur les actions de sensibilisation communes et coordonnées avec des associations et les pouvoir publics

Dans cet objectif d'environnement de confiance, les plateformes apportent leur soutien aux campagnes menées par des acteurs publics et les associations.

À titre d'exemple, elles œuvrent notamment au soutien de grandes journées :

- La journée nationale de lutte contre le harcèlement scolaire, portée par le Ministère de l'Éducation Nationale, et les acteurs du secteur.

- Le Safer Internet Day, journée mondiale pour un Internet plus sûr opérée en France par Internet Sans Crainte, le programme national de sensibilisation du Safer Internet France, qui associe différents soutiens associatifs.

Ce soutien se concrétise généralement par un relai important sur les plateformes, un soutien financier ou un crédit publicitaire.

Elles peuvent aussi soutenir des projets de développement d'outils de sensibilisation ou de campagnes de sensibilisation portés par des associations.

Dans l'idée de rendre plus efficace et juste le système de modération et d'aider à faire retirer des contenus manifestement illicite, les plateformes ont mis en place un programme de ***« Trusted Flaggers »*** (signaleur de confiance) qui permettent aux associations concernées d'avoir un canal direct avec les équipes de modération et ainsi remonter des contenus et tendances problématiques.

Pour cela, elles entretiennent des relations de travail et de suivi avec les associations. En effet, un dialogue régulier existe entre elle afin d'entretenir la connaissance et la confiance réciproque.

Ces partenariats sont primordiaux, ils permettent la prise de conscience collective de ces sujets et des actions à mener. Les défis posés par la violence et la haine en ligne chez les mineurs sont complexes, et nécessitent une approche multidimensionnelle impliquant tous les acteurs concernés.

Les associations jouent un rôle de catalyseur en fournissant des données, des témoignages et des analyses approfondies sur les conséquences de comportements à risque en ligne.

Les campagnes de sensibilisation, conférences et des événements permettent aujourd'hui d'attirer l'attention du plus grand nombre sur ces questions urgentes et mobiliser l'opinion publique.

Les plateformes sont également parties prenantes d'initiatives lancées par le gouvernement, c'est l'exemple récent du laboratoire pour la protection de l'enfance en ligne annoncé par le Président Emmanuel Macron à l'occasion du Paris Peace Forum.

Elles peuvent aussi nouer des relations partenariales avec des chercheurs dont l'expertise pourrait nourrir leurs outils. Par exemple : en sciences de l'information et de la communication, en droit, en sociologie, en psychologie, en informatique, en sciences de gestion qui travaillent sur les pratiques numériques des jeunes.

2.2 Toucher les utilisateurs et adultes référents (enjeux d'accompagnement)

Les parents et adultes référents sont le premier relais de communication avec les mineurs. Toutefois, toucher les adultes et parents peut s'avérer complexe : méconnaissance des pratiques numériques des plus jeunes, des plateformes qu'ils utilisent et des ressources d'information existantes, manque de disponibilité, difficultés à nouer un dialogue sur les sujets de prévention (cyber harcèlement, pédocriminalité, consultation de contenus pour adultes, etc.). Il s'agit pourtant d'un axe essentiel de la sensibilisation, et il est primordial que l'ensemble des utilisateurs des réseaux sociaux soient investis dans la protection des mineurs en ligne. Il est donc important d'organiser des moments de rencontre avec des professionnels de l'éducation, de l'enfance, et des usages numériques.

Les organisations suivantes ont ainsi l'habitude d'intervenir auprès des parents et adultes référents sur ces sujets, lors de réunions d'information ou ateliers :

- Associations familiales
- CNAF et UNAF
- Association e-Enfance/3018
- Internet Sans Crainte
- Respect Zone
- Initiatives telles qu'Internet Matters
- Associations d'éducation populaire comme les Cemea
- Associations de parents d'élèves

Cela peut également être l'occasion pour les acteurs publics d'intervenir auprès des usagers :

- Collectivités
- Centres sociaux, centres d'animation, fédérations
- Autorités indépendantes

Dans cette optique, les entreprises peuvent jouer un rôle crucial pour sensibiliser les adultes (parents, membres de la famille, adultes référents…) aux bonnes pratiques à mettre en place sur les réseaux sociaux afin de protéger les plus jeunes. En effet, s'il est peut-être difficile ou contraignant pour eux d'assister sur leur temps libre à des événements de sensibilisation dédiés, des actions menées dans le cadre de leur vie professionnelle, sur le temps de travail, pourraient faciliter l'éducation d'un large public et encourager les prises de conscience.

Cette sensibilisation peut par exemple prendre la forme de webinaires ou formations à destination des salariés, afin de les former sur les risques et les mesures de précaution à prendre. Ces sessions pourraient aborder des sujets tels que la protection de la vie privée en ligne, le contrôle des paramètres de confidentialité, l'importance du consentement et de l'autorisation parentale pour les mineurs, ainsi que les signes révélateurs de comportements dangereux ou inappropriés (cyber harcèlement, contact avec des adultes inconnus, accès aux contenus pornographiques, violents ou addictifs…).

Les entreprises pourront également relayer des ressources et des guides sur les bonnes pratiques à mettre en place (documents de référence, infographies pédagogiques…), accessibles via leurs plateformes en ligne ou intranets. En collaborant avec des experts de la sécurité en ligne, des acteurs publics et des organisations spécialisées dans la protection des mineurs, les entreprises pourront renforcer leur crédibilité sur ces sujets, et leur efficacité dans la sensibilisation des parents aux enjeux des réseaux sociaux. Enfin, elles pourront relayer et promouvoir des campagnes de sensibilisation, en diffusant des messages et des vidéos informatifs pour atteindre un large public et encourager le partage de bonnes pratiques au sein de la communauté. Par exemple, elles pourront diffuser les ressources mises à disposition par les pouvoirs publics, telles que le site jeprotègemonenfant.gouv, élaboré par l'État en partenariat avec les plateformes et les associations, qui ont participé à la construction du site, l'élaboration de son contenu, ainsi qu'à son amplification.

AFNOR SPEC 2305

LEXIQUE

Lexique

Compte malveillant

Désigne un compte créé sur une plateforme dans le but de nuire à autrui ou à la l'intégrité de la plateforme, ou encore dans le but de commettre une fraude. Le compte malveillant agit souvent sous un faux compte, c'est-à-dire un compte tentant de dissimuler son identité ou une information à propos de son identité (telle que l'âge) pour tromper les autres utilisateurs à des fins malveillantes.

Donnée biométrique

Désigne une information portant sur une caractéristique physique ou biologique permettant d'identifier une personne de manière unique. Il s'agit d'une catégorie particulière de données à caractère personnel (voir définition ci-dessous).

Donnée sensible ou catégorie particulière de donnée à caractère personnel

Désigne, selon l'article 9 du RGPD, une donnée à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, une donnée biométrique aux fins d'identifier une personne physique de manière unique, une donnée concernant la santé ou concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique.

DSA

Désigne le Règlement sur les services numériques, en anglais *Digital Services Act*, c'est-à-dire le 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.

La différence entre la vérification et l'estimation de l'âge

La vérification de l'âge consiste à effectuer un contrôle de l'âge exact d'un individu ou que son âge se situe au-dessus ou au-dessous d'un âge (souvent 18 ans), à partir de sa date de naissance, supposant généralement la communication d'un document officiel tel qu'une pièce d'identité. L'estimation de l'âge consiste à attribuer à un individu un âge approximatif, avec une faible marge d'erreur tolérée, généralement via un système d'intelligence artificielle capable d'analyser les traits du visage d'un individu à partir d'une photo.

Pédopiégeage (ou « grooming »)

Désigne, selon la Directive européenne 2011/93 du 13 décembre 2011, le fait pour un adulte de proposer, au moyen des technologies de l'information et de la communication, une rencontre à un mineur qui n'a pas atteint la majorité sexuelle, dans le but de commettre un abus sexuel.

Performance des algorithmes

Désigne le niveau de précision d'un système d'intelligence artificielle (tel qu'une technologie de détection de contenus), c'est-à-dire la proportion de solutions exactes ou inexactes obtenues par un système d'IA. Cette appréciation renvoie aux contrôles à mettre en place pour que le système d'IA soit suffisamment précis sur le plan statistique afin de garantir que le traitement de données personnelles respecte le principe de loyauté.

Règles de la communauté

Désigne les règles de bonne conduite à respecter sur un réseau social et les comportements qui ne sont pas tolérés. Ces règles constituent en général une extension ou une annexe des conditions générales d'utilisation ; elles sont acceptées par l'utilisateur lors de son inscription.

RGPD

Désigne le Règlement Général sur la Protection des Données, c'est-à-dire le Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE.

Safety et Privacy by Design

Désigne l'ensemble des mesures techniques, organisationnelles et humaines, mises en œuvre par une plateforme dès les premières étapes de la conception des services qu'elle propose, afin de protéger la vie privée et la sécurité de ses utilisateurs.

Signaleur de confiance

Entité pouvant effectuer des signalements sur les réseaux sociaux, et dont les notifications doivent être traitées en priorité selon le DSA. Ce statut est attribué dans chaque pays à des entités ou organisations en raison de leur expertise et de leurs compétences.

Système d'intelligence artificielle (ou IA)

Désigne, selon la CNIL, un procédé logique et automatisé reposant généralement sur un algorithme et en mesure de réaliser des tâches bien définies.

Technologie de détection de contenu

Désigne un outil automatisé, basé sur l'intelligence artificielle ou du moins des algorithmes, permettant de scanner tout ou partie d'une plateforme pour détecter des contenus ou des comportements suspects. Il s'agit d'une technologie d'aide à la modération, facilitant le travail des équipes de modération.

Vérification d'identité

Signifie la vérification de l'identité complète d'un individu à partir d'un document officiel d'identité.

ANNEXE JURIDIQUE

Cadre juridique et obligations légales applicables aux plateformes

Annexe juridique - Cadre juridique et obligations légales applicables aux plateformes

Note : Il est rappelé au lecteur que cette section, consacrée aux règles juridiques applicables, ne constitue qu'un aperçu non exhaustif des principales réglementations applicables aux réseaux sociaux. Il convient pour chaque organisation d'approfondir les obligations auxquelles elle est soumise selon son propre modèle d'affaires et le droit local applicable.

Par ailleurs, la protection des mineurs en ligne est un sujet sur lequel plusieurs gouvernements ont adopté une politique législative volontariste. Cette AFNOR SPEC a donc vocation à être mise à jour en fonction de l'avancement des diverses réglementations.

Enfin, cette AFNOR SPEC n'a pas vocation à se substituer au conseil d'un professionnel du droit dans la détermination des obligations applicables à un modèle d'affaires spécifique.

1 Droits et libertés fondamentales

Cela fait plus d'une décennie que les réseaux sociaux sont devenus des outils incontournables de communication, d'information et de partage. Cependant, leur utilisation soulève également des enjeux importants en matière de droits et libertés fondamentales, dont notamment le droit à la liberté d'expression (1.1), le droit au respect de la vie privée (1.2), et le droit à l'égalité (1.3).

Chaque plateforme devrait mener des études d'impact pour déterminer comment assurer le respect des droits et libertés fondamentales dans leurs produits et services. Ainsi, les textes internationaux et la jurisprudence en matière de droits et libertés fondamentales, et de droits de l'Homme, devraient servir d'orientation pour créer les règles de conduite à suivre sur la plateforme.

1.1 La liberté d'expression

Cette liberté est protégée par l'article 10 de la Convention européenne de sauvegarde des droits de l'Homme (CEDH), l'article 11 de la Charte des droits fondamentaux de l'Union européenne (CFUE) et l'article 11 de la Déclaration des droits de l'Homme et du Citoyen de 1789.

La jurisprudence européenne a eu l'occasion de préciser que la liberté d'expression est *« l'une des conditions de base pour le progrès des sociétés démocratiques et pour le développement de chaque individu »* (Cour EDH, 07/12/1976, Handyside c. Royaume-Uni, § 49). Elle comporte deux dimensions : la liberté *« de recevoir ou de communiquer des informations »* (article 10 de la CESDH).

La liberté d'expression vaut notamment pour les informations ou idées *« qui heurtent, choquent ou inquiètent l'État ou une fraction quelconque de la population. Ainsi le veulent le pluralisme, la tolérance et l'esprit d'ouverture sans lesquels il n'est pas de « société démocratique »* » (Cour EDH, 07/12/1976, Handyside c. Royaume-Uni, § 49).

La possibilité pour les individus de s'exprimer sur Internet constitue un outil sans précédent d'exercice de la liberté d'expression. Le numérique accroît les capacités effectives de jouir de ses libertés, *« sans considération de frontière »* (article 10 de la CESDH). Ce rôle éminent d'internet dans la liberté d'expression est reconnu par les cours suprêmes, tant au niveau européen (CJUE, 08/09/2016, GS Media BV c/ Sanoma Media Netherlands BV e.a) que national (CC, décision DC n° 2009-580 DC, 10/06/2009).

Dans une démocratie, les réseaux sociaux offrent un espace de liberté d'expression sans précédent, où les utilisateurs peuvent s'exprimer, réagir, commenter, et partager leurs opinions sur tous les sujets et sans être contraints par des frontières physiques. Les possibilités offertes par les plateformes constituent indéniablement d'immenses progrès ; cependant, elles s'accompagnent de dérives et de risques importants, parmi lesquels :

- la provocation à la commission d'actes de terrorisme et leur apologie ;
- l'apologie, la négation ou la banalisation de crimes contre l'humanité ;
- la diffusion de contenus pédopornographiques ;
- l'exposition de mineurs à des contenus pornographiques ;
- le cyberharcèlement, qui peut revêtir diverses formes (l'intimidation, la propagation de rumeurs, la moquerie, la menace, la pornodivulgation ou « revenge porn ») ;
- l'incitation à la haine, à la violence ou à la discrimination en raison de l'origine, de la religion, du sexe, de l'orientation sexuelle, de l'identité de genre ou d'un handicap ;
- l'incitation à la violence, notamment aux violences sexuelles et sexistes ;
- la diffusion d'atteintes à la réputation : les propos diffamatoires (la diffamation est le fait d'imputer publiquement un fait précis qui porte atteinte à l'honneur et à la considération d'une personne) ; les dénonciations calomnieuses (le fait d'accuser une personne d'un acte qu'elle n'a pas commis, mais pour lequel elle pourrait être sanctionnée) ; les injures ;
- la violation de la vie privée, notamment le doxing (acte de révéler au public des informations permettant d'identifier une personne en ligne, sans l'autorisation de la victime) ou du droit à l'image ;
- l'usurpation d'identité.

Au regard de ces risques, la liberté d'expression ne saurait être absolue. Elle est encadrée et limitée, dès lors que ces limites sont (i) prévues par la loi, (ii) proportionnées au but poursuivi et (iii) nécessaires dans une société démocratique, notamment pour protéger la sécurité nationale, l'ordre public, les droits et la réputation d'autrui (article 10 de la CEDH).

Elle peut notamment être tempérée par d'autres libertés comme le droit au respect de la vie privée ou la sécurité publique. Mais *« le filtre ne joue plus en amont, au stade de l'accès aux médias, mais en aval, au stade de la sélection des contenus par l'internaute lui-même. Il s'agit là d'un progrès considérable »* (Conseil d'État, Étude annuelle 2014, Le numérique et les droits fondamentaux, p. 145).

La jurisprudence française et européenne qui s'est bâtie au fil des années est venue préciser tant la portée de la liberté d'expression que ses limites[12)].

La désinformation et les fausses informations (*« fake news »*) qui circulent sur Internet constituent aujourd'hui une dérive majeure de l'étendue de la liberté d'expression permise, notamment, sur les réseaux sociaux. En réponse à ce phénomène, le droit français s'est considérablement étoffé, créant diverses infractions :

- Les fausses informations diffusées en période électorale (article L. 163-2 du Code électoral).
- Les fausses nouvelles (article 27 de la loi du 29 juillet 1881) et les fausses nouvelles diffusées à des fins électorales (article L. 97 du Code électoral).
- Les fausses alertes (article 322-14 du Code pénal).
- La diffamation (article 29 de la loi de 1881).
- Le dénigrement (article 1240 du Code civil).
- La publicité trompeuse (articles L. 121-1 et 121-1-1 du Code de la consommation).

1.2 Le droit au respect de la vie privée

Le droit au respect de la vie privée est également reconnu comme un droit fondamental et constitutionnel[13)]. Il est protégé par l'article 9 du Code civil au niveau français, et par l'article 8 de la CEDH et 7 de la CDFUE au niveau européen.

Le droit au respect de la vie privée revêt plusieurs aspects et s'applique aussi bien dans le monde réel que sur Internet. Ainsi, il recouvre notamment (sans que cette liste ne soit exhaustive) :

- la confidentialité des correspondances privées, y compris par voie électronique ;
- la protection de l'intimité, telle que les éléments concernant la vie familiale et amoureuse, ou les préférences sexuelles ;
- la protection du droit à l'image, qui permet d'autoriser - ou non - la diffusion publique de son image.

Pour sanctionner les violations du droit à la vie privée, le Code pénal français dispose d'un chapitre entier, allant des articles 226-1 à 226-32.

Enfin, des réglementations spécifiques se sont développées, notamment, pour protéger les citoyens européens des nouveaux usages résultant de l'économie numérique, et pour créer des obligations spécifiques pour les organismes traitant de plus en plus massivement aux données personnelles. Ces réglementations, dont le RGPD constitue le fer de lance, sont étudiées à la Section 4.

---

12) Pour un compte-rendu complet de l'application de l'article 10 de la CEDH par la Cour européenne des droits de l'Homme, consulter son guide sur l'article 10 : https://www.echr.coe.int/documents/d/echr/Guide_Art_10_FRA

13) Décision du Conseil constitutionnel n°99-416, 99-419 et 99-422 rendues en 1999.

1.3 Le principe d'égalité

Le principe d'égalité est un principe à valeur constitutionnelle au titre de l'article 1er de la Constitution française, qui implique l'égalité de traitement et l'égalité devant la loi.

C'est sur ce principe que repose l'interdiction de toute discrimination fondée sur la race, le sexe, l'origine ethnique, la religion, l'orientation sexuelle, l'état de santé ou toute autre caractéristique personnelle.

Le législateur français a créé des infractions spécifiques liées aux discriminations, dont en particulier les discours de haine qui recouvrent l'incitation à la haine, à la violence ou à la discrimination (article 24 de la loi du 29 juillet 1881). Il s'agit, en résumé, des propos, des écrits, des images ou de tout autre moyen d'expression qui incitent à la discrimination à la haine ou à la violence envers une personne ou un groupe de personne en raison de leur origine, leur appartenance ou leur non-appartenance à une ethnie, une nation, une race, une religion, leur sexe, leur orientation sexuelle, ou leur handicap.

Ces discours peuvent prendre différentes formes, comme les injures, les menaces, les propos racistes, antisémites, homophobes, etc. En l'absence de régulation, les réseaux sociaux peuvent involontairement servir de caisse de résonance des discours haineux.

Il est intéressant de noter que les peines encourues sont plus lourdes lorsque l'incitation à la haine est publique[14)], ce qui est le cas lorsqu'elle est proférée sur un réseau social public, c'est-à-dire au-delà d'un cercle restreint d'utilisateurs.

La viralité propre au fonctionnement de certains réseaux sociaux peut créer un risque de diffuser à plus grande échelle ce type de comportements ; c'est pourquoi les plateformes doivent se doter de dispositifs spécifiques pour prévenir ces effets délétères pour leur communauté.

2 La régulation des plateformes

2.1 Introduction sur le régime de responsabilité générale des « hébergeurs »

En droit français, les règles légales applicables aux plateformes hébergeant les réseaux sociaux sont principalement énoncées dans la loi n°2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique (« **LCEN** »), qui transpose la directive 2000/31/CE du 8 juin 2000 sur le commerce électronique.

Ces textes sont en voie d'être mis à jour avec l'entrée en vigueur du Règlement (UE) 2022/2065 du 19 octobre 2022 sur les services numériques, également connu sous le nom de ***« Digital Services Act »*** ou ***« DSA »***, le 17 février 2024 (ou le 25 août 2023 pour les très grandes plateformes).

14) Une incitation publique à la haine constitue un délit passible d'une peine d'emprisonnement d'un an et de 45 000 euros d'amende alors qu'une incitation non publique constitue une contravention sanctionnée par une amende de 1 500 euros.

L'objectif de cette réglementation française et européenne est d'établir des régimes harmonisés de responsabilité pour les acteurs de l'Internet, dont les réseaux sociaux font partie.

Ces derniers sont ainsi considérés comme des ***« hébergeurs de contenus »***, c'est-à-dire des intermédiaires techniques bénéficiant d'un régime de responsabilité allégé. Concrètement, un hébergeur de contenus est une entité qui offre un espace de stockage et de mise à disposition des contenus publiés par les destinataires de leurs services.

La réglementation en vigueur n'attribue pas aux hébergeurs la responsabilité des contenus qu'ils hébergent, ni ne les oblige à mettre en place un système de surveillance générale des contenus. En revanche, ils sont tenus de mettre en œuvre des mesures efficaces pour lutter contre les contenus ***« manifestement illicites »*** qui seraient portés à leur connaissance.

Ils ne deviennent responsables de ces contenus que s'ils n'agissent pas rapidement après avoir reçu une notification d'un tiers les informant du caractère illicite du contenu concerné.

2.2 Loi française pour la confiance dans l'économie numérique (« LCEN »)

À qui s'applique cette réglementation ?

La loi n°2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique (**« LCEN »**)[15] est la loi transposant la Directive européenne 2000/31/CE du 8 juin 2000 qui a consacré le régime de responsabilité des hébergeurs. Son contenu sera prochainement mis à jour par le DSA.

La LCEN définit les hébergeurs de contenus comme des personnes physiques ou morales *« qui assurent, même à titre gratuit, pour mise à disposition du public par des services de communication au public en ligne, le stockage de signaux, d'écrits, d'images, de sons ou de messages de toute nature fournis par des destinataires de ces services »* (article 6-I-2).

Cette large définition englobe les réseaux sociaux, quelle que soit leur taille, leurs fonctionnalités, et leurs usages.

15) La LCEN : https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000801164#:~:text=%2D1.,de%20ces%20moyens%20sans%20surco%C3%BBt

Quelles sont leurs obligations ?

La LCEN a été régulièrement amendée pour prendre en compte les enjeux sociétaux du moment, notamment :

- La loi n°2018-1202 du 22 décembre 2018[16)] relative à la lutte contre la manipulation de l'information ;
- La loi n°2020-766 du 24 juin 2020[17)] visant à lutter contre les contenus haineux sur Internet ;
- La loi n°2022-1159 du 16 août 2022[18)] portant diverses dispositions d'adaptation au droit de l'Union européenne en matière de prévention de la diffusion de contenus à caractère terroriste en ligne.

Les règles de modération des contenus sont principalement édictées aux articles 6 et suivants de la loi[19)].

En principe, les réseaux sociaux ne sont pas responsables (sur le plan civil et pénal) des images, des messages et des activités menées par leurs utilisateurs sur la plateforme, s'ils se contentent de les héberger et sans avoir effectivement connaissance de contenus illicites.

Cependant, elles doivent mettre en place un dispositif de signalement facilement accessible pour tous les contenus, permettant aux utilisateurs, le cas échéant, de les signaler. Cela se matérialise généralement par un bouton ou une icône facilement identifiable, pouvant être activé sur chaque contenu ou information postée, ainsi qu'un formulaire de signalement dans lequel l'utilisateur peut indiquer la raison pour laquelle il considère le contenu comme illicite. La plateforme a ensuite l'obligation de traiter les signalements qui lui sont remontés.

S'il est établi par la plateforme que le contenu est *« manifestement »* illicite (pour plus d'informations sur la notion de contenus manifestement illicite, voir le point 1.1 du Groupe 2 « Détection, Modération et Signalement »), il doit être retiré dans les meilleurs délais. La loi ne spécifie pas explicitement de délai, mais la jurisprudence française a généralement retenu un délai allant de 48 heures à 72 heures, y compris les jours fériés et les week-ends.

---

16) La loi n°2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information : https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037847559

17) La loi n°2020-766 du 24 juin 2020 visant à lutter contre les contenus haineux sur Internet : https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000042031970/

18) La loi n°2022-1159 du 16 août 2022 portant diverses dispositions d'adaptation au droit de l'Union européenne en matière de prévention de la diffusion de contenus à caractère terroriste en ligne : https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000046186784

19) Articles 6 et suivants de la loi : https://www.legifrance.gouv.fr/loda/id/LEGISCTA000006117685

Certains contenus sont considérés comme particulièrement graves et doivent donc être retirés encore plus rapidement. C'est le cas par exemple des contenus terroristes (Article 6-1 de la LCEN), qui doivent être retirés dans les 24 heures suivant la notification d'une autorité compétente (généralement Pharos[20)], le portail de signalement des contenus illicites sur Internet, rattaché au Ministère de l'Intérieur). Le projet de loi SREN[21)] ***« Sécuriser et réguler l'espace numérique »***, présenté par le Gouvernement, envisage de soumettre les réseaux sociaux à la même obligation de rapidité en ce qui concerne les contenus pédopornographiques.

Lorsque le contenu n'est pas « manifestement » illicite, par exemple dans le cas des demandes fondées sur le droit d'auteur ou le droit à l'image d'une personne, le réseau social se doit de mener une enquête auprès du signaleur et du signalé afin d'établir l'illicéité éventuelle du contenu. Ce débat contradictoire organisé par le réseau social doit être réalisé selon les principes de neutralité et de non-immixtion, à défaut de quoi il encourt le risque d'être requalifié en éditeur.

En effet, le contrôle ou la surveillance des contenus ou des utilisateurs, ainsi que la partialité dans les décisions de retrait de contenus peuvent être à l'origine d'une requalification d'un hébergeur en éditeur si la politique de la plateforme dépasse ce qui est strictement attendu d'un hébergeur selon les critères légaux tels que précisés par la jurisprudence. Cela a été le cas notamment pour :

- **Airbnb**, requalifiée en éditeur de contenus dans une décision de la Cour d'appel de Paris en date du 3 janvier 2023[22)], notamment parce que la plateforme imposait à ses utilisateurs de respecter certaines normes ou valeurs de la « communauté Airbnb » et exerçait un pouvoir de contrôle sur le contenu des annonces publiées sur la plateforme ;

- **Ticketbis**, plateforme de revente de billets sportifs, requalifiée en éditeur par la Cour de cassation dans un arrêt du 1er juin 2022[23)], pour avoir optimisé la présentation des offres à la vente et en les promouvant à partir de ses connaissances et de son contrôle des données stockées, elles jouaient un rôle actif ;

- la marketplace **eBay** a également été requalifiée en éditeur, par une décision de la cour d'appel de Paris du 1er juillet 2021[24)], notamment parce que les utilisateurs étaient soumis à un règlement assorti de sanctions, eBay ayant le droit de retirer les annonces non conformes au règlement et fournissant des recommandations pour améliorer la performance et la visibilité des annonces postées par les utilisateurs.

---

20) Pharos : https://www.internet-signalement.gouv.fr/PharosS1/

21) La loi SREN « Sécuriser et réguler l'espace numérique » : https://www.senat.fr/dossier-legislatif/pjl22-593.html

22) La décision de la Cour d'appel de Paris en date du 3 janvier 2023 : https://www.doctrine.fr/d/CA/Paris/2023/CAPD63336A8A6C3ABF2C0EE

23) L'arrêt du 1er juin 2022 : https://www.courdecassation.fr/en/decision/6297029565ec7ea9d4f0dcd5

24) La décision de la Cour d'appel de paris du 1er juillet 2021 : https://www.courdecassation.fr/en/decision/63be61c013ef607c90ab61e5

2.3 Le règlement « Platform to Business » (P2B)

À qui s'applique cette réglementation ?

Le Règlement européen 2019/1150 du 20 juin 2019 dit ***« Platform-to-Business »*** ou ***« P2B »*** s'applique aux ***« services d'intermédiation en ligne »***[25] et vise à réguler leurs relations avec les entreprises qui utilisent ces services. Les types de services concernés comprennent les moteurs de recherche, les magasins d'applications ou encore les *marketplaces*. Les réseaux sociaux sur lesquels des professionnels ont la possibilité d'offrir des biens et des services sont également concernés, par exemple lorsqu'ils offrent des services payants pour améliorer le référencement de contenus ou la publicité.

Quelles sont leurs obligations ?

Le texte impose en premier lieu une obligation de transparence aux réseaux sociaux. Leurs conditions générales doivent indiquer les principaux paramètres utilisés pour le classement de biens et des services et ***« les raisons justifiant l'importance relative de ces principaux paramètres par rapport aux autres paramètres »*** (article 5). Le réseau social n'a pas l'obligation de détailler précisément le fonctionnement de ses algorithmes de référencement.

Tout traitement différencié réservé à une entreprise utilisatrice (meilleur référencement, meilleures conditions de rémunération) doit être détaillé dans les conditions générales du service.

Les réseaux sociaux étant libres de modifier leurs conditions générales à leur convenance, il est prévu une obligation d'information préalable d'au moins 15 jours avant l'entrée en vigueur en cas de modification substantielle des conditions. En l'absence de cette notification, les nouvelles conditions ne sont pas opposables à l'utilisateur.

En cas de décision de suspension ou de bannissement d'un utilisateur, le réseau social devra fournir une motivation détaillée et circonstanciée justifiant cette décision. Les réseaux sociaux conformes au Règlement P2B pourront donc s'inspirer des procédures déjà mises en place pour se préparer à l'entrée en vigueur du DSA (Digital Services Act).

2.4 Le règlement sur les services numériques (DSA)

À qui s'applique cette réglementation ?

Le Règlement européen 2022/2065 du 19 octobre 2022 relatif à un marché unique des services numériques (dit *« Digital Services Act »* ou *« DSA »*)[26] est un règlement européen visant à harmoniser le régime de responsabilité des plateformes, renforcer la modération des contenus ainsi que la protection des utilisateurs et en particulier des mineurs. Le texte s'applique à tous les fournisseurs de ***« services intermédiaires »*** qui ciblent un public européen, peu importe le lieu où est situé leur siège social.

---

25) https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32019R1150

26) https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32022R2065

Les réseaux sociaux sont considérés comme des services intermédiaires, car ils stockent des contenus partagés par leurs utilisateurs. Les obligations sont renforcées pour les « très grandes plateformes » qui comptent plus de 45 millions d'utilisateurs actifs, soit environ 10 % de la population européenne. La liste des plateformes concernées est publiée et régulièrement mise à jour par la Commission européenne. À l'inverse, les micros et petites entreprises (moins de 50 salariés et 10 millions de chiffre d'affaires) sont exemptées de certaines obligations.

Le DSA est entré en vigueur le 17 février 2023, avec pour seule obligation la publication du nombre moyen mensuel d'utilisateurs actifs afin d'établir la liste des très grandes plateformes. Hormis ce point, le texte entre en application le 17 février 2024, à l'exception des très grandes plateformes de plus de 45 millions d'utilisateurs mensuels pour lesquelles il est applicable depuis le 25 août 2023.

Quelles sont leurs obligations ?

Au niveau français, le DSA reprend et enrichit les dispositions existantes de la LCEN. Parmi les nouvelles mesures, les réseaux sociaux devront nommer un ***« point de contact unique »*** chargé d'assurer la communication avec les autorités compétentes.

Le DSA introduit la notion de « *signaleurs de confiance* », qui sont des entités accréditées dont les signalements devront être traités en priorité par les plateformes. Pour accélérer et augmenter la confiance dans le processus de modération, les coordinateurs nationaux pour les services numériques[27)] attribueront le statut de signaleurs de confiance à des entités répondant aux conditions prévues par l'article 22 du DSA.

Par ailleurs, les réseaux sociaux devront également améliorer leur système de traitement des réclamations (par exemple en cas de suspension de compte ou de décision prise après un signalement), en établissant une structure à deux niveaux. Concrètement, le premier niveau de décision sera assuré par les équipes dédiées à la modération, tandis que le second niveau de décision devrait revenir au service juridique, qui devra tâcher de revoir la proposition de décision de manière impartiale. Les décisions prises par le réseau social, même au premier niveau, devront être établies avec un exposé des motifs expliquant précisément la motivation de la décision ainsi que, le cas échéant, des dispositions contractuelles (telles que les conditions générales ou les règles de la communauté) enfreintes par l'utilisateur du réseau social. Une attention particulière devra être portée au traitement des demandes des mineurs sur les réseaux sociaux, pour qui les mécanismes de signalement doivent être facilités. Les motifs devront être rédigés de manière intelligible en fonction de l'âge de l'utilisateur.

Il sera interdit pour toutes les plateformes de diffuser de la publicité ciblée auprès des mineurs, lorsqu'elle repose sur du profilage au sens du RGPD, de même que la publicité basée sur des données sensibles telles que les opinions politiques, la religion ou l'orientation sexuelle.

27) Pour la France, le Projet de loi visant à sécuriser et réguler l'espace numérique, en cours d'adoption, désigne l'Autorité de régulation de la communication audiovisuelle et numérique (Arcom) comme coordinateur national des services numériques.

Les algorithmes utilisés par les plateformes pour recommander du contenu ou sélectionner et diffuser des publicités devront être expliqués pour que les utilisateurs comprennent leur mécanisme général ainsi que les critères utilisés pour les cibler.

Les interfaces trompeuses ***« dark patterns »*** et autres stratégies d'incitation (***« nudge »***) visant à induire les utilisateurs à effectuer des choix qui pourraient leur être défavorables, seront interdites.

Enfin, les réseaux sociaux ont l'obligation de publier annuellement des rapports contenant des données et des chiffres précis tels que le nombre de suspension et de suppression de comptes, le nombre de signalements et de contenus effectivement retirés, le nombre d'injonctions reçues de la part d'autorités compétentes, etc.

En cas de non-conformité, les amendes maximales encourues peuvent atteindre jusqu'à 6 % du chiffre d'affaires mondial annuel de l'entreprise, pris au niveau du groupe le cas échéant.

Compte tenu de leur influence sur Internet et de leur rôle dans la société en général, des obligations plus importantes sont imposées aux très grandes plateformes : analyse des risques systémiques, réalisations d'audits, mise à disposition de leurs algorithmes à la Commission européenne et aux autorités nationales, protection renforcée des mineurs, etc.

2.5 La loi française visant à encadrer l'exploitation commerciale de l'image d'enfants de moins de 16 ans sur les plateformes en ligne

À qui s'applique cette loi ?

Depuis avril 2021, la loi n°2020-1266 du 19 octobre 2020 visant à encadrer l'exploitation commerciale de l'image d'enfants de moins de 16 ans sur les plateformes en ligne (dite ***« loi du travail des enfants youtubeurs influenceurs sur internet »***) encadre le travail des enfants sur les plateformes de vidéos en ligne.

Dès lors qu'un réseau social propose une fonctionnalité permettant du partage de vidéos (et ce quel que soit le format : live, vidéos préenregistrés etc.), la loi est applicable à ce réseau social.

Il peut être considéré que l'image d'un mineur de moins de 16 ans est commercialisée dès lors que des revenus sont tirés de l'audience de ses vidéos ou de partenariats conclus avec des annonceurs, diffusés en cours de vidéo.

Quelles sont leurs obligations ?

Outre les obligations de déclaration auprès de l'administration incombant aux parents, les plateformes sont tenues de mettre en place certaines mesures pour protéger ces mineurs dont la notoriété leur permet de devenir un vecteur marketing.

Les réseaux sociaux concernés :

- sont incités à adopter des chartes pour favoriser l'information des mineurs sur les conséquences de la diffusion de leur image sur leur vie privée ainsi que sur les risques psychologiques et juridiques, en lien avec les associations de protection de l'enfance. L'Arcom (Autorité de régulation de la communication audiovisuelle et numérique) a publié le 28 novembre 2022 la charte ***« Studer »***. Les réseaux sociaux concernés sont incités à la signer. Elle comporte des engagements non contraignants sur l'information des mineurs, la détection des contenus dangereux et l'exercice du droit à l'effacement (voir ci-dessous).

- sont obligés de répondre dans les meilleurs délais à la demande d'effacement de l'enfant de ses vidéos, même sans le consentement de ses parents. Sauf développements contraires, ce droit d'origine légale devrait écarter tout argument d'un annonceur souhaitant le maintien d'une vidéo en ligne au titre d'un contrat qu'il a conclu avec l'enfant et qui prévoyait une durée minimale de diffusion.

2.6 La loi française sur la majorité numérique

À qui s'applique cette loi ?

La loi n°2023-566 du 7 juillet 2023 visant à instaurer une majorité numérique et à lutter contre la haine en ligne a modifié la LCEN (étudiée dans la section 2.2) pour renforcer la protection des mineurs en ligne.

Elle est applicable aux hébergeurs au sens de la LCEN et plus spécifiquement aux ***« fournisseurs de services de réseaux sociaux en ligne »***.

Au vu des risques de non-conformité de la loi avec le droit de l'Union européenne, l'entrée en application de la loi (qui sera faite de manière progressive) est repoussée à une date fixée par décret, après avis de la Commission européenne.

Quelles sont leurs obligations ?

Pas de mineurs de moins de 15 ans sur les réseaux sociaux, sauf exception

Les réseaux sociaux doivent refuser l'inscription des mineurs de moins de 15 ans, sauf consentement exprès d'un titulaire de l'autorité parentale.

Le respect de cette obligation impose de manière indirecte aux réseaux sociaux de vérifier :

- l'âge de tous leurs utilisateurs, ce afin de s'assurer que des mineurs de moins de 15 ans ne puissent pas s'inscrire sur leur plateforme ;

- l'autorisation du titulaire de l'autorité parentale.

Concrètement, les réseaux sociaux devront mettre en place une solution technique, intégrée à leur service au moment de l'inscription, qui devra être conforme à un référentiel élaboré par l'Arcom. Ce référentiel n'a, pour l'heure, pas été publié.

La sanction prévue en cas de manquement à ces obligations est une amende pouvant aller jusqu'à 1 % du chiffre d'affaires mondial de l'année précédente.

Des solutions techniques existent d'ores et déjà pour vérifier ou estimer l'âge des utilisateurs et recueillir un consentement parental. Il n'est toutefois pas sûr que ces solutions soient acceptées en l'état par le régulateur français, notamment pour des questions de sécurité et de confidentialité des données personnelles. Un autre projet de loi **(*« Sécuriser et réguler l'espace numérique »*)** envisage de confier à l'Arcom l'élaboration d'un référentiel technique définissant des exigences minimales pour ces solutions de vérification de l'âge dans le cadre de l'accès aux sites à caractère pornographique, pris après avis de la CNIL. Il devrait être publié fin 2023 ou courant 2024.

Retirer les contenus contenant du cyberharcèlement

La loi élargit le champ des raisons pour lesquelles un contenu peut être considéré comme manifestement illicite au sens de l'article 6 de la LCEN, ajoutant les contenus révélant du harcèlement conjugal ou moral, du chantage, des atteintes à la vie privée (diffusion de contenus intimes ou *« revenge porn »*) ou des « deep fakes » (technique de création d'un contenu visuel ou audio truqué basée sur l'intelligence artificielle).

Les réseaux sociaux sont également dans l'obligation de répondre aux réquisitions judiciaires dans un délai de 10 jours, voire de 8 heures en cas d'urgence (par exemple en cas de soupçons de menaces, d'appel au lynchage, de diffusion de vidéos représentant des violences, de la barbarie). Enfin, ils devront rendre visibles des messages de prévention contre le harcèlement en ligne et renvoyer vers des structures d'accompagnement face au cyberharcèlement, notamment le numéro 3018, le numéro pour les jeunes victimes de harcèlement et de violences numériques.

3 La régulation des contenus

3.1 Le régime général de lutte contre les contenus manifestement illicites

Conformément à ce qui a été exposé dans la section 2 ci-dessus, l'hébergeur ne prend pas activement connaissance des contenus publiés sur sa plateforme et se contente d'assurer le stockage ou la mise à disposition de ces contenus. En raison de son rôle passif et purement technique, il n'est pas responsable des contenus illicites hébergés, à condition qu'il n'en ait pas connaissance. Ainsi, l'hébergeur n'a pas une obligation générale de surveiller en permanence les contenus présents sur ses services. Sa responsabilité peut être engagée s'il est notifié de contenus manifestement illicites qu'il n'a pas retirés dans les délais appropriés.

Selon le régime de responsabilité général des hébergeurs découlant de la LCEN et du DSA, les plateformes ont l'obligation de mettre en place des moyens efficaces pour supprimer les contenus ***« manifestement illicites »*** qui sont portés à leur connaissance.

Ces obligations se matérialisent par l'obligation de retirer les contenus illicites plus ou moins rapidement : parfois la loi prescrit un délai fixe (les hébergeurs doivent retirer les contenus terroristes dans l'heure suivant une injonction judiciaire), ou se réfère à un *« prompt »* délai (tel est le cas pour les infractions de droit commun - l'appréciation du délai étant laissé aux juges du fond). De façon générale, en moyenne, les hébergeurs ont jusqu'à 72 heures pour agir, au-delà desquelles ils s'exposent à ne pas avoir agi *« promptement ».* Néanmoins, le délai retenu par les tribunaux dépend généralement de la gravité et de la complexité de l'infraction en cause. Il sera généralement attendu des plateformes qu'elles retirent dans l'heure les contenus extrêmement violents, de sorte qu'ils ne puissent être consultés par de larges audiences, et en particulier les jeunes.

Cette approche vise à préserver la liberté d'expression sur Internet et la neutralité des plateformes. Ces dernières ne doivent pas se substituer à l'autorité judiciaire pour juger de la licéité d'un contenu selon le droit applicable. C'est pourquoi leur responsabilité est limitée aux actions qu'elles entreprennent une fois informées de la présence de contenus **manifestement** illicites sur leurs services. Les politiques de modération des plateformes doivent donc prendre en compte ce régime de responsabilité.

À ce régime de responsabilité général s'ajoutent des obligations spécifiques de régulation des contenus, notamment dans les cas où la protection d'un intérêt public majeur - protection des mineurs, lutte contre le terrorisme, etc. - nécessite la mise en place de moyens renforcés.

Pour plus d'information sur la notion de ***« contenus manifestement illicites »***, se référer à la Section 1.1 Partie II.

3.2 Le règlement européen sur la lutte contre les contenus terroristes (*« Terrorist Content Online »* ou *« TCO »*)

À qui s'applique cette réglementation ?

Le règlement européen 2021/784 du 29 avril 2021[28] (***« Règlement TCO »***) a été mis en place dans le but de lutter contre la radicalisation dans l'Union européenne, en apportant une réponse commune et harmonisée au sein de l'Union. Ce règlement s'applique à tous les hébergeurs proposant leurs services dans l'un des États membres de l'Union européenne, quel que soit leur lieu d'établissement, dès lors qu'ils diffusent des informations au public.

Les réseaux sociaux se trouvent donc particulièrement concernés par ce texte.

Comme indiqué à la section 2.2 ci-dessus, le législateur français a adopté une loi pour transposer le Règlement TCO et adapter la LCEN[29].

---

28) https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32021R0784

29) Il s'agit de la loi n°2022-1159 du 16 août 2022 portant diverses dispositions d'adaptation au droit de l'Union européenne en matière de prévention de la diffusion de contenus à caractère terroriste en ligne.

Quelles sont leurs obligations ?

Le texte impose une obligation renforcée de diligence en ce qui concerne le retrait des contenus terroristes et instaure une procédure d'injonction de retrait dans l'heure des contenus terroristes postés en ligne.

Ainsi, les réseaux sociaux ont l'obligation de retirer ou de bloquer, dans l'heure suivant une injonction des autorités compétentes, les contenus terroristes qui leur sont signalés. Ces contenus doivent être conservés pendant 6 mois, sans être disponibles au public, afin de les utiliser éventuellement dans le cadre d'une procédure judiciaire.

La plateforme doit également informer ses utilisateurs dans ses conditions générales des mesures qu'elle met en place pour lutter contre les contenus terroristes. Lorsqu'un contenu est retiré, elle doit informer l'utilisateur à l'origine du contenu de la décision de retrait et des motifs qui y sont liés.

Les réseaux sociaux, ainsi que les utilisateurs, bénéficient de la possibilité de faire un recours accéléré (dans un délai de 48 heures) devant le tribunal administratif, lequel aura 72 heures pour statuer, notamment dans les cas où du matériel terroriste est utilisé à des fins éducatives. Des rapports de transparence doivent être établis par les réseaux sociaux sur les injonctions reçues concernant les contenus terroristes.

En cas de non-respect de ces obligations, des sanctions pouvant atteindre jusqu'à 4 % du chiffre d'affaires annuel peuvent être imposées aux personnes morales. De plus, le réseau social peut également faire l'objet d'une mesure de déréférencement par les fournisseurs d'accès à Internet et par les moteurs de recherche.

Cette nouvelle procédure d'injonction s'ajoute aux procédures déjà existantes prévues, notamment, par la LCEN.

3.3 Le futur règlement européen sur la lutte contre les abus sexuels concernant des enfants (*« Child Sexual Material »* ou *« CSAM »*)

Àqui s'applique cette réglementation ?

À ce jour, il s'agit seulement d'une proposition de règlement, émanant de la Commission européenne en date du 11 mai 2022. Cette proposition est en cours de négociation au sein des institutions européennes.

Le champ d'application de ce règlement comprendrait principalement les prestataires de service d'hébergement, les fournisseurs de services de communication interpersonnelle, (tels que les services de messagerie), et les boutiques d'applications logicielles. Les réseaux sociaux font partie des acteurs les plus significativement impactés par ce projet de réglementation.

Quelles sont leurs obligations ?

Les réseaux sociaux devront effectuer une évaluation des risques liés à l'utilisation abusive de leur plateforme ou de leur service de messagerie à des fins d'abus sexuels sur les enfants et/ou de sollicitation d'enfants (pédo piégeage ou « *grooming* »).

En aval de cette analyse de risque, le texte impose aux réseaux sociaux de mettre en place les mesures d'atténuation raisonnables et adaptées pour le réduire *a minima*. Le texte mentionne des mesures opérationnelles telles que l'utilisation des outils de modération automatique des messages, l'augmentation du personnel dédié ou encore la mise en place d'outils de vérification de l'âge.

S'agissant de la détection, l'autorité nationale de coordination pourra obtenir une injonction judiciaire pour contraindre les plateformes de détecter et retirer ou bloquer les contenus à caractère sexuel de mineurs ou sollicitations de mineurs (ensemble désignés comme « *contenus CSAM* »). Cette injonction, limitée dans le temps, pourra porter tant sur des contenus nouveaux que ceux déjà connus.

Le texte prévoit également la création d'un centre européen de lutte contre la pédopornographie en ligne. Son rôle consistera notamment à fournir aux opérateurs de réseaux sociaux les technologies pertinentes et apprécier la validité des mesures d'atténuation mentionnées ci-dessus.

Les hébergeurs ont également une obligation de signalement des contenus CSAM dont ils ont connaissance, autrement que par une injonction. Le signalement se fera, le cas échéant, au Centre de l'UE.

Ils seront soumis à une obligation de conservation des informations traitées à des fins de lutte contre les contenus CSAM (données de contenu et données à caractère personnel), en particulier pour les besoins des procédures pénales.

Les opérateurs de réseaux sociaux seront également tenus de fournir des rapports de transparence comprenant des chiffres détaillés sur le nombre d'injonctions reçues, le taux d'erreur des technologies employées, le temps moyen pour retirer des contenus CSAM, etc.

Le texte détaille également les garanties que devront présenter les technologies de détection qui seront mises en œuvre pour lutter contre les contenus CSAM. Ces exigences incluent d'efficacité, le taux d'erreur le plus bas possible, l'interdiction de les utiliser à d'autres fins que celles prévues par le règlement, etc.

Des sanctions allant jusqu'à 6 % du chiffre d'affaires annuel mondial sont prévues en cas de manquement aux obligations énoncées dans ce règlement.

4 La protection des données personnelles

En tant qu'outils de communication et de partage, les plateformes de réseaux sociaux ont souvent besoin de collecter d'importants volumes de données personnelles pour fonctionner. Ces traitements sont généralement inhérents aux services qu'ils proposent : partage de contenus, de photos et de vidéos, *marketplaces*, services de rencontre, services de streaming, etc. Ainsi, les règles applicables à la protection des données personnelles sont particulièrement importantes pour les réseaux sociaux.

4.1 Le règlement général sur la protection des données personnelles (RGPD)

À qui s'applique cette réglementation ?

Le Règlement (UE) 2016/679 du 27 avril 2016 (connu sous le nom de **« Règlement Général sur la Protection des Données »** ou **« RGPD »)**, s'applique à toute organisation publique ou privée, quelle que soit sa taille, qui utilise des données personnelles (salariés, utilisateurs, prestataires) dans le cadre de son activité.

Le RGPD se distingue par son extraterritorialité. D'une part, il est applicable aux organisations situées au sein de l'Union européenne. D'autre part, il est également applicable aux organisations situées en dehors de l'Union européenne qui offrent des biens ou des services à des personnes situées dans les États membres de l'Union européenne ou qui suivent le comportement des utilisateurs européens.

Ainsi, le champ d'application du RGPD est donc très large et a vocation à régir les organisations offrant des services aux citoyens européens dans la quasi-totalité des cas.

Quelles sont leurs obligations ?

Justifier l'utilisation des données personnelles

Chaque utilisation de données personnelles doit être justifiée par au moins l'une des **bases légales** suivantes énoncées à l'article 6 du RGPD) :

- Le consentement de l'utilisateur.
- L'existence d'un contrat entre l'utilisateur et la plateforme.
- Une obligation légale.
- La sauvegarde des intérêts vitaux d'une personne.
- L'exécution d'une mission de service public ou relevant de l'exercice de l'autorité publique dont est investi le responsable de traitement.
- Les intérêts légitimes de l'organisation.

Par ailleurs, le RGPD établit une liste de **données considérées comme sensibles**, dont le traitement est en principe interdit, sauf s'il existe une justification légale permettant leur traitement. Il s'agit des données qui révèlent *« l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques, l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique »* (article 9 du RGPD).

Par exception, il est possible de traiter ces types de données sensibles en se fondant sur au moins l'un des éléments suivants :

- Le consentement de l'utilisateur.
- La réglementation en droit du travail et de la sécurité sociale.
- La sauvegarde des intérêts vitaux d'une personne.
- Le caractère public de ces données.
- L'exercice d'un droit en justice.
- L'existence d'un motif d'intérêt public important.

L'usage d'un réseau social implique souvent de laisser les utilisateurs exercer leur liberté d'expression, en leur permettant de s'exprimer sur des sujets aussi variés que la politique, la religion, ou leur préférences personnelles. Dès lors, les plateformes peuvent être amenées à traiter les données particulières listées à l'article 9 du RGPD. Dans la mesure où ces informations sont fournies volontairement par l'utilisateur qui les rend publiques, et qu'elles ne sont pas sollicitées par la plateforme, le traitement est justifié par l'article 9 e) du RGPD.

Les plateformes en ligne étant liées à leurs utilisateurs par des contrats (en général des conditions générales d'utilisation), la base légale retenue pour l'utilisation des données est, dans un grand nombre de cas, la base contractuelle. Pour autant, cette base contractuelle ne peut pas être utilisée pour toutes les activités menées par la plateforme. En effet, les régulateurs adoptent une vision de plus en plus restrictive s'agissant des utilisations de données personnelles qui sont liées à l'exécution d'un contrat.

Une illustration de cette approche plus restrictive a été donnée par le régulateur irlandais **« Data Protection Commission »** dans une décision du 31 décembre 2022[30)]. Cette décision concernait l'utilisation des données personnelles des utilisateurs à des fins de publicité personnalisée et d'amélioration du service.

Selon les régulateurs européens, l'utilisation des données personnelles à des fins de personnalisation et de publicité comportementale dépasse ce qui est strictement nécessaire pour fournir les services proposés sur les réseaux sociaux aux utilisateurs.

Il en ressort que l'utilisation des données des utilisateurs dans le cadre de l'exécution des contrats doit se limiter aux fonctionnalités élémentaires du service, telles que la mise place d'une messagerie, la fourniture éventuelle de moyens de paiement pour régler un achat ou la mise en place d'un support client. Les autres utilisations doivent être justifiées en ayant recours à une autre base légale, telle que le consentement ou l'intérêt légitime.

30) La décision du 31 décembre 2022 : https://www.dataprotection.ie/sites/default/files/uploads/2023-04/Meta%20FINAL%20Decision%20%28ADOPTED%29%20-%20IN-18-5-7%20-%2031-12-22%20%28Redacted%29.pdf

En ce qui concerne le recours à l'intérêt légitime, celui-ci est généralement pertinent pour les opérations de modération et de détection des infractions en ligne, ainsi que la lutte contre la fraude en ligne. En effet, le Considérant 47 du RGPD reconnaît que les traitements de données personnelles à des fins de prévention de la fraude constitue un intérêt légitime du responsable de traitement. Pour cela, l'organisation devrait effectuer une évaluation de l'intérêt légitime en question afin de s'assurer qu'il ne porte pas atteinte d'une manière disproportionnée aux intérêts, droits et libertés des utilisateurs du service.

Documenter l'utilisation des données personnelles

Le RGPD impose aux plateformes en ligne de constituer un ***« registre »*** recensant leurs activités de traitement de données personnelles, dans le but de responsabiliser des entreprises (article 30 du RGPD). Une exception est prévue pour les organisations de moins de 250 salariés, qui peuvent décider de documenter uniquement les traitements présentant un risque pour les droits et libertés des personnes concernées, notamment l'utilisation de données sensibles.

Ce registre doit contenir des informations générales comme les coordonnées du DPO, les finalités du traitement des données, les durées de conservation, les transferts internationaux, etc. Il doit pouvoir être mis à la disposition des autorités de contrôle, sur demande.

Par ailleurs, le RGPD impose aux plateformes de mener des analyses d'impact relatives à la protection des données (AIPD) lorsqu'une utilisation spécifique de données présente un risque élevé sur les droits et libertés des personnes concernées (article 35 du RGPD).

Selon un faisceau d'indices dégagé par le Comité européen de la protection des données (CEPD), un risque élevé est présumé lorsqu'au moins deux des neuf critères suivants sont réunis :

- La plateforme effectue une évaluation ou un scoring de ses utilisateurs.
- Des décisions automatiques ayant un effet légal ou similaire sont mises en œuvre.
- Une surveillance systématique est mise en place.
- Des données sensibles sont collectées.
- Des données sont collectées à grande échelle.
- Des jeux de données différents font l'objet de recoupements.
- Des données de personnes vulnérables sont traitées (enfants, personnes âgées, etc.).
- Une technologie innovante est utilisée.
- La personne risque d'être exclue du bénéfice d'un droit ou d'un contrat.

Les plateformes peuvent réaliser les AIPD en interne, en utilisant l'outil élaboré par la CNIL[31)], ou en faisant appel à tout prestataire spécialisé ou à leur conseil juridique.

Par ailleurs, la CNIL a publié une liste des traitements pour lesquels une AIPD est requise[32)]. Parmi ceux-ci, l'autorité inclut :

- tout *« traitement reposant sur une analyse comportementale visant à détecter des comportements « interdits » sur un réseau social »* ;
- tout *« traitement visant à personnaliser les publicités en ligne »* ;
- tout *« dispositif de signalement de mineurs en danger »*.

Assurer un recours humain en cas de prise de décision automatique

Sur les réseaux sociaux, il est courant que des opérations automatisées, qu'elles soient basées sur l'intelligence artificielle ou non, soient mises en place pour collecter des données personnelles relatives aux utilisateurs, que cela soit à des fins de sécurité, de modération, ou encore à des fins commerciales et/ou publicitaires. Ces opérations automatisées peuvent également recourir au profilage, c'est-à-dire au traitement des données visant à analyser ou prédire le comportement d'un individu, ses préférences, ses habitudes de consommation, sa situation financière, etc.

Le RGPD vise à limiter les risques associés à ces opérations, afin d'éviter que les plateformes ne prennent des décisions erronées, injustes ou ayant des effets indésirables sur les personnes tels que le refus de services, la discrimination ou l'enfermement dans des stéréotypes.

Dans ce type de situations, les individus disposent de droits spécifiques lorsque des décisions entièrement automatisées sont prises à leur encontre et qui pourraient avoir des conséquences négatives pour eux. Ils doivent notamment être informés de l'existence et des conséquences de ces décisions, ainsi que de sa logique sous-jacente (article 13 du RGPD). Ensuite, ils ont le droit de demander une intervention humaine (article 22 du RGPD), notamment pour obtenir un nouvel examen de la situation et éventuellement contester la décision prise.

Ces principes pourraient s'appliquer dans les domaines suivants des réseaux sociaux :

- Modération et sécurité en ligne, lorsque des algorithmes fournissent un support automatisé aux équipes de modérateurs, par exemple en classant les signalements des utilisateurs, en recherchant proactivement des usages abusifs du service, et/ou en retirant automatiquement des contenus abusifs.
- Profilage commercial des utilisateurs, reposant sur des traitements de données pour comprendre leurs attentes et les inciter à consommer des produits et services.
- Système de recommandation de contenus reposant sur une analyse des préférences personnelles des utilisateurs ou de l'historique de leur navigation.

---

31) L'outil pour la réalisation des AIPD élaboré par la CNIL : https://www.cnil.fr/fr/outil-pia-telechargez-et-installez-le-logiciel-de-la-cnil

32) https://www.cnil.fr/sites/cnil/files/atoms/files/liste-traitements-aipd-requise.pdf

Dans une logique d'efficacité et en anticipation de l'entrée en vigueur du DSA, le recours à un interlocuteur humain est souvent assuré dans un premier temps par le personnel dédié au service client (assistance, service clientèle, support etc.) puis, en cas de contestation ou de difficulté particulière, par le personnel du service juridique ou, le cas échéant, un médiateur extérieur.

En cas de non-respect du RGPD, les autorités de contrôle nationales peuvent prononcer des amendes administratives pouvant atteindre jusqu'à 20 millions d'euros ou 4 % du chiffre d'affaires mondial annuel de l'entreprise, selon le montant le plus élevé.

Respecter les droits des individus

Le RGPD octroie de nouveaux droits aux individus et leur accorde ainsi un meilleur contrôle sur les données personnelles qui les concernent. Ces droits, détaillés aux articles 12 à 22 du RGPD, peuvent se résumer ainsi :

- **Droit à l'information** : Les utilisateurs ont le droit d'être informés de manière claire et transparente sur la manière dont leurs données personnelles sont collectées et utilisées. Les plateformes doivent notamment publier une politique de confidentialité qui contient toutes les informations requises.

- **Droit d'accès** : Les utilisateurs ont le droit de demander quelles données personnelles sont détenues à leur sujet et d'obtenir une copie de ces données.

- **Droit de rectification** : Les utilisateurs ont le droit de demander la correction de données personnelles inexactes ou incomplètes les concernant.

- **Droit à l'effacement (ou « droit à l'oubli »)** : Les utilisateurs ont le droit de demander la suppression de leurs données personnelles dans certaines circonstances, par exemple, si les données ne sont plus nécessaires aux fins pour lesquelles elles ont été collectées. En France, la loi ajoute que lorsque la personne était mineure au moment de la collecte des données, elle a de ce seul fait le droit d'en obtenir rapidement l'effacement (article 51 de la loi Informatique et Libertés).

- **Droit de limitation du traitement** : Les utilisateurs peuvent demander la restriction du traitement de leurs données personnelles dans certaines situations, notamment si les données sont inexactes ou si le traitement est illégal.

- **Droit à la portabilité des données** : Les utilisateurs ont le droit de recevoir leurs données personnelles dans un format structuré, couramment utilisé et lisible par machine, et de les transmettre à une autre entreprise si nécessaire et si cela est faisable.

- **Droit d'opposition** : Les utilisateurs ont le droit de s'opposer au traitement de leurs données personnelles à des fins de marketing direct ou pour des raisons liées à leur situation particulière.

- **Prise de décision automatisée et profilage** : Les utilisateurs ont le droit de ne pas être soumis à des décisions automatisées, y compris le profilage, qui produisent des effets juridiques significatifs ou ont un impact similaire sur eux sans intervention humaine.

- **Droit de retirer le consentement** : Si le traitement des données repose sur le consentement de l'utilisateur, ce dernier a le droit de retirer son consentement à tout moment.

4.2 La loi Informatique et Libertés

À qui s'applique cette loi ?

Comme pour le RGPD, la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (dite **« loi Informatique et Libertés »**)[33], est applicable à toute personne morale ou physique dès lors qu'elle utilise des données personnelles relatives à des tiers hors d'un cadre domestique ou personnel.

Elle s'applique aux réseaux sociaux ayant un établissement sur le sol français (quel que soit la nationalité de ses utilisateurs) ainsi qu'aux réseaux sociaux qui s'adressent à un public français (un utilisateur étant suffisant pour caractériser le public français).

Quelles sont leurs obligations ?

La loi française Informatique et Libertés a été largement remaniée pour intégrer les obligations issues du RGPD. En revanche, le législateur français a la possibilité d'aller plus loin dans la protection des utilisateurs en ligne, et particulièrement des mineurs.

Cette loi prévoit ainsi que :

- la CNIL publie des recommandations et des lignes directrices s'agissant du respect des droits des mineurs ;
- le traitement de données personnelles des mineurs de moins de 15 ans dans le cadre de services de la société de l'information est licite sous réserve d'obtenir le consentement du ou des titulaires de l'autorité parentale ;
- les informations et communications fournies aux mineurs doivent apparaître dans un langage clair et aisément compréhensible. Une bonne pratique peut consister à mettre à la disposition des utilisateurs un résumé des informations requises par le RGPD (souvent matérialisées par la politique de protection des données personnelles). Ce résumé devrait être bien plus concis que son document de référence, user d'un vocabulaire simple et moins juridique, et pourrait même prendre la forme d'un schéma, d'un dessin, de graphiques ou de vidéos, pour autant que le format soit facile à appréhender pour un jeune ;
- le droit à l'effacement des données pour les mineurs est renforcé. Il doit être fait dans les *« meilleurs délais »* et ce droit s'applique sans restriction dès lors que la personne était mineure au moment de la collecte des données. Tout réseau social devrait donc traiter les demandes d'effacement de données personnelles de leurs utilisateurs mineurs en priorité.

33) Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

4.3 La directive e-Privacy et le Code des Postes et des Communication Électroniques

À qui s'applique cette directive européenne ?

La Directive européenne 2000/31 du 8 juin 2000, amendée à plusieurs reprises, régit la protection de la vie privée dans le domaine des communications électroniques, comme les échanges au sein d'un service de messagerie instantanée.

Selon son article 3, la directive s'applique au ***« traitement des données à caractère personnel dans le cadre de la fourniture de services de communications électroniques accessibles au public sur les réseaux de communications publics*** *dans la Communauté, y compris les réseaux de communications publics qui prennent en charge les dispositifs de collecte de données et d'identification »*. Outre le traitement de données personnelles, l'entreprise doit fournir un service de communications électroniques, lequel est accessible au public via les réseaux présents au sein de l'Union européenne.

Quelles sont leurs obligations ?

La directive établit le principe de confidentialité de ces communications et interdit l'interception, la surveillance ou l'enregistrement de ces communications sans le consentement des utilisateurs, sauf dans certaines exceptions strictement définies.

Cette directive englobe également :

- la réglementation des cookies et des traceurs, qui sont soumis au consentement préalable de l'utilisateur, sauf lorsqu'ils sont strictement nécessaires à la fourniture du service ;

- les règles applicables au marketing direct et à la prospection commerciale : en principe, les entreprises doivent obtenir le consentement préalable des destinataires avant d'envoyer des communications de marketing direct, sauf dans le cas d'une relation client existante et sous réserve du droit d'opposition effectif des utilisateurs.

En tant que directive européenne, sa mise en œuvre dans les États membres n'est pas actuellement harmonisée. Chaque État membre est libre de définir à la fois les instruments juridiques pour transposer la directive et les interprétations de fond du texte.

En France, la directive e-Privacy est partiellement transposée dans le Code des postes et des communications électroniques, dont l'article L. 32-3[34)] protège le secret des correspondances. Quant aux dispositions sur les cookies, les traceurs et la prospection commerciale, la CNIL a émis plusieurs séries des lignes directrices et recommandations pratiques pour aider les acteurs à se conformer aux principes énoncés dans la directive[35)].

La Directive e-Privacy devrait prochainement être mise à jour par un futur règlement e-Privacy, en vue d'harmoniser les règles applicables au niveau de l'Union européenne. Une proposition de règlement e-Privacy a été publiée en janvier 2017 par la Commission européenne. La négociation du texte ayant fait l'objet de vifs débats et de désaccords persistants, les négociations au niveau européen sont toujours en cours.

5 Les recommandations des régulateurs nationaux

5.1 Les recommandations de la CNIL

À qui s'appliquent ces normes ?

La CNIL a publié des recommandations pratiques sur la protection des mineurs. Celles-ci sont applicables aux entreprises proposant des produits ou services fournis en ligne et susceptibles d'être consommés par des mineurs.

Les recommandations qui retiennent notre attention dans le cadre de ce document sont :

- celles portant sur les **droits numériques des mineurs** ;
- celles portant sur le **contrôle de l'âge en ligne**, en date du 26 juillet 2022.

S'agissant du contrôle de l'âge, ces recommandations s'appliquent tout particulièrement aux acteurs dont les services sont soumis à une condition d'âge (par exemple : les jeux d'argent, la vente de boissons alcoolisées, etc.) et ayant besoin de s'assurer de l'âge de leurs utilisateurs. Les sites pornographiques ont, quant à eux, l'obligation expresse de vérifier l'âge de leurs utilisateurs, afin d'en interdire l'accès aux mineurs.

---

34) L'article 32-3 de la Directive e-Privacy protégeant le secret des correspondances :
https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000034312035

35) Concernant les cookies et traceurs :
- Délibération n° 2020-091 du 17 septembre 2020 portant adoption de lignes directrices relatives à l'application de l'article 82 de la loi du 6 janvier 1978 modifiée aux opérations de lecture et écriture dans le terminal d'un utilisateur (notamment aux « cookies et autres traceurs ») et abrogeant la délibération n° 2019-093 du 4 juillet 2019 ;
- Délibération n° 2020-092 du 17 septembre 2020 portant adoption d'une recommandation proposant des modalités pratiques de mise en conformité en cas de recours aux « cookies et autres traceurs ».

Concernant la prospection commerciale : https://www.cnil.fr/fr/la-prospection-commerciale

Quelles sont leurs obligations ?

L'article 8 du RGPD établit une protection renforcée pour les mineurs lorsque leurs données personnelles sont utilisées dans le cadre de services de la société de l'information (dont font partie les réseaux sociaux) sur la base du consentement.

Chaque État membre est libre de fixer l'âge à partir duquel un enfant est considéré comme capable de donner son consentement dans le cadre d'un service en ligne, entre 13 et 16 ans. En France, cet âge minimum a été fixé à **15 ans** (article 45 de la loi Informatique et Libertés). Dans l'hypothèse où la personne concernée par le traitement est un mineur, le réseau social doit obtenir un double consentement : celui de l'enfant et celui du titulaire de l'autorité parentale. Cette obligation pose des difficultés de mise en œuvre, c'est pourquoi le texte précise que le responsable de traitement doit mettre en place des mesures *« raisonnables »*, et *« compte tenu des moyens technologiques disponibles ».*

Pour accompagner les réseaux sociaux dans le respect de leurs obligations, la CNIL a publié une série de 8 recommandations sur les droits numériques des mineurs[36]. Elles ont vocation à indiquer aux acteurs concernés comment mettre en place le recueil du consentement des mineurs, leur délivrer une information intelligible ou encore protéger la vie privée des enfants pour éviter la surveillance généralisée de leurs activités en ligne.

Les acteurs devraient aussi mettre en place des moyens pour les encourager et les aider à exercer directement leurs droits en ligne, et notamment pour leur permettre de se défendre face aux menaces de cyberharcèlement.

Par ailleurs, la CNIL préconise que les services en ligne fournis à des mineurs soient paramétrés par défaut avec une confidentialité renforcée.

Dans ses recommandations sur la vérification de l'âge en ligne, la CNIL recommande aux acteurs de veiller à ce que les dispositifs soient efficaces, transparents et respectueux de la vie privée. Ils doivent aussi être suffisamment fiables pour garantir que seuls les adultes ont accès aux contenus ou services interdits aux mineurs. Pour renforcer la confidentialité d'un tel dispositif, l'autorité propose notamment de recourir à un système de double anonymat, mis en œuvre par des acteurs différents : d'une part, un tiers indépendant en charge du contrôle de l'âge, et d'autre part, le site ou la plateforme consultée devant vérifier l'âge de ses visiteurs. La CNIL promeut un modèle où les rôles sont séparés et où la confidentialité des données est assurée tout au long de la chaîne (par exemple, le tiers vérificateur indépendant n'a pas besoin de savoir quel site est consulté). Le tiers de confiance pourrait revêtir un rôle de *« gestionnaire d'attributs »*, placés sous le contrôle des individus.

[36] https://www.cnil.fr/fr/les-droits-numeriques-des-mineurs

5.2 Les recommandations de l'Arcom

À qui s'appliquent ces normes ?

L'une des missions de l'Arcom consiste en la supervision des moyens mis en œuvre par les plateformes pour lutter contre la diffusion des contenus préjudiciables, notamment à des fins de protection des publics (dont mineurs). À ce titre, l'Arcom a publié des recommandations et participé à de nombreuses campagnes d'information et de sensibilisation. L'autorité a notamment publié des lignes directrices pour contribuer à la lutte contre la dissémination des contenus haineux en ligne, en date du 28 novembre 2022, en application de l'article 42 de la loi du 24 août 2021 confortant le respect des principes de la République.

Les acteurs soumis à ce dispositif sont les opérateurs de plateformes en ligne dont la fréquentation dépasse les 10 millions de visiteurs uniques par mois. Des obligations supplémentaires sont imposées à celles dont la fréquentation dépasse 15 millions de visiteurs uniques par mois.

Quelles sont leurs obligations ?

Voici ci-dessous un condensé des lignes directrices de l'Arcom du 28 novembre 2022 :

- Il est proposé aux plateformes, pour être conformes à l'article 6-4 de la LCEN, de mettre en place *« un dispositif permettant d'accuser réception immédiatement des demandes et injonctions »* émises par les autorités compétentes.

- Les plateformes devraient porter une attention singulière aux injonctions provenant des autorités concernant une *« atteinte grave ou imminente à la vie ou à l'intégrité physique, ou qui visent expressément une personne particulièrement vulnérable »*.

- Dans un souci d'efficacité de la coopération avec les autorités publiques, les plateformes doivent rendre publique et facilement accessible l'adresse électronique de contact où adresser les demandes.

- Dans un souci de transparence, il convient de mettre à la disposition des utilisateurs les politiques de modération, lesquelles doivent être accessibles *« rapidement »* - c'est-à-dire en un ou deux clics par exemple. Le format doit être *« pratique et convivial »*.

- Dans leurs politiques de modération, les plateformes doivent porter une attention particulière au respect de la liberté d'expression, ainsi qu'au pluralisme des médias, et notamment lorsqu'elles utilisent des moyens automatisés.

- Dans le cadre des signalements des utilisateurs, l'autorité recommande de : (i) accuser réception *« sans délai »* des signalements, (ii) informer l'auteur du signalement des suites qui y sont données ainsi que des voies de recours disponibles contre la décision, (iii) dans la décision, insérer un lien renvoyant directement vers le mécanisme de recours, (iv) faire examiner les recours par des collaborateurs qualifiés, (v) laisser un délai de 6 mois à compter de la notification de la décision pour former un recours contre celle-ci.

Par ailleurs, il est également intéressant de mentionner certaines initiatives prises en dehors de l'Union européenne, qui exercent une influence certaine sur le paysage normatif européen dans le domaine de la protection de l'enfance.

À ce sujet, le Royaume-Uni est un précurseur, avec l'adoption par l'*Information Commissioner's Office* (ICO), du code des enfants ***« UK Age Appropriate Design Code »***[37] (ou ***« Children's Code »***), entré en vigueur en septembre 2021. Il impose aux services susceptibles d'être utilisés par des mineurs, y compris les réseaux sociaux, une série de 15 principes pour protéger la vie privée et le bien-être des enfants. Les obligations sont multiples : réalisation d'analyses d'impact, paramétrage de confidentialités restrictif par défaut, limitation de la géolocalisation, prohibition des systèmes de *« nudge »* ou encore limitation du partage des données avec des tiers.

En Californie, le ***« California Age-Appropriate Design Code Act »***[38] entrera en vigueur le 1er juillet 2024. Inspiré de son équivalent britannique, il reprend en substance les mêmes obligations.

En Australie, le ***« Online Safety Act »***[39] de 2021 impose à certaines industries, notamment aux réseaux sociaux, d'élaborer des codes contraignants. Le ***« Social Media Services Code »***[40] pris en application du Online Safety Act, entre en vigueur le 16 décembre 2023. Il oblige les réseaux sociaux à prendre des mesures pour garantir la confidentialité et la protection des mineurs, avec un niveau d'exigence adapté selon la taille et les fonctionnalités offertes par le réseau social.

6 Le droit pénal

Infractions liées à la criminalité et à la pédocriminalité en ligne

À qui s'appliquent ces règles ?

Toute société opérant sur le sol français ou s'adressant à un public français. Au titre de leur qualité d'hébergeur, les plateformes opérant des réseaux sociaux ont une responsabilité civile et pénale spéciale, décrite ci-dessous.

Quant aux destinataires des services, les utilisateurs, ils sont responsables personnellement des contenus et des informations qu'ils postent sur ces services. Des contenus illicites peuvent donner lieu à des condamnations pénales.

---

37) Uk Age Appropriate Design Code : https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/introduction-to-the-childrens-code/#:~:text=The%20Children's%20code%20(or%20the,to%20protect%20children's%20data%20online

38) California Age-Appropriate Design Code Act : https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/introduction-to-the-childrens-code/#:~:text=The%20Children's%20code%20(or%20the,to%20protect%20children's%20data%20online

39) Online Safety Act : https://www.esafety.gov.au/newsroom/whats-on/online-safety-act

40) Social Media Services Code : https://www.esafety.gov.au/sites/default/files/2023-05/eSafety_summary_Social_media_service_providers.pdf

Qui porte la responsabilité ?

La caractérisation d'un délit ou d'un crime comprend généralement deux éléments : un élément matériel (la réalisation de l'infraction) et un élément intentionnel (la volonté de la commettre).

Parce qu'ils ne peuvent pas contrôler et avoir connaissance de tous les contenus hébergés sur leur plateforme, les réseaux sociaux bénéficient en principe d'un régime de responsabilité allégé sur le même modèle que leur régime de responsabilité civile (voir partie sur la responsabilité civile des réseaux sociaux).

En effet, comme évoqué dans la section 3.1 ci-dessus, les plateformes ne peuvent voir leur responsabilité pénale engagée à raison des contenus hébergées à la demande de leurs utilisateurs (c'est-à-dire que ces derniers publient), dès lors qu'elles n'avaient pas eu connaissance de leur caractère manifestement illicite ou qu'elles ont retiré de tels contenus rapidement après en avoir eu connaissance. La loi ajoute explicitement que les hébergeurs n'ont pas d'obligation de rechercher des infractions commises via leurs services.

Le texte précise que cette exonération de responsabilité ne s'applique pas si la personne ayant posté le contenu illicite a agi sous l'autorité ou le contrôle de la plateforme.

Ce n'est qu'après avoir été notifié de la présence d'un contenu manifestement illicite sur la plateforme, et en l'absence du prompt retrait du contenu, que la responsabilité pénale d'une plateforme pourrait être engagée. Ainsi, pour les infractions commises par leurs utilisateurs, l'intentionnalité du réseau social (et par conséquent sa responsabilité pénale potentielle) s'entend donc comme son absence de réaction suite à la notification de l'infraction sur la plateforme.

Lorsqu'une infraction est commise par une personne morale, les amendes prononcées peuvent être multipliées par cinq en vertu de l'article 131-8 du Code pénal. La responsabilité des dirigeants peut également être engagée s'ils ont personnellement commis les actes incriminés dans le cadre de leurs fonctions de représentation de l'entreprise, ou si l'un de leurs salariés a commis une faute grave.

Quelles sont les infractions ?

Un grand nombre d'infractions peuvent être commises par l'intermédiaire des réseaux sociaux à l'encontre desquelles la LCEN prévoit un ensemble de dispositions.

S'agissant spécifiquement de la protection des mineurs, le code pénal s'est adapté afin de renforcer la protection des mineurs sur Internet. On peut distinguer deux types de situations :

- Les cas où des adultes approchent des mineurs.
- Les cas où des mineurs accèdent à des contenus destinés aux adultes.

Les cas où des mineurs sont approchés par des adultes (« pédopiégeage » ou « grooming ») :

Le pédopiégeage (ou « *grooming* » en anglais) consiste, pour un adulte, à solliciter un enfant n'ayant pas atteint la majorité sexuelle en vue de commettre un abus sexuel. Le plus souvent, cela consiste pour l'adulte à instaurer une relation de confiance avec l'enfant afin d'obtenir une faveur sexuelle. Ce type de comportement peut se dérouler sur les réseaux sociaux, en particulier lorsqu'il est possible de rester anonyme ou de dissimuler identité, et lorsque le service est accessible aux mineurs.

Ce type de pratique, ainsi que les infractions sexuelles spécifiquement prévues pour protéger les mineurs sont détaillées aux articles 227-21-1 à 227-28-3[41)] du code pénal.

Sans être exhaustif, les infractions suivantes sont les plus susceptibles d'être commise par l'intermédiaire d'un réseau social à l'issue d'une pratique de pédopiégeage :

- Faire des propositions sexuelles à un mineur de 15 ans ou moins en utilisant un moyen de communication électronique, la peine étant plus lourde si la sollicitation a été suivie d'une rencontre (Article 227-22-1)[42)].
- Inciter un mineur à commettre un acte de nature sexuelle (Article 227-22-2)[43)]).
- Filmer et/ou transmettre l'image d'un mineur en train de réaliser des actes sexuels (Article 227-23)[44)].
- Solliciter auprès d'un mineur des photos ou des vidéos à caractère sexuel le représentant (Article 227-23-1)[45)].

Le principal responsable de la commission de ces infractions est le pédocriminel qui utilise le réseau social comme intermédiaire pour contacter des mineurs. En principe, les opérateurs de réseaux sociaux ainsi que leurs salariés ne sont *a priori* pas les personnes concernées par ces infractions. Toutefois, leur responsabilité pourrait être engagée dans l'hypothèse où de multiples signalements concernant le profil d'un pédocriminel ne seraient pas traités avec diligence, ou en l'absence de signalement auprès des autorités compétentes.

En outre, pour renforcer la protection des enfants sur Internet, il est courant que les plateformes mettent en place volontairement des outils de détection et de signalement proactif de contenus portant sur des abus sexuels commis contre les mineurs et/ou sur de la sollicitation de mineur. Ces outils peuvent, par exemple, recourir à des technologies de comparaison d'images par rapport à une base d'images déjà existantes, correspondant à des contenus pédopornographiques, et recensées sous forme de hash. Ce système permet de détecter proactivement ce type de contenus afin de les retirer le plus rapidement possible et signaler leurs auteurs.

Ce type de technologies employées à des fins de lutte contre les contenus CSAM seront régies par le futur règlement CSAM, étudié à la section 3.3. Toutefois, la réglementation européenne a d'ores et déjà prévu un règlement spécifique pour encadrer l'usage de ces technologies dans les services de messageries, constituant une dérogation au principe du secret des correspondances édicté dans la directive 2002/58, visée à la section 4.3.

---

41) Articles 227-21-1 à 227-28-3 : https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006070719/LEGISCTA000043405084/#LEGISCTA000043405084

42) Article 227-22-1 : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000043409180

43) Article 227-22-2 : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000043405753

44) Article 227-23 : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000043409170

45) Article 227-23-1 : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000043405758

Ainsi, le règlement européen 2021/1232 du 14 juillet 2021 est un règlement temporaire qui permet aux fournisseurs de services de communications interpersonnelles non fondés sur la numérotation (tels que les services de messageries électroniques) de mettre en place des technologies automatisées de détection d'abus sexuels commis contre des enfants en ligne sur leurs services, dans les limites et selon les conditions prévues par le texte. Ce règlement est applicable jusqu'au 31 décembre 2023.

À défaut d'emploi de technologies de détection de façon volontaire, la lutte contre les contenus CSAM repose principalement sur le signalement par les autres utilisateurs. C'est pourquoi le règlement CSAM envisage de créer un cadre juridique complet et harmonisé (ce qui implique d'écarter le secret des correspondances lorsque la protection des mineurs sur les plateformes le nécessite), afin de lutter contre cette utilisation opaque des réseaux sociaux par les pédocriminels.

Sur le retrait des contenus à caractère pédopornographique, ceux-ci doivent être promptement retirés. Le projet de loi ***« Sécuriser et réguler l'espace numérique »*** vise à réduire les délais de réaction des réseaux sociaux : les contenus devront être retirés sous 24 heures suivant le signalement une autorité compétente (comme Pharos), sous peine d'un an de prison et de 250 000 euros d'amende, et jusqu'à 4 % du chiffre d'affaires mondial d'une personne morale.

Les hypothèses de complicité, de recel ou de négligence de la plateforme sont en pratique beaucoup plus difficiles à caractériser, et à ce jour, il n'existe pas de jurisprudence allant dans ce sens.

Les cas où des mineurs peuvent accéder à des contenus pour adultes :

Conformément à une jurisprudence constante consacrée par la loi du 30 juillet 2020, les plateformes offrant au public des contenus pornographiques ne peuvent plus se contenter de demander une déclaration de majorité à leurs utilisateurs ; ils doivent concrètement vérifier leur âge et bloquer l'accès à leurs services aux mineurs.

L'article 227-24[46)] du Code pénal sanctionne le simple fait que le mineur puisse accéder à ces services, y compris en fournissant une déclaration selon laquelle il a au moins 18 ans.

Dans ce cas, ce n'est pas la personne ayant publié les contenus dont la responsabilité peut être recherchée, mais plutôt la plateforme centralisant la diffusion de ces contenus. De plus, il s'agit d'une infraction *« formelle »*. Il importe peu qu'un mineur ait effectivement accédé à un contenu pornographique : il suffit qu'il ait la capacité de le faire, en l'absence d'un mécanisme valable de contrôle de l'âge, pour que la plateforme puisse être condamnée. Il est donc nécessaire pour les services de communication au public en ligne de prendre des mesures pour vérifier l'âge de leurs utilisateurs, lorsqu'ils diffusent des contenus pornographiques.

Le projet de loi ***« Sécuriser et réguler l'espace numérique »*** envisage de confier à l'Arcom l'élaboration d'un référentiel, pris après avis de la CNIL, déterminant les caractéristiques techniques pour ces solutions de vérification de l'âge. La solution privilégiée par la CNIL[47)] repose sur des technologies impliquant un tiers de confiance qui joue le rôle d'interface entre le site qui nécessite une information (la plateforme qui doit vérifier la majorité) et un site qui détient cette information (opérateur téléphonique par exemple).

---

46) Article 227-24 du Code pénal : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000044394218

47) https://www.cnil.fr/fr/verification-de-lage-en-ligne-trouver-lequilibre-entre-protection-des-mineurs-et-respect-de-la-vie

BIBLIOGRAPHIE ET RESSOURCES

Bibliographie et ressources

Par Public

Pour les mineurs

- Stop la violence : des enquêtes interactives pour apprendre à repérer et agir face au cyberharcèlement : https://www.stoplaviolence.net/ (Internet Sans Crainte)
- Les réseaux sociaux connaissent-ils la couleur de mon slip ? Données personnelles sur les réseaux sociaux : comment gérer son e-réputation ? : https://www.youtube.com/watch?v=G7qfSL2hYss (vidéo Internet Sans Crainte)
- Mon fil d'actu tourne-t-il en rond ? Comment fonctionne l'info sur les réseaux : https://www.youtube.com/watch?v=NNZ0ztngm-Q (vidéo Internet Sans Crainte)
- Fakes news ou comment s'en prémunir - vidéo pédagogique (Respect Zone) - https://www.youtube.com/watch?v=ZlytJLxqphM
- Comprendre ses droits quand on est cyberharcelé (Respect Zone) https://www.youtube.com/watch?v=epUncz0warc&t=151s
- Un jeu quiz pour tester ses connaissances, et s'informer sur les dangers éventuels et les bons usages numériques tout en s'amusant https://e-enfance.org/quiz/
- L'application 3018, pour les victimes ou témoins de cyberharcèlement : comment ça marche ? https://www.youtube.com/watch?v=fYW2YvEt9iI
- Les différentes formes de harcèlement sur les réseaux sociaux https://www.youtube.com/watch?v=9UwQqNuCh_M

Pour les parents

- Les écrans, les réseaux... et vos ados ! un guide interactif pour accompagner ses ados pas à pas dans leur vie numérique : https://www.internetsanscrainte.fr/ressources/les-ecrans-les-reseaux-et-vos-ados
- Le guide des paramètres indispensables des réseaux sociaux des ados : https://www.internetsanscrainte.fr/dossiers/reseaux-sociaux (Internet Sans Crainte)
- Comment expliquer à mon enfant qu'il n'a pas l'âge pour être sur un réseau social ? vidéo 2 min (Internet Sans Crainte) : https://www.youtube.com/watch?v=STIeYgckJbc
- Comment lui parler de ses amis en ligne ? vidéo 2 min (Internet Sans Crainte) : https://www.youtube.com/watch?v=on_Hvhn6AyU
- Comment aider mon ado à se déconnecter de la nuit ? vidéo 2 min : https://www.youtube.com/watch?v=ZHZLSIby7H4 (Internet Sans Crainte)

- Le compte Instagram sur la parentalité numérique d'Internet Sans Crainte : https://www.instagram.com/internetsanscrainte/?hl=fr
- La plateforme pour aider les parents à mettre en place une charte numérique à la maison (Internet Sans Crainte) : https://www.faminum.com/
- Comprendre ses droits quand on est cyberharcelé (Respect Zone) https://www.youtube.com/watch?v=epUncz0warc&t=151s
- Fake news ou comment s'en prémunir - vidéo pédagogique (Respect Zone) - https://www.youtube.com/watch?v=ZlytJLxqphM
- Reconnaître les signaux-faibles de mal-être chez les jeunes, application Respect Zone (https://www.respectzone.org/application-respectzone/)
- Jeu quiz à faire avec les enfants pour tester leurs connaissances, les sensibiliser et les informer sur les bons usages numériques tout en s'amusant https://e-enfance.org/quiz/
- Qu'est-ce que le cyberharcèlement et comment agir ? https://e-enfance.org/informer/cyber-harcelement/
- Site ressources pour les parents, les professionnels et les jeunes avec conseils et informations sur les risques dans l'univers numérique, les bonnes pratiques, l'accompagnement à la parentalité numérique etc. https://e-enfance.org/
- L'application 3018, pour les victimes ou témoins de cyberharcèlement : comment ça marche ? https://www.youtube.com/watch?v=fYW2YvEt9iI
- Comment amorcer une discussion sur le cyberharcèlement avec ses enfants ou adolescents ? https://www.youtube.com/watch?v=bYmAgrN1YUM&t=26s
- Les différentes formes de harcèlement sur les réseaux sociaux https://www.youtube.com/watch?v=9UwQqNuCh_M
- Pour comprendre le lexique sur Internet et sur les réseaux sociaux https://e-enfance.org/glossaire/

Pour les adultes et les éducateurs

- Présentation de la plateforme Seriously, permettant de lutte contre les discours de haine en ligne par le biais de l'argumentation : https://www.seriously.ong/
- https://yakamedia.cemea.asso.fr/univers/comprendre/numerique-media-et-education-citoyennete/seriously-anti-haine-lgbt
- https://yakamedia.cemea.asso.fr/univers/comprendre/numerique-media-et-education-citoyennete/seriously-anti-racisme
- https://yakamedia.cemea.asso.fr/univers/comprendre/numerique-media-et-education-citoyennete/seriously-antisemitisme

- Animation d'un débat avec la plateforme Seriously : https://yakamedia.cemea.asso.fr/univers/animer/activites-autour-des-medias-et-du-numerique/medias-de-linformation/mener-des-ateliers-debats-avec-seriously

- Lutte contre le cyberharcèlement : Fiche d'animation d'un ciné- débat sur le cybersexisme : https://yakamedia.cemea.asso.fr/univers/animer/activites-autour-des-medias-et-du-numerique/cinema/animer-un-cine-debat-sur-le-cyber-sexisme

- Une application pour sensibiliser les jeunes aux questions de cyberharcèlement par le biais de jeux de rôle, quizz et mises en situation : https://yakamedia.cemea.asso.fr/univers/agir/activites-autour-des-medias-et-du-numerique/medias-internet/sengager-contre-le-cyberharcelement-une-application-mobile

- Proposition de trois mini parcours pour aller vers un usage responsable et distancié des réseaux sociaux https://eduscol.education.fr/3481/ressources-pour-des-usages-responsables-sur-internet

- "Réseaux-sociaux où en êtes-vous?" https://yakamedia.cemea.asso.fr/univers/animer/activites-autour-des-medias-et-du-numerique/medias-internet/reseaux-sociaux-ou-en-etes-vous-testez-et-approfondissez-vos-connaissances

- "Comment gérez-vous vos publications sur les réseaux-sociaux ?" https://yakamedia.cemea.asso.fr/univers/animer/activites-autour-des-medias-et-du-numerique/medias-internet/comment-gerez-vous-vos-publications-sur-les-reseaux-numeriques-testez-et-approfondissez-vos

- "Information ou intox, comment faites-vous la différence ?" https://yakamedia.cemea.asso.fr/univers/animer/activites-autour-des-medias-et-du-numerique/medias-de-linformation/information-ou-infox-comment-faites-vous-la-difference-testez-et-approfondissez-vos-connaissances

- Comprendre ses droits quand on est cyber-harcelé (Respect Zone) https://www.youtube.com/watch?v=epUncz0warc&t=151s

- "Comment modérer les discours de haine en ligne ? " - MOOC gratuit 6h - (Respect Zone) https://www.facingfactsonline.eu/course/view.php?id=36

- Fakes news ou comment s'en prémunir - vidéo pédagogique (Respect Zone) - https://www.youtube.com/watch?v=ZlytJLxqphM

- Reconnaître les signaux-faibles de mal-être chez les jeunes, application Respect Zone (https://www.respectzone.org/application-respectzone/)

- Jeu quiz à faire avec les enfants pour tester leurs connaissances, les sensibiliser et les informer sur les bons usages numériques tout en s'amusant https://e-enfance.org/quiz/

- Un site ressources pour les parents, les professionnels et les jeunes avec conseils et informations sur les risques dans l'univers numérique, les bonnes pratiques, l'accompagnement à la parentalité numérique etc. https://e-enfance.org/

- Kit de sensibilisation : Supports de prévention, vidéos de prévention et affiches sur les dangers d'Internet https://e-enfance.org/nos-interventions/kit-sensibilisation/

- La valise pédagogique des super-héros du Net (ressources et outils pédagogiques pour sensibiliser et éduquer les enfants aux bons usages numériques) https://valisepedagogique.e-enfance.org/