

Point d'étape sur les moyens mis en œuvre par les plateformes en ligne pour lutter contre la diffusion de contenus haineux en France

Merci de répondre pour chacun des services de plateforme en ligne de l'opérateur dont la fréquentation est supérieure à 10 millions de visiteurs uniques mensuels en France.

Des éléments chiffrés sont demandés pour l'année 2022. L'opérateur peut, en sus, fournir des chiffres plus détaillés. Des chiffres portant sur les premiers mois de 2023 intéresseraient également l'Arcom.

L'opérateur est également invité à fournir toute autre information qu'il estimerait utile sur les tendances observées s'agissant de ces indicateurs.

Les informations fournies par l'opérateur seront rendues publiques, sauf informations confidentielles qui doivent être clairement identifiées comme telles et dont la confidentialité doit être justifiée.

Madame, Monsieur,

Nous vous remercions pour l'opportunité qui est donnée à Meta Platforms Ireland Limited (Meta Ireland) de pouvoir contribuer à ce point d'étape sur les moyens mis en œuvre pour lutter contre la diffusion de contenus haineux en France, conformément à l'article 6-4 de la LCEN. Comme vous le savez, Meta Ireland est actuellement engagée dans sa mise en conformité avec le règlement (UE) 2022/65 (le « **Règlement sur les services numériques** ») actualisant et renforçant le marché unique de services numériques de l'UE. Les réponses de Meta Ireland aux questions soulevées pour ses services de plateforme en ligne Instagram et Facebook sont présentées ci-dessous.

SIGNALEMENTS

- 1. Avez-vous rencontré des difficultés pour inclure les motifs de haine en ligne dans le dispositif de signalement mis à disposition des utilisateurs de votre service ? Merci de préciser.**

Avant la promulgation de l'article 6-4 de la LCEN, il était déjà possible pour les utilisateurs de signaler un contenu potentiellement en violation des Standards de la Communauté de Facebook et des Règles de la Communauté d'Instagram, qui établissent ce qu'il n'est pas autorisé de publier sur les plateformes de Meta. Plusieurs de ces types de violations recoupent la définition des contenus haineux mentionnée au premier alinéa de l'article 6-4 de la LCEN (« **Contenus Haineux** »).

Chaque trimestre, nous rendons compte de l'application de nos règles sur ce sujet, ainsi que d'une douzaine d'autres catégories, dans notre Rapport de transparence, accessible à ce lien :

<https://transparency.fb.com/data/community-standards-enforcement/hate-speech/facebook/>.

Depuis 2016, nous avons étoffé notre équipe en charge de la sécurité et de l'intégrité, représentant près de 40 000 personnes dans le monde. Elle traite notamment des enjeux de sûreté et de modération des contenus. Nous avons mis en place des outils d'intelligence artificielle et automatisés de pointe pour détecter et supprimer les contenus qui enfreignent nos politiques.

En ce qui concerne la lutte contre les discours haineux, par exemple, entre avril et juin 2022, nous avons supprimé 17,5 millions de contenus contraires à nos standards sur Facebook et Instagram dans le monde. Plus de 90% d'entre eux ont été identifiés et supprimés avant d'être signalés par un utilisateur.

En 2022, afin de se mettre en conformité avec l'article 6-4 de la LCEN, Meta Ireland a missionné une équipe spécialisée d'ingénieurs et des membres de l'équipe des opérations réglementaires (« *Scaled Regulatory Operations* ») afin de créer un formulaire de signalement légal, qui est disponible à la fois dans l'application pour les utilisateurs et dans nos Pages d'aide pour les auteurs de signalement qui ne sont pas connectés, permettant à toute personne (utilisateur et non-utilisateur) de signaler des contenus spécifiques qui peuvent potentiellement être considérés comme un contenu haineux, y compris les autorités judiciaires et administratives et les signaleurs de confiance qui peuvent s'identifier en tant que tels.

Le formulaire de signalement légal permet à toute personne d'identifier la catégorie de contenu haineux au nom de laquelle elle signale le contenu. Pour les utilisateurs connectés, ce formulaire de signalement est accessible via un lien disponible à côté de chaque contenu sur les plateformes Facebook/Instagram. Il est également possible d'accéder au formulaire via les Pages d'aide de chaque plateforme, sur une page ayant vocation à aider toute personne à comprendre comment soumettre des signalements légaux. Toute personne en France (utilisateur connecté et non connecté) peut accéder à cette page d'aide et signaler des contenus potentiellement en violation de l'article 6-4 de la LCEN.

Voici les liens vers les Pages d'aide concernées:

- <https://facebook.com/help/614105866347860?ref=cr>
- <https://facebook.com/help/instagram/633340764651750?ref=cr>

2. Parmi les signalements que vous avez reçus concernant la version française de votre service en 2022 et qui portaient sur un motif correspondant à la définition des contenus haineux au sens de l'article 6-4 la loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique, combien d'entre eux provenaient :

- 1) Des utilisateurs de la plateforme :**
- 2) Des signaleurs de confiance avec lesquels vous travaillez le cas échéant :**
- 3) Des autorités publiques (administratives ou judiciaires) :**

En 2022, Meta Ireland a reçu 12 424 signalements via le formulaire de signalement légal pour signaler des contenus haineux au sens de l'article 6-4 de la LCEN. Il s'agit du nombre cumulé de signalements reçus à la fois sur Facebook et sur Instagram. Les 12 424 rapports concernaient 10 179 contenus individuels.

Parmi ces signalements, il n'y avait aucun signalement provenant de signaleurs de confiance ou d'autorités publiques. Comme indiqué ci-dessus, le formulaire de signalement légal permet aux signaleurs de confiance ou aux autorités judiciaires ou administratives de s'auto-identifier comme tels. À ce jour, les seuls signaleurs qui se sont identifiés dans ces catégories étaient en réalité soit des particuliers, soit des personnes se livrant à du *spamming*.

EXERCICE DE LA MODÉRATION

3. Combien d'actions de modération portant sur des contenus correspondant à la définition des contenus haineux au sens de l'article 6-4 susmentionné avez-vous effectué en 2022, réparties par type ?

En 2022, 4807 contenus signalés via le formulaire de signalement légal ont été supprimés ou rendus inaccessibles en France. Parmi ceux-ci, 4788 ont été supprimés au niveau mondial pour violation de nos Standards de la Communauté et 19 ne violaient pas nos Standards de la Communauté mais ont été bloqués en France en raison d'une violation d'une disposition de l'article 6-4 de la LCEN. La répartition de ces 19 violations est la suivante :

- 10 pour « Incitation à la discrimination, à la haine ou à la violence à l'égard d'une personne ou d'un groupe de personnes » (article 24, alinéas 7 et 8 de la loi française sur la liberté de la presse) ;
- 5 pour « Apologie des crimes, des crimes de guerre, des crimes contre l'humanité, des crimes de réduction en esclavage ou d'exploitation d'une personne réduite en esclavage » (article 24, alinéa 5 de la loi française sur la liberté de la presse) ;
- 2 pour « diffusion de message à caractère violent, incitant au terrorisme, pornographique, y compris des images pornographiques impliquant un ou plusieurs animaux, ou de nature à porter gravement atteinte à la dignité humaine ou à inciter des mineurs à se livrer à des jeux les mettant physiquement en danger » (article 227-24 du Code pénal français) ; et
- 2 pour « Injure commise envers une personne ou un groupe de personnes » (article 33, alinéas 3 et 4 de la loi française sur la liberté de la presse).

Il convient également de noter que Meta a pris des mesures contre plus de 66 millions de contenus pour discours haineux dans le monde sur Facebook et Instagram en 2022 à la suite de violations des Standards de la Communauté. Cela comprend les contenus ayant fait l'objet d'une mesure après avoir été signalés par une personne, ainsi que les contenus qui ont été trouvés de manière automatique. Meta n'autorise pas les discours haineux sur Facebook et Instagram. Nous définissons les discours haineux comme des discours violents ou déshumanisants, des déclarations d'infériorité, des appels à l'exclusion ou à la ségrégation fondés sur des caractéristiques protégées ou des insultes. Ces caractéristiques comprennent la race, l'origine ethnique, l'origine nationale, l'appartenance religieuse, l'orientation sexuelle, la caste, le sexe, le genre, l'identité de genre et le handicap ou la maladie grave.

De plus amples informations sont disponibles dans notre rapport de transparence :

<https://transparency.fb.com/data/community-standards-enforcement/hate-speech/facebook/>

4. Combien de mesures avez-vous pris en 2022 à l'encontre d'utilisateurs abusant des outils de signalement, réparties par type ?

Nous avons reçu un petit nombre de signalements par le biais du formulaire de signalement légal fournissant des adresses électroniques qui n'appartiennent presque certainement pas au signaleur. En effet, ces signaleurs étaient situés en-dehors des territoires français et utilisaient un VPN pour accéder au formulaire. De plus, les adresses électroniques qu'ils renseignaient appartenaient à des personnalités (souvent politiques) étrangères pour signaler un contenu de mauvaise qualité, c'est-à-dire du contenu qui n'est manifestement pas illicite ou qui n'enfreint pas nos règles. Pour protéger le(s) véritable(s) propriétaire(s) de ces comptes de messagerie électronique contre le spam, nous

avons examiné les signalements et pris des mesures en fonction des résultats de notre examen, mais aucune notification des résultats de l'examen n'a été envoyée aux adresses électroniques fournies.

5. Quelle est la part des actions visées aux questions 3 d'une part, 4 d'autre part, prises après décision humaine (i.e. actions qui ne résultent pas uniquement d'un processus automatique) ?

Toutes les mesures prises en vertu du formulaire de signalement légal ont été prises par des humains car ce formulaire est directement géré par notre équipe de modération dédiée aux contenus illicites. L'examen humain a été possible dans ces circonstances étant donné le volume relativement faible de signalements. Ce ne sera pas toujours le cas en ce qui concerne les canaux de signalement à volume élevé.

6. À la suite des actions visées aux questions 3 d'une part, et 4 d'autre part, quel est le taux de recours internes provenant des utilisateurs accédant depuis la France à vos services, et quels ont été les résultats de ceux-ci (pourcentage de confirmation de la décision initiale, pourcentage d'infirmerie de la décision initiale) ?

Meta Ireland offre la possibilité de demander un nouvel examen des décisions prises relativement à un contenu signalé comme contenu haineux, soumis par le biais du formulaire de signalement légal. 625 demandes de réexamen ont été reçues suite à un signalement via ce formulaire en 2022. En raison de la faible qualité et de l'absence de motifs juridiques des demandes de réexamen qui ont été reçues (y compris un certain nombre de spam), aucune de nos décisions initiales n'a dû être infirmée.

7. Quels sont le nombre, la localisation et la ou les langues de travail des personnes affectées au traitement des signalements et des recours provenant des utilisateurs de la version française de votre service en matière de haine en ligne (donner des informations valant au mois de décembre 2022) ?

Les signalements au titre de l'article 6-4 de la LCEN sont examinés en deux étapes par des équipes de professionnels et de juristes formés, qui couvrent à la fois les plateformes Facebook et Instagram.

Tout d'abord, le contenu signalé via le formulaire de signalement légal est examiné par les membres de notre équipe en charge des Opérations mondiales (*Global Operations team*). Cette équipe est composée de salariés à temps plein et de membres d'entreprises partenaires avec lesquelles nous travaillons dans le monde. Chaque rapport est examiné par un membre individuel de cette équipe afin de déterminer si le contenu signalé enfreint les Standards /Règles de la Communauté de Meta Ireland (par opposition à l'examen du contenu pour une éventuelle illicéité, qui, comme indiqué ci-dessous, est traitée par des équipes distinctes). S'il est constaté que le contenu enfreint les Standards /Règles de la Communauté, le contenu est supprimé au niveau mondial.

Par ailleurs, si le contenu signalé ne viole pas nos Standards /Règles de la Communauté, nous examinons sa licéité sur la base des informations fournies dans le signalement. Plus précisément, nous déterminons si le contenu signalé enfreint les dispositions pertinentes de l'article 6-4 de la LCEN. Si le contenu signalé est considéré comme illicite en vertu de l'article 6-4 de la LCEN, nous désactivons l'accès à ce contenu en France. Nous communiquons également avec le signaleur pour lui fournir des mises à jour de son signalement, demander des informations supplémentaires si nécessaire et l'informer de notre décision une fois que nous avons terminé notre examen.

Chaque signalement est examiné par un membre individuel de notre équipe de sous-traitants, qui est chargé d'identifier et de bloquer les contenus manifestement illicites. Tout cela se fait dans le respect des directives élaborées par nos juristes internes et nos conseils juridiques externes. Si le signalement

nécessite une analyse plus approfondie, il est mis en file d'attente pour examen par l'un des salariés de l'équipe en charge des opérations réglementaires (*Scaled Regulatory Operations*). S'il y a encore un doute sur la licéité du contenu signalé, le signalement est ensuite transmis à nos juristes internes.

Les langues couvertes par ces équipes incluent le français, l'anglais, l'allemand et le turc. Pour des raisons de sécurité et de confidentialité, nous ne pouvons pas divulguer publiquement les informations relatives à la localisation et au nombre de salariés travaillant sur la modération de contenus et nous avons donc partagé ces informations de manière confidentielle avec l'ARCOM.

Le flux de travail concernant l'article 6-4 de la LCEN est soumis à un processus de contrôle qualité pour les sous-traitants et les salariés afin d'assurer un traitement cohérent et de qualité et identifier tout besoin supplémentaire de formation ou d'orientation.

Il est important de noter que les agents dédiés à l'examen du flux de travail concernant l'article 6-4 de la LCEN ne représentent qu'un faible pourcentage des effectifs mondiaux en charge de la modération de tous les signalements de violations des Standards de la Communauté/Règles de la Communauté.

COOPÉRATION AVEC LES AUTORITÉS PUBLIQUES

8. En 2022, combien de suspicions d'infraction ont fait l'objet d'une transmission de votre part aux autorités publiques compétentes, en particulier au ministère public, et pour quels motifs (donner, le cas échéant, le seul motif principal) :

Depuis 2013, nous publions des rapports de transparence semestriels concernant la nature et l'étendue des demandes gouvernementales que nous recevons pour les données des utilisateurs. Ces rapports donnent à notre communauté une visibilité sur le nombre de demandes que nous recevons et sur la manière dont nous appliquons nos politiques et répondons aux demandes de données. Entre janvier et juin 2022, Meta Ireland a reçu 12 769 demandes de données du gouvernement et des autorités publiques françaises. Parmi ces demandes, 6 702 étaient des demandes dans le cadre de procédures juridiques et 6 067 étaient des demandes de divulgation d'urgence. Des données ont été produites et transmises dans près de 84.47 % des cas. À l'échelle mondiale, Meta a reçu 237 414 demandes de la part de gouvernements.

Entre juillet et décembre 2022, Meta Ireland a reçu 12 682 demandes de données du gouvernement et des autorités françaises. Parmi ces demandes, 5818 étaient des demandes dans le cadre de procédures juridiques et 6871 des demandes de divulgation d'urgence. Des données ont été produites dans 85,3 % des cas. Au niveau mondial, Meta a reçu 238 388 demandes gouvernementales.

De plus amples renseignements sont disponibles dans notre rapport de transparence :

<https://transparency.fb.com/data/government-data-requests/country/FR/>.

9. Quels procédures et moyens humains et technologiques avez-vous mis en œuvre pour répondre aux autorités administratives ou judiciaire dans les meilleurs délais ?

Meta Platforms Ireland s'efforce d'assurer la sécurité de sa communauté, en ligne comme hors ligne, y compris en répondant aux demandes d'informations des autorités gouvernementales. Meta reçoit des demandes de données d'utilisateurs de la part d'organismes gouvernementaux du monde entier.

Nous partageons des détails sur ces demandes et nos réponses dans notre Rapport relatif aux demandes gouvernementales de données utilisateur (« *Government Requests for User Data Report* »). Meta exige que les organismes gouvernementaux qui font des demandes se conforment à la loi applicable, à nos politiques (y compris notre Politique de confidentialité) et, dans certains cas, qu'ils aient recours à la procédure mise en place par un traité d'entraide judiciaire.

Nous disposons d'une équipe en charge des réponses aux demandes des autorités (« *Law Enforcement Response Team* », LERT) dédiée et formée, qui examine et évalue chaque demande d'autorités publiques pour des données utilisateur individuellement, que la demande ait été soumise en cas d'urgence ou dans le cadre d'un processus juridique initié par les autorités gouvernementales. Cette équipe s'assure que toutes les demandes sont conformes à la loi applicable et à nos politiques, y compris la Politique de confidentialité de Meta.

Meta a publié des ressources spécifiques et dédiées à ce sujet – y compris le portail « [Informations pour les services de police](#) » et le centre de transparence. Ensemble, ces informations publiquement accessibles expliquent les politiques et procédures de Meta en ce qui concerne les demandes d'information d'autorités gouvernementales. Notre portail « *Informations pour les services de police* » fournit des directives opérationnelles pour les services de police qui cherchent à obtenir des données auprès de Meta Platforms Ireland Limited. Meta dispose également d'un portail « *Law Enforcement Online Request System* » permettant la soumission, le suivi et le traitement des demandes des représentants des autorités de police.

PRÉPARATION DE LA MISE EN ŒUVRE DU RSN

10. Quelles sont les actions menées par votre entreprise pour préparer le service à la mise en application du règlement européen sur les services numériques (RSN) ?

Nous soutenons pleinement l'ambition du RSN d'actualiser et de renforcer le marché unique des services numériques de l'UE et nous continuons à travailler avec les régulateurs européens pour nous conformer aux nouvelles règles.

Tout au long des discussions, nous avons salué le besoin de nouvelles règles qui fournissent plus de clarté, de responsabilité et de transparence sur les rôles et responsabilités des plateformes en ligne. Un tel cadre nous permet de maintenir la sécurité sur Internet, de renforcer la cohérence et la fiabilité et de créer un environnement plus sain pour les adultes et les jeunes, tout en préservant la liberté d'expression et d'autres droits fondamentaux. Toutefois, nous devons nous assurer que la loi soit adaptée à tout le monde et tienne compte de la nature unique des différents types de services en ligne.

Nous travaillons actuellement à la mise en conformité. Notre objectif reste de fournir aux utilisateurs la meilleure expérience possible de nos produits.

Compte tenu des investissements déjà substantiels de Meta dans les domaines de l'intégrité et de la sécurité, et de son approche globale sur un large éventail de questions relatives, y compris la transparence, et le signalement de contenus, nous avons pour objectif, dans la mesure du possible, d'adapter les systèmes et processus existants de Meta pour nous conformer aux nouvelles exigences réglementaires.

Pour bon nombre des obligations du RSN, Meta dispose de systèmes ou de processus qui peuvent correspondre en partie ou en totalité, avec les obligations considérées. Dans ces cas, nous visons à

utiliser nos mécanismes existants comme base de référence tout en déterminant si des changements sont nécessaires pour répondre aux obligations du RSN. Lorsque les systèmes et les processus existants ne sont pas suffisants, nous travaillons à développer des innovations pour répondre aux exigences spécifiques.

11. Avez-vous fait face à des difficultés ou problématiques (de tous ordres : compréhension, méthode, calcul des indicateurs, modification du produit, etc.) dans la mise en œuvre de l'article 6-4 susmentionné qui seraient susceptibles d'être de nouveau rencontrées dans le cadre de celle du RSN, ou qui vous auraient permis d'anticiper ce dernier ? Lesquelles ?

Les signalements sous forme de spam soumis via nos formulaires d'appel représentent un problème que nous avons rencontré lors de l'ouverture de ce nouveau canal de signalement légal en France. Nous nous sommes servis de cette expérience pour mieux nous préparer aux appels au titre du RSN, car nous travaillons actuellement au développement d'une nouvelle fonctionnalité pour bloquer les signaleurs de spam et les empêcher d'abuser de nos formulaires de contact en soumettant des demandes de réexamen sans signalement initial. Cette nouvelle mesure nous aidera à réduire les volumes de demandes d'appels non valides en vertu du RSN.

Nous souhaitons également noter que le processus de signalement légal sera maintenu pour faciliter le signalement de contenus illicites pour les utilisateurs français dans le cadre du RSN.

12. Travaillez-vous avec des tiers établis en France que vous reconnaissez comme signaleurs de confiance en matière de haine en ligne ? Si oui, lesquels ? Le cas échéant, pour quelles raisons avez-vous choisi de collaborer avec eux ?

Nous avons mis en place un système permettant à certaines autorités gouvernementales et organisations de la société civile de nombreux États membres de nous signaler des contenus. Les organisations de la société civile sont des partenaires essentiels afin de permettre à Meta de mieux comprendre l'impact de nos plateformes et le contexte des diverses communautés dans lesquelles nous exerçons nos activités dans le monde entier. Nous avons mis en place un programme de partenariat de confiance pour favoriser ce partenariat avec la société civile, renforcer la capacité de surveillance des réseaux sociaux par les organisations locales et améliorer nos politiques, nos processus de mise en application et nos produits pour contribuer à assurer la sécurité des utilisateurs sur nos plateformes. Notre réseau de partenaires de confiance comprend plus de 400 organisations non gouvernementales, agences humanitaires, défenseurs des droits de l'homme et chercheurs de 113 pays.

Nous travaillons en partenariat avec des organisations expertes qui représentent les voix et les expériences des utilisateurs à risque du monde entier et sont à même de soulever des questions et des préoccupations concernant le contenu sur Facebook et Instagram pour (i) traiter les tendances relatives aux contenus problématiques et prévenir les préjudices, (ii) favoriser la sécurité et la sûreté en ligne ; et (iii) aider à l'élaboration de politiques efficaces et transparentes.

En plus de signaler des contenus, les partenaires de confiance fournissent des commentaires cruciaux sur nos Politiques en matière de contenus et leur mise en application afin de garantir la sécurité des utilisateurs. L'expertise de nos partenaires de confiance et leur expertise régionale nous aident à renforcer nos politiques en nous permettant de prendre en compte un éventail de perspectives pour

éclairer nos efforts de modération de contenus. Ces consultations nous permettent d'élaborer des politiques sur mesure qui contribuent à assurer la sécurité de notre communauté en temps de crise.

Meta Ireland a travaillé au fil des ans avec 17 partenaires de confiance en France qui sont membres de notre programme de Partenaires de confiance. Parmi ceux-ci, SOS Homophobie, le CRIF, la LICRA ou encore e-Enfance ont été ajoutés à ce programme au fil des ans. Ils ont tous été soigneusement sélectionnés pour leur expertise dans la lutte contre la haine ainsi que l'utilisation des technologies numériques. En général, lors de la sélection de nos partenaires de confiance, nous recherchons des organisations qui ont de l'expérience dans la surveillance des réseaux sociaux, qui souhaitent en savoir plus sur nos Politiques en matière de contenus, qui démontrent leur engagement à assurer la sécurité des communautés en ligne et qui représentent des groupes marginalisés qui sont touchés de manière disproportionnée par les contenus préjudiciables.

Comme vous le savez, l'application du RSN à nos services sera principalement assurée par la Commission européenne (pour les très grandes plateformes en ligne) et par la *Irish Media Commission* (pour les autres services qui ne sont pas qualifiés de très grandes plateforme). Toutefois, les États membres conserveront certains pouvoirs formels.

Toutefois, étant donné que chaque État membre désignera ses propres signaleurs de confiance et qu'il existe peu d'indications sur les critères que les États membres doivent utiliser à cet effet, il nous est difficile d'estimer le volume de rapports que nous pourrions recevoir, et par quels canaux nous pourrions les recevoir. Cela représente un certain défi lorsque nous essayons d'anticiper l'affectation adéquate des ressources.

Lorsque nous examinons la manière de mettre en œuvre cette disposition d'un point de vue national, il est essentiel que le nombre de signaleurs de confiance reste dans une limite raisonnable, conformément à ce que le RSN indique également, afin que les contenus signalés puissent être classés par ordre de priorité. Dans le cas contraire, la valeur des signaleurs de confiance serait réduite et leur impact moins pertinent.

Les considérants du RSN indiquent que ces signalements pourraient être soumis par différents canaux, qui seront également difficiles à trier et à gérer, y compris à des fins de transparence. Il serait utile de permettre aux entreprises d'établir un point de référence unique pour les contenus signalés par les signaleurs de confiance, et de faciliter un processus de familiarisation à nos systèmes, politiques et canaux de signalement existants.