



Oversight Board

# 2021

RAPPORT  
ANNUEL 2021  
DU CONSEIL DE  
SURVEILLANCE



# Sommaire

Préface	2
Avant-propos des Coprésidents	4
Synthèse analytique	6
Application des normes internationales des droits de l’homme à la modération du contenu à une échelle mondiale : Article 19	9
Comment le Conseil tient compte des appels d’utilisateurs	10
Chronologie des événements clés	12

## Sélection des cas

Vue d’ensemble	14
Cas soumis au Conseil	17
Cas pris en considération par le Comité de sélection des cas	21

## Décisions

Vue d’ensemble	22
Tableau : décisions publiées en 2021	25
Décision phare : premières décisions	28
Décision phare : décision de suspension du compte de l’ancien président des États-Unis, Donald Trump	33
Décision phare : garantir le respect de la liberté d’expression	37
Décision phare : protéger les utilisateurs des contenus préjudiciables	40
Normes internationales des droits de l’homme dans le processus décisionnel du Conseil	45
Questions posées par le Conseil à Meta dans le cadre de nos décisions	51

Commentaires du public	54
------------------------	----

## Recommandations

Vue d’ensemble	57
Comment Meta a répondu à nos recommandations	60
Des engagements à l’action : obtenir des résultats pour les utilisateurs	63
Tenir Meta pour responsable du cross-check	66
Engagement et portée	68

## À suivre

À partir de 2022	70
------------------	----

## Annexe

Comment Meta a réagi à nos recommandations et les a appliquées	72
--	----



# Préface

Le Conseil de surveillance a été créé pour améliorer la façon dont Meta traite les individus et les communautés du monde entier en appliquant les normes de contenu publié sur Facebook et Instagram, et ce de manière à faire valoir la liberté d'expression de chacun ainsi que les normes pertinentes des droits de l'homme. Pour ce faire, nous proposons un contrôle indépendant de la modération du contenu de Meta, en prenant des décisions exécutoires sur les problématiques de contenu les plus complexes. Nous offrons des recommandations de politique qui incitent Meta à améliorer ses règles et à agir de façon raisonnée, transparente et en traitant équitablement tous ses utilisateurs.

Depuis 2018, la communauté Facebook et Instagram n'a cessé de croître pour compter désormais plus de 2 milliards d'utilisateurs. Connecter les individus les uns aux autres a généré des avantages considérables : les rapprocher des êtres qui leur sont chers, renforcer les communautés existantes et en créer des nouvelles. Cependant, à côté de ces avantages, se dressent d'énormes obstacles pour la société. Des contenus représentent une menace pour la sécurité et les libertés individuelles. Certains acteurs diffusent de fausses informations qui mettent en danger la démocratie et la société. Les discours incitant à la haine et le contenu faisant la promotion de ce que Meta désigne sous l'appellation « personnes et organisations dangereuses » sont susceptibles de diviser les communautés et de nuire à la valeur des réseaux sociaux dont le but principal consiste à rassembler les personnes.

En novembre 2018, dans le but d'entamer une discussion sur la procédure à suivre pour créer un système de gouvernance en charge de la modération du contenu à l'échelle internationale, Meta a présenté sa vision d'une nouvelle approche permettant aux personnes de faire appel des décisions relatives au contenu auprès d'un organe indépendant dont les décisions seraient exécutoires. Créer une entité indépendante, juridiquement et financièrement distincte de l'entreprise, traduit notre conviction toujours plus profonde que Meta ne devrait pas prendre autant de décisions relatives à l'expression et à la sécurité en ligne par elle-même.

Le Conseil de surveillance est né de cette idée : promouvoir les droits et les intérêts des utilisateurs en créant de la transparence et en offrant une plus grande responsabilité, cohérence et équité concernant l'approche de Meta en matière de décisions relatives au contenu.

En septembre 2019, une charte a été publiée ; elle définit l'indépendance institutionnelle, les principes fondateurs et l'objectif du Conseil. La charte donne au Conseil la capacité de prendre des décisions exécutoires pour autoriser ou interdire certains contenus sur Facebook et Instagram. Ces décisions doivent ensuite être implémentées par Meta dans les 7 jours qui suivent leur publication. Elle a également permis au Conseil de publier des recommandations non contraignantes dont le but consiste à améliorer

la manière dont Meta traite les personnes et les communautés dans le monde entier. En parallèle à la charte, un ensemble de statuts définissant les procédures opérationnelles du Conseil a été publié.

Le Conseil de surveillance est constitué de trois éléments interdépendants : les Membres du Conseil, la Fiducie et l'Administration. Les fiduciaires sont responsables de la protection de l'indépendance du jugement du Conseil et assurent le bon fonctionnement de celui-ci. Les Membres du Conseil, parmi lesquels figurent quatre Coprésidents, sélectionnent des cas et formulent des décisions les concernant, font des recommandations à Meta et dirigent nos travaux d'implémentation. L'Administration est constituée d'une équipe d'employés à temps plein qui assistent les Membres du Conseil dans leurs missions.

Dirigeants provenant d'horizons variés, nos Membres du Conseil sont expérimentés en matière de problématiques complexes, dont celles qui relèvent des droits de l'homme. Parmi eux figurent des universitaires, des dirigeants issus de la société civile, d'anciens juges ainsi que des titulaires de mandats des Nations Unies et d'organismes régionaux de protection des droits de l'homme. Nos Membres vivent dans 27 pays et parlent au moins 29 langues. La parité hommes-femmes des Membres du Conseil est parfaitement respectée. Il en va de même pour la parité hommes-femmes au sein de la Fiducie. Le personnel de l'Administration, même si peu nombreux, parle également 28 langues, ce qui renforce l'approche globale du Conseil. Près de 60 % de nos employés sont des femmes, contre 40 % d'hommes.

Conformément aux engagements que nous avons pris en matière de transparence et de responsabilité, nous avons publié plusieurs rapports de transparence trimestriels. Désormais, avec la publication de ce Rapport annuel, nous mettons publiquement à disposition : une synthèse complète des cas soumis au Conseil, une synthèse des décisions et des recommandations du Conseil, ainsi qu'une vue d'ensemble des commentaires publics. Enfin, et pour la première fois, ce Rapport comprend une analyse de la manière dont les décisions du Conseil ont tenu compte et respecté les normes internationales des droits de l'homme qui sont en cause dans plusieurs cas représentatifs.

# Avant-propos des Coprésidents



Catalina Botero-Marino, Jamal Greene, Michael McConnell, Helle Thorning-Schmidt  
*Coprésidents du Conseil de surveillance*

L'idée à l'origine du Conseil était simple à verbaliser, mais complexe à mettre en place : les sociétés de réseaux sociaux ne doivent pas prendre par elles-mêmes des décisions déterminantes sur la modération du contenu. Lors de notre première année d'activité, nous avons commencé à concrétiser cette idée.

Nous avons créé un système d'appel indépendant accessible aux milliards d'utilisateurs du monde entier. Nous avons mis en place un processus de commentaires publics pour donner à chacun la possibilité de s'exprimer dans le cadre de nos prises de décisions. Nous avons appris, en nous appuyant sur nos différentes nationalités, nos différents parcours et nos différents points de vue, à délibérer de cas qui ne donnent pas lieu à des réponses faciles. Dans chaque cas, nous avons étudié le contenu qui faisait l'objet du litige, conformément au Standard de la communauté Facebook pertinent ou à la règle de la communauté Instagram associée, ainsi qu'à la lumière des valeurs de Meta. Ce faisant, nous avons examiné si ces principes avaient été appliqués de manière appropriée et avec cohérence, si les utilisateurs ont fait l'objet d'une notification et d'une procédure adaptées, et si les standards et les règles sont compatibles avec les

normes internationales des droits de l'homme que Meta s'est elle-même engagée à respecter.

Le principe fondateur du Conseil ? La liberté d'expression est un élément essentiel de la société démocratique et doit être respectée pour protéger les droits de l'homme. Notre engagement pour soutenir la liberté d'expression est cohérent avec le Pacte international relatif aux droits civils et politiques, qui stipule que « Toute personne a droit à la liberté d'expression ». Toutes nos décisions reflètent le fait que les normes internationales des droits de l'homme sont une source importante d'autorité pour le Conseil.

Le premier Rapport annuel, qui couvre la période allant d'octobre 2020 (date à partir de laquelle nous avons commencé à accepter les appels) à décembre 2021, décrit les progrès réalisés concernant la manière dont Meta traite les utilisateurs et les autres groupes de personnes concernés à travers le monde. L'émergence d'une pandémie mondiale a bien évidemment compliqué la création d'une institution préparée pour ces tâches.

Visiblement, il y avait une forte demande sans cesse croissante parmi les utilisateurs de Facebook et

d'Instagram pour une procédure d'appel auprès d'une organisation indépendante de Meta concernant les décisions de modération du contenu prises par Meta. Les utilisateurs ont soumis **plus d'un million d'appels** au Conseil au cours de cette période. La grande majorité des appels consistaient à restaurer le contenu publié sur Facebook ou Instagram concernant des publications supposées enfreindre les règles de Meta en matière de harcèlement, de discours incitant à la haine ou de violence et d'incitation. En avril 2021, les utilisateurs ont également obtenu la possibilité de faire appel auprès du Conseil de surveillance concernant le contenu qu'ils souhaitaient voir supprimé de Facebook ou d'Instagram.

Nous **avons formulé des décisions avec des explications publiques complètes sur 20 cas importants** en 2021, sur des problématiques allant du discours incitant à la haine aux fausses informations relatives à la COVID-19. Ce faisant, nous avons adopté une approche basée sur les droits de l'homme pour analyser des décisions de modération du contenu et **reçu près de 10 000 commentaires publics** qui ont contribué à formuler nos premiers jugements. Dans de nombreux cas, le travail du Conseil a généré une décision volontaire de l'entreprise d'annuler les décisions préjudiciables de modération du contenu.

Nous avons également fait **86 recommandations auprès de Meta** afin que l'entreprise gagne en transparence concernant ses politiques. Les réponses de Meta à nos décisions et recommandations en matière de politique commencent à améliorer sa façon de tenir compte des utilisateurs. Meta donne désormais aux personnes qui utilisent Facebook en anglais et qui enfreignent ses règles sur les discours incitant à la haine plus de détails sur ce qu'elles ont fait de mal. L'entreprise diffuse de nouveaux messages dans certaines régions indiquant aux personnes si l'automatisation ou l'examen manuel a entraîné la suppression de leur contenu publié. En outre, elle s'est engagée à fournir de nouvelles informations sur les requêtes gouvernementales et son intérêt médiatique

dans ses rapports sur la transparence. Meta s'est également engagée à traduire ses Standards de la communauté en plusieurs langues parlées en Inde. Autrement dit, à l'issue de ce processus de traduction, plus de 400 millions d'utilisateurs en plus seront en mesure de lire les règles de Facebook dans leur langue d'origine.

Alors que Meta s'est engagée à implémenter la plupart des recommandations que nous avons effectuées en 2021, notre prochaine tâche consiste à veiller à ce que l'entreprise concrétise ses promesses en actions pour optimiser l'expérience des utilisateurs de Facebook et d'Instagram. Ainsi, ce rapport applique une nouvelle approche pilotée par les données pour suivre la manière dont l'entreprise implémente chacune de nos recommandations. Nous sommes également en quête de nouvelles données de Meta pour mieux comprendre l'impact précis de nos propositions sur les utilisateurs.

À l'issue de notre première année de prises de décision, le Conseil a commencé à poser les bases d'un projet qui peut se révéler un véritable moteur du changement au sein de Meta : aider l'entreprise à mieux servir les personnes et les communautés concernées par la portée extraordinaire des plateformes de réseaux sociaux de l'entreprise. En 2022, forts de cet excellent départ sur lequel nous appuyer, nous ajouterons des nouveaux Membres au Conseil, développerons notre mission dans de nouvelles régions et continuerons à accompagner Meta sur la bonne voie. Nous sommes convaincus que les premiers succès du Conseil de surveillance dans l'établissement des responsabilités de Meta prouvent la viabilité du Conseil et offrent un cadre d'autoréglementation pour développer et améliorer nos opérations à venir. De cette façon, le Conseil fera partie d'un effort collectif des entreprises, gouvernements, universitaires et personnes issues de la société civile pour constituer un meilleur avenir numérique, plus sûr, dont chacun pourra tirer parti à l'échelle mondiale.

# Synthèse analytique

**1,1m+**  
de cas soumis au  
Conseil par les  
utilisateurs et Meta.

## PLUS DE LA MOITIÉ

de nos décisions étaient  
associées aux Pays du Sud.



Les décisions de modération  
du contenu de Meta ont été  
analysées selon les

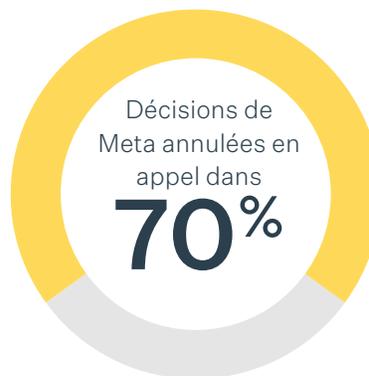
**normes  
internationales des  
droits de l'homme**

Plus de   
**8 appels d'utilisateur sur 10**

sollicitant une restauration de contenu  
concernaient les règles de Meta sur le  
harcèlement, les discours incitant à la haine  
ou la violence et l'incitation.

**20** Décisions  
publiées,

allant de fausses informations  
sur la COVID aux discours  
incitant à la haine.



des cas –  
annulation de ses  
décisions de modération  
du contenu à 14 reprises  
et confirmation de ses  
décisions à 6 reprises. .

Près de  
**10 000**



commentaires publics de personnes et organisations du monde entier ont contribué à formuler nos décisions.

**86** recommandations effectuées auprès de Meta.

Parmi les recommandations du Conseil, Meta s'est engagée à en tenir certaines :



**Fournir plus d'informations** aux utilisateurs lors de la suppression de publications dans lesquelles figurent un discours incitant à la haine ;



Fournir davantage **de rapports sur les requêtes gouvernementales**



Traduire ses règles dans **des langues parlées par +400 millions d'individus**



Adopter un nouveau **Protocole de politique de crise** pour régir sa réponse aux situations de crise



Diffuser de nouveaux messages dans certaines régions indiquant aux utilisateurs si **l'automatisation** ou **l'examen manuel** est à l'origine de la suppression de leur contenu



Ce rapport opte pour une **approche inédite pilotée par les données** pour suivre l'implémentation et veiller à ce que Meta honore ses engagements pris vis-à-vis des recommandations.



de nos 86 recommandations, Meta a procédé à leur implémentation ou indiqué un progrès en cours, les recommandations sur la **transparence** étant les plus susceptibles d'être concernées par l'une de ces deux catégories.

**NOS AMBITIONS POUR 2022, et au-delà, sont de :**

Discuter avec Meta du **développement de notre portée**, dont la révision des appels d'utilisateurs déposés contre ses décisions dans des domaines tels que les groupes et les comptes ;



**Développer la portée des parties prenantes** en Asie, Amérique Latine, au Moyen-Orient et en Afrique.

## Comment le Conseil définit les responsabilités de Meta

Le Conseil est encouragé par les tendances de la première année dans son engagement auprès de Meta, mais cette dernière doit améliorer de toute urgence sa transparence. Nous sommes ravis d'indiquer que l'entreprise a pleinement répondu à ses engagements sur les décisions, a accepté d'implémenter plus de la moitié des recommandations de politique du Conseil et répond de plus en plus aux questions du Conseil.

Néanmoins, le Conseil continue de faire face à de grandes préoccupations, notamment concernant la transparence de Meta et la fourniture d'informations concernant certains cas et certaines recommandations de politique.

### D'octobre 2020 à décembre 2021, Meta...

- ✔ a implémenté 100 % de nos décisions et s'est engagée à implémenter la plupart de nos recommandations ;
- ✔ a répondu à une majeure partie des questions posées dans le cadre de nos décisions avant la fin 2021, atteignant 94 % au T4 2021 ;
- ✔ a prouvé une amélioration de sa façon de donner suite à nos recommandations, la part des réponses de Meta jugées « complètes » ou « assez complètes » s'étant accrue à chaque trimestre de 2021 ;
- ✘ ne s'est pas montrée entièrement coopérative en ce qui concerne les informations transmises au Conseil à propos de son système cross-check ;
- ✘ a égaré une exception de politique importante et ne l'a pas appliquée pendant trois ans, comme décrit dans notre décision relative à l'isolement d'Öcalan ;
- ✘ Q a reconnu que dans 51 cas sur les 130 sélectionnés par le Conseil, sa décision d'origine était infondée.

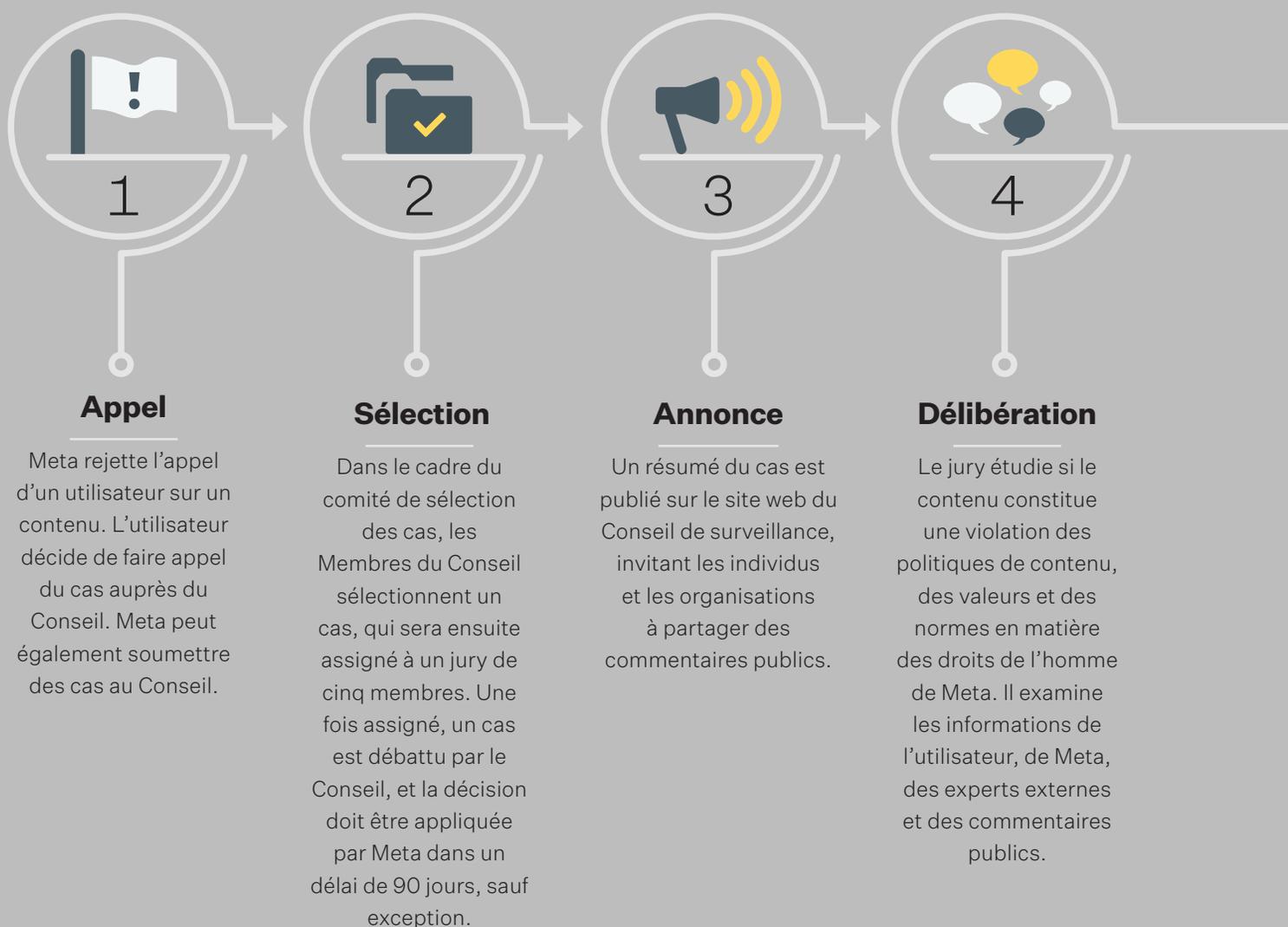


# Application des normes internationales des droits de l'homme à la modération du contenu à une échelle mondiale : Article 19

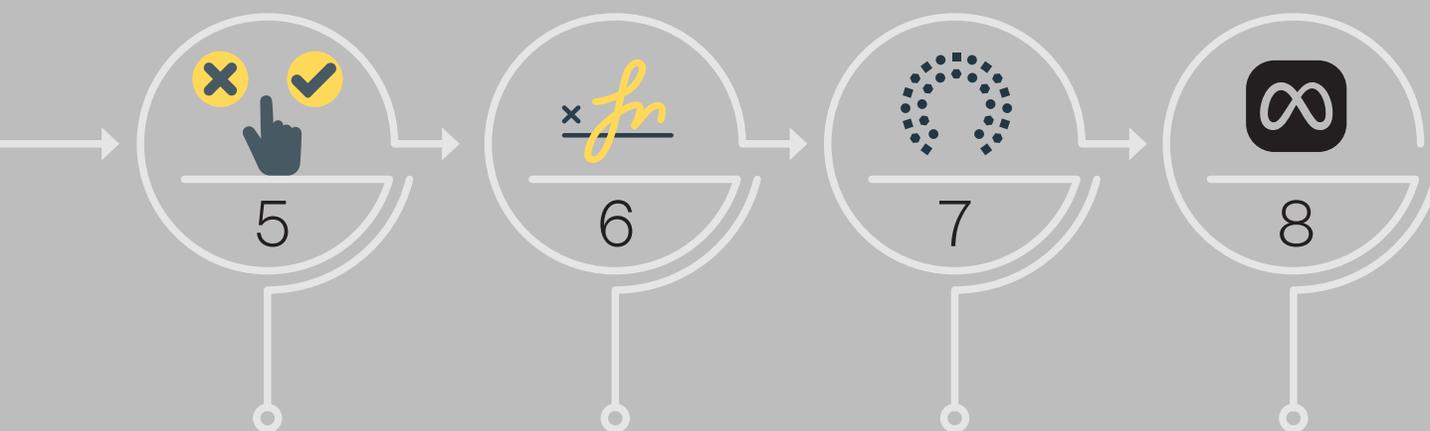
Un thème déterminant de la mission du Conseil en 2021 réside en la conviction que Meta prendra des décisions de modération du contenu de façon plus juste et raisonnée si elle base ces décisions sur les normes internationales des droits de l'homme qu'elle s'est elle-même engagée à appliquer. À cette fin, la charte du Conseil de surveillance définit que nous devons « accorder une attention particulière à l'impact de la suppression de contenu au regard des normes relatives aux droits de l'homme protégeant la liberté d'expression ». Ces normes incluent l'article 19 du Pacte international relatif aux droits civils et politiques (PIDCP), qui stipule que tandis que « toute personne a droit à la liberté d'expression... l'exercice de [ce droit] peut... être soumis à certaines restrictions, qui doivent toutefois être expressément fixées par la loi et qui sont nécessaires ». L'article propose un test en trois parties pour évaluer les restrictions en matière d'expression :

- 1 La restriction est-elle conforme au principe de **légalité** ? Nous examinons si les règles sur lesquelles Meta se base pour prendre sa décision sont accessibles et suffisamment claires aux yeux des utilisateurs pour garantir leur compréhension et leur application. Il est essentiel que les règles soient claires de sorte que ceux et celles en charge de les appliquer puissent prendre des décisions justes et cohérentes.
- 2 La restriction proposée a-t-elle un **objectif légitime** ? Le Conseil s'est penché sur les objectifs en matière de restrictions de l'expression reconnus à l'Article 19 du PIDCP pour évaluer si la règle sur laquelle une décision est fondée respecte un objectif compatible avec les droits en vigueur.
- 3 La restriction était-elle **nécessaire et proportionnelle** ? La suppression du contenu était-elle le moyen le moins intrusif d'atteindre l'objectif ?

# Comment le Conseil tient compte des appels d'utilisateurs<sup>1</sup>



<sup>1</sup> Graphique : Conseil de surveillance. Source : Comme indiqué dans l'outil de gestion des cas du Conseil de surveillance.



### Décision

Le jury parvient à une décision quant à l'autorisation du contenu : approuver ou contredire Meta.

### Approbation

Une décision préliminaire est envoyée pour révision à tous les membres du Conseil. Une décision doit être signée par la majorité pour être publiée.

### Publication

Notre décision est publiée sur le site web du Conseil de surveillance.

### Réponse

En 2021, Meta devait appliquer nos décisions dans un délai de sept jours après publication et répondre à toutes les recommandations dans un délai de 30 jours. Depuis, ce délai a été prolongé à 60 jours.

# Chronologie des événements clés



21 SEPTEMBRE

**RÉPONSE AUX RÉVÉLATIONS SUR  
CROSS-CHECK**

En nous basant sur des rapports de la presse, nous avons commencé à vérifier si Meta était totalement honnête envers nous au sujet de son système cross-check.

15 JUIN

**ACCEPTATION DU PREMIER AVIS  
CONSULTATIF EN MATIÈRE DE POLITIQUES**

Acceptation de l'avis consultatif en matière de politiques sur le partage d'informations privées sur le lieu de résidence. Les avis consultatif en matière de politiques nous permettent de suggérer à Meta des améliorations de ladite politique.

21 OCTOBRE.

**RAPPORTS DU PREMIER  
TRIMESTRE**

Publication du premier rapport de transparence trimestriel, rien au sujet du manque d'honnêteté de l'équipe de Meta à propos de cross-check.

8 JUILLET

**DÉCISION SUR L'ISOLEMENT  
D'ÖCALAN**

Publication de la décision sur l'isolement d'Öcalan. Meta n'a rien perdu et n'a pas appliqué une exception de politique importante pendant trois ans.

26 MAI

**DÉCISIONS SUR LES  
MANIFESTATIONS DE SOUTIEN À  
NAVALNY**

Publication de la première décision où nous avons renversé Meta sur la base de ses responsabilités envers les droits de l'homme, même si la suppression du contenu était conforme aux règles de Facebook.

# Sélection des cas



**1 152 181**

cas ont été soumis au Conseil d'octobre 2020 à décembre 2021,

→ DONT **47** de la part de Meta

**2 649**



cas : nombre moyen de cas reçus par jour par le Conseil

Plus de  **8 appels d'utilisateur sur 10**

sollicitant une restauration de contenu concernaient les règles de Meta sur le harcèlement, les discours incitant à la haine ou la violence et l'incitation.



des cas soumis l'ont été depuis les États-Unis, le Canada et l'Europe.

DANS **51**

CAS SUR LES **130**

sélectionnés par le Conseil, Meta a reconnu que sa décision d'origine sur le contenu était mauvaise.

# Vue d'ensemble

Les utilisateurs Facebook et Instagram peuvent remettre en question les décisions de Meta en faisant appel du contenu éligible auprès du Conseil. Entre octobre 2020, moment où le Conseil a commencé à accepter les appels d'utilisateurs, et décembre 2021, terme de la portée de ce rapport, nous avons reçu plus de 1,1 millions de demandes de la part des utilisateurs pour réexaminer de manière indépendante les décisions de modération du contenu de Meta. Meta a également fait remonter 47 cas au Conseil. En moyenne, d'octobre 2020 à décembre 2021, le Conseil a reçu 2 649 cas par jour.

Le volume de cas soumis indique l'importance de la mission du Conseil aux yeux des utilisateurs. En 2020 comme en 2021, nous avons intentionnellement donné la priorité aux cas qui avaient le potentiel d'avoir un impact considérable sur de nombreux utilisateurs à l'échelle internationale, qui ont été extrêmement importants dans le débat public ou qui ont soulevé des questions essentielles concernant les politiques de Meta. Pour traiter de problèmes propres aux habitants de pays spécifiques, nous avons également sélectionné des cas de différentes régions du monde entier. En outre, nous avons sélectionné plusieurs cas ayant soulevé des implications majeures concernant l'application des normes internationales des droits de l'homme à la modération du contenu à l'échelle mondiale.

Le Conseil a également **écouté les utilisateurs** en donnant la priorité aux cas concernant des problématiques visées de façon récurrente par

leurs appels. Autrement dit, malgré que nous ne sélectionnions qu'un cas qui relève d'une problématique spécifique pour révision, le Conseil est ainsi en mesure de traiter des problèmes communs partagés par un bien plus grand nombre de personnes, dont celles dont les cas ne sont pas sélectionnés. Par exemple, lorsque nous avons vu Meta supprimer de nombreuses publications faisant référence à des personnalités nazies, même si le contenu ne faisait ni l'apologie ni l'éloge d'une quelconque « personne ou organisation dangereuse », nous avons sélectionné le cas Citation nazie pour étudier un problème significatif ayant un impact sur un grand nombre d'utilisateurs.

Afin de s'assurer que notre processus de sélection des cas s'appuie sur un large éventail d'avis, le Conseil désigne périodiquement un nouveau Comité de sélection des cas constitué de 5 membres pour traiter une variété de préoccupations largement partagées par les utilisateurs. À ce jour, et conformément à nos critères primordiaux de sélection, les comités ont donné la priorité aux cas où les systèmes automatiques de Meta ont potentiellement modéré le contenu, notamment dans les pays où les utilisateurs ont signalé un manque d'examen manuel dans la langue d'origine ; aux cas et au contenu impliquant l'éloge ou la promotion présumée de « personnes ou d'organisations dangereuses » ; et aux cas ayant soulevé d'importants problèmes liés à la liberté de la presse, notamment pour les journalistes s'exposant au risque de « préjudice imminent » dans les zones dites de conflit.



## ENSEIGNEMENTS TIRÉS

En faisant ses sélections, le Conseil a été confronté à plusieurs problématiques et obstacles d'ordre opérationnel :

### **S'adapter aux droits de libre expression des utilisateurs pour retirer leur propre contenu après la sélection des cas**

Un exemple : peu de temps après avoir annoncé nos premiers cas en décembre 2020, un utilisateur a choisi de supprimer un contenu lié à des tweets de l'ancien Premier ministre malaisien, Dr Mahathir Mohamad. En raison de cette intervention de l'utilisateur, le Conseil n'a plus eu la possibilité d'examiner le cas. Dans une situation similaire datant de fin 2021, un utilisateur a supprimé le contenu faisant référence à un journaliste travaillant en Afghanistan pendant et après le retrait des troupes américaines. Dans les deux cas, même si ces suppressions de contenu ont empêché le Conseil de s'engager sur des sujets relevant de préoccupations communes, nous respectons le

droit des utilisateurs à supprimer leur contenu pour n'importe quelle raison. Comme nous pouvons nous attendre à ce que de telles situations surviennent de temps à autre, chaque fois qu'un cas est retiré à la suite de l'intervention d'un utilisateur, nous l'annoncerons immédiatement et en expliquerons les circonstances.

### **Offrir une transparence optimale au contenu supprimé par le « système de suspension » opaque de Meta**

Au cours du deuxième trimestre 2021, deux cas ont été attribués à un jury, mais pas annoncés du fait que Meta a appliqué des mesures de modération. Un cas concernant les origines de la COVID-19 n'a pas été annoncé après le retrait, par Meta, de la page qui hébergeait le contenu ; en cause, un trop grand nombre de « suspensions » prononcées à son égard. Dans un deuxième cas, concernant une erreur sur le genre, Meta a également supprimé la page qui hébergeait la publication suite à un trop grand nombre de « suspensions » prononcées.

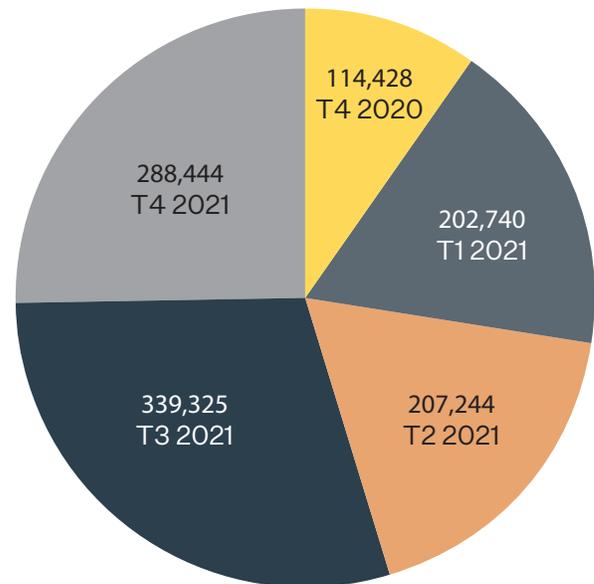
# Cas soumis au Conseil<sup>2</sup>

Depuis que nous avons commencé à accepter des cas en octobre 2020, le volume de cas soumis au Conseil n'a eu de cesse d'augmenter chaque trimestre jusqu'au T3 2021, avec une légère baisse au T4 2021. Cela s'explique par l'importance à l'échelle internationale des problématiques liées à la modération du contenu pour les utilisateurs Facebook et Instagram.

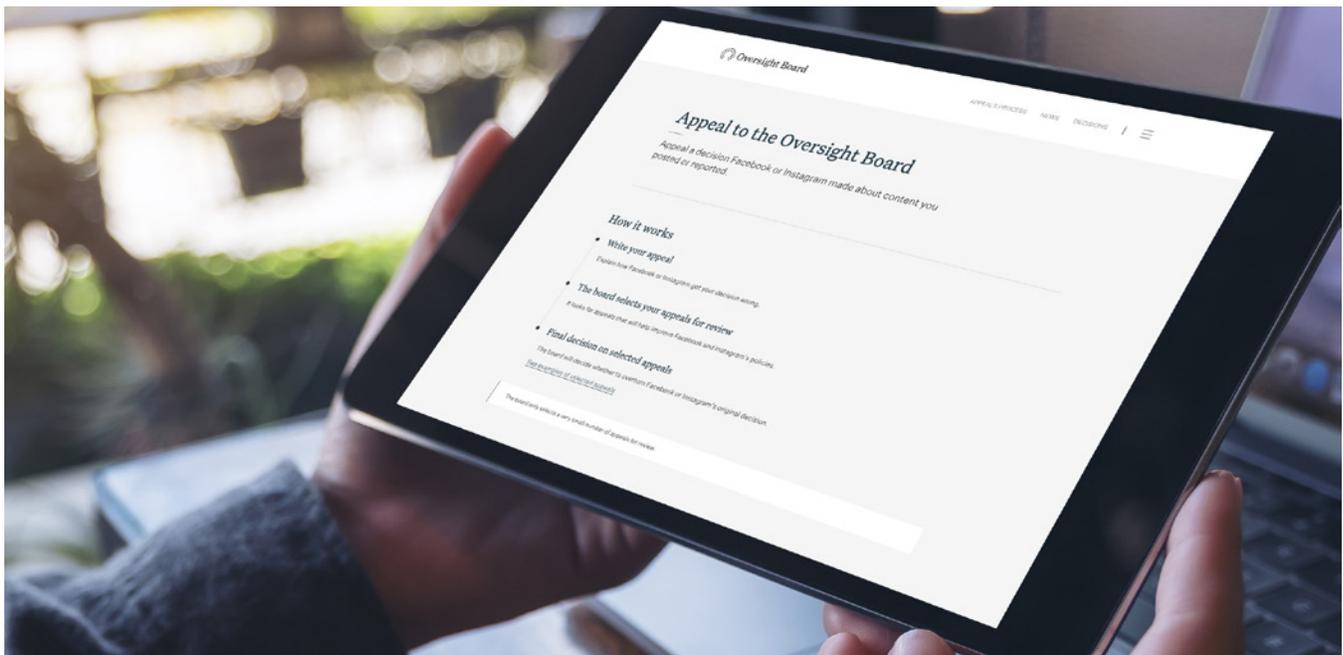
Entre le T2 et le T3 2021, le nombre d'appels a augmenté de 64 %. Cette croissance est vraisemblablement imputable aux améliorations de Meta concernant la manière dont les utilisateurs peuvent déposer un appel auprès du Conseil via l'application mobile de Facebook, ainsi qu'à la visibilité accrue du Conseil. En une semaine à compter du lundi 30 août 2021, les utilisateurs ont établi un record des cas soumis en une semaine, en déposant plus de 32 000 appels distincts auprès du Conseil.

Sur les cas soumis, 99 % étaient associés au contenu Facebook,<sup>3</sup> tandis que seul 1 % concernait le contenu Instagram. À compter du deuxième trimestre 2021, une fois que les utilisateurs ont été habilités à faire appel auprès du Conseil pour supprimer le contenu publié par toute *autre* personne, 94 % des appels

## Estimation du nombre de cas soumis au Conseil par trimestre



Graphique : Conseil de surveillance. Source : Comme indiqué dans l'outil de gestion des cas du Conseil de surveillance.

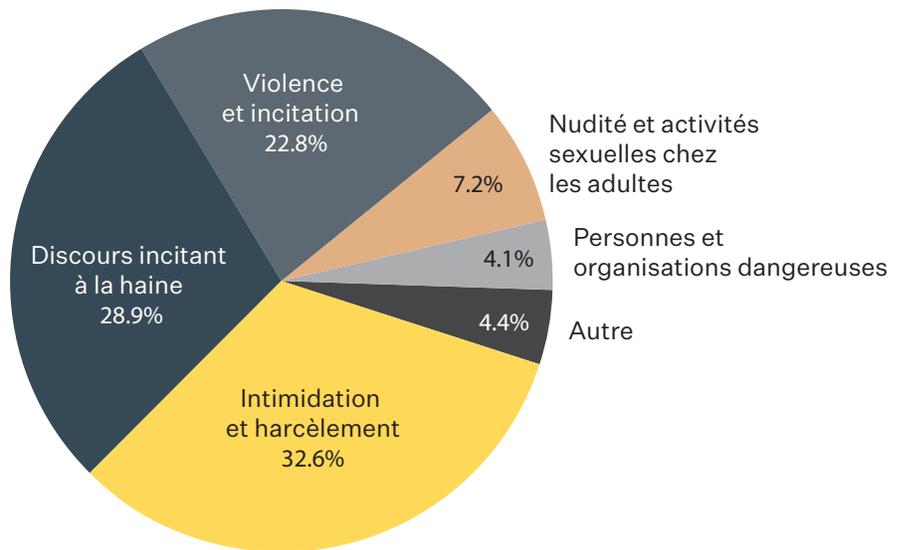


<sup>2</sup> Due to limitations in the Case Management Tool, through which cases are submitted to the Board, and the data currently available to us, submitted cases were sometimes counted manually by the Board's Case Selection Team as they appeared in the CMT at the time. Breakdowns of these numbers may also be based on samples. As such, these numbers should be taken as an estimate.

<sup>3</sup> This figure only covers Q2-Q4 2021 as data was only available for this period.

## Estimation du nombre de cas soumis au Conseil par Standard de la communauté (oct. 20 - déc. 2021)

(ne comprend que les appels déposés par des utilisateurs pour restaurer des contenus sur Facebook et Instagram)



Graphique : Conseil de surveillance.  
Source : Comme indiqué dans l'outil de gestion des cas du Conseil de surveillance.

d'utilisateurs avaient pour objectif de *restaurer* du contenu sur Facebook ou Instagram, tandis que 6 % concernaient des appels pour *supprimer* du contenu des plateformes.

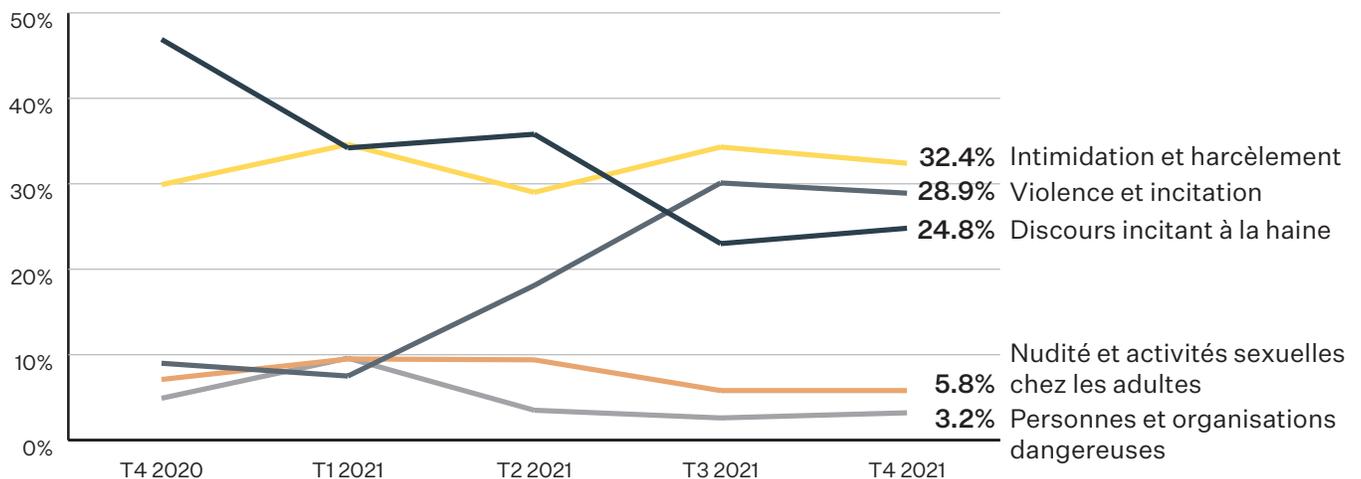
Une majorité écrasante des appels d'utilisateurs pour restaurer du contenu (84 %) était associée à trois Standards de la communauté Facebook : « Intimidation et harcèlement », « Discours incitant à la haine » et « Violence et incitation ». Les cas associés aux standards « Nudité et activités sexuelles chez les adultes » et « Personnes et organisations dangereuses »

constituaient la majeure partie des autres appels visant à restaurer du contenu.

Pour offrir un contexte plus large à ces chiffres, du T4 2020 au T4 2021, les Standards de la communauté en vertu desquels Meta « a traité » la plupart des éléments de contenu (par la suppression de contenus, mais également par d'autres actions comme l'insertion d'écrans d'avertissement) étaient « Nudité et activités sexuelles chez les adultes », suivi de « Contenu violent et graphique » et de « Discours incitant à la haine ». Les Standards de la communauté pour lesquels Meta a reçu le plus d'appels pour la suppression de contenu publié

## Estimation du nombre de cas soumis au Conseil par Standard de la communauté et par trimestre

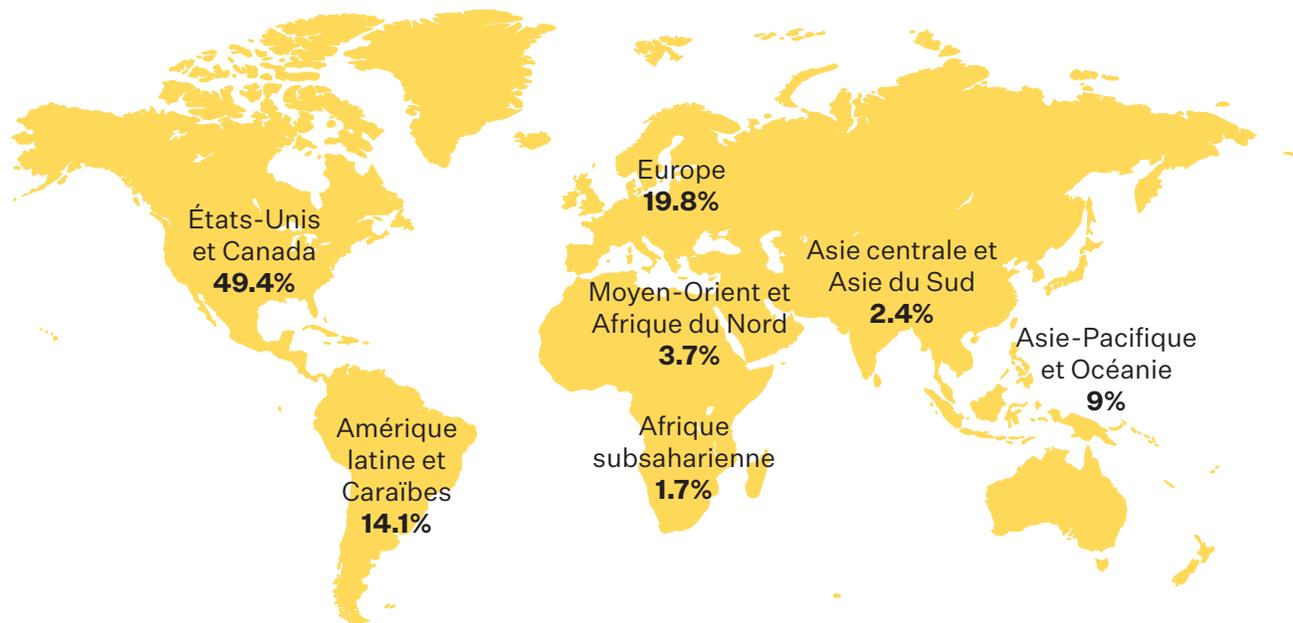
(ne comprend que les appels déposés par des utilisateurs pour restaurer des contenus sur Facebook et Instagram)



Graphique : Conseil de surveillance. Source : Comme indiqué dans l'outil de gestion des cas du Conseil de surveillance.

4 Source: Meta's Community Standards Enforcement Reports (CSER). The Community Standards mentioned here exclude policy areas outside of the Board's scope and those not mentioned in the CSER.

## Estimation du nombre de cas soumis au Conseil par région sélectionnée par l'utilisateur (oct. 2020 - déc. 2021)



Map: Oversight Board. Source: As seen in the Oversight Board Case Management Tool.

au cours de cette période furent néanmoins « Discours incitant à la haine », suivi de « Intimidation et harcèlement » et de « Nudité et activités sexuelles chez les adultes ». Parallèlement, au cours de cette période, « Intimidation et harcèlement » suivi de « Discours incitant à la haine » furent les deux Standards de la communauté pour lesquels le Conseil a reçu le plus d'appels d'utilisateurs pour la restauration de contenu.

Visualisées trimestre par trimestre, certaines tendances intéressantes ont émergé des données réunies à propos des appels d'utilisateurs demandant la restauration de contenu sur les deux plateformes. Tandis que le Standard de la communauté « Intimidation et harcèlement » représentait toujours environ un tiers de ces appels, d'autres Standards de la communauté ont évolué considérablement au cours de l'année. La part des cas associés aux règles de Facebook sur le standard « Violence et incitation » a triplé d'année en année, passant de 9 % au T4 2020 à 29 % au T4 2021. Sur la même période, la part des cas associée au standard « Discours incitant à la haine » a quasiment été réduite de moitié, passant de 47 % au T4 2020 à près de 25 % au T4 2021.

Plus des deux tiers des utilisateurs ayant fait appel provenaient des pays du Nord. Ainsi, 49 % du total des appels provenaient des États-Unis et du Canada et

20 % d'Europe, tandis que 14 % des appels provenaient d'Amérique latine et des Caraïbes, 9 % d'Asie-Pacifique et d'Océanie, et seulement 4 % provenaient du Moyen-Orient et d'Afrique du Nord, 2 % d'Asie centrale et d'Asie du Sud et 2 % d'Afrique subsaharienne.

Nous reconnaissons que cette distribution ne reflète pas la répartition des utilisateurs Facebook et Instagram dans le monde entier. En 2019, par exemple, seuls six des 20 pays comptant le plus d'utilisateurs de Facebook se trouvaient en Europe et en Amérique du Nord, tandis que l'Inde compte le plus grand nombre d'utilisateurs de Facebook et d'Instagram de tous les pays. Le nombre plus faible d'appels d'utilisateurs émanant en dehors de l'Europe et des États-Unis et du Canada pourrait également indiquer que nombre d'utilisateurs Facebook et Instagram dans le reste du monde ne savent pas qu'ils ont la possibilité de faire appel des décisions de modération du contenu de Meta auprès du Conseil. Par ailleurs, nous ne pensons pas que la distribution des données relatives aux appels sur cette carte reflète la distribution réelle des problèmes liés à la modération du contenu dans le monde entier. De fait, nous avons des raisons de croire que les utilisateurs en Asie, en Afrique et au Moyen-Orient rencontrent plus, et non moins, de problèmes avec les plateformes de Meta que les autres parties

du monde. Nos décisions jusqu'à présent concernant des publications en provenance d'Inde et d'Éthiopie ont soulevé des préoccupations quant au fait de savoir si Meta a investi suffisamment de ressources dans la modération du contenu dans des langues autres que l'anglais.

Outre le développement de notre portée dans des régions en dehors de l'Europe, des États-Unis et du Canada, nous tenons également compte et sélectionnons plusieurs cas en dehors des pays du Nord pour traiter les préoccupations des utilisateurs à l'échelle internationale. Plus de la moitié des 278 cas considérés pour sélection par le Comité de sélection des cas provenaient de pays en dehors des États-Unis, du Canada et de l'Europe, tandis que plus de la moitié des 20 décisions que nous avons publiées en 2021 concernaient les pays du Sud.

Outre les appels issus d'utilisateurs Facebook et Instagram, Meta peut, elle aussi, nous soumettre des cas importants et complexes.

Sur les 47 cas soumis par Meta au Conseil en décembre 2021, 41 concernaient un contenu partagé sur Facebook tandis que 6 impliquaient un contenu

sur Instagram. 29 des 47 cas mentionnés par Meta concernaient un contenu toujours diffusé sur Facebook ou Instagram et n'avaient pas été jugés comme enfreignant les règles de l'entreprise.

Les 18 cas restants impliquaient un contenu supprimé par Meta de Facebook ou Instagram pour une infraction supposée des Standards de la communauté suivants : Violence et incitation (six cas), Discours incitant à la haine (trois cas), Individus et organisations dangereuses (trois cas), Exploitation sexuelle des enfants, Maltraitance et nudité (trois cas), Intimidation et harcèlement (deux cas) et Nudité et activités sexuelles chez les adultes (un cas).

Près des deux tiers des cas soumis par Meta au Conseil étaient associés à un contenu provenant d'Europe, des États-Unis et du Canada ; 30 des 47 cas provenaient ainsi de ces régions. Meta a soumis 17 cas couvrant le reste du monde : quatre en provenance d'Amérique latine et des Caraïbes, quatre en provenance du Moyen-Orient et d'Afrique du Nord, trois en provenance d'Asie centrale et d'Asie du Sud, trois en provenance d'Asie-Pacifique et d'Océanie et trois en provenance d'Afrique subsaharienne.

## Cas mentionnés par Meta



# Cas pris en considération par le Comité de sélection des cas

Pour traiter les problèmes auxquels sont confrontés les utilisateurs Facebook et Instagram du monde entier, le Conseil sélectionne un éventail diversifié de cas dans de nombreuses langues. Les 278 cas pris en compte par le Comité de sélection des cas en décembre 2021 couvraient plus de 70 pays, allant des Îles Fidji au Tchad, en passant par Trinité-et-Tobago. Ces cas représentaient un contenu dans 38 langues différentes.

Une fois que le Comité de sélection des cas du Conseil sélectionne des cas pour examen par le Conseil, Meta reconnaît parfois que sa décision d'origine sur une partie du contenu était infondée. C'est ce qu'on appelle également une « erreur d'application ».

Le graphique sur cette page présente, par trimestre, le nombre de cas que Meta a identifié comme erreurs

d'application. Aux T4 2020 et T2 2021, Meta a conclu que, dans un tiers des cas sélectionnés, sa décision d'origine était infondée. Au T1 2021, ce taux s'élevait à 44 %, tandis qu'au T3 2021, il s'élevait à 46 %. Le Conseil n'a pas sélectionné de cas au T4 2021.

Sur les 130 cas sélectionnés par le Conseil jusqu'à décembre 2021, Meta a identifié 51 occasions où sa décision d'origine sur le contenu était infondée. Bien qu'il ne s'agisse que d'un petit échantillon et que le Conseil recherche volontairement des cas complexes, il est interpellant de constater que Meta a jugé sa décision infondée seulement dans moins de 4 cas sur 10. Ce taux d'erreurs élevé soulève des questions plus larges à la fois sur la précision de la modération du contenu de Meta et sur la procédure d'appel que Meta applique avant que les cas n'arrivent au Conseil.

## Erreurs d'application par nombre de cas sélectionnés

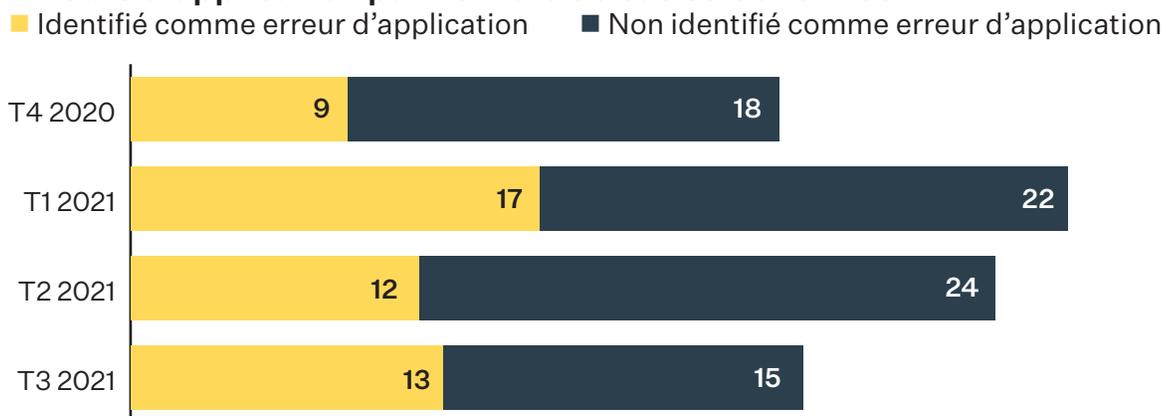


Chart: Oversight Board. Source: As seen in the Oversight Board Case Management Tool.

# Décisions



**20**

Décisions publiées en 2021



Décisions de Meta infirmées à **14** reprises



Confirmées à **6** reprises



**16** Cas d'utilisateurs

**4** Cas de Meta



**17** Cas de Facebook

**3** Cas d'Instagram

**83**

ont été nécessaires, en moyenne, pour formuler une décision et l'implémenter



**PLUS DE LA MOITIÉ**

de nos décisions étaient associées aux Pays du Sud.



**313**

questions posées à Meta, dans le cadre de nos révisions de cas.

Les décisions de modération du contenu de Meta ont été analysées selon les

**normes internationales des droits de l'homme**



# Vue d'ensemble

Le Conseil de surveillance existe pour prendre des décisions exécutoires et indépendantes sur les problématiques de contenu les plus complexes auxquelles Facebook et Instagram font face. Nos 20 premières décisions intégrales et les recommandations de politique associées ont soulevé de manière répétée des problèmes majeurs sur la manière dont l'entreprise traite les utilisateurs dans le monde entier. Le Conseil infirme non seulement les décisions de Meta le cas échéant, mais recommande également des changements de politique pour améliorer le traitement des utilisateurs et offrir une plus grande transparence. Collectivement, ces actions défendent les utilisateurs dans les cas où les politiques de l'entreprise sont peu claires, incohérentes ou non appliquées. Les principaux thèmes couverts par nos décisions sont les suivants :

- **Protection de la liberté d'expression et d'autres droits de l'homme.** Bien que Meta, une entreprise servant de réseau principal de communication pour des milliards de personnes, juge la parole « primordiale », la liberté d'expression doit être limitée, le cas échéant, pour protéger de tout préjudice réel les utilisateurs et tout autre individu. Concilier ces préoccupations est la plus importante, et la plus difficile, des responsabilités du Conseil.
- **Tenir compte de la signification contextuelle.** Nos décisions reposaient souvent sur des questions de contexte. Parmi celles-ci figuraient la

signification d'un mot particulier ou de la publication elle-même, ainsi que les circonstances dans lesquelles le contenu était publié, comme un conflit violent ou des contestations politiques.

- **Être clair avec les utilisateurs.** Nous avons débattu d'un cas où Meta n'avait pas informé un utilisateur sur le Standard de la communauté qu'il avait enfreint, un cas où l'entreprise n'avait pas examiné l'appel de l'utilisateur et un cas où Meta avait égaré un élément de directives de politique pendant trois ans. Trois décisions concernant des contenus Instagram ont également souligné que le lien entre les règles de Facebook et celles d'Instagram n'est pas clair pour les utilisateurs.
- **Traiter les utilisateurs avec équité, peu importe où ils se trouvent.** Nos décisions ont souligné que les règles de Meta pour Facebook et Instagram ne sont pas disponibles dans toutes les langues des utilisateurs. Nous avons également soulevé des préoccupations quant au fait de savoir si Meta investissait suffisamment de ressources dans la modération du contenu dans des langues autres que l'anglais.
- **S'assurer que les règles de Meta sont appliquées correctement.** Nos décisions précisent clairement que Meta doit améliorer la précision des aspects automatiques et humains de la modération du contenu et être plus transparente avec le Conseil et le public sur les raisons des erreurs.

# ENSEIGNEMENTS TIRÉS

## Réviser le calendrier pour refléter les réalités opérationnelles

Lorsque nous avons formulé nos premières décisions, il est rapidement apparu que le délai initial de 90 jours à compter de la dernière décision de Meta concernant un cas donné ne suffirait pas. En mars, afin de veiller à ce que tous les cas bénéficient du même temps pour la phase de délibération, nous avons reporté le point de départ de ce délai de 90 jours prévu pour la décision et l'implémentation au moment de l'attribution du cas à un jury. Comme le Conseil a développé ses propres procédures de prise de décisions et commencé à poser à Meta plus de questions, le point de départ du délai de 90 jours a été à nouveau modifié en novembre pour commencer à l'annonce d'un cas.

## Affiner nos procédures et standards

Alors que nous délibérons de nos premiers cas, nous avons dû faire face à plusieurs considérations complexes. Par exemple, lorsque nous examinons un élément de contenu, devons-nous considérer toutes les informations qui nous ont été fournies via notre procédure d'examen de cas ou seulement évaluer si la décision de Meta était une réponse raisonnable en tenant compte des circonstances de la révision et des informations disponibles au moment même ? En outre, comment le Conseil doit-il appliquer les normes internationales des droits de l'homme, avant tout conçues pour les États, aux décisions prises par les plateformes de réseaux sociaux ? La section des droits de l'homme de ce chapitre aborde cette problématique plus en détail.

# TABLEAU : DÉCISIONS PUBLIÉES EN 2021

	PLATEFORME	SOURCE	STANDARD DE LA COMMUNAUTÉ	PAYS	DÉCISION DU CONSEIL
<b>Publication du Myanmar à propos des musulmans</b> Numéro du cas : 2020-002-FB-UA		Utilisateur		Myanmar, France, Chine	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Arméniens en Azerbaïdjan</b> Numéro du cas : 2020-003-FB-UA		Utilisateur		Arménie, Azerbaïdjan	<b>Décision de suppression de contenu</b> de Meta confirmée
<b>Symptômes du cancer du sein et nudité</b> Numéro du cas : 2020-004-IG-UA		Utilisateur		Brésil	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Citation nazie</b> Numéro du cas : 2020-005-FB-UA		Utilisateur		États-Unis	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Revendication d'un traitement contre la COVID</b> Numéro du cas : 2020-006-FB-FBR		Meta		France	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Protestation contre la France en Inde</b> Numéro du cas : 2020-007-FB-FBR		Meta		Inde, France	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Suspension du compte de l'ancien président des États-Unis, Donald Trump</b> Numéro du cas : 2021-001-FB-FBR		Meta		États-Unis	<b>Décision de suppression de contenu</b> de Meta confirmée
<b>Représentation de Zwarte Piet</b> Numéro du cas : 2021-002-FB-UA		Utilisateur		Pays-Bas	<b>Décision de suppression de contenu</b> de Meta confirmée



Personnes et organisations dangereuses



Violence et incitation



Discours incitant à la haine



Nudité et activités sexuelles chez les adultes



Bullying and Harassment

	PLATEFORME	SOURCE	STANDARD DE LA COMMUNAUTÉ	PAYS	DÉCISION DU CONSEIL
<b>Préoccupations au Pendjab par rapport au RSS en Inde</b> Numéro du cas : 2021-003-FB-UA		Utilisateur		Inde	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Manifestations de soutien à Navalny en Russie</b> Numéro du cas : 2021-004-FB-UA		Utilisateur		Russie	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Mème « Deux boutons »</b> Numéro du cas : 2021-005-FB-UA		Utilisateur		Arménie, Turquie, États-Unis	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Isolement d'Öcalan</b> Numéro du cas : 2021-006-IG-UA		Utilisateur		Turquie, États-Unis	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Le bot du Myanmar</b> Numéro du cas : 2021-007-FB-UA		Utilisateur		Myanmar	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Confinements liés à la COVID au Brésil</b> Numéro du cas : 2021-008-FB-FBR		Meta	N/A	Brésil	<b>Décision</b> de Meta de garder le contenu confirmée
<b>Partage d'une publication d'Al Jazeera</b> Numéro du cas : 2021-009-FB-UA		Utilisateur		Israël, Égypte	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Manifestations en Colombie</b> Numéro du cas : 2021-010-FB-UA		Utilisateur		Colombie	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Insultes en Afrique du Sud</b> Numéro du cas : 2021-011-FB-UA		Utilisateur		Afrique du Sud	<b>Décision de suppression de contenu</b> de Meta confirmée



Bullying and Harassment



Personnes et organisations dangereuses



Discours incitant à la haine

	PLATEFORME	SOURCE	STANDARD DE LA COMMUNAUTÉ	PAYS	DÉCISION DU CONSEIL
<b>Collier de wampum</b> Numéro du cas : 2021-012-FB-UA		Utilisateur		Canada	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Ayahuasca</b> Numéro du cas : 2021-013-IG-UA		Utilisateur		Brésil	<b>Décision de suppression de contenu</b> de Meta infirmée
<b>Crimes présumés à Raya Kobo</b> Numéro du cas : 2021-014-FB-UA		Utilisateur		Éthiopie	<b>Décision de suppression de contenu</b> de Meta confirmée



Discours incitant à la haine



Regulated Goods

## Premières décisions



« Jusqu'à ce jour, c'était à [Mark] Zuckerberg de prendre les décisions en matière de contenu... Aujourd'hui est un jour historique. C'est la première fois que la modération du contenu se fait en dehors de Facebook. De plus, Facebook doit respecter nos décisions. »

Helle Thorning-Schmidt  
*Coprésidente du Conseil de surveillance*



Le 28 janvier 2021, le Conseil de surveillance a publié cinq décisions, basées sur les délibérations menées depuis la sélection de nos premiers cas parmi 20 000 appels. Les Membres du Conseil ont mené les délibérations selon une procédure décrite dans nos [Statuts](#) : chaque cas a été attribué à un jury mixte de cinq membres, dont au moins un membre provient de la région concernée par le contenu. Après la formulation d'une décision par le jury, une ébauche a été transmise à tous les Membres du Conseil pour commentaires, puis soumise à un vote au sein de l'intégralité du Conseil. Les cas couvraient des contenus sur quatre continents, à savoir : l'Asie, l'Europe, l'Amérique du Nord et l'Amérique du Sud.

Sur les cinq premiers cas, quatre ont infirmé et un a confirmé la décision de Meta consistant à supprimer une publication. Les cas ont exploré une vaste plage de problématiques. Dans un cas, les Membres du Conseil ont étudié si, dans le contexte d'un conflit armé, Meta avait raison de supprimer une publication autrement permise, car elle contenait une insulte haineuse. Dans un autre cas, ils ont étudié si une publication accusée de diffuser une fausse information sur la COVID-19 contribuait à un préjudice imminent. Une autre décision définissait un important précédent en soulignant l'incapacité des systèmes d'application automatisés de Meta à distinguer les images autorisées de seins nus de femmes dans le cadre d'une campagne de sensibilisation au cancer du sein.



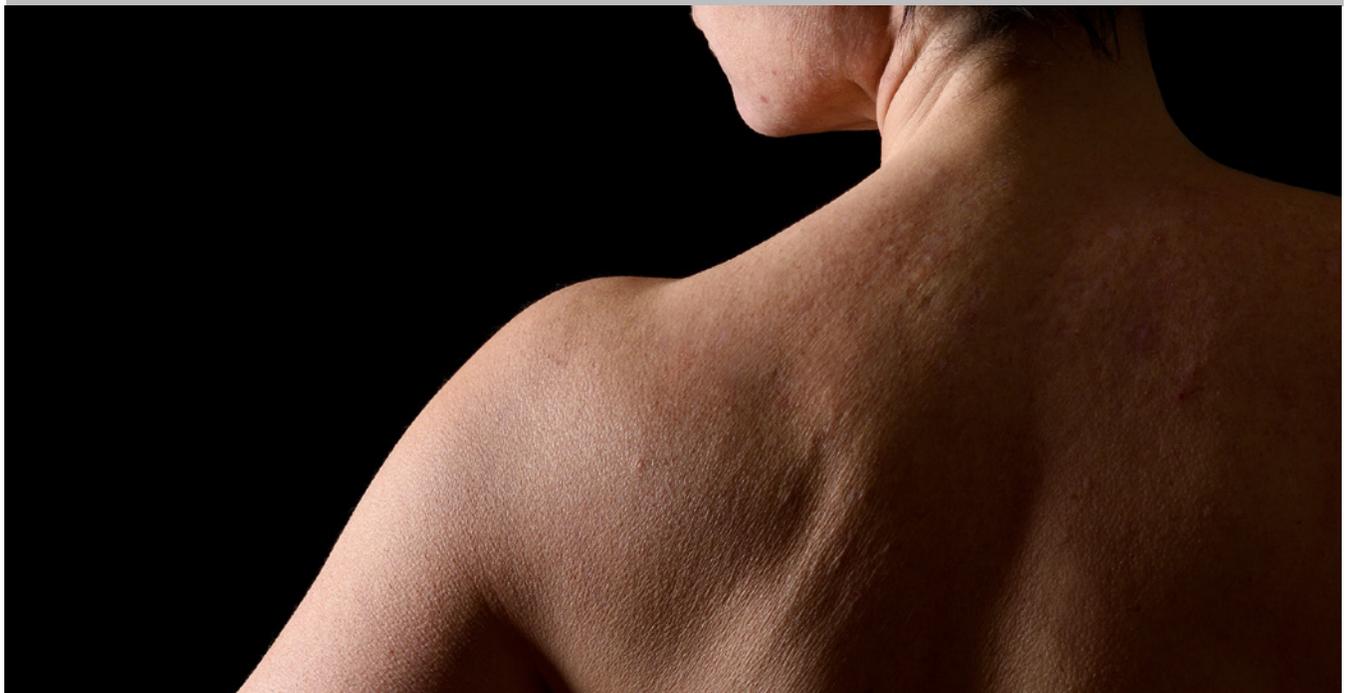
## Symptômes du cancer du sein et nudité

DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
28 janvier 2021	Brésil	Nudité et activités sexuelles chez les adultes

En octobre 2020, un utilisateur au Brésil a publié une photo sur Instagram accompagnée d'un titre, en portugais, dont le but consistait à accroître la sensibilisation au cancer du sein. La publication a été supprimée par un système automatisé appliquant le Standard de la communauté Facebook sur la nudité et les activités sexuelles chez les adultes, qui s'applique également à Instagram, comme l'a expliqué l'entreprise. Nous avons infirmé la décision de Meta, car le Standard de la communauté inclut une exception claire qui autorise la nudité lorsque l'utilisateur cherche à « accroître la sensibilisation à une cause pour des raisons éducatives et médicales », et autorise spécifiquement la publication des photos de seins nus de femmes pour « faire avancer la sensibilisation au cancer du sein ».

Une fois le contenu supprimé, l'utilisateur a fait appel de cette décision auprès de Meta. Dans des déclarations publiques, Meta avait précédemment indiqué qu'elle ne pourrait pas toujours offrir aux utilisateurs la possibilité de faire appel en raison d'une réduction temporaire de sa capacité à réaliser des examens en raison de la COVID-19. De plus, Meta a également stipulé que les appels ne recevront pas tous un avis favorable pour un examen manuel.

Nous avons rejeté l'argument de Meta selon lequel le cas « n'avait plus lieu d'être », car ayant restauré la publication, l'utilisateur et Meta n'étaient plus en désaccord sur le maintien de la diffusion du contenu. Nous avons expliqué que le Conseil avait l'autorité d'examiner les cas d'utilisateurs, même si Meta choisit ultérieurement de corriger son erreur et de restaurer le contenu. Enfin, le Conseil a recommandé que les utilisateurs soient en mesure de faire appel des décisions prises par des systèmes automatisés avec la réalisation d'un examen manuel lorsque leur contenu est considéré comme ayant enfreint le Standard de la communauté Facebook sur la nudité et les activités sexuelles chez les adultes.





## Citation nazie

DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
28 janvier 2021	États-Unis	Personnes et organisations dangereuses

En octobre 2020, un utilisateur aux États-Unis a publié une citation en anglais, indûment attribuée à Joseph Goebbels, Ministre de la propagande nazie du IIIe Reich. La citation affirmait qu'il était inutile de faire appel aux intellectuels, car ils ne se convertiraient pas et, en tout état de cause, céderaient à l'homme le plus fort sur la place publique. Elle soutenait ainsi que les arguments devaient solliciter les émotions et l'instinct. Meta a supprimé la publication pour infraction du Standard de la communauté Facebook sur les personnes et organisations dangereuses. L'utilisateur, pour sa part, a confirmé au Conseil que son intention était de dresser une comparaison entre le sentiment exprimé dans la citation et la présidence de Donald Trump.

En examinant le cas, le Conseil a déterminé que puisque la citation ne soutenait pas l'idéologie du parti nazi ou les actes de haine et de violence du régime, elle n'était pas en infraction avec le Standard de la communauté Facebook. Nous avons remarqué que, dans ce cas, l'approche de Meta exigeant des modérateurs qu'ils examinent le contenu sans tenir compte du contexte dans son ensemble a entraîné la restriction inutile du droit d'expression de l'utilisateur. En outre, le Conseil a recommandé que Meta explique et donne des exemples d'applications de termes clés utilisés dans sa politique sur les personnes et organisations dangereuses, y compris ses définitions de « faire l'éloge », « soutenir » et « représentation », qui devraient s'aligner sur les définitions utilisées dans les standards internes d'implémentation de Facebook.



## Revendication d'un traitement contre la COVID

DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
28 janvier 2021	France	Violence et incitation

Le Conseil a envisagé une autre décision, considérant que la fausse information relative à la santé contribue au risque de préjudice imminent. En octobre 2020, un utilisateur a publié une déclaration en français concernant une décision prise par une agence de santé publique française, l'Agence Nationale de Sécurité du Médicament, de ne pas autoriser l'utilisation de l'hydroxychloroquine combinée à l'azithromycine pour lutter contre la COVID-19. L'utilisateur faisait référence à cette association médicamenteuse comme un « médicament inoffensif » et la revendiquait comme un « traitement ». Meta a supprimé la publication conformément à sa politique sur la violence et l'incitation qui incluait une règle spécifique sur la fausse information et le préjudice imminent. Le Conseil a infirmé la décision de Meta en argumentant que Meta n'avait pas démontré que ce contenu avait contribué à un préjudice imminent, selon le standard repris dans ses propres politiques. Nous avons tenu compte du fait que la publication relevait de la politique publique plutôt que la recommandation de traitements à des individus et que les médicaments en question n'étaient pas disponibles en France sans ordonnance. Nous avons recommandé que Meta adopte des mesures moins intrusives pour traiter les publications qui critiquent les positions des autorités de santé et exposent à des préjudices éventuels, mais pas imminents, en ce compris la mise à disposition de liens vers des sites web d'agences de santé publique.

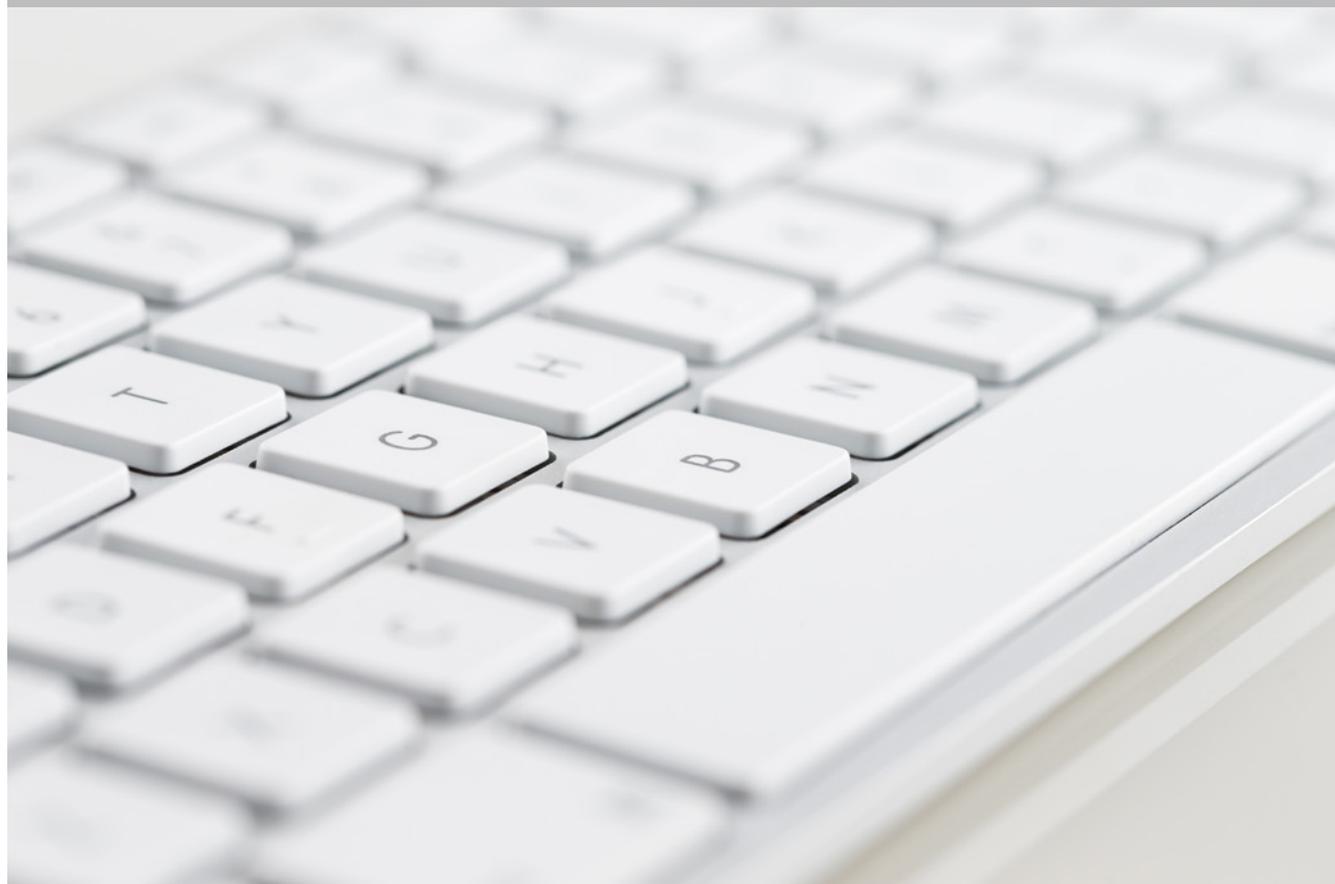


## Publication du Myanmar à propos des musulmans

DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
28 janvier 2021	Myanmar, France, Chine	Discours incitant à la haine

Une décision concernait une publication par un utilisateur au Myanmar, ayant partagé deux photos largement partagées d'un enfant en bas âge syrien d'origine kurde qui s'était noyé en tentant de rejoindre le continent européen en septembre 2015. La publication s'interrogeait sur le manque de réaction des musulmans en général concernant le traitement des musulmans Ouïgours en Chine alors que des caricatures du prophète Mahomet ont provoqué des assassinats en France. Le texte d'accompagnement, en birman, a été traduit par Meta comme suit : « [I]l y a en effet un problème psychologique chez les musulmans », tandis que les traducteurs du Conseil ont suggéré « [c]es hommes musulmans ont un problème d'état d'esprit ». Comme le Conseil l'a indiqué en infirmant la décision de Meta de supprimer la publication, en le prenant dans son contexte, le Conseil est d'avis que le texte est plutôt à comprendre comme « un commentaire sur l'incohérence manifeste entre les réactions des musulmans sur les événements en France et en Chine. »

Même si nous avons reconnu que le discours incitant à la haine en ligne au Myanmar a été lié à des accusations de crimes possibles contre l'humanité et de génocide de la minorité ethnique musulmane, les Rohingyas, le Conseil a estimé que la publication « ne prônait pas la haine ni n'incitait intentionnellement à nuire. » En un mot, la décision de Meta selon laquelle la publication enfreignait le Standard de la communauté de Facebook sur les discours incitant à la haine était injustifiée.





## Arméniens en Azerbaïdjan

DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
28 janvier 2021	Arménie, Azerbaïdjan	Discours incitant à la haine

La seule de nos cinq premières décisions qui ait confirmé la décision de Meta concernait la publication de novembre 2020 des photos historiques décrites comme montrant des églises à Bakou, capital de l'Azerbaïdjan. Le texte d'accompagnement en russe prétendait que les Arméniens avait construit Bakou et que cette héritage, y compris les églises, a été détruit. Dans sa publication, l'utilisateur a inclus des hashtags appelant à la fin de l'agression et du vandalisme azéris. L'utilisateur a utilisé le terme (« taziks ») pour décrire les Azéris, qui selon l'utilisateur, sont des nomades et n'ont « aucune histoire comparé aux Arméniens ».

Le terme « taziks », qui signifie « cuvette », semblait avoir été utilisé comme jeu de mots par rapport au terme russe « aziks », un terme dégradant pour les Azéris qui figure sur la liste interne d'insultes de Meta. Dans cette décision, nous avons remarqué que « le contexte dans lequel le terme a été utilisé montre clairement qu'il a été utilisé pour déshumaniser sa cible ». Ainsi, nous avons conclu que la suppression de cette publication dans ce contexte était cohérente avec les responsabilités de Meta en matière de droits de l'homme. Nous avons par ailleurs relevé que si elle n'est pas contrée, « une accumulation de ce type de contenu peut créer un environnement dans lequel des actes de discrimination et de violence sont plus probables ». Nous avons recommandé que Meta informe immédiatement l'utilisateur du motif de violation, dans le cas présent un simple terme, afin qu'il puisse publier à nouveau le reste du message s'il le souhaite.

### À l'annonce de notre premier round de décisions, le Conseil a observé ce qui suit :

*Nous sommes convaincus que les premières décisions rendues par le Conseil de surveillance témoignent de notre engagement à responsabiliser [Meta], en défendant les intérêts des utilisateurs et des communautés du monde entier et en redéfinissant l'approche en matière de modération de contenu. Nous sommes au début d'un processus qui prendra du temps et nous avons hâte de vous montrer nos progrès à travers les nombreuses décisions sur les cas que prendra le Conseil.*



# Décision de suspension du compte de l'ancien président des États-Unis, Donald Trump

**DATE DE PUBLICATION**

5 mai 2021

**PAYS**

États-Unis

**STANDARD DE LA COMMUNAUTÉ**

Personnes et organisations dangereuses

À 16:21 EST le 6 janvier 2021, après qu'une foule ait pris d'assaut le bâtiment du Capitole de Washington, D.C. et que l'insurrection battait son plein, M. Trump a publié une vidéo d'une minute sur Facebook et Instagram :

*Je comprends votre douleur. Je sais que vous vous sentez blessés. On nous a volé cette élection. C'était une victoire écrasante et tout le monde le sait, surtout nos adversaires. Mais il faut rentrer maintenant. Nous devons retrouver la paix. Il faut que la loi et l'ordre règnent. Nous devons respecter nos forces de l'ordre. Nous ne voulons aucun blessé. C'est une période difficile.*

Et de poursuivre :

*Une période inédite, où de tels événements surviennent, où ils ont pu nous la voler à tous, à moi, à vous, à notre pays. C'était une élection frauduleuse, mais nous ne pouvons pas entrer dans leur jeu. Nous devons retrouver la paix. Nous devons retrouver la paix. Alors rentrez chez vous. Nous vous aimons. Vous êtes exceptionnels. Vous avez vu ce qu'il s'est passé. Vous avez vu la manière dont les autres sont traités, à quel point ils sont mauvais et diaboliques. Je sais ce que vous ressentez. Mais rentrez chez vous et rentrez en paix.*

Une heure et 20 minutes plus tard, Meta retirait la vidéo pour violation de son Standard de la



Credit: Tyler Merbler

communauté sur les personnes et organisations dangereuses. À 18:07 EST, alors que la police sécurisait le Capitole, M. Trump a publié une déclaration écrite sur Facebook :

*Ce sont des choses et des événements qui arrivent quand une immense victoire électorale est si peu cérémonieuse et si vicieusement arrachée à de grands patriotes qui ont été mal et injustement traités pendant si longtemps. Rentrez chez vous en paix et avec amour. Souvenez-vous de ce jour pour toujours !*

Huit minutes plus tard, Meta a supprimé cette publication de Facebook pour avoir violé son Standard de la communauté sur les personnes et organisations dangereuses. L'entreprise a également bloqué la fonction de publication de M. Trump sur Facebook et Instagram pendant 24 heures.

Le 7 janvier, après un examen approfondi des publications de M. Trump, de ses récentes communications en dehors de Facebook et des compléments d'information sur la gravité de la violence perpétrée au Capitole, Meta a étendu la suspension « jusqu'à nouvel ordre et pour au moins les deux prochaines semaines jusqu'à ce que la transition pacifique du pouvoir soit terminée. »

## Sollicitation du Conseil par Meta

Deux semaines plus tard, le 21 janvier, au lendemain de l'investiture du président Joe Biden, Meta a fait part au Conseil de sa décision de suspendre jusqu'à une date indéterminée l'accès de M. Trump à ses comptes Facebook et Instagram. L'entreprise a également sollicité le Conseil pour des observations ou des recommandations sur les suspensions à suivre lorsque l'utilisateur concerné est un dirigeant politique.

La suspension indéfinie par Meta du compte Instagram et de la page Facebook de M. Trump avait, sans surprise, fait l'objet d'une attention forte de la part du public, de la sphère politique et de la presse, et lancé un débat survolté autour de l'équité, de la complexité et des limites des politiques régissant la modération du

contenu sur les plateformes des réseaux sociaux. Elle a également renforcé l'attention, a remarqué le Conseil, « sur l'intérêt d'une surveillance indépendante en ce qui concerne les décisions les plus importantes que prennent des entreprises telles que Facebook. ». Notre rôle pivot visant à essayer de résoudre cette tension a suscité un commentaire de la part du *Washington Post* :

*Les représentants et les législateurs du monde entier, à la recherche de nouveaux cadres de réglementation de l'industrie des réseaux sociaux, observent le Conseil de près. S'il est considéré comme un succès, cela peut réduire les appels à réglementation. Si c'est un échec, celui-ci peut précipiter les demandes visant à créer des garde-fous juridiques plus contraignants pour la modération du contenu dans de nombreux pays. »<sup>5</sup>*

Le *New York Times* a brièvement résumé les enjeux.<sup>6</sup>

*Cette décision a des conséquences majeures non seulement pour les politiques américaines, mais également pour la façon dont les réseaux sociaux sont réglementés, et pour l'éventuelle émergence d'un nouveau type de pouvoir d'entreprise transnational à un moment où quasiment aucun pouvoir ne semble légitime.*

Dans sa prise de décision, le Conseil s'est penché sur la question de savoir si la suspension par Meta de l'accès de M. Trump à la publication de contenu sur Facebook et Instagram était conforme aux propres politiques de Meta, à ses valeurs et aux normes internationales des droits de l'homme.

Le Conseil a pris en compte plusieurs sources d'informations : un argumentaire décisionnel fourni par Meta indiquant le raisonnement derrière sa décision ; les réponses de Meta aux questions spécifiques posées par le Conseil ; une Déclaration de créateur du contenu soumise au Conseil au nom de M. Trump par le biais du Centre américain pour le droit et la justice et d'un administrateur de page ; et, enfin, une étude indépendante.

5 "Facebook's New 'Supreme Court' Overturns Firm in First Rulings," Elizabeth Dwoskin, Craig Timberg, *The Washington Post*, January 28, 2021

6 "Trump Wants Back on Facebook. This Star-Studded Jury Might Let Him." Ben Smith, *The New York Times*, Jan. 24, 2021

## Explication de Meta

Dans les cas où Meta doit prendre une décision d'urgence suscitant un intérêt général, elle s'attache à partager sa décision et son raisonnement avec le public, le plus souvent par le biais de son Newsroom. » Concernant ce cas, Meta a stipulé qu'elle a supprimé les deux contenus publiés le 6 janvier 2021, car ils enfreignaient le Standard de la communauté sur les personnes et organisations dangereuses. Plus particulièrement, le contenu a été supprimé pour avoir enfreint sa « politique interdisant tout soutien, éloge ou représentation d'événements violents désignés ». Le contenu, selon Meta, violait également la partie de cette politique interdisant « l'éloge d'individus engagés dans des actes de violence organisée ».

Meta a fait remarquer que dans sa vidéo, bien que M. Trump ait appelé les personnes présentes à « rentrer chez [elles] en paix », il a également réitéré ses allégations de fraude électorale et suggéré un objectif commun par les mots « I know how you feel » (Je sais ce que vous ressentez). Compte tenu de l'instabilité permanente au moment de ses commentaires et de la teneur générale de ses propos, Meta avait conclu que les termes « We love you. You're very special » (Nous vous aimons. Vous êtes exceptionnels), étaient destinés à faire l'éloge des personnes ayant enfreint la loi en prenant d'assaut le Capitole. L'entreprise a également estimé que la seconde publication loue les actes perpétrés, car M. Trump a qualifié les personnes qui ont envahi le Capitole de « great patriots » (grands patriotes) et a invité le monde à se souvenir de ce jour pour toujours (« remember this day forever »).

Quant à sa décision ultérieure d'étendre à une durée indéterminée la suspension de 24 heures bloquant la fonctionnalité de publication de M. Trump, Meta a indiqué, après avoir examiné plus en profondeur l'évolution de la situation et les détails entourant les actes de violence commis au Capitole, qu'elle était arrivée à la conclusion que l'exclusion de 24 heures n'était pas suffisante pour lutter contre « le risque que M. Trump utilise sa présence sur Facebook et Instagram pour contribuer à d'éventuelles nouvelles formes de violence ». L'entreprise a cité un bulletin national d'alerte antiterroriste publié par le Département de la sécurité intérieure mettant en garde contre un « climat de menaces accrues aux États-

“

« Le jugement du cas Trump ne peut pas satisfaire tout le monde. Mais ce texte de 38 pages est une contribution importante à la réflexion à mener sur la manière de considérer la libre expression à une ère où règne le chaos informationnel. »

Alan Rusbridger  
membre du Conseil



Unis qui, selon le département, persistera dans les semaines qui suivront l'investiture présidentielle. »

Après l'investiture de M. Biden le 20 janvier, Meta a laissé le blocage en place « indéfiniment », ne stipulant aucune période déterminée ni aucun critère spécifique de restauration, laissant la durée du blocage à son entière discrétion.

Selon l'entreprise, la décision de la suspension indéterminée des comptes de M. Trump s'est appuyée sur les normes internationales des droits de l'homme, plus particulièrement sur la base de l'Article 19 du PIDCP et de l'Observation générale n° 34 du Comité des droits de l'homme des Nations Unies, lesquels autorisent des restrictions nécessaires et proportionnées de la liberté d'expression dans des situations d'urgence publique qui menacent la vie de la nation.

Bien que Meta ait un « intérêt médiatique » qui autorise le contenu qui enfreint ses politiques de rester sur la plateforme si l'entreprise considère le contenu « intéressant et dans l'intérêt du public », Meta a expliqué qu'elle n'avait pas appliqué cette autorisation aux publications posant problème dans ce cas.

**Le Conseil confirme la décision initiale de suspendre le compte, mais renverse la décision de laisser la durée indéfinie.**

Le 5 mai, nous avons publié notre décision et diffusé un ensemble de recommandations. Compte tenu de la gravité des infractions et du risque continu de

violence, la décision de Meta visant à suspendre les comptes de M. Trump le 6 janvier 2021 et à étendre cette suspension le 7 janvier était justifiée. Le Conseil approuve le fait que les deux publications enfreignaient les Standards de la communauté Facebook et les Règles de la communauté Instagram. Les règles de Facebook interdisent la publication de contenu soutenant ou faisant l'éloge de groupes, dirigeants ou individus impliqués dans des activités telles que le terrorisme, la violence ou les activités criminelles organisées.

Lors de sa décision, le Conseil a noté que :

*Au moment de la publication des contenus, les violences se poursuivaient au Capitole. Les deux publications soutenaient ou faisaient l'éloge des personnes prenant part aux violences. Les propos « We love you. You're very special » (Nous vous aimons. Vous êtes exceptionnels) de la première publication ainsi que les termes « great patriots » (grands patriotes) et « remember this day forever » (souvenez-vous de ce jour pour toujours) de la seconde publication équivalaient à soutenir ou louer les personnes impliquées dans les actes de violence et les événements survenus au Capitole ce jour-là.*

Cependant, la décision a également indiqué qu'il n'était pas approprié pour Meta d'imposer une « une suspension sine die, une sanction réputée indéterminée et arbitraire ». Elle a ajouté qu'une sanction de durée indéfinie « ne se fonde sur aucune base » présente dans les Standards de la communauté. Par ailleurs, le sens imprécis de la sanction imposée « enfreint les principes de la liberté d'expression... ». En outre, la décision stipule que « l'on ignore encore quels standards déclencheraient cette sanction ou seraient employés pour la maintenir ou la lever. »



« Nous disons à [Meta] qu'ils ne peuvent pas inventer de sanctions. Ils doivent respecter leurs propres règles. »

Helle Thorning-Schmidt  
Coprésidente du Conseil



Nous avons rejeté la demande de Meta pour que le Conseil « approuve des restrictions sine die imposées et levées sans critères explicites. » Nous avons insisté sur le fait que « [Meta] doit reconsidérer la décision arbitraire imposée le 7 janvier dernier et décider d'une sanction appropriée dans les six mois suivant la publication de la présente décision. La sanction doit être fondée sur la gravité de l'infraction et l'éventualité d'un préjudice futur. Elle doit également être conforme aux règles de [ses plateformes] en matière d'infractions graves, lesquelles doivent à leur tour être claires, nécessaires et proportionnées. »

## Meta répond à la décision du Conseil

Le 4 juin, Meta a répondu à la décision du Conseil, y compris à notre appel à réviser sa sanction indéterminée et décider d'une réponse proportionnelle cohérente avec les règles appliquées aux autres utilisateurs Facebook. La société a annoncé un nouvel ensemble de protocoles d'application « à appliquer dans des cas exceptionnels tels que celui-ci ». Conformément à ces protocoles, Meta a annoncé une nouvelle suspension du compte de M. Trump pendant deux ans, à compter du 7 janvier 2021. « Compte tenu de la gravité des circonstances qui ont conduit à la suspension du compte de M. Trump », Meta a confirmé « nous sommes convaincus que ses actions ont constitué une grave infraction de nos règles, ce qui mérite la plus grande sanction possible conformément aux nouveaux protocoles d'application ». Comme la date d'échéance de l'expiration de la sanction arrivait à grands pas, Meta a promis de :

*consulter des experts pour évaluer si le risque lié à la sécurité publique a diminué. Nous évaluerons les facteurs externes, notamment les faits de violence, les restrictions concernant les rassemblements pacifiques et les autres marqueurs de troubles civils. Si nous déterminons qu'il y a toujours un risque sérieux lié à la sécurité publique, nous étendrons la restriction à une période définie et continuerons de réévaluer la situation jusqu'à ce que le risque diminue. Une fois la suspension éventuellement levée, un ensemble strict de sanctions pouvant augmenter rapidement en sévérité sera déclenché si M. Trump commet d'autres infractions à l'avenir, jusqu'à la radiation définitive de ses pages et comptes.*

# Garantir le respect de la liberté d'expression

Sur les 20 décisions prises par le Conseil en 2021, un nombre significatif a exigé que Meta restaure le contenu dont nous avons estimé qu'il a été supprimé par erreur. Les actions de Meta dans ces cas ont été désignées comme incohérentes avec les règles, les valeurs et les engagements de l'entreprise en matière des droits de l'homme à l'échelle internationale. Dans les cinq cas représentatifs suivants, nous avons conclu que la décision de Meta visant à restreindre l'expression des utilisateurs n'était pas justifiée dans ces circonstances.



## Préoccupations au Pendjab par rapport au RSS en Inde

DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
29 avril 2021	Inde	Personnes et organisations dangereuses

En novembre 2020, un utilisateur a partagé une publication vidéo de l'entreprise de médias en ligne de langue pendjabie, Global Punjab TV. Cette publication comprenait une interview de 17 minutes du professeur Manjit Singh, qui est décrit comme « un activiste social et défenseur de la culture pendjabie ». Dans un texte d'accompagnement, l'utilisateur mentionnait que l'organisation nationaliste hindoue, Rashtriya Swayamsevak Sangh (RSS), et le parti au pouvoir en Inde, Bharatiya Janata Party (BJP) menaçaient de tuer des Sikhs, un groupe religieux minoritaire en Inde. L'utilisateur soutenait ensuite que le Premier ministre Modi en personne faisait peser la menace d'un « génocide des sikhs » sur conseil du président du RSS.

Après son signalement par un utilisateur, un modérateur humain l'a supprimée, comme il estimait qu'elle enfreignait le Standard de la communauté Facebook relatif aux personnes et organisations dangereuses. Cela a déclenché une restriction automatique sur le compte de l'utilisateur. L'utilisateur a ensuite fait appel auprès de l'entreprise. Meta a informé l'utilisateur qu'elle ne pouvait pas examiner cet appel, citant une réduction temporaire de capacité liée à la COVID-19. Suite à la sélection du cas par le Conseil, Meta a restauré avec retard le contenu, concédant que sa décision initiale était mauvaise.

Comme le Conseil l'a rappelé dans sa décision, rien dans la vidéo n'enfreignait les politiques de Meta. La publication mettait en avant les préoccupations des voix des minorités et de l'opposition en Inde prétendument discriminées par le gouvernement. En réponse à l'allégation par Meta d'un manque de capacité pour réviser l'appel de l'utilisateur en raison de la COVID-19, nous avons répété notre requête formulée dans le cadre de notre décision *Symptômes du cancer du sein et nudité* afin que tous les cas soient soumis en appel à Meta avant de parvenir au Conseil.

## Rapport sur la rapidité d'implémentation et de réponse de Meta à la suite de nos décisions.

- Conformément à nos statuts, Meta est tenue d'implémenter nos décisions **dans les 7 jours** qui suivent la publication.
- Pour les 20 décisions que nous avons publiées en 2021, Meta a restauré ou supprimé le contenu dans les 7 jours, hormis dans des cas où le contenu avait déjà été restauré.
- Pour quasiment toutes nos décisions nécessitant une action, Meta a restauré ou supprimé la publication le jour même où nous avons publié notre décision.
- Via notre Comité d'implémentation, actuellement constitué de cinq Membres du Conseil, nous continuons de presser Meta à fournir une plus grande transparence concernant ses procédures visant à identifier dans des contextes parallèles des contenus identiques à ceux ayant fait l'objet de nos décisions et à agir par rapport à ceux-ci. Cela garantirait que nos décisions soient traitées en dehors des cas particuliers et généralisées aux contenus concernés dans des contextes similaires.

### Mème « Deux boutons »

DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
20 mai 2021	États-Unis, Arménie, Turquie	Discours incitant à la haine, Contenu cruel et insensible

Le 24 décembre 2020, un utilisateur de Facebook aux États-Unis a publié un commentaire incluant une adaptation à un contexte politique turc du mème « lutte au quotidien », également appelé le mème aux deux boutons. Le mème (publié en 2014) représente un personnage de bande dessinée tentant d'appuyer sur un des deux boutons accompagnés de déclarations potentiellement contradictoires. Cette adaptation représentait le dessin en deux parties du mème aux deux boutons original, le visage du personnage de bande dessinée ayant toutefois été remplacé par un drapeau turc, accompagnés des légendes suivantes en anglais : « Le génocide arménien est un mensonge » et « Les Arméniens étaient des terroristes qui l'ont bien mérité ». Alors qu'un modérateur de contenu a estimé que le mème enfreignait le standard de la communauté de Facebook relatif au discours incitant à la haine, un autre a trouvé qu'il enfreignait le standard de la communauté relatif au contenu cruel et indélicat. Facebook l'a supprimé et en informé l'utilisateur. L'utilisateur a déposé un appel auprès de Meta, puis auprès du Conseil.

À nouveau, le contexte était la clé. Meta a basé sa décision pour retirer la publication sur la déclaration « Les Arméniens étaient des terroristes qui l'ont bien mérité ». Cependant, nous avons observé que, lorsque comprise dans son contexte, la publication était en réalité une critique de la position négationniste du gouvernement turc sur le génocide arménien. Ainsi, la publication était couverte par une exception au Standard de la communauté sur les discours incitant à la haine qui permet le partage de discours incitant à la haine « pour les condamner ou sensibiliser. » Nous avons également conclu que le contenu était couvert par l'exception en matière de satire de Facebook. Par ailleurs, nous avons appelé l'entreprise à inclure cette exception dans ses Standards de la communauté accessibles au public.





## Isolement d'Öcalan

DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
8 juillet 2021	Turquie, États-Unis	Personnes et organisations dangereuses

La décision de juillet 2021 du Conseil a mis en avant l'importance de la révision indépendante des actions de Meta en découvrant que celle-ci avait égaré une exception de politique interne pendant trois ans.

Le 25 janvier 2021, un utilisateur Instagram aux États-Unis a publié une photo d'Abdullah Öcalan, un membre fondateur du parti des travailleurs kurdes (PKK).

Une légende en anglais encourageait les lecteurs à parler de la détention d'Öcalan en prison sur l'île d'Imrali en Turquie et la nature inhumaine de son isolement solitaire. Le PKK et Öcalan sont tous deux désignés comme « entités dangereuses » en vertu de la politique de Facebook sur les personnes et organisations dangereuses, qui interdit également « l'éloge » ou le « soutien » des entités désignées.

Après qu'un modérateur a supprimé la publication pour avoir enfreint la politique sur les personnes et organisations dangereuses, l'utilisateur a fait appel de cette décision. Après le rejet de cet appel par Meta, l'utilisateur a fait appel auprès du Conseil de surveillance.

Un aspect important de cette décision qui n'a émergé qu'après la sélection par le Conseil du cas a souligné l'importance de notre révision indépendante des actions de Meta. L'entreprise a déclaré qu'elle avait localisé un ensemble d'instructions internes, établies en 2017, qui autorisaient expressément la discussion des conditions de détention d'individus désignés. Cette politique avait été élaborée, en partie, pour répondre aux préoccupations sur les conditions de l'emprisonnement d'Öcalan et pour laisser les individus en parler sur la plateforme. En annulant la décision de Meta de supprimer la publication, nous avons transmis notre préoccupation concernant le manque de transparence de Meta. « [Sa] politique consistant à supprimer par défaut le contenu témoignant du « soutien » à des individus désignés, tout en cachant au public des exceptions clés, a permis à cette erreur de passer inaperçue pendant une longue période. »

### Quels Standards de la communauté nos décisions ont-elles examinées le plus ?



**Discours incitant à la haine**

**9 décisions**



**Personnes et organisations dangereuses**

**5 décisions**



**Violence et incitation**

**2 décisions**



## Partage d'une publication d'Al Jazeera

DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
14 septembre 2021	Israël, Égypte	Personnes et organisations dangereuses

Le 10 mai 2021, un utilisateur Facebook en Égypte a partagé une publication par la page arabe vérifiée d'Al Jazeera constituée d'un texte en arabe et d'une photo. La photo représente deux hommes en uniforme militaire, à la figure masquée et portant un bandeau avec l'insigne des Brigades al-Qassam.

Le texte était le suivant : « Les chefs de la résistance de la Salle commune accordent à l'occupant jusqu'à 18h pour retirer ses soldats de la mosquée Al-Aqsa et du quartier Sheikh Jarrah. Sinon, l'auteur de cet avertissement est excusé. Abu Ubaida, porte-parole militaire des Brigades al-Qassam. » L'utilisateur a partagé la publication d'Al Jazeera et a ajouté une légende en arabe d'un seul mot : « Ooh » en arabe. Aussi bien les Brigades al-Qassam que leur porte-parole Abu Ubaida sont considérés comme dangereux en vertu du Standard de la communauté Facebook relatif aux personnes et organisations dangereuses. Meta a supprimé la publication conformément à cette politique, mais l'a restaurée après sélection de ce cas par le Conseil.

Meta n'a pas été en mesure d'expliquer pourquoi deux réviseurs avaient initialement considéré que le contenu violait sa politique, notant que les modérateurs ne sont pas tenus de consigner les motifs qui sous-tendent leurs décisions en matière de contenu individuel. La décision du Conseil a souligné que si la politique de Facebook relative aux personnes et organisations dangereuses interdisait de « diffuser des informations ou des ressources, y compris des communications officielles, au nom d'une entité désignée », elle contenait également une exception relative aux contenus publiés en qualité de « reportage d'actualité ». Dans le cas présent, le contenu était une réimpression d'un reportage largement repris par Al Jazeera sur un conflit armé survenu en mai 2021 entre les forces israéliennes et des groupes militants palestiniens en Israël et à Gaza, un territoire palestinien gouverné par le Hamas. La seule modification apportée consistait en l'ajout du commentaire superficiel « ooh ». Nous en avons donc conclu que la publication tombait sous le coup d'une exception de Facebook, à savoir le reportage d'actualité.

Ce cas concernait également le traitement équitable des utilisateurs, car cette publication, qui était une reprise d'une actualité provenant d'un média légitime, a été traitée différemment du contenu publié par l'organisme de presse lui-même. Comme nous l'avons précisé dans notre décision, de même que les organes de presse sont autorisés à publier un article pour la première fois, les individus ont tout autant le droit de le republier. Comme nous l'avons également fait remarquer, si la page d'Al Jazeera a bénéficié du système cross-check que Meta applique à certains comptes très médiatisés, tel n'a pas été le cas pour l'utilisateur ayant partagé ce contenu.



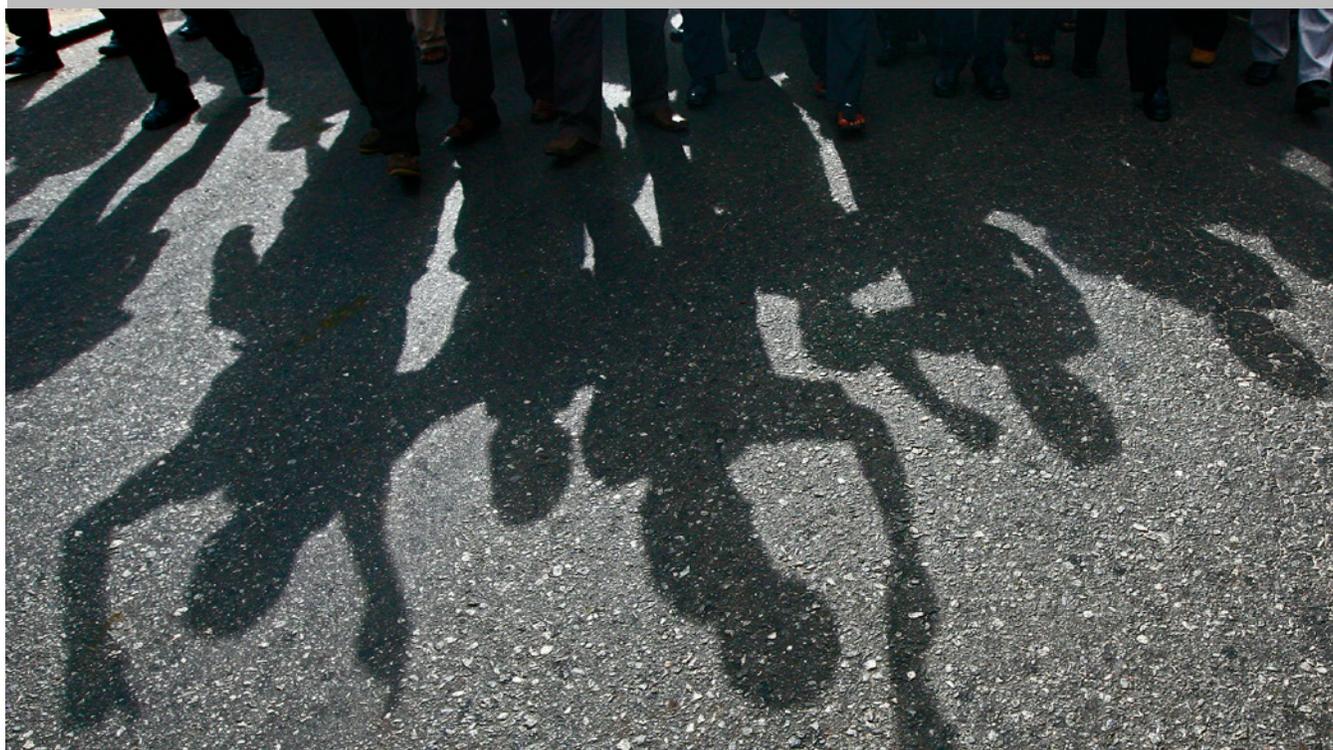
## Manifestations en Colombie

DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
27 septembre 2021	Colombie	Discours incitant à la haine

En juin 2021, la page Facebook d'un média régional colombien a partagé une vidéo montrant des manifestants marchant derrière une bannière arborant le slogan « SOS COLOMBIA ». La bande audio de la vidéo contenait des paroles scandées et des chants en espagnol, et les manifestants dénonçant le président colombien, Ivan Duque, prononçaient des insultes homophobes. C'est ce qui a incité Meta à supprimer la vidéo (provenant de Tik-Tok) pour violation du Standard de la communauté relatif aux discours incitant à la haine. Ce dernier n'autorise pas les contenus « désignant ou ciblant négativement des personnes par des insultes... lesquelles sont utilisées dans le but d'injurier. » Cette politique interdit spécifiquement les « qualificatifs insultants » qui font référence de manière négative à des caractéristiques relevant de la vie privé, comme l'orientation sexuelle.

En faisant appel de la décision de Meta de supprimer la vidéo, l'utilisateur a déclaré qu'il était un journaliste local rendant compte de l'actualité dans sa province. L'utilisateur a également affirmé qu'en partageant cette vidéo, l'organe de presse n'avait nullement l'intention de nuire, mais simplement de montrer un groupe de jeunes, manifestant pacifiquement et revendiquant des droits dans un langage qui leur est propre.

En décidant que Meta devait restaurer la vidéo, le Conseil a reconnu que si le contenu violait a priori le Standard de la communauté Facebook régissant les discours incitant à la haine (le partage d'une insulte peut en effet contribuer à générer un environnement favorisant l'intimidation et l'exclusion des personnes LGBT), Meta avait commis une erreur lors de sa suppression, en n'identifiant pas l'« intérêt médiatique » de la publication. Cette disposition peut permettre à des contenus contraires aux standards de rester en ligne, selon leur niveau d'intérêt pour le public et leur risque de préjudices.



# Protéger les utilisateurs des contenus préjudiciables

Dans certaines décisions essentielles rendues en 2021, le Conseil a confirmé la décision de Meta de supprimer du contenu, estimant qu'il peut être justifié, dans certaines circonstances, de limiter la liberté d'expression des utilisateurs. Ces décisions ont également clairement montré que le retrait de contenu par Meta pouvait être compatible avec les responsabilités de l'entreprise en matière de droits de l'homme, en particulier lorsque la suppression du droit d'expression d'un utilisateur était nécessaire et proportionnée pour respecter les droits d'autrui.



## Description de Zwarte Piet

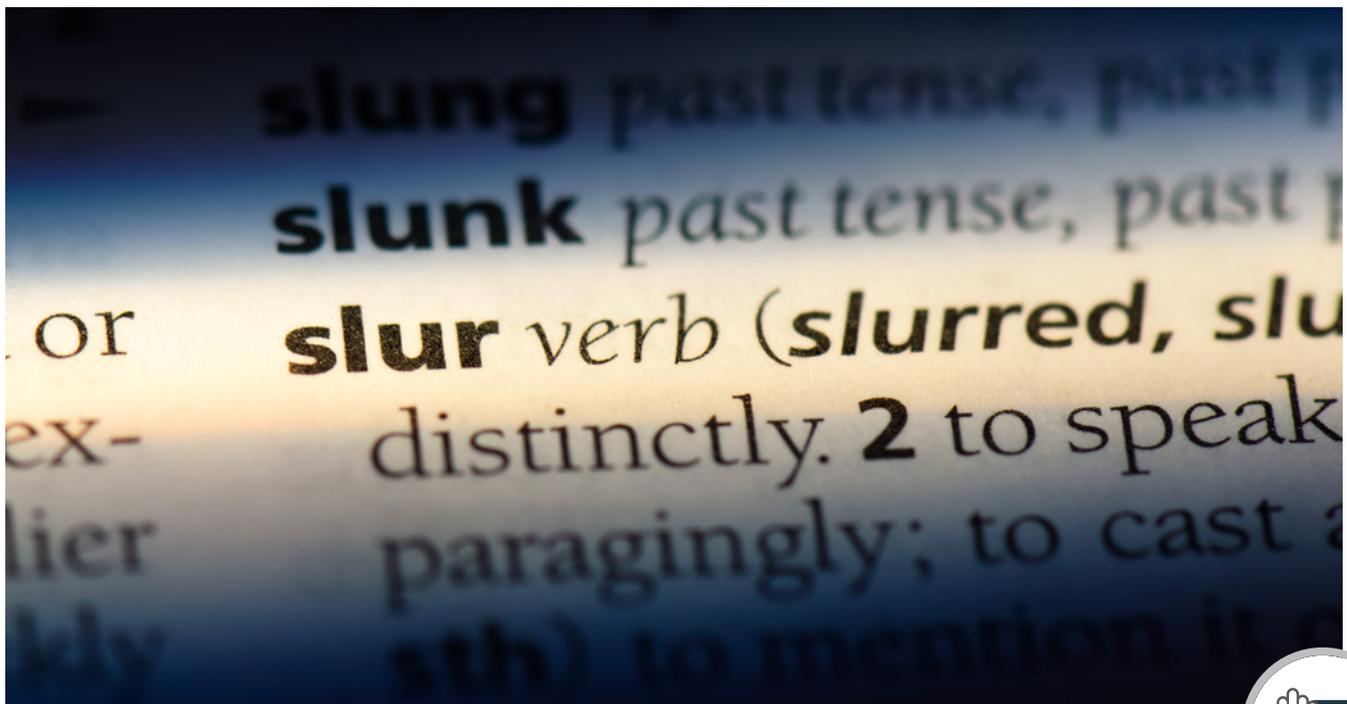
DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
13 avril 2021	les Pays-Bas	Discours incitant à la haine

Le 5 décembre 2020, un utilisateur de Facebook aux Pays-Bas a partagé une publication sur son fil d'actualité, laquelle comportait un texte en néerlandais et une vidéo de 17 secondes mettant en scène un jeune enfant rencontrant trois adultes. L'un d'eux était déguisé en « Sinterklaas », ou Saint Nicolas, tandis que les deux autres sont apparus le visage recouvert de noir, portant des perruques afro sous des chapeaux et des vêtements colorés de style renaissance. Ils représentaient une caricature néerlandaise connue sous le nom de « Zwarte Piet », également traduit en français par « père Fouettard ».

Nous avons confirmé la décision de Meta de supprimer la publication pour violation du Standard de la communauté sur les discours incitant à la haine. Le Conseil a en majorité estimé que le contenu comportait des « caricatures intimement liées à des stéréotypes négatifs et racistes et considérées par certaines parties de la société néerlandaise comme favorisant un racisme systémique aux Pays-Bas ». Pour parvenir à cette décision, la majorité du Conseil s'est référée à des études montrant que « les enfants noirs se sentaient effrayés et en danger chez eux et avaient peur d'aller à l'école pendant la fête de Sinterklaas. »

Bien qu'une partie minoritaire des membres du Conseil n'ait pas trouvé de preuves suffisantes à établir un lien direct entre ce contenu et le préjudice supposé être amoindri par la suppression de la vidéo, le Conseil a estimé, à la majorité, que le fait de laisser un tel contenu sur la plateforme contribuerait à générer un environnement discriminatoire pour les personnes noires, source d'humiliation et de harcèlement. La majorité a souligné que :

*La politique [de Meta] était claire et garantissait à grande échelle la dignité, la sécurité et la voix des personnes noires sur la plateforme. Pour atteindre cet objectif, il est acceptable de limiter la voix des personnes qui partagent des images de « blackface » dans des contextes où le racisme ne fait l'objet d'aucune condamnation.*



## Insultes en Afrique du Sud

DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
28 septembre 2021	Afrique du Sud	Discours incitant à la haine

En mai 2021, un utilisateur de Facebook a posté une publication en anglais dans un groupe public dont l'objectif revendiqué était d'« éveiller les esprits ». La publication traitait du « multiracialisme » en Afrique du Sud, et affirmait que la pauvreté, le sans-abrisme et la non-propriété s'étaient accentués pour les personnes noires dans le pays depuis 1994. L'utilisateur affirmait que les personnes blanches détenaient et contrôlaient la majorité des richesses, et que si les personnes noires riches pouvaient être propriétaires de certaines entreprises, elles n'en avaient pas le contrôle. La publication conseillait aux personnes ayant des opinions divergentes de « se faire soigner », décrivant un groupe de personnes noires utilisant des insultes raciales qui, selon Meta, violaient sa politique régissant les discours incitant à la haine.

Nous avons confirmé la décision de Meta de supprimer ce contenu, estimant que l'un des termes utilisés s'avérait « particulièrement haineux et néfaste dans le contexte sud-africain ». Nous avons souligné le fait qu'une discussion ouverte sur les inégalités socio-économiques est essentielle pour faire progresser l'égalité. Néanmoins, nous avons considéré qu'il était possible pour l'utilisateur de s'engager dans une telle discussion et de faire appel aux émotions de son public sans recourir à cette insulte. En outre, une insulte en particulier était étroitement liée à l'histoire de l'apartheid dans le pays, et son emploi reste associé à une démarche d'humiliation et d'ostracisation des personnes noires. La décision a réaffirmé l'importance pour Meta de fournir aux personnes dont le contenu a été supprimé en vertu de la politique régissant les discours incitant à la haine une explication spécifique de la règle enfreinte.



## Crimes présumés à Raya Kobo

DATE DE PUBLICATION	PAYS	STANDARD DE LA COMMUNAUTÉ
14 décembre 2021	Éthiopie	Discours incitant à la haine

Dans une décision rendue en décembre 2021, il a été constaté qu'une publication contenant des rumeurs invérifiables sur un groupe ethnique violait les Standards de la communauté Meta, et que sa suppression était conforme aux principes des droits de l'homme.

Fin juillet 2021, un utilisateur de Facebook a publié sur son fil d'actualité, en amharique, la langue éthiopienne, des allégations selon lesquelles le Front de libération du peuple du Tigré (FLPT) aurait tué et violé des femmes et des enfants et pillé des biens dans le district éthiopien de Raya Kobo et dans d'autres villes de la région d'Amhara du pays. Dans la publication, l'utilisateur affirmait que des civils tigréens avaient aidé le FLPT à perpétrer ces atrocités. La publication se terminait ainsi : « nous garantirons notre liberté en nous battant ». En outre, le contenu, qui a été visionné par des milliers de locuteurs de l'amharique pendant les 24 heures de sa mise en ligne, contenait des commentaires de certains de ces téléspectateurs, y compris des appels à la vengeance.

Un modérateur de contenu de l'équipe de révision des contenus en amharique de Meta a déterminé que la publication violait le Standard de la communauté Facebook relatif aux discours incitant à la haine et l'a supprimée. La politique proscriit les « discours violents » ciblant une personne du fait de sa race, de son ethnie ou de son origine nationale. L'utilisateur a fait appel de la décision auprès de Meta, qui, après une seconde révision par un autre modérateur de l'équipe de révision des contenus en amharique, a confirmé que la publication avait violé les politiques de Facebook.

Suite à la sélection du cas par le Conseil pour révision, Meta a identifié sa suppression comme une « erreur d'application » et a restauré le contenu concerné. Pour prendre la décision de restaurer la publication, Meta s'est appuyée sur une évaluation selon laquelle les déclarations contenues dans ladite publication ne constituaient pas un discours incitant à la haine.

Nous n'étions pas d'accord, estimant que l'explication de Meta était « incorrecte et insuffisamment détaillée ». Nous avons appliqué à cette publication le Standard de la communauté relatif à la violence et à l'incitation, estimant qu'elle violait l'interdiction de Facebook concernant les « fausses informations et rumeurs invérifiables présentant un risque d'actes de violence ou de dommages physiques imminents ». Le Conseil a estimé que « les rumeurs alléguant la complicité d'un groupe ethnique dans des atrocités de masse étaient dangereuses et augmentaient considérablement le risque d'hostilités imminentes pendant un conflit violent en cours », tel que celui qui se déroule actuellement en Éthiopie.

# Normes internationales des droits de l'homme dans le processus décisionnel du Conseil

Il est de notre responsabilité de veiller à ce que l'approche de Meta en matière de modération de contenu soit fondée sur le respect des droits de l'homme. Notre Charte et les propres engagements de Meta en matière de droits de l'homme en témoignent. Notre travail ne consiste pas seulement à déterminer si les décisions relatives au contenu de Meta suivent les règles et les valeurs de l'entreprise en la matière, nous devons également nous assurer que ces décisions sont conformes aux normes internationales des droits de l'homme.

Les normes internationales des droits de l'homme fournissent un cadre universel pour la protection de la liberté d'expression et des autres droits du genre humain. Elles fournissent également un cadre cohérent



« Il est grand temps que nous ayons une conversation sur la façon dont nous procédons pour créer une technologie qui respecte, de par sa conception, les droits de l'homme. »

Julie Owono  
Membre du Conseil



à l'échelle mondiale qui responsabilise Meta vis-à-vis des répercussions de son activité au plan mondial. La politique d'entreprise de Meta en matière de droits de l'homme définit ses engagements en la matière, y compris dans sa façon d'élaborer ses politiques de contenu. La politique indique également que les principes directeurs des Nations unies relatifs aux entreprises et aux droits de l'homme (UNGP) ont nourri l'élaboration du Conseil.

Nous pensons qu'en positionnant les droits de l'homme au cœur de notre travail, et en incitant Meta à les respecter, nous ferons de Facebook et d'Instagram de meilleurs environnements pour les utilisateurs. Au fil du temps, nous espérons que nos décisions permettront de rendre l'approche de Meta en matière de modération de contenu encore plus cohérente avec les normes internationales des droits de l'homme.

## Un nouveau défi

L'application par une société comme Meta des normes internationales des droits de l'homme à la modération de contenu est un défi relativement nouveau. Aucune autre entité ne s'est engagée dans le type d'exercice auquel se livre actuellement le Conseil.

Utiliser un cadre juridique largement créé par des États pour des États afin de décrypter les responsabilités qu'une entreprise active dans le secteur des réseaux sociaux peut avoir vis-à-vis des droits de l'homme,

soulève de nombreuses questions. Dans quelle mesure la modération de contenu par Meta doit-elle être soumise aux mêmes exigences que celles appliquées aux gouvernements lorsqu'ils réglementent la liberté d'expression en d'autres lieux ? Comment cette approche doit-elle tenir compte des défis liés à la modération de contenu à grande échelle ? Comment prendre connaissance des préjudices pouvant résulter des choix de conception de Meta, y compris l'amplification algorithmique de contenu ? Nos décisions ont permis d'analyser la façon dont les normes internationales des droits de l'homme peuvent être utilisées pour définir les responsabilités de Meta dans ce domaine, lesquelles peuvent différer de celles applicables aux États. Dans nos décisions portant sur la *Description de Zwarte Piet* et le cas des *Arméniens en Azerbaïdjan* par exemple, nous avons confirmé la décision de Meta de supprimer le contenu, tout en reconnaissant que les normes internationales des droits de l'homme ne permettraient probablement pas

à un État de sanctionner un discours similaire par des mesures pénales ou administratives.

## Bâtir une organisation axée sur les droits de l'homme

En tant qu'organisme dont les décisions affectent des personnes dans le monde entier, il est vital que nous respections les droits de l'homme dans tous les aspects de notre travail. La charte du Conseil exige de ce dernier qu'il accorde une attention particulière aux droits de l'homme lors de ses prises de décisions. Nous cherchons à suivre les UNGP en faisant preuve d'engagement et de dialogue dans notre travail, en le rendant légitime, accessible, équitable, transparent, compatible en matière de droits, ainsi qu'en le transformant en une source d'apprentissage continue. En 2020, le Conseil a commandé un rapport de suivi auprès de l'organisation à but non lucratif Business for Social Responsibility (BSR), qui a déclaré que le

### Les organisations réagissent aux travaux du Conseil

Les universitaires, la société civile et les institutions nationales et internationales ont commencé à analyser les travaux et les décisions du Conseil. Voici quelques-unes des principales étapes :

- En décembre 2020, le rapporteur spécial des Nations unies sur les questions relatives aux minorités, Fernand de Varennes, nous a exhortés à prendre en considération les droits des minorités, en affirmant que le Conseil constituait « un projet novateur et ambitieux de réglementation de la liberté d'expression en ligne, en particulier en ce qui concerne les discours incitant à la haine, un point essentiel pour garantir une protection efficace des minorités vulnérables dans le monde entier. »
- En mars 2021, un analyste en recherche juridique de la Law Library of Congress a examiné la façon dont les cinq premières décisions du Conseil ont appliqué la législation internationale des droits de l'homme aux entreprises du secteur des réseaux sociaux.
- En septembre 2021, des organisations, dont Access Now, Article 19 et 7amleh ont salué la décision du Conseil sur la publication partagée par Al Jazeera, déclarant qu'il s'agissait d'un « pas dans la bonne direction » et affirmant que le cas était « emblématique de la surapplication systématique, arbitraire et non transparente » par Facebook du Standard de la communauté relatif aux personnes et organisations dangereuses.
- En décembre 2021, la Commission américaine sur la liberté religieuse internationale a mentionné notre travail de mise en application des normes relatives aux droits de l'homme à la modération de contenu.
- « Lawfare » a également mis en place un blog relatif au Conseil de surveillance, lequel compile l'ensemble des commentaires et analyses se rapportant à nos décisions jusqu'à présent.
- Nos décisions sont également reprises par les bases de données juridiques internationales Westlaw et Lexis/Lexis+.

Conseil avait accompli des « progrès notables » par rapport aux recommandations formulées dans un rapport précédent.

## Comment les droits de l'homme sous-tendent notre conception de la liberté d'expression

Parmi les sources d'autorité guidant les décisions du Conseil figurent les normes internationales des droits de l'homme, l'article 19 du PIDCP étant parmi les plus cités. Le PIDCP est un traité mondial sur les droits de l'homme auquel Meta se réfère dans sa politique d'entreprise en la matière. De son propre chef, l'entreprise s'engage ainsi à respecter les droits de l'homme tels qu'ils sont décrits dans ce manuel. Le travail du rapporteur spécial des Nations unies sur la liberté d'opinion et d'expression, en particulier son [rapport de 2018 sur la modération du contenu](#), dans lequel il explore la façon dont les entreprises privées peuvent mieux respecter les droits de l'homme, s'est avéré particulièrement utile pour le Conseil.

L'article 19, paragraphe 2, du PIDCP stipule que « Toute personne a droit à la liberté d'expression ; ce droit comprend la liberté de rechercher, de recevoir et de répandre des informations et des idées de toute espèce, sans considération de frontières, sous une forme orale, écrite, imprimée ou artistique, ou par tout autre moyen de son choix. » Toutefois, conformément aux normes internationales des droits de l'homme, le Conseil ne considère pas que le droit à la liberté d'expression soit absolu. Bien que ce droit soit important, il peut être restreint dans certaines circonstances. L'article 19, paragraphe 3, du PIDCP prévoit un test en trois parties destiné à évaluer toute restriction de la liberté d'expression, en tenant compte de critères tels que la légalité, l'objectif légitime, la nécessité et la proportionnalité. Ces principes sont similaires, mais pas identiques, aux normes en matière de liberté d'expression que l'on trouve dans le droit national et constitutionnel de nombreuses nations où Meta exerce ses activités, y compris le premier amendement de la Constitution des États-Unis. Nous tenons compte également des restrictions en matière de liberté d'expression requises par la législation internationale des droits de l'homme lorsqu'elles sont pertinentes, y compris l'interdiction de tout appel à la haine qui constituerait une incitation à l'hostilité,

à la discrimination ou à la violence (article 20, paragraphe 2, PIDCP).

Dans notre analyse du *caractère légal* des décisions, nous examinons le degré de limpidité des règles de Meta pour les personnes qui utilisent Facebook et Instagram. Les règles doivent être faciles à trouver et à comprendre pour les utilisateurs. Dans de nombreuses décisions, nous avons exhorté Meta à publier certaines parties de ses règles qui restent inaccessibles au public, y compris les listes de personnes et d'organisations considérées comme « dangereuses » par l'entreprise.

Dans notre analyse de l'*objectif légitime* retenu dans les décisions, nous examinons si les règles de Meta sur Facebook et Instagram poursuivent des objectifs de restriction de la liberté d'expression qui sont considérés comme légitimes au regard de la législation internationale des droits de l'homme. La plupart de nos décisions identifient le respect des « droits d'autrui » comme un objectif légitime sous-tendant les politiques de contenu de Meta qui restreignent la liberté d'expression.

Dans notre analyse des critères de *nécessité et de proportionnalité*, nous examinons si l'action menée par Meta dans un cas particulier s'avérait nécessaire pour atteindre l'objectif légitime, et si les mesures imposées, telles que le retrait de contenu, étaient les moins restrictives pour y parvenir.

Dans nos 20 premières décisions, nous avons exprimé certaines préoccupations quant au fait que certaines des règles de contenu de Meta sont trop vagues, trop générales ou peu claires, ce qui a donné lieu à des recommandations visant à clarifier ces règles ou à rendre publiques des orientations internes secrètes sur l'interprétation de ces dernières. Le Conseil a attaché une importance particulière aux contenus politiques et aux contenus sensibilisant aux violations des droits de l'homme, dont la caricature politique, ainsi qu'à l'expression artistique et à l'évocation des thématiques de santé. Nous nous sommes également demandé si la suppression de contenu constituait toujours une réponse proportionnée à la publication d'un contenu susceptible d'être en lien avec un préjudice, sans pour autant en être la cause directe. Par exemple, dans deux décisions relatives à la COVID-19 (*Guérison alléguée de la COVID* et *Confinements au Brésil en raison de la COVID*), nous avons demandé si d'autres mesures

autres que la suppression des publications (comme des écrans d'avertissement, des messages d'alerte ou la rétrogradation du contenu) pouvaient atténuer les risques tout en protégeant la liberté d'expression.

Au delà du test en trois parties ci-dessus, nous avons également intégré les principes d'égalité et de discrimination dans notre processus décisionnel, en reconnaissant que différentes personnes, lorsqu'elles sont confrontées à différentes situations, peuvent être différemment impactées par l'approche de Meta en matière de modération de contenu. Cela a permis de formuler d'une part des recommandations pour que Meta puisse évaluer son impact sur les droits de l'homme dans les zones touchées par des conflits et dans des langues telles que l'arabe et l'amharique, et d'autre part des incitations à améliorer la modération du birman. L'attention portée par le Conseil aux questions d'égalité et de non-discrimination dans le cas *Symptômes du cancer du sein et nudité* a également montré comment la dépendance de Meta à l'égard de l'automatisation, laquelle a conduit à la suppression injustifiée d'informations de santé spécifiques aux femmes, pouvait avoir des impacts disproportionnés et discriminatoires sur les utilisateurs. Pour des raisons similaires, nous sommes attentifs à l'importance de la non-discrimination d'un point de vue politique.

## Les droits face aux règles – manifestations pro-Navalny en Russie

Dans la plupart de nos décisions, notre analyse des droits de l'homme a donné lieu au même résultat que notre interprétation des règles de Meta. Cependant, lors d'une étape importante pour le Conseil, la décision relative aux Manifestations pro-Navalny en Russie a été le premier cas pour lequel nous avons annulé la décision de Meta sur la base de ses responsabilités en matière de droits de l'homme, et ce alors même que le retrait était conforme aux règles de Meta.

Le 24 janvier 2021, un utilisateur en Russie a publié des photos, une vidéo et un texte décrivant les manifestations qui avaient éclaté la veille en soutien au leader de l'opposition Alexei Navalny à Saint-Petersbourg, à Moscou et dans d'autres communautés du pays. Un autre utilisateur (le contre-manifestant) a

répondu en écrivant qu'il ne savait pas ce qu'il s'était passé à Saint-Petersbourg, mais que les manifestants moscovites étaient tous des écoliers « diminués mentalement » et qu'ils avaient été « instrumentalisés sans scrupule ». Un autre utilisateur, qui a affirmé avoir participé à la manifestation de Saint-Petersbourg, a conclu en traitant le contre-manifestant de « robot lâche ». Le « contre-manifestant » a alors signalé le commentaire du manifestant à Facebook pour cause d'intimidation et de harcèlement.

Meta a retiré la publication au motif qu'elle violait son Standard en matière d'intimidation et de harcèlement, en vertu duquel un particulier peut demander à Meta de supprimer une publication contenant un « commentaire négatif sur sa personnalité ». Décidant de répondre favorablement à la demande du protestataire de restaurer la publication, le Conseil a fait l'observation suivante :

*Ce cas met en lumière les antagonismes entre les politiques de protection des personnes contre l'intimidation et le harcèlement et la nécessité de préserver la liberté d'expression. C'est tout particulièrement vrai lorsque les contestations politiques se produisent dans un pays où il existe des plaintes recevables sur l'absence de mécanismes efficaces de protection des droits de l'homme.*

Comme nous le constatons dans notre décision, même si la suppression du contenu était conforme au Standard de la communauté Meta en matière d'intimidation et de harcèlement, elle constituait une restriction inutile et disproportionnée de la liberté d'expression au regard des normes internationales des droits de l'homme.

Le Conseil a en outre plaidé en faveur du droit d'expression de l'utilisateur en estimant que si la décision initiale de Meta de supprimer le contenu « pouvait être réputée conforme à une application stricte des Standards de la communauté », cette application par trop stricte l'avait amenée « à ne pas tenir compte du contexte global et à restreindre de manière disproportionnée la liberté d'expression ».

Ce cas a montré que lorsqu'un conflit survient entre les politiques de contenu de Meta et ses responsabilités en matière de droits de l'homme, **le Conseil est prêt à donner la priorité aux droits de l'homme.**

## Études de cas sur des standards spécifiques

Les 20 premières décisions du Conseil ont tenu compte d'un large éventail de normes internationales régissant les droits de l'homme. Chaque décision publiée en 2021 fait référence à l'article 19 du PIDCP, aux principes directeurs des Nations unies relatifs aux entreprises et aux droits de l'homme et à l'observation générale 34 du Comité des droits de l'homme des Nations unies sur les libertés d'opinion et d'expression. L'application de ces normes et d'autres normes internationales relatives aux droits de l'homme à nos décisions a permis d'affiner notre compréhension et notre analyse des questions sous-jacentes qui motivent de nombreuses décisions en matière de modération de contenu. Elle nous a également aidés à parvenir à des conclusions davantage fondées sur des principes. Bien que chaque décision fasse référence à plusieurs normes régissant les droits de l'homme, nous fournissons ci-dessous quelques exemples de normes qui ont contribué de manière cruciale à nos prises de décision.

### Principes directeurs des Nations unies relatifs aux entreprises et aux droits de l'homme

Dans une décision concernant le retrait par Meta d'une publication qui montrait un collier de wampum, une forme d'expression artistique indigène d'Amérique du Nord, portant l'indication « Kill the Indian / Save the Man », nous avons exhorté Meta à s'acquitter de ses responsabilités en matière de droits de l'homme envers les communautés marginalisées, et ce en vertu du principe 17 des UNGP. Le contexte culturel de cette mention a été fourni dans un commentaire public de l'Association of American Indian Affairs, qui a identifié son auteur comme étant Richard Henry Pratt, le fondateur du premier pensionnat indien fédéral aux États-Unis. Prise dans son contexte, cette mention constituait un exemple clair de « contre-discours », une forme d'expression qui fait délibérément référence au discours incitant à la haine pour résister, et non pour soutenir ou promouvoir l'oppression et la discrimination.

En réponse à ce cas, nous avons exhorté Meta à s'acquitter de ses responsabilités en matière de droits de l'homme envers les communautés marginalisées,

en vertu du principe 17 des UNGP. Meta doit donc faire preuve de diligence raisonnable en matière de droits de l'homme pour s'assurer que ses systèmes fonctionnent de manière équitable et n'intensifient pas ce contexte d'oppression historique et permanent. Cela inclut l'identification, la prévention et l'atténuation des impacts négatifs de la modération du contenu sur la liberté d'expression des peuples autochtones luttant contre la discrimination. Enfin, nous avons remarqué que Meta améliorerait sa gestion prioritaire des risques pour les groupes marginalisés et la fourniture de preuves attestant de cette évolution.

### Observation générale n° 34

Dans le cas du *Bot du Myanmar*, qui portait sur le contenu publié à la suite d'un coup d'État non démocratique dans le pays, nous avons précisé à Meta que la critique d'un utilisateur à l'égard d'un gouvernement devait être différenciée du discours discriminatoire ciblant des personnes à cause de leur nationalité. Pour ce faire, le Conseil s'est appuyé sur l'Observation générale n°34 du Comité des droits de l'homme des Nations unies, qui fournit des orientations sur la façon dont les États doivent interpréter l'article n°19 du PIDCP, lequel protège la liberté d'expression au niveau mondial. La décision prise pour le cas *Bot du Myanmar* fait spécifiquement référence aux parties de l'observation générale évoquant l'importance du discours politique, et en particulier des échanges sans entrave au sujet des institutions publiques. Si la liberté d'expression peut être limitée dans certaines circonstances, notamment pour protéger les personnes des préjudices occasionnés par les discours discriminatoires et incitant à la haine, les gens devraient pouvoir s'exprimer en toute liberté au sujet des États-nations et des institutions publiques.

### Droit à un recours (article 2, PIDCP)

La décision relative au cas *Isolement d'Öcalan* mentionnait la déclaration du rapporteur spécial des Nations unies sur la liberté d'expression, selon laquelle le recours « devrait comporter un processus transparent et accessible pour faire appel des décisions de la plateforme, tandis que les entreprises pourraient fournir une réponse motivée qui devrait également être accessible au public. »

En l'espèce, les actions de Meta n'ont cependant pas respecté ces normes. Initialement, l'entreprise a

informé l'utilisateur qu'il ne pouvait déposer un appel auprès de l'entreprise en raison de la COVID-19. Nous avons également exprimé notre inquiétude « face à ce qui pourrait constituer un nombre important de suppressions qui n'auraient pas dû se produire ». En effet, Meta avait égaré pendant trois ans un élément des directives de la politique qui autorisait clairement le contenu de l'utilisateur. L'entreprise a affirmé qu'il était techniquement impossible de déterminer le nombre d'éléments de contenu qui avaient été supprimés à l'époque où les réviseurs n'avaient pas accès aux directives. Concernant ces préoccupations, et d'autres, nous avons critiqué Meta pour « ne pas avoir respecté le droit de recours, en violation de sa politique d'entreprise en matière de droits de l'homme. »

## Une tâche difficile, mais utile

Nous reconnaissons que l'utilisation des normes internationales des droits de l'homme pour interpréter les responsabilités en la matière d'une entreprise du secteur des réseaux sociaux est un exercice nouveau et difficile.

Pourtant, bien que cette tâche soit difficile, elle est utile. L'application des droits de l'homme internationaux à la modération de contenu peut nous aider à garantir le respect par Meta de la liberté d'expression ainsi que d'autres droits de l'homme, et à fournir un cadre cohérent et mondial permettant de demander des comptes aux entreprises du secteur des réseaux sociaux.

Constituer un corpus de travail montrant à quoi ressemble la modération de contenu basée sur des droits est un aspect central de la mission du Conseil. En 2022, nous nous appuyerons sur la valeur jurisprudentielle de nos décisions antérieures pour continuer à garantir le respect des droits de l'homme pour les personnes qui utilisent les plateformes de Meta.

# Questions posées par le Conseil à Meta dans le cadre de nos décisions

Pour nous aider à prendre des décisions et inciter Meta à faire preuve de la plus grande transparence possible, nous envoyons des questions à l'entreprise avant de délibérer sur les cas. En posant des questions spécifiques et en incluant les détails dans notre décision finale, nous pouvons demander des comptes à Meta et fournir aux chercheurs de nouvelles informations sur le fonctionnement de l'entreprise.

Concernant les 20 décisions que nous avons publiées en 2021, nos questions couvraient de nombreux domaines différents. Dans de nombreux cas, nos questions ont exploré des préoccupations spécifiques mises en avant par les utilisateurs dans leur déclaration au Conseil, ainsi que des problèmes soulevés dans les commentaires publics. Un grand nombre de nos questions concernaient la manière dont Meta traitait l'utilisateur qui avait fait appel du contenu, par exemple la raison invoquée pour laquelle son contenu avait été supprimé et si la société avait examiné son appel initial. Concernant d'autres cas, nous avons demandé plus de détails sur les règles, les systèmes automatisés et les équipes régionales de Meta.

313 questions ont été posées à Meta dans le cadre de nos 20 premières décisions. Parmi toutes ces questions, Meta en a traitées 270, a répondu partiellement à 24 autres et en a laissé 19 sans réponse. Nous présentons ces chiffres dans un objectif de transparence et ne prétendons aucunement que Meta avait de bonnes raisons de ne pas répondre.

Au cours de notre première année, Meta s'est généralement montrée réactive face aux demandes de renseignements du Conseil, répondant à la grande majorité (86 %) de nos questions. Lorsque Meta n'a pas répondu à nos questions, nous en avons fait mention à la fois dans nos rapports de transparence trimestriels et dans certaines décisions. Une description détaillée des questions auxquelles Meta n'a pas répondu et les raisons invoquées à cet égard figurent plus loin dans cette section.

Après avoir reçu quelques réponses tardives à nos questions lors des premiers cas, nous avons travaillé avec Meta pour convenir d'un système à plusieurs

## Réponses de Meta aux questions du Conseil de surveillance

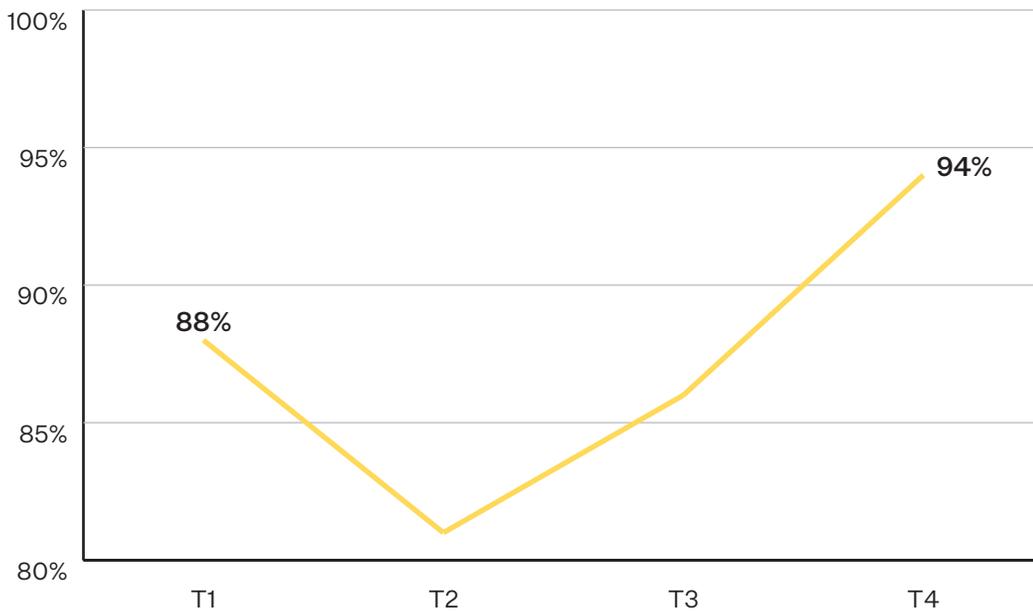


niveaux dans lequel le délai de réponse de l'entreprise dépendait du nombre de questions posées par le Conseil. Nous apprécions également le fait que Meta se soit rendue disponible pour organiser des réunions d'information sur des questions liées à ses règles, y compris le système cross-check.

### La part des questions auxquelles Meta a répondu a augmenté pour les décisions publiées au T4 2021

La part des questions auxquelles Meta a répondu pour les décisions publiées entre le T1 et le T3 2021 s'est établie à 88 % pour le T1, contre 81 % pour le T2 et 86 % pour le T3. Pour les trois décisions publiées au T4 2021, cette proportion est cependant passée à 94 %, Meta ayant répondu en totalité à 51 de nos 54 questions, et en partie à trois questions. Le T4 a été le premier où aucune des réponses de Meta n'a été classée dans la catégorie « Sans réponse ». Ces progrès sont encourageants, et nous espérons que cette tendance se poursuivra en 2022.

## Pourcentage de questions auxquelles Meta a répondu par trimestre



### Détails sur les questions auxquelles Meta n'a pas répondu

Meta n'a pas répondu à 19 des 313 questions que nous lui avons posées dans le cadre de nos décisions publiées en 2021. Veuillez trouver ci-dessous plus de détails sur ces questions, y compris les cas auxquels elles se rapportaient, leur contenu, et la raison invoquée par Meta pour son absence de réponse.

- Meta n'a pas répondu à **deux questions** dans le cas de la *publication du Myanmar à propos des musulmans*. La première question demandait si Meta avait déjà mis en place des sanctions à l'encontre de l'utilisateur et du groupe en question pour violation du Standard sur les discours incitant à la haine. Meta a répondu que les informations sur le comportement antérieur de l'utilisateur sur la plateforme étaient sans intérêt au regard de la décision du Conseil. La deuxième question demandait à Facebook Market Team la traduction en anglais des commentaires faits sur la publication. Meta a répondu qu'elle ne traduisait pas automatiquement les commentaires dans le cadre de son processus de révision de contenu, et qu'elle ne disposait pas de traductions manuelles desdits commentaires.
- Meta n'a pas répondu à **une question** au sujet du cas *Arméniens d'Azerbaïdjan*. Le Conseil a demandé pourquoi l'une des publications précédentes de l'utilisateur avait été supprimée.

Meta a répondu que les informations sur le comportement antérieur de l'utilisateur sur la plateforme étaient sans intérêt au regard de la décision du Conseil.

- Meta n'a pas répondu à **une question** concernant le cas *Citation nazie*. Le Conseil a demandé si l'utilisateur disposait d'un historique de ses publications interdites ou suspendues. Meta a répondu que les informations sur le comportement antérieur de l'utilisateur sur la plateforme étaient sans intérêt au regard de la décision du Conseil.
- Meta n'a pas répondu à **une question** concernant le cas *Protestation contre la France en Inde*. Le Conseil demandait si Meta avait déjà mis en place des sanctions à l'encontre de l'utilisateur ou du groupe en question pour violation du Standard de la communauté sur la violence et l'incitation. Meta a répondu que les informations sur le comportement antérieur de l'utilisateur sur la plateforme étaient sans intérêt au regard de la décision du Conseil.
- Meta n'a pas répondu à **deux questions** concernant le cas *Préoccupations au Pendjab par rapport au RSS en Inde*. La première question demandait quelle était la langue spécifique du contenu qui avait conduit Meta à le supprimer en vertu du Standard de la communauté relatif aux personnes et organisations dangereuses. Meta a répondu qu'elle n'était pas en mesure d'identifier la langue spécifique qui avait conduit à la conclusion

erronée que le contenu violait la politique relative aux personnes et organisations dangereuses. La deuxième question consistait à demander quel devait être le nombre de « suspensions » et d'infractions à la politique relative aux personnes et organisations dangereuses pour que le compte d'un utilisateur soit soumis à des restrictions de la part de Meta. Meta a répondu que ces informations ne pouvaient être raisonnablement exigées pour la prise de décision, conformément à l'esprit de la charte.

- Meta n'a pas répondu à **sept questions** concernant le cas *Suspension du compte de l'ancien président des États-Unis*. Les questions auxquelles Meta n'a pas répondu portaient sur les sujets suivants ; la façon dont le fil d'actualité et d'autres fonctionnalités de Facebook influent sur la visibilité du contenu de M. Trump ; si Meta avait fait des recherches, ou prévoyait d'en faire sur les décisions de conception en relation avec les événements du 6 janvier 2021 ; par ailleurs, des informations étaient également demandées sur les contenus en infraction provenant des partisans des comptes de M. Trump. Le Conseil a également posé des questions relatives à la suspension d'autres personnalités politiques et au retrait d'autres contenus ; a demandé si Meta avait été contactée par des dirigeants politiques ou leur personnel au sujet de la suspension des comptes de M. Trump ; et si la suspension ou la suppression des comptes avait un impact sur la capacité des annonceurs à cibler les comptes des partisans. Meta a déclaré que ces informations ne pouvaient être raisonnablement exigées pour la prise de décision conformément à l'esprit de la charte ; qu'il était techniquement impossible de les fournir ; qu'elles étaient couvertes par le secret professionnel de la relation avocat/client ; et/ou qu'elles ne pouvaient pas ou ne devaient pas être fournies en raison de considérations juridiques, de confidentialité, de sécurité ou de protection des données.
- Meta n'a pas répondu à **trois questions** concernant le cas *Isolement d'Öcalan*. La première question demandait si Meta pouvait déterminer combien de contenus avaient été retirés à tort lors de l'absence d'application d'une politique. La deuxième question demandait combien de contenus mentionnant Öcalan avaient été supprimés ou laissés en ligne

par Meta au cours des cinq dernières années. Pour les deux questions, Meta a répondu qu'il était techniquement impossible d'obtenir cette information. La troisième question demandait, entre autres, pourquoi les modérateurs considéraient la publication comme un appel à l'action pour soutenir Öcalan et le Parti des travailleurs kurdes (PKK). Meta a répondu qu'elle n'était pas en mesure de fournir une description détaillée des arguments ayant conduit à la suppression de la publication, car l'entreprise n'exige pas de ses réviseurs de contenu qu'ils consignent les motifs sous-tendant chacune de leurs décisions en matière de contenu.

- Meta n'a pas répondu à **deux questions** concernant le cas *Insultes en Afrique du Sud*. La première question demandait à Meta de fournir des données métriques sur trois termes offensants utilisés dans la publication. Meta a répondu qu'il était techniquement impossible de fournir les informations demandées. La deuxième question demandait si l'équipe Meta en charge des politiques publiques en Afrique du Sud comprenait des personnes originaires du pays. Meta a refusé de fournir les informations demandées, invoquant les statuts du Conseil.

Sur les 19 questions auxquelles Meta n'a pas répondu, environ deux tiers d'entre elles (13) concernaient des cas où le contenu avait été supprimé en vertu du Standard de la communauté relatif aux personnes et organisations dangereuses. Il s'agissait des cas suivants : *Citation nazie, Préoccupations au Pendjab par rapport au RSS en Inde, Suspension du compte de l'ancien président des États-Unis et Isolement d'Öcalan*. Cinq des questions restantes concernaient des cas où le contenu avait été supprimé en vertu de la politique relative aux discours incitant à la haine, tandis qu'une question concernait le contenu supprimé en vertu de la politique de Meta relative à la violence et l'incitation.

# Commentaires du public

Dans le cadre de notre processus décisionnel, les personnes et les organisations peuvent soumettre des commentaires publics. Au cours de notre première année, ces commentaires ont permis aux gens de s'exprimer au sujet de nos décisions, et ainsi de fournir une expertise essentielle au sujet de la langue, la culture, la politique et les droits de l'homme. Cette contribution est essentielle pour atteindre notre objectif, à savoir améliorer la façon dont Meta traite les personnes et les communautés dans le monde entier. À maintes reprises, les commentaires du public ont influé sur nos décisions et nos recommandations auprès de Meta.

Pour les 20 décisions publiées en 2021, nous avons reçu 9 986 commentaires publics provenant de personnes et d'organisations du monde entier. L'écrasante majorité d'entre eux (97 %) ont été soumis lors de notre décision concernant le cas *Suspension du compte de l'ancien président des États-Unis*, mais nous tenons à souligner la qualité et la valeur de l'expertise souvent reçue, même lorsque le nombre absolu de commentaires pouvait s'avérer limité. Certaines questions sont extrêmement complexes et nécessitent une

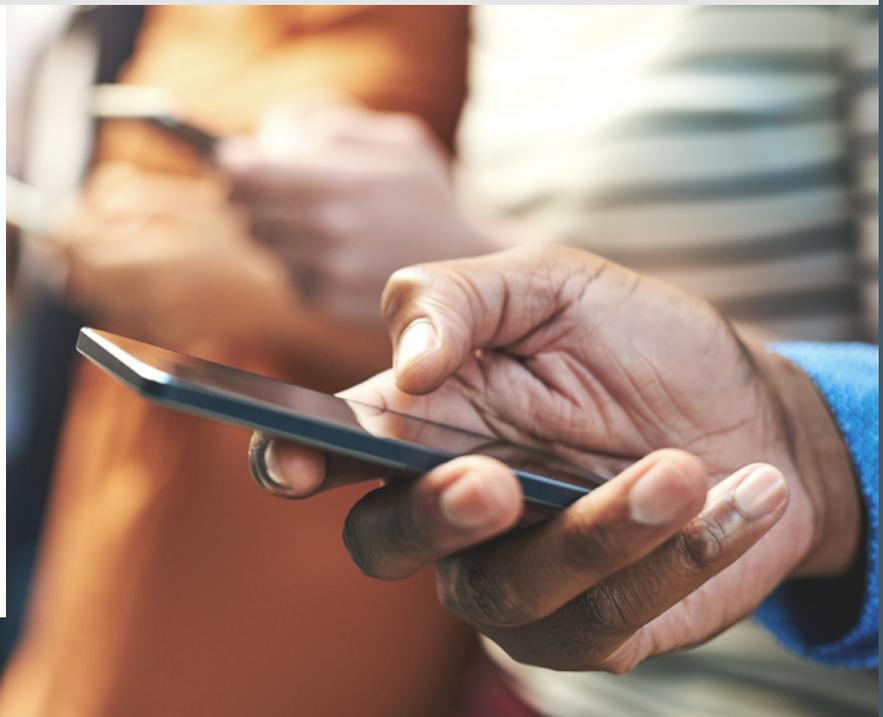
expertise spécialisée supplémentaire qui peut être fournie par les commentaires du public.

96 % des commentaires du public provenaient des États-Unis et du Canada, en raison notamment des nombreuses contributions américaines suscitées par la décision prise pour le cas *Suspension du compte de l'ancien président des États-Unis*. 95 % des commentaires du public provenaient de particuliers, les 5 % restants étant le fait d'organisations.

Dans un tiers des cas, les personnes ont soumis leurs commentaires de manière anonyme. Il est intéressant de constater que beaucoup ont choisi de faire usage de cette option, afin de partager leur expertise tout en protégeant leur droit à la vie privée. Les trois décisions pour lesquelles plus de la moitié des commentaires publiés ont été soumis de manière anonyme (*Préoccupations au Pendjab par rapport au RSS en Inde*, « *Même aux deux boutons* » et *Manifestations en Colombie*) sont toutes liées à des contenus critiquant des gouvernements actuellement au pouvoir.

## Décisions du Conseil de surveillance qui ont reçu le plus de commentaires du public en 2021 :

1. Suspension du compte de l'ancien président des États-Unis
2. Arméniens en Azerbaïdjan
3. Confinements liés à la COVID au Brésil
4. Partage d'une publication d'Al Jazeera
5. Symptômes du cancer du sein et nudité



## Comment les commentaires du public ont influencé nos décisions

Les commentaires du public ont joué un rôle essentiel dans les 20 décisions publiées par le Conseil en 2021. Les commentaires ont incité les utilisateurs à s'exprimer au sujet des cas concernés, fournissant ainsi des informations linguistiques, culturelles et contextuelles qui ont enrichi nos décisions. Au cours de l'année 2021, le Conseil a remarqué une amélioration de la qualité des commentaires soumis par le public, un plus grand nombre de commentaires étant explicitement cités en référence ou mentionnés dans les décisions publiées ultérieurement dans l'année.



« Grâce aux commentaires déposés par le public, le Conseil peut se tenir aux côtés de la société civile et faire entendre sa voix, et cette fois, Meta doit lui répondre. »

Afia Asantewaa Asare-Kyei  
Membre du Conseil



### Fournir une expertise linguistique

Dans plusieurs cas, les commentaires du public ont fourni un contexte linguistique essentiel. Dans le cas *Bot du Myanmar*, les commentaires du public s'accordaient avec l'avis des experts linguistiques du Conseil pour affirmer qu'en birman, le même mot s'emploie pour désigner un État et les personnes de cet État. Ce contexte était essentiel pour permettre au Conseil de conclure que les injures dans la publication ciblaient la « Chine » en tant qu'État et non les « Chinois » en tant que peuple. Le contenu devait donc être restauré.

Dans le cas *Insultes en Afrique du Sud*, les commentaires du public ont également aidé le Conseil à comprendre la signification locale des

insultes utilisées dans la publication. Cela nous a permis d'en déduire que l'une des insultes était un mot particulièrement haineux et néfaste dans le contexte sud-africain. Par conséquent, cela nous a confortés dans notre décision de ne pas restaurer la publication sur Facebook.

### Offrir un contexte plus large

Dans notre décision sur les *confinements liés à la COVID au Brésil*, les commentaires du public ont souligné la politisation des mesures visant à contrer la propagation de la COVID-19 dans le pays. Cela a contribué à la mise en place par le Conseil d'une recommandation selon laquelle Meta devait, compte tenu de la pandémie, privilégier la vérification des faits concernant le partage par les autorités publiques d'informations sanitaires potentiellement fausses, et ce en tenant compte du contexte local.

### Exprimer les préoccupations des utilisateurs

Les commentaires publics ont également permis aux personnes de soulever des questions plus larges liées aux cas. Les commentaires soumis pour le cas *Partage d'une publication d'Al Jazeera*, par exemple, comportaient des allégations selon lesquelles Meta supprimait ou rétrogradait de manière disproportionnée les contenus d'utilisateurs palestiniens et les contenus en arabe, notamment en comparaison du traitement réservé aux publications faisant preuve de violence anti-arabe ou anti-palestinienne en Israël. Ces commentaires ont contribué à la rédaction dans ce cas d'une recommandation exhortant Meta à vérifier si la modération des contenus en arabe et en hébreu avait été effectuée sans parti pris.

### Évaluation des engagements de Meta en matière de droits de l'homme

Le Conseil a reçu trois commentaires publics de rapporteurs spéciaux de l'ONU, ainsi que d'autres experts en droits de l'homme du monde entier. Dans le cas *Isolement d'Öcalan*, le rapporteur spécial des Nations unies sur les droits de l'homme et la lutte contre le terrorisme a noté que les règles

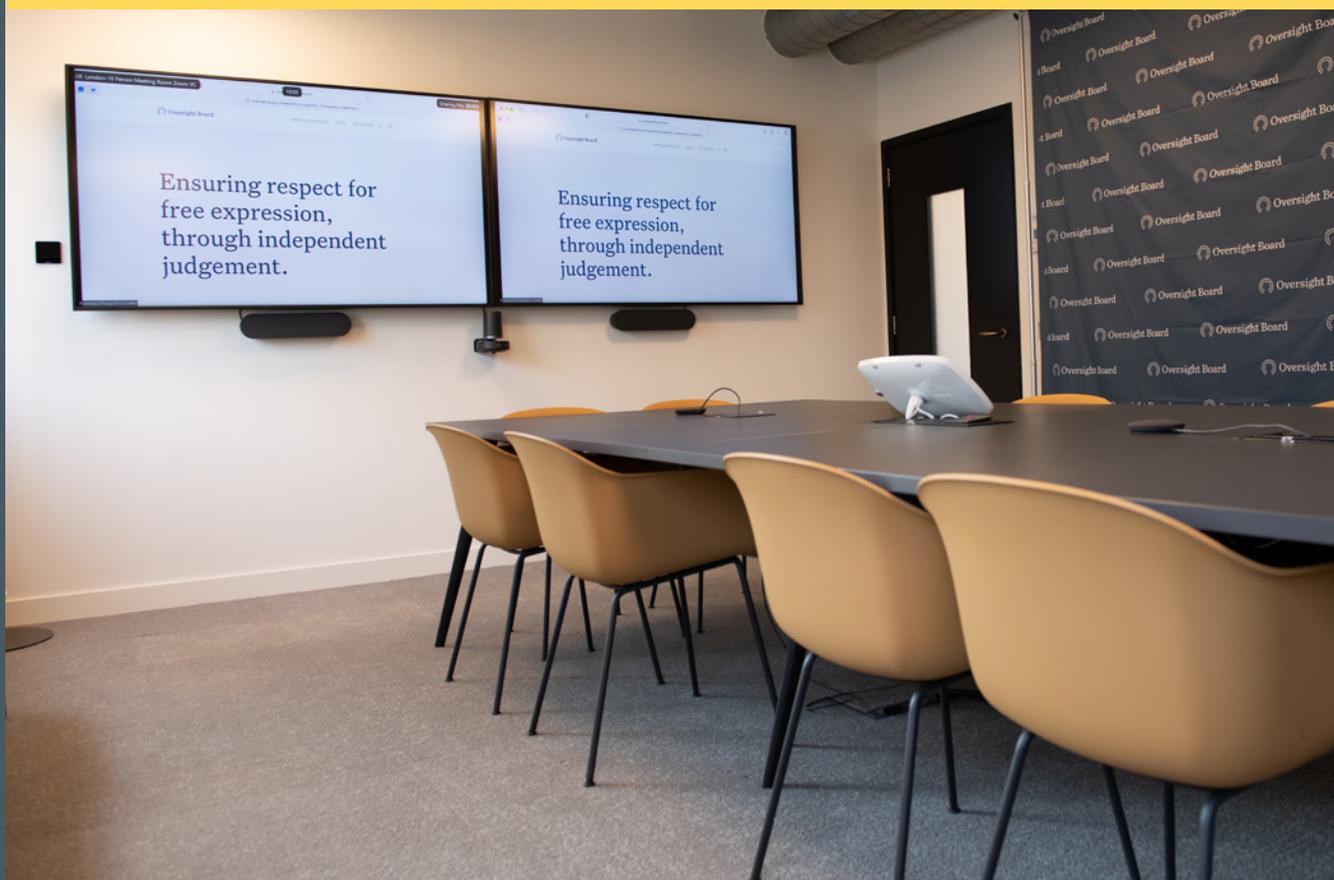
et Standards de la communauté Meta sur les personnes et organisations dangereuses « ne sont pas suffisamment conformes au droit international et peuvent dans les faits porter atteinte à certains

droits fondamentaux, y compris, mais sans s’y limiter, la liberté d’expression, d’association, de participation aux affaires publiques et de non-discrimination. »

## ENSEIGNEMENTS TIRÉS

Depuis que nous avons annoncé nos premiers cas en décembre 2020, nous avons été constamment à l’écoute des utilisateurs et avons amélioré notre processus dédié aux commentaires publics.

En réponse aux rétroaction des intervenants, nous avons fourni plus de détails dans les résumés des cas et ajouté des questions « guides » pour mettre en avant les domaines dans lesquels nous aimerions particulièrement recevoir des commentaires des intervenants. Afin d’aider le plus grand nombre de personnes et d’organisations possible à s’engager dans notre travail, nous avons prolongé la fenêtre initiale d’envoi des commentaires publics de sept à 10, puis 14 jours. Nous avons également augmenté le nombre limite de pages pour permettre l’envoi de commentaires plus détaillés. Bien que nous ayons toujours accepté les commentaires rédigés dans la langue de la publication ainsi qu’en anglais, nous prévoyons en 2022 d’offrir aux utilisateurs la possibilité de soumettre leurs commentaires dans un nombre plus important de langues.



# Recommandations



# 86

Recommandations auprès de Meta en 2021

Suite à nos recommandations, Meta s'est engagée à :



## Être plus spécifique

avec les utilisateurs lors de la suppression de publications dans lesquelles figurent un discours incitant à la haine



Diffuser de nouveaux messages dans certaines régions indiquant aux utilisateurs si la suppression de leur contenu relève d'une **procédure automatisée** ou d'un **examen manuel**



Traduire ses règles dans **des langues parlées par plus de 400 millions de personnes**



Adopter un nouveau **Protocole de politique de crise** pour ajuster sa réponse face aux situations de crise.



Fournir de nouvelles informations **sur les demandes gouvernementales**



Ce rapport opte pour une **approche inédite pilotée par les données** pour suivre l'implémentation et veiller à ce que Meta honore ses engagements pris en matière de recommandations



de nos 86 recommandations, Meta a prouvé leur implémentation ou signalé leur progression, les recommandations sur la **transparence** étant les plus susceptibles d'être concernées par l'une ou l'autre des catégories.



Les réponses de Meta aux recommandations se sont **améliorées** au fil du temps.

# Vue d'ensemble

Dans les 20 décisions que nous avons publiées en 2021, nous avons formulé 86 recommandations à l'intention de Meta sur la manière dont elle peut améliorer les politiques et les procédures qu'elle applique aux contenus publiés par ses milliards d'utilisateurs.

Bien que nos recommandations n'aient qu'une valeur consultative, Meta doit y répondre publiquement. En 2021, Meta disposait de 30 jours à compter de la publication d'une décision pour répondre à nos recommandations. Ce délai a depuis été étendu à 60 jours.

Nos recommandations ont à plusieurs reprises exhorté Meta à respecter certains principes fondamentaux en matière de transparence : rendre vos règles facilement accessibles dans les langues de vos utilisateurs ; expliquer aux gens aussi clairement que possible comment vous prenez et appliquez vos décisions ; et, lorsque les gens enfreignent vos règles, leur dire exactement ce qu'ils ont fait de mal.

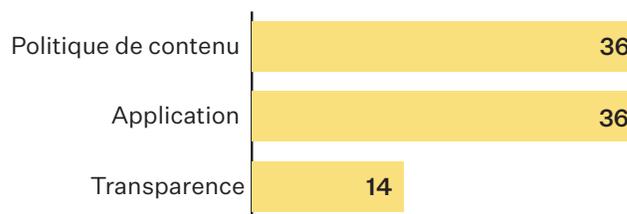
## Des engagements précoces, un impact croissant

En 2021, nous avons vu Meta prendre des engagements en réponse à nos recommandations, dont le corpus réuni représente un véritable levier pour permettre à l'entreprise d'améliorer sa façon de traiter les utilisateurs Facebook et Instagram dans le monde entier.

### Expliquer aux utilisateurs quelles sont vos règles et ce qu'ils ont fait de mal :

- En 2021, la recommandation que nous avons faite le plus souvent à Meta a accompagné six décisions : lorsque vous supprimez du contenu, **dire aux personnes quelle règle spécifique elles ont violée**. Conséquence de la réponse positive apportée par Meta, les personnes qui utilisent Facebook en anglais reçoivent maintenant un message spécifique lorsque leur contenu est supprimé pour cause de discours incitant à la haine. Meta teste actuellement ce message spécifique pour l'appliquer aux contenus en arabe, en espagnol et en portugais, ainsi qu'aux publications

## Quels types de recommandations le Conseil a-t-il formulé en 2021 ?



supprimées pour cause d'intimidation et de harcèlement.

- Dans nos trois décisions portant sur le contenu d'Instagram, nous avons exhorté Meta à informer les utilisateurs que **les Standards de la communauté Facebook s'appliquent à Instagram**. L'entreprise a déclaré qu'elle « implémentait pleinement » cette recommandation.
- En réponse à notre recommandation d'expliquer et d'illustrer par des exemples les termes clés de sa **Politique relative aux personnes et organisations dangereuses**, Meta a ajouté des définitions et des exemples pour les termes « louange », « soutien substantiel » et « représentation ». Ces clarifications sont cruciales pour permettre aux utilisateurs de comprendre ce qui est autorisé ou pas.

### Être transparent avec les utilisateurs sur la façon dont vous modérez leur contenu :

- En réponse à une recommandation émise lors de la décision relative au cas *Symptômes du cancer du sein et nudité*, Meta envoie dans certains endroits de nouveaux messages indiquant aux utilisateurs si la suppression de leur contenu est d'origine automatique ou manuelle.
- Après adoption d'une autre de nos propositions, Meta a accepté de fournir des informations au sujet des contenus supprimés pour violation des Standards de la communauté **suite à un rapport officiel d'un gouvernement**. Cela permettra d'accroître la transparence, tant au niveau des pressions exercées par les gouvernements sur Meta que de la réponse apportée par l'entreprise.

- En réponse à une autre recommandation, Meta a entamé les préparatifs nécessaires pour signaler dans son Rapport trimestriel d'application des Standards de la communauté les cas où s'applique sa clause d'**intérêt médiatique**. L'entreprise prévoit d'implémenter cette mesure d'ici la fin de 2022.

### Traiter les utilisateurs de manière équitable, où qu'ils se trouvent :

- Suite à une autre recommandation, Meta **a traduit les Standards de la communauté Facebook en punjabi et en urdu**. Elle s'est également engagée à traduire les règles de Facebook en marathi, telugu, tamoul, gujarati et dans d'autres langues en début d'année 2022. Une fois terminé, **plus de 400 millions de personnes supplémentaires**, principalement en Asie du Sud, disposeront des Standards de la communauté Facebook dans leur langue maternelle.
- Après réception de revendications indiquant que Meta avait supprimé de manière disproportionnée des publications d'utilisateurs palestiniens, nous avons exhorté l'entreprise à vérifier si **la modération des contenus en arabe et en hébreu avait été effectuée sans parti pris**. Meta a accepté de le faire.

### Améliorer la façon dont vous faites respecter vos règles, et les appliquer de manière cohérente,

- En réponse à une recommandation formulée dans notre décision relative au cas *Symptômes du cancer*

*du sein et nudité*, laquelle portait sur la suppression à tort d'une publication de sensibilisation à la maladie, Meta a mis à jour ses modèles de détection automatique de la nudité pour tenir compte des contextes de santé.

- Suite à notre décision concernant le cas *Isolement d'Öcalan*, Meta a tout d'abord supprimé le contenu, car elle estimait que l'utilisateur avait enfreint les règles relatives aux individus dangereux. Cependant, elle a ensuite révélé qu'elle avait trouvé **une exception à cette politique restée inappliquée depuis trois ans**, laquelle autorisait clairement la publication. L'exception autorisait spécifiquement les personnes à discuter des conditions de détention des personnes figurant sur sa liste d'individus appartenant à des « organisations dangereuses ». À la suite de notre recommandation pour ce cas, Meta affirme qu'elle a maintenant restauré cette directive. Cela signifie qu'à l'avenir, les utilisateurs auront une meilleure connaissance de la façon dont Meta modère le contenu lié aux individus dangereux sur ses plateformes.
- En réponse à une recommandation que nous avons émise au sujet du cas *Suspension du compte de l'ancien président des États-Unis*, et après consultation de plus de 50 experts mondiaux extérieurs à l'entreprise, Meta a adopté un Protocole de politique de crise régissant sa réponse face aux crises. Cela fournira un cadre plus cohérent et transparent sur la façon dont l'entreprise gère les situations de crise.

## ENSEIGNEMENTS TIRÉS

Malgré les problèmes liés aux réponses de Meta à nos premières recommandations, Meta et le Conseil ont pris des mesures pour améliorer le processus de recommandations au cours de l'année 2021. Bien que ce travail produise déjà des résultats, il reste des questions importantes. Comment pouvons-nous rendre nos recommandations plus pertinentes ? Comment pouvons-nous collaborer avec Meta dans ce domaine sans compromettre notre indépendance ? Et, surtout, comment pouvons-nous nous assurer que Meta honore ses engagements, par des actions qui peuvent être évaluées par le Conseil et perçues par le public dans le monde entier ? En réponse à ces questions, nous avons établi une nouvelle approche, plus rigoureuse, pour contrôler la façon dont Meta implémente nos recommandations. Elle est décrite ultérieurement dans ce chapitre.

# Comment Meta a répondu à nos recommandations

Le tableau ci-dessous présente les réponses de Meta aux recommandations formulées par le Conseil en 2021 :

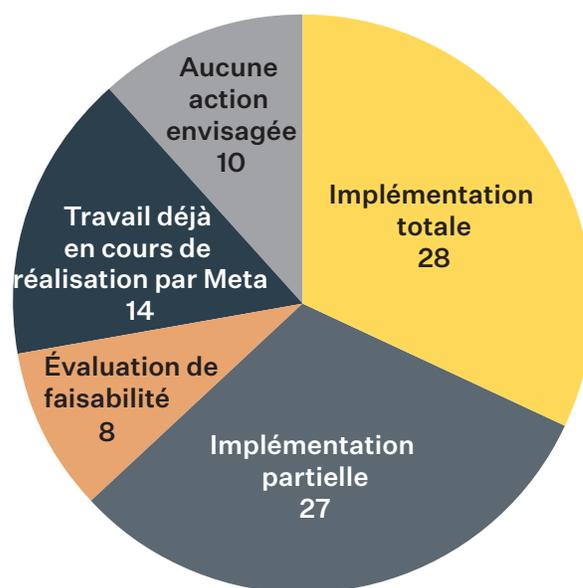
Meta s'est engagée à implémenter la plupart de nos recommandations, 55 des 87 réponses de Meta étant classées dans la catégorie « Application totale » ou « Application partielle ».

Pour huit recommandations, Meta a répondu qu'elle « évaluait la faisabilité », tandis que pour 14 autres le travail était déjà en cours. Meta a également déclaré qu'elle ne prendrait aucune mesure supplémentaire en réponse à 10 recommandations. Ces chiffres reflètent les réponses apportées dans le rapport de mise à jour du T4 soumis par Meta au Conseil, lequel a été publié en mars 2022. Le nombre de réponses de Meta (87) ne correspond pas au nombre total de recommandations que nous stipulons ici (86) en raison de différences dans le décompte de la première série de recommandations du Conseil.

## Amélioration du processus de recommandations

Au cours de l'année 2021, nous avons constaté des progrès en matière de recommandations. Le Conseil a amélioré sa façon de formuler les recommandations et Meta a parfait ses réponses. Malgré certains problèmes soulevés par les réponses de Meta à nos premières recommandations, les deux parties ont réitéré leur approche. Nous sommes entrés en 2022 avec une compréhension plus claire du processus de recommandations et de la manière dont nous pouvons l'améliorer.

En février 2021, Meta a répondu aux recommandations que nous avons formulées dans nos cinq premières décisions. Après s'être engagée à adopter la plupart de nos propositions, dans certains cas, l'entreprise a recadré ou reformulé nos recommandations. Dans d'autres cas, Meta n'a répondu qu'à certaines parties de nos propositions ou n'a pas semblé pleinement comprendre nos intentions.



Tout au long de la première année de travail, nous avons également apporté des modifications à nos processus afin de renforcer les recommandations du Conseil.

Dans la décision relative au cas *Description de Zwarte Piet*, nous avons créé dans nos décisions une nouvelle section intitulée « déclaration consultative sur les politiques », laquelle numérote clairement nos recommandations et stipule que « le Conseil demande que [Meta] fournisse une réponse individuelle à chacune d'entre elles telles qu'elles sont rédigées. » À partir d'octobre, nous avons également commencé à utiliser nos rapports de transparence trimestriels pour noter les cas où Meta avait déformé ou mal compris nos recommandations.

Nous reconnaissons également le fait que Meta a de son côté pris des mesures en 2021 pour améliorer son taux de réponse à nos recommandations.

Meta a rendu ses catégories de réponse plus spécifiques au fil du temps, en remplaçant « Engagement à agir », par « Application totale » ou « Application partielle », par exemple. L'entreprise a également demandé à l'organisation BSR (Business for Social Responsibility) d'examiner les différentes façons d'interagir avec nous pour implémenter nos recommandations.



## Rapport sur l'intemporalité de l'implémentation et de la réponse à nos recommandations par Meta

- En vertu de nos statuts, lesquels s'appliquaient en 2021, Meta devait répondre publiquement à nos recommandations dans les 30 jours suivant la publication d'une décision par le Conseil.
- En 2021, Meta a répondu à nos recommandations en respectant ce délai.

Dans l'ensemble, l'entreprise a fourni en fin d'année des réponses plus complètes qu'en début d'année à nos recommandations.

Deux exemples de réponses aux recommandations formulées à la fin de l'année 2021 illustrent ces progrès. En répondant à une recommandation émise lors de notre décision concernant le cas *Collier de wampum*, Meta a décrit deux expériences qu'elle avait menées dans des domaines significatifs du point de vue de notre proposition. Cela montre comment le processus d'élaboration et de réponse aux recommandations augmente la transparence et révèle de nouvelles informations sur les processus internes de Meta. Dans une réponse distincte à une recommandation émise lors de notre décision concernant le cas *Ayahuasca*, Meta a noté qu'elle évaluerait la faisabilité dans son Forum sur les politiques de produits, lequel publie le compte rendu de ses discussions. Nous accueillerons favorablement davantage de réponses de ce type de la part de Meta.

Meta est également devenue plus transparente avec le Conseil au sujet des preuves concrètes de son impact. Lors d'une récente évaluation de faisabilité, Meta a découvert qu'en informant les utilisateurs sur la sous-catégorie de l'infraction d'incitation à la haine qui avait provoqué la suppression de leur contenu, elle leur permettait de mieux comprendre comment et pourquoi les règles étaient appliquées. Meta a indiqué qu'elle avait constaté une augmentation statistiquement significative de la perception de sa transparence et de sa légitimité sur plusieurs marchés. À la lumière de ces résultats, Meta expérimente actuellement la possibilité de fournir aux utilisateurs des notifications

plus spécifiques pour les infractions ayant trait à l'intimidation et au harcèlement.

À plusieurs reprises en 2021, Meta a évoqué certaines difficultés quant au niveau de ressources nécessaires pour répondre à nos recommandations. Dans son premier rapport de transparence trimestriel sur le Conseil, lequel a été publié en juillet 2021, l'entreprise a déclaré :

*« [L']ampleur et la portée des recommandations du Conseil vont au-delà de l'orientation choisie pour la politique que nous avons initialement envisagée lorsque nous avons créé le Conseil, et plusieurs nécessitent des investissements s'étalant sur plusieurs mois ou années. »*

Meta a ensuite exposé ces préoccupations dans son rapport de transparence trimestriel pour le Conseil publié en novembre 2021 :

*« [L]e rythme et le volume des recommandations ne nous laissent pas suffisamment de temps pour évaluer initialement les recommandations et les implémenter. [...] [E]n moyenne, nos équipes évaluent et traitent la plupart du temps un nombre de recommandations compris entre cinq et 35. La majorité de ces recommandations nécessitent plus d'une douzaine de personnes pour en évaluer la faisabilité, ce qui est difficilement réalisable en 30 jours. Cette difficulté est encore accrue par la nécessité d'intégrer les recommandations aux priorités et projets existants, comme ceux inscrits dans les « feuilles de route » de nos équipes Integrity et des équipes associées en charge des produits.*

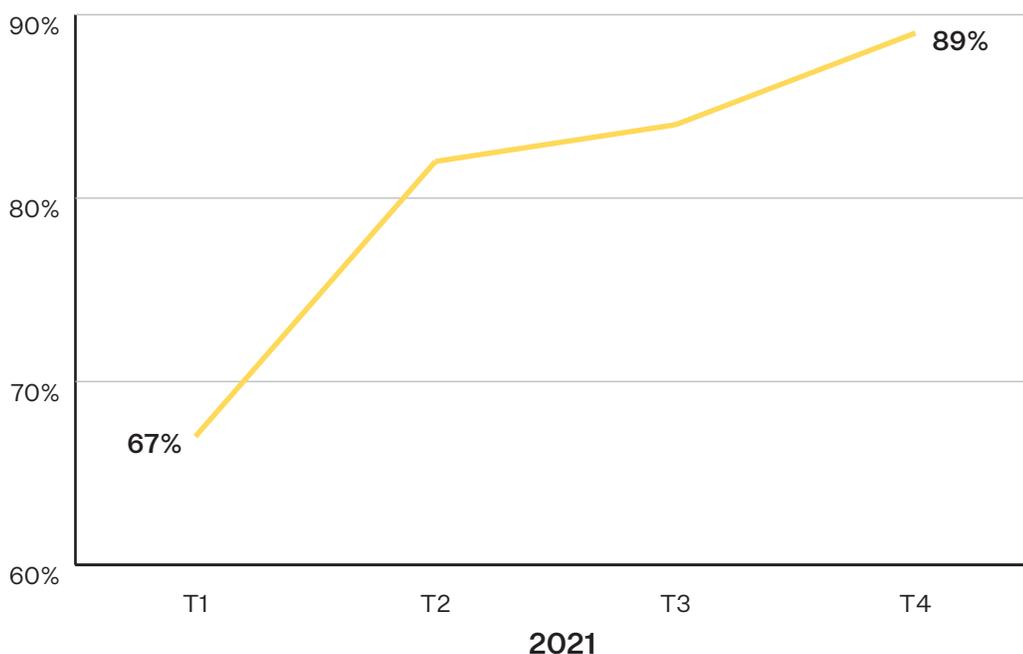
En 2021, nos recommandations ont incité Meta à apporter des changements ambitieux à la façon dont les utilisateurs font l'expérience de Facebook et Instagram. Certains de ces changements pourraient l'obliger à procéder à des modifications techniques importantes. Nous savons qu'il faut d'abord du temps pour évaluer si ces changements sont réalisables, puis pour nous fournir une réponse complète. En réponse à ces préoccupations, nous avons par la suite accepté de prolonger le délai de réponse de Meta à nos recommandations de 30 à 60 jours.

Sans compter les cas *Suspension du compte de l'ancien président des États-Unis* et *Isolement d'Öcalan*, pour lesquels Meta a spécifiquement demandé des directives détaillées sur les politiques, les décisions publiées en 2021 comprenaient, en moyenne, environ trois recommandations chacune. Nous nous sommes donc efforcés de faire en sorte que les recommandations du Conseil soient à la fois applicables et implémentées, et nous nous sommes engagés avec Meta sur les ressources nécessaires pour y parvenir.

## Amélioration des réponses de Meta à nos recommandations

Ce rapport contient une nouvelle analyse du Conseil sur les réponses de Meta aux recommandations

### Quel est le taux de réponses « Complètes » ou « Assez complètes » de Meta à nos recommandations ?



effectuées au cours de l'année 2021. Notre évaluation de la « réponse » consiste à estimer si Meta a fourni une réponse complète à la recommandation du Conseil. Ces statistiques reflètent les mises à jour trimestrielles effectuées par Meta pour le Conseil jusqu'au T3 2021.

Pour évaluer chaque réponse, nous avons posé trois questions : (1) Toutes les parties de notre recommandation sont-elles traitées ? (2) Fournit-elle un engagement à agir ? (3) Comporte-t-elle un calendrier ? Lorsque Meta répond à l'un de ces critères, nous considérons qu'elle a fourni une réponse « Assez complète », et lorsqu'elle répond à au moins deux de ces critères, nous considérons qu'elle a fourni une réponse « Complète ».

Lorsque nous examinons les réponses de Meta aux recommandations publiées au cours de différents trimestres, nous pouvons constater que la part des réponses jugées « Complètes » ou « Assez complètes » a augmenté à chaque période, passant de 67 % au T1 à 89 % au T4 2021.

Bien que des améliorations soient encore possibles, cela semble suggérer que les changements apportés au processus de recommandations en 2021 ont conduit à la fourniture de réponses plus complètes de la part de Meta.

# Des engagements à l'action : obtenir des résultats pour les utilisateurs



« Il est important de veiller à ce que les engagements pris par Meta, dont certains s'inscrivent dans la durée, ne restent pas lettre morte. »

Thomas Hughes  
Directeur du Conseil de surveillance



Alors que Meta s'est engagée à implémenter la plupart de nos recommandations en 2021, nous devons à présent nous assurer qu'elle tient ses promesses. La priorité absolue du Conseil est de s'assurer que Meta transforme ses engagements en actions concrètes qui améliorent la façon dont l'entreprise traite ses utilisateurs dans le monde entier.

Le nombre de nos recommandations ayant augmenté au cours de l'année 2021, nous avons élargi notre travail de vérification des implémentations. Le présent rapport définit une approche rigoureuse, fondée sur des données, pour suivre la façon dont Meta implémente nos recommandations et évaluer plus clairement leur impact sur les utilisateurs.

## Renforcer notre capacité à demander des comptes à Meta

À partir de la mi-2021, nous avons commencé à élargir notre capacité à suivre la façon dont Meta implémentait nos recommandations. En juillet, nous avons créé et doté en personnel l'équipe d'implémentation et de surveillance des cas afin de soutenir le Conseil. Cette équipe, qui s'est étoffée depuis, surveille et évalue les réponses et les actions de Meta, afin de comprendre l'impact de nos recommandations sur les utilisateurs Facebook et Instagram.

À la mi-2021, nous avons également créé un groupe de travail sur l'implémentation, lequel se compose de membres du Conseil et de cadres supérieurs de Meta. Ce groupe s'est réuni à plusieurs reprises pour discuter du processus de recommandations et l'améliorer, et obtenir de la part de Meta des réponses à nos questions sur la façon dont il implémente nos propositions. Plus tard dans l'année, nous avons également créé un comité d'implémentation, actuellement composé de cinq membres du Conseil, qui siège aux côtés de nos comités de sélection des cas et d'adhésion. Il s'agissait d'une volonté délibérée de placer l'implémentation au même niveau que les fonctions les plus stratégiques de notre organisation. Le Comité d'implémentation a dirigé les efforts visant à affiner les recommandations du Conseil et à s'assurer qu'elles se concentrent sur des impacts spécifiques et mesurables.

## Une nouvelle approche de l'implémentation, fondée sur les données

Ce rapport marque une nouvelle étape dans notre démarche visant à demander des comptes à Meta. Dans l'annexe de ce rapport, nous fournissons, pour la première fois, notre évaluation de la façon dont Meta a implémenté chacune de nos 86 recommandations effectuées en 2021. Ces statistiques reflètent les mises à jour de Meta jusqu'au T3 2021.

Pour mesurer les progrès accomplis par Meta en matière d'implémentation, nous avons examiné si certains critères avaient été respectés pour une recommandation donnée. Ceux-ci varient en fonction de la recommandation. Par exemple, si nous avons suggéré à Meta d'ajouter des détails supplémentaires aux Standards de la communauté accessibles au public ou de publier un rapport de diligence raisonnable sur les droits de l'homme dans l'optique d'un sujet spécifique, la publication externe de ces éléments constituerait la preuve de leur implémentation. Pour d'autres recommandations, Meta devrait fournir des

données qui ne sont pas disponibles publiquement pour prouver l'implémentation. Il pourrait s'agir par exemple de présenter les données de notification envoyées par Meta aux utilisateurs afin de prouver que l'entreprise informe ces derniers sur les règles spécifiques qu'ils ont enfreintes. Par la suite, nous mesurerons l'implémentation de nos recommandations par Meta à l'aide de quatre catégories, en mettant à jour nos évaluations sur une base trimestrielle :

- « *Implémentation prouvée* » – Meta a fourni suffisamment de données pour que nous puissions dire que cette recommandation a été appliquée.
- « *Progrès signalés* » – Meta s'est engagée à implémenter cette recommandation, mais n'a pas encore finalisé toutes les actions nécessaires.
- « *Meta a signalé l'implémentation d'une action, ou décrit un travail en cours de réalisation, mais n'a publié aucune information attestant de cette implémentation* » – Meta affirme qu'elle a implémenté cette recommandation, mais elle n'a pas fourni de preuves suffisantes pour que nous puissions en attester.
- « *Recommandation omise, refusée ou redéfinie* » – Meta ne donnera pas suite à notre proposition.

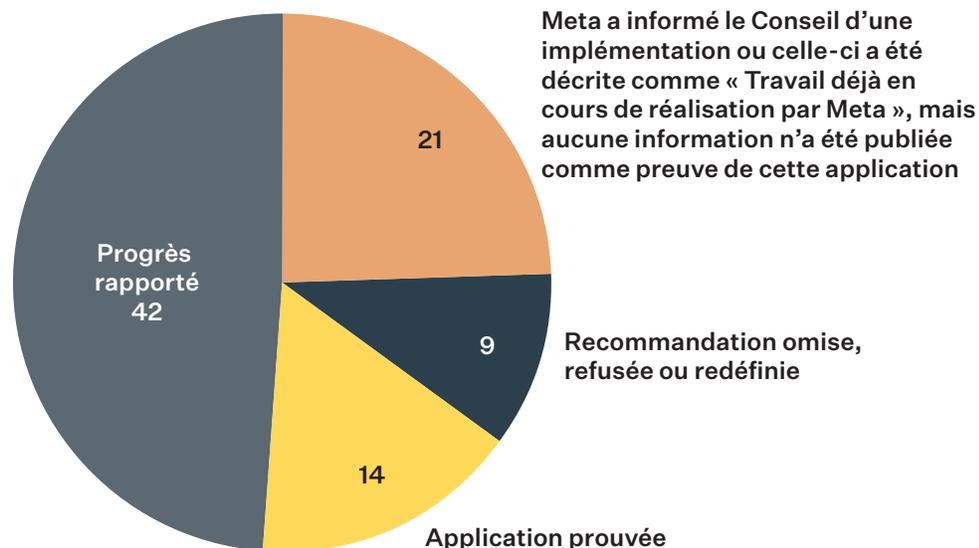
Cette nouvelle approche fondée sur des données signifie que notre évaluation de l'implémentation d'une recommandation par Meta peut parfois différer

des rapports produits par l'entreprise. Nous estimons cependant que ce type de validation menée de manière indépendante est essentiel pour demander des comptes à Meta et s'assurer que les utilisateurs ressentent l'impact de nos recommandations. Le tableau ci-dessous présente en détail une évaluation de la façon dont nos recommandations ont été implémentées par Meta :

Il montre qu'en 2021, Meta a prouvé l'implémentation de 14 de nos recommandations. Le fait que le Conseil ne formule des recommandations que depuis une période relativement courte, associé au niveau d'ambition élevé des propositions que nous avons faites durant la première année, signifie qu'il faudra du temps à Meta pour implémenter nos recommandations et attester du résultat final. En tout, 42 recommandations ont été classées dans la catégorie « Progrès signalés », laquelle comprend un grand nombre des recommandations à fort impact présentées dans l'introduction de cette section. Nous nous attendons à ce que de nombreuses recommandations de cette catégorie passent dans la catégorie « Implémentation prouvée » au cours de l'année 2022.

Pour 21 recommandations, si Meta a indiqué avoir implémenté notre proposition, ou l'a décrite comme un travail déjà en cours, elle n'a pas publié suffisamment d'informations pour nous permettre de valider cette implémentation. À l'avenir, nous encourageons

## Application de nos recommandations par Meta



Meta à partager plus de données et à accompagner de davantage de preuves ses déclarations sur l'implémentation de nos recommandations. Nous invitons également Meta à donner plus largement accès aux données de l'entreprise pour permettre la validation des progrès réalisés. Neuf recommandations sont tombées dans la catégorie « Recommandation omise, refusée ou redéfinie », ce qui signifie que Meta ne donnera aucune suite à ces propositions.

Lorsque nous avons examiné les différents types de recommandation associés aux statuts « Implémentation prouvée » ou « Progrès signalés », nous avons constaté que 79 % des recommandations de transparence et 67 % des recommandations de politique de contenu appartenaient à ces catégories, contre 58 % pour les recommandations d'application. Le taux plus faible pour les recommandations

d'application peut s'expliquer par le fait que ces propositions exigent souvent de la part de Meta qu'elle effectue des changements de plus grande ampleur sur Facebook et Instagram.

## Obtenir des résultats pour les utilisateurs Facebook et Instagram

Bien que la validation de la façon dont Meta implémente nos recommandations soit une tâche extrêmement complexe, nous avons choisi de donner la priorité à ce domaine et d'investir dedans. La raison de ces investissements est simple : ce n'est qu'en aidant Meta à concrétiser ses engagements que nous pouvons améliorer la façon dont l'entreprise traite les utilisateurs Facebook et Instagram sur le long terme, et fait face à ses responsabilités en matière de droits de l'homme.



# Tenir Meta pour responsable du cross-check

Le 13 septembre 2021, le *Wall Street Journal* a annoncé que le programme cross-check de Meta était beaucoup plus vaste dans son ampleur et ses ambitions que ce que la société avait précédemment indiqué. L'un des premiers articles d'une série s'appuyant sur des documents de l'entreprise obtenus auprès d'un lanceur d'alerte stipulait :

*Mark Zuckerberg a déclaré publiquement que [Meta] permettait à ses plus de trois milliards d'utilisateurs de disposer d'un droit d'expression à égalité avec les élites de la politique, de la culture et du journalisme, et que ses standards en matière de comportement s'appliquaient à tous, indépendamment du statut ou de la notoriété. [Mais] en privé, la société a conçu un système qui exemptait les utilisateurs très en vue de tout ou partie des règles.<sup>7</sup>*

L'article indiquait que le cross-check était passé d'un programme « initialement conçu comme une mesure de contrôle qualité des actions prises à l'encontre de comptes très médiatisés, notamment des célébrités, des politiciens et des journalistes » à un programme qui protégeait « des millions d'utilisateurs VIP » des processus d'application normaux de l'entreprise. Selon les documents du lanceur d'alerte, certains profils de prestige ont été placés sur une « liste blanche ». L'article rapporte qu'une partie de ce processus permettait à des employés individuels de Meta d'ajouter certaines entités (par exemple, des profils, des pages) à des listes qui auraient permis à des « documents violant les règles » de rester affichés indéfiniment, « en attendant de faire l'objet d'hypothétiques révisions de la part des employés de Facebook. »<sup>8</sup>

Le 21 septembre 2021, nous avons publié notre réponse suite à ces révélations :

*Au Conseil de surveillance, nous posons des questions sur le cross-check depuis un certain temps. Dans notre décision relative aux comptes de l'ancien président américain Donald Trump, nous avons averti qu'un*

*manque d'informations publiques claires sur les cross-check et sur la clause d'« intérêt médiatique » de [Meta] pourrait contribuer à renforcer l'impression selon laquelle [la société] est indûment influencée par des considérations politiques et commerciales.*

Parmi les recommandations que nous avons envoyées à Meta dans le cadre de cette décision, l'une d'entre elles l'avait exhorté à « réaliser un rapport sur les taux d'erreur relatifs et la cohérence thématique des déterminations effectuées par le biais du processus cross-check par rapport aux procédures d'application ordinaires. » Dans sa réponse, Meta a refusé d'implémenter cette proposition au motif que « le suivi de ces informations n'est pas réalisable. » Meta avait étayé cette affirmation par un lien renvoyant à un article de blog de 2018 qui indiquait : « Nous supprimons les contenus de Facebook, quels que soient leurs auteurs, lorsqu'ils enfreignent nos standards ». Le *Wall Street Journal* a toutefois rapporté qu'un examen interne effectué en 2019 par Meta avait conclu que ce billet de blog était « mensonger »... »<sup>9</sup>

Le Conseil a conclu que « À la lumière de récents développements, nous étudions à quel point [Meta] a réellement communiqué toutes les réponses relatives au cross-check, notamment concernant les listes blanches. Le Conseil a contacté [Meta] pour... mieux clarifier les informations qu'elle nous a précédemment transmises. Nous devrions recevoir un compte-rendu dans les prochains jours. »

Dans ce compte-rendu, les équipes de Meta reconnaissent « qu'elles n'auraient pas dû dire que le cross-check n'intervenait que dans «un faible nombre de décisions» » dans sa réponse à notre recommandation relative au cas de l'ancien président des États-Unis Donald Trump. Le 28 septembre 2021, soit une semaine après notre publication sur le blog, en date du 21 septembre, Meta a répondu aux questions du Conseil sous forme de demande d'avis consultatif en matière de politiques. Meta nous demandait d'étoffer notre avis de recommandations spécifiques visant à

7 “Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That’s Exempt,” Jeff Horwitz, *The Wall Street Journal*, Sept. 13, 2021

8 Ibid.

9 Ibid.

améliorer cross-check. Lors de l'annonce de cette saisine, Meta a déclaré :

*Récemment, le Conseil a montré un intérêt accru pour notre système cross-check. Cette saisine dépasse le compte-rendu que nous avons déjà fourni. Nous demandons proactivement au Conseil de nous faire part de son avis en suivant un processus formel et transparent... Nous demandons notamment au Conseil de nous aider à mieux définir les critères que nous utilisons pour déterminer ce qui justifie une deuxième vérification via cross-check, et dans notre gestion du programme.<sup>10</sup>*

Début octobre, avant de publier le premier rapport de transparence trimestriel du Conseil, nous avons annoncé notre prochaine rencontre avec Frances Haugen, lanceuse d'alerte et ancienne employée de Meta, pour discuter des importants problèmes découlant de ses actes de divulgation des procédures et politiques de modération du contenu de Meta.

*Au cours des dernières semaines, de nouvelles informations concernant l'approche de [Meta] en matière de modération de contenu ont été révélées, à la suite des actions d'une ancienne employée de [Meta], Frances Haugen. En raison du caractère sérieux des affirmations de madame Haugen, nous l'avons invitée à s'entretenir avec le Conseil dans les prochaines semaines, une invitation qu'elle a acceptée. Les membres du Conseil sont heureux d'avoir l'occasion de discuter avec madame Haugen de son expérience et de recueillir des informations qui nous permettront, par le biais de nos décisions sur les cas et de nos recommandations, d'inciter [Meta] à faire preuve de plus de transparence et de responsabilité.*

Conclusion de notre blog :

*Les choix des entreprises comme [Meta] ont des conséquences bien réelles sur la liberté d'expression et les droits de milliards de personnes dans le monde. Dans ce contexte, la transparence concernant les règles est essentielle.*

*En tant que Conseil, nous continuerons à poser des questions difficiles à Facebook et à inciter l'entreprise à s'engager pour une transparence, une responsabilité et une équité plus grandes. En fin de compte, il s'agit de la seule façon de garantir aux utilisateurs et aux utilisatrices un traitement équitable.<sup>11</sup>*

Dans notre premier rapport de transparence trimestriel, qui sera publié plus tard ce mois-ci, nous expliquons que, selon nous, les équipes internes de Meta chargées de nous fournir des informations n'ont pas été « entièrement sincères » au sujet de cross-check. « En certaines occasions », nous avons déclaré « [que les équipes] n'ont pas fourni d'informations pertinentes au Conseil... [et] dans d'autres cas, que les informations qu'elles ont fourni étaient incomplètes ». Par conséquent, nous avons défini que la future relation du Conseil avec Meta dépendrait de notre capacité à « croire que les informations fournies par Meta sont exactes, complètes et dépeignent la totalité du sujet étudié ». Suivant ce même objectif, le 21 octobre 2021, nous avons annoncé que nous avons accepté la demande de Meta de rédiger un avis consultatif en matière de politiques, suite à l'examen de son programme cross-check, et de fournir des recommandations en vue de son amélioration.

<sup>10</sup> "Requesting Oversight Board Guidance on Our Cross-Check System," Meta Blog Post by Nick Clegg, VP, Global Affairs, September 28, 2021 (Updated October 21, 2021)

<sup>11</sup> "Oversight Board to meet with Frances Haugen" Oversight Board Blog Post October 2021

# Engagement et portée

Au cours de notre première année d'existence, nous nous sommes engagés auprès d'individus et d'organisations du monde entier, en expliquant le rôle du Conseil, en encourageant la participation à nos prises de décision et en définissant l'importance pour nos utilisateurs, d'une approche globale, basée sur des principes de modération de contenu.



Julie Owono



Nighat Dad



Suzanne Nossel

## Encourager les utilisateurs à donner leur avis

Notre engagement externe a encouragé les individus et les organisations à prendre part à notre travail et à partager des commentaires publics sur des cas en cours. Nous avons organisé 19 sessions « Office Hours » expliquant comment publier des commentaires publics, qui ont réuni plus de 500 participants. Pour encourager des contributions venant du monde entier, ces sessions se sont tenues sur différents fuseaux horaires, et certaines ont fait l'objet d'une interprétation dans des langues telles que l'arabe, l'hébreu, le portugais et l'espagnol. Nous avons également organisé 12 événements et tables rondes dans les deux semaines suivants l'annonce du cas, afin d'encourager les commentaires publics sur le cas de *la suspension du compte de l'ancien président des États-Unis, Donald Trump*. Ceux-ci ont rassemblé plus de 1 000 personnes.

## Animer le débat sur la modération du contenu et les droits de l'homme

Notre approche de l'application des droits de l'homme à la modération de contenu n'a pas uniquement été façonnée par des délibérations internes entre Membres du Conseil, mais également par des discussions publiques avec des représentants du monde

universitaire, de la presse, de la société civile et des institutions internationales.

En juin, nous avons organisé des événements dans le cadre de la RightsCon, notamment une séance sur l'alignement de la modération de contenu sur les droits de l'homme. En septembre, lors de l'Assemblée générale des Nations unies, la Membre du Conseil Julie Owono a participé à deux comités d'experts des Nations Unies contre la désinformation et les propos haineux et pour la transparence. En décembre, Maina Kiai, Membre du Conseil et ancienne Rapporteuse spécial des Nations Unies s'est exprimée, lors d'un événement, sur les leçons apprises par le Conseil sur la protection des droits de l'homme et sur la gouvernance des plateformes numériques. Prenant la parole aux côtés de Peggy Hicks, directeur du Conseil des droits de l'homme des Nations Unies, elle a dit : « Au sein du Conseil de surveillance, nous avons conscience de ne combattre qu'une partie du problème. Des experts de bien d'autres domaines doivent également agir ». Nous avons également organisé un événement dans le cadre du Forum sur la gouvernance d'Internet, en décembre dernier. Celui-ci a réuni le directeur du Conseil, Thomas Hughes ; la Membre du Conseil, Afia Asantewaa Asare-Kyei et le fiduciaire du Conseil, Cherine Chalaby, autour de questions sur le droit numérique et l'auto-réglementation.

## Rendre global le débat sur la modération du contenu

Le médias sociaux concernent des événements, et des millions de personnes, dans le monde. Pourtant, trop souvent, le débat contre la modération du contenu ne prend en compte que les utilisateurs européens et nord-américains. Pour ouvrir le débat à tous, nous avons favorisé l'organisation d'événements s'intéressant aux défis de la modération du contenu rencontrés dans les pays du Sud.

En mai, les Membres du Conseil Catalina Botero-Marino et Jamal Greene ont participé à un dialogue interaméricain sur les conséquences des décisions du Conseil pour les pays du Sud, notamment en Amérique latine. En août, le Membre du Conseil Endy Bayuni a participé à un événement avec FORUM-ASIA, sur la cybersécurité centrée sur l'humain. En septembre, le Membre du Conseil Julie Owono a discuté des leçons de la modération du contenu en Afrique de l'Ouest et en Afrique centrale, lors d'un événement à laquelle participait la Commission de protection des données personnelles du Sénégal.



« Ce qui est proposé dans les pays du Nord peut porter préjudice aux communautés des pays du Sud. »

Nighat Dad  
Membre du Conseil



Enfin, en novembre, les Membres du Conseil Nighat Dad, Ronaldo Lemos et Maina Kiai se sont exprimés au Forum de Paris sur la Paix, et Nighat Dad a notamment fait remarquer que « Certaines réglementations observées dans les pays du Sud sont copiées-collées des pays du Nord et ne prennent pas en compte le contexte local et la réalité de notre région. Ce qui est proposé dans les pays du Nord peut porter préjudice aux communautés des pays du Sud. »



# À suivre : à partir de 2022

En 2021, le Conseil de surveillance a livré 20 décisions et 86 recommandations à Meta. Mises en commun, elles prouvent notre engagement pour soutenir l'entreprise, pour la développer et la guider plus de transparence. En ce début d'année 2022, au-delà d'accueillir de nouveaux membres du Conseil, nous cherchons à approfondir notre impact, à améliorer la façon dont Meta traite les personnes et les communautés du monde entier.

## Élargir notre portée

Nous dialoguons avec Meta sur l'augmentation de la portée du Conseil, notamment l'étude des appels formulés par des utilisateurs face aux décisions de l'entreprise en tant que groupes et comptes, contrairement à de simples publications d'individus, et nous espérons atteindre des progrès dans ce domaine, dès l'année prochaine. Nous

travaillons aussi sur notre rôle dans l'évaluation des programmes de modération du contenu de Meta pour le récent « Metaverse ». En 2021, l'entreprise a dépensé 10 milliards d'USD dans le Metaverse, ce qui indique qu'il s'agira d'un domaine de travail majeur dans les années à venir.

## Avoir un impact global

L'impact de Meta est mondial, la portée du travail du Conseil doit également l'être. En 2021, plus de deux tiers des utilisateurs ayant fait appel venaient d'Europe, des États-Unis et du Canada (49 % des États-Unis et du Canada et 20 % d'Europe). En revanche, seuls 14 % venaient d'Amérique latine et des Caraïbes ; 9 % d'Asie-Pacifique et d'Océanie ; 4 % du Moyen-Orient et d'Afrique du Nord ; 2 % d'Asie centrale et d'Asie du Sud et 2 % d'Afrique subsaharienne. Cette répartition reflète un déséquilibre que nous nous efforçons de résoudre.

En 2022, nous étendons notre portée et notre engagement en tant qu'intervenant pour travailler avec des décideurs politiques et des leaders au sein de la société civile en Asie, en Amérique latine, au Moyen-Orient et en Afrique. Nous construisons un réseau mondial de consultants régionaux pour encourager les personnes à faire appel et à envoyer des commentaires publics dans leurs régions respectives. Nous avons également pour projet d'augmenter le nombre de langues autres que l'anglais dans lesquelles nous acceptons des commentaires publics, et la langue du contenu en question.

## Publier des avis consultatif en matière de politiques

Au-delà de l'étude de cas individuels pour supprimer ou restaurer du contenu, le Conseil peut aussi accepter des demandes d'avis consultatif en matière de politiques de la part de Meta. Ces avis nous permettent d'étudier les politiques de l'entreprise dans un domaine en particulier et d'émettre des recommandations en vue d'améliorations. En 2021,

nous avons accepté deux demandes de ce type de la part de Meta sur le partage d'informations privées sur le lieu de résidence et le système cross-check de l'entreprise. Notre avis sur le partage d'informations privées sur le lieu de résidence a été publié en février 2022 et celui sur le système cross-check sera publié plus tard dans l'année.

## Sécuriser un meilleur accès aux données

En 2022, dans le cadre de notre travail pour garantir que Meta respecte ses engagements, le Conseil cherche à accéder aux données et aux recherches de Facebook. Cela permettra de mieux informer les cas

sélectionnés par le Conseil, nos recommandations et notre impact. Nous planifions également d'embaucher des scientifiques des données dans le

cadre de nos efforts pour comprendre et augmenter notre impact.

## Partager ce que nous avons appris jusqu'ici

Au cours de notre première année d'exercice, nous avons développé, affiné et mis en place une nouvelle approche de modération du contenu en appliquant une structure basée sur les normes internationales des droits de l'homme. Nous avons ainsi établi un ensemble de bonnes pratiques et développé une expérience solide.

En 2022, nous voulons partager ce que nous avons appris grâce à nos efforts communs pour améliorer les médias sociaux. Nous continuerons également de

développer notre rôle en tant que source de bonnes pratiques d'auto-réglementation dans un plus vaste paysage de réglementation et de modération du contenu en évolution. Nous partagerons notre expertises en matière de solutions d'auto-réglementation lors de discussions sur les politiques technologiques, et participerons à des discussions sur les innovations en matière de modération du contenu avec d'éventuels partenaires du secteur technologique.

## Créer une institution pour une réussite à long terme

Enfin, le coprésident et les Membres du Conseil, les fiduciaires et l'administration continueront de construire une institution hautement performante et durable, ainsi qu'un plan stratégique sur plusieurs années. Celui-ci comprendra des processus autonomes pour l'identification, la sélection et

l'intégration des Membres du Conseil, ainsi que les performances et la conduite. Nous continuerons également de garantir la gestion efficace et responsable des ressources fiables, notamment en développant un modèle de financement à long terme.

# Conclusion

Les informations de notre premier Rapport annuel montrent que le Conseil de surveillance a déjà un impact tangible sur la façon dont Meta travaille et offre ses services à ses utilisateurs. Chaque décision prise, chaque expérience accumulée et chaque conclusion établie représentent les premières étapes d'un parcours qui ne fait que commencer.

Le terme anodin « modération du contenu » renferme une vérité : le contenu posté et partagé sur les médias sociaux peut avoir un impact significatif, positif comme négatif, sur les vies des individus, en ligne et hors ligne. Alors que les médias sociaux ont, dans une mesure importante, rempli leur promesse initiale de réunir des milliards de personnes, ils leur ont également offert de nouvelles façons de se faire du tort.

Outre le résultat de décisions sur n'importe quel contenu, le Conseil aura un impact de plus en plus évident sur des utilisateurs qui bénéficieront d'une application claire, juste et transparente des propres politiques de Meta. Nous espérons que le travail du Conseil offrira une structure fiable à d'autres entreprises de médias sociaux et éditeurs de plateformes, régulateurs et législateurs du monde entier, aux prises avec les tensions et les défis inhérents à la modération du contenu.

Au sein du Conseil de surveillance, nous présentons ce premier Rapport annuel comme un compte-rendu exact de ce travail. Nous sommes fiers de ce que nous avons accompli jusqu'à présent, et lucides quant à la conséquente quantité de travail qui s'annonce. Ce document reflète, par-dessus tout, notre engagement collectif pour faire progresser les valeurs de transparence et de responsabilité qui, nous l'espérons, feront des plateformes de médias sociaux des endroits plus sûrs pour leurs utilisateurs. Nous continuerons de pousser Meta à prendre ses responsabilités et à améliorer continuellement la façon dont elle traite les individus et les communautés à travers le monde.

# Annexe

## Comment Meta a réagi à nos recommandations et les a appliquées<sup>12</sup>

Identifiant de la recommandation	Recommandation du Conseil de surveillance	Catégorie	Évaluation du Conseil quant	
			à la réponse de Meta	à l'implémentation
<b>Arméniens en Azerbaïdjan</b>				
1	Garantir que les raisons de toute mise en application des Standards de la communauté à leur rencontre soient toujours communiqués aux utilisateurs, y compris la règle spécifique que Facebook fait appliquer. Ce faisant, Facebook encouragerait une expression conforme à ses Standards de la communauté plutôt que d'adopter une posture antagoniste envers les utilisateurs. Dans ce cas, l'utilisateur était informé du fait que la publication enfreignait les Standards de la communauté en matière de discours incitant à la haine, mais n'avait pas été averti du fait que la publication enfreignait le standard, car elle contenait une insulte ciblant l'origine nationale. Facebook satisfait au principe de légalité dans le cas présent, mais son manque de transparence pouvait laisser croire que l'entreprise avait décidé de supprimer la publication, car l'utilisateur abordait un sujet controversé ou exprimé un point de vue avec lequel Facebook n'était pas d'accord.	A	●	
<b>Symptômes du cancer du sein et nudité</b>				
1	Améliorer la détection automatique d'images avec texte superposé afin d'assurer que les publications de sensibilisation aux symptômes du cancer du sein ne soient pas signalées à tort pour examen.	A	●	
2	Garantir que les utilisateurs soient notifiés des raisons de la mise en application des politiques relatives au contenu à leur rencontre, en leur fournissant notamment la règle spécifique des Standards de la communauté sur laquelle Facebook s'est basée pour prendre sa décision.	A	✘	
3	Dire aux utilisateurs quand l'automatisation est utilisée pour appliquer des mesures de modération contre leur contenu, y compris en leur fournissant des descriptions accessibles de ce que cela signifie.	A	●	
4	Garantir que les utilisateurs puissent faire appel à un examen manuel en cas de décisions prises par les systèmes automatisés lorsque le contenu est considéré comme ayant enfreint le standard de la communauté Facebook relatif à la nudité et aux activités sexuelles chez les adultes. Lorsque Facebook cherche à prévenir l'exploitation sexuelle des enfants ou la diffusion d'images intimes non consensuelles, elle doit appliquer ses standards de la communauté relatifs à l'exploitation sexuelle des adultes et à l'exploitation sexuelle, à la maltraitance et à la nudité d'enfants au lieu de recourir à des politiques d'hypermodération de la nudité chez les adultes. Les appels doivent toujours être disponibles dans ces cas-là de sorte que les suppressions à tort de contenus de nudité adulte consensuelle puissent être annulées.	A	✘	
5	Implémenter une procédure d'examen interne pour analyser continuellement un échantillon statistiquement représentatif des décisions de suppression de contenu automatisées afin d'annuler les erreurs de modération et d'en tirer des leçons.	A	✘	

12 Les évaluations désignées dans ce tableau prennent en compte les réactions de Meta aux recommandations du Conseil au T3 2021

**CATÉGORIE:** A– Application, T–Transparence, CP–Politique de contenu

**LA RÉPONSE DE META:** ● Complète, ● Assez complète, ✘ Incomplète

**IMPLÉMENTATION**

■ Progrès rapporté

□ Progrès rapporté

□ Meta a informé le Conseil d'une implémentation ou celle-ci a été décrite comme « Travail déjà en cours de réalisation par Meta », mais aucune information n'a été publiée comme preuve de cette application

■ Recommandation omise, refusée ou redéfinie

Identifiant de la recommandation	Recommandation du Conseil de surveillance	Catégorie	Évaluation du Conseil quant	
			à la réponse de Meta	à l'implémentation
<b>Breast cancer symptoms and nudity</b>				
6	Élargir les rapports de transparence pour dévoiler des données sur le nombre de décisions de suppression automatisées par standard de la communauté et sur la proportion de ces décisions qui ont dû être annulées à la suite d'un examen manuel.	T	●	■
7	Revoir l'explication « en bref » des Règles de la communauté Instagram afin de préciser que l'interdiction de nudité d'adultes n'est pas absolue.	CP	✘	■
8	Revoir la « version détaillée » des Règles de la communauté Instagram afin de préciser que les mamelons de femme visibles peuvent être affichés pour sensibiliser au cancer du sein ;	CP	●	■
9	Préciser que les Règles de la communauté Instagram sont interprétées en phase avec les Standards de la communauté Facebook et que ces derniers prévalent en cas d'incohérences entre les deux politiques.	CP	●	■
<b>Citation nazie</b>				
1	Garantir que les raisons de la mise en application des Standards de la communauté à leur encontre soient toujours communiquées aux utilisateurs, y compris la règle spécifique que Facebook fait appliquer (p. ex. en cas de soutien à une organisation incitant à la haine).	A	●	■
2	Expliquer et donner des exemples de l'application de termes clés utilisés dans la politique relative aux personnes et organisations dangereuses, en ce compris la signification des termes « faire l'éloge », « soutenir » et « représentation ». Ce point doit s'aligner sur les définitions utilisées dans les standards internes de mises en œuvre de Facebook. Le Standard de la communauté doit fournir des directives plus claires aux utilisateurs sur la manière dont ils peuvent préciser leur intention lorsqu'ils discutent de personnes ou d'organisations désignées comme dangereuses.	CP	●	■
3	Fournir une liste publique des organisations et des personnes dites « dangereuses » en vertu du Standard de la communauté sur les personnes et organisations dangereuses. Des exemples doivent à tout le moins être fournis. Ceci permettrait aux utilisateurs de mieux comprendre la politique et de se comporter en conséquence.	CP	●	■
<b>Revendication d'une cure contre le COVID</b>				
1	Le Conseil recommande à Facebook de présenter un Standard de la communauté clair et accessible en matière de fausses informations sanitaires afin de consolider et clarifier les règles existantes en les regroupant sur un seul emplacement (y compris en définissant des termes clés comme les fausses informations). Cette élaboration de règles devrait s'accompagner d'« hypothèses détaillées qui mettent en évidence les nuances existant dans l'interprétation et l'application de [ces] règles » permettant de clarifier les choses pour les utilisateurs (voir rapport A/HRC/38/35, paragraphe 46 (2018)). Facebook devrait mener une évaluation de l'impact sur les droits de l'homme en compagnie des parties prenantes pertinentes dans le cadre de son processus de modification des règles (UNGP, principes 18 et 19).	CP	●	■
2	Pour garantir que les mesures de mise en application en matière de fausses informations sanitaires sont les moyens les moins intrusifs de protéger la santé publique, le Conseil recommande à Facebook de : clarifier les préjudices particuliers que l'entreprise cherche à empêcher et faire preuve de transparence concernant la manière dont elle évalue le préjudice potentiel d'un contenu particulier ; procéder à une évaluation de son éventail existant d'outils pour gérer les fausses informations sanitaires ; prendre en considération le potentiel à développer des autres outils qui s'avèrent moins intrusifs que la suppression de contenu ; publier son éventail d'options de mise en application dans les Standards de la communauté, en classant ces options de la plus à la moins intrusive en se basant sur la manière dont elles empiètent sur la liberté d'expression ; expliquer quels facteurs, y compris les critères factuels, que la plateforme utilisera pour sélectionner l'option la moins intrusive afin de faire appliquer ses Standards de la communauté pour protéger la santé publique ; et clarifier quel option de mise en application s'applique à chaque règle dans les Standards de la communauté.	A	●	■

**CATÉGORIE:** A– Application, T–Transparence, CP–Politique de contenu

**LA RÉPONSE DE META:** ● Complète, ● Assez complète, ✘ Incomplète

**IMPLÉMENTATION**

■ Progrès rapporté

□ Progrès rapporté

□ Meta a informé le Conseil d'une implémentation ou celle-ci a été décrite comme « Travail déjà en cours de réalisation par Meta », mais aucune information n'a été publiée comme preuve de cette application

■ Recommandation omise, refusée ou redéfinie

Identifiant de la recommandation	Recommandation du Conseil de surveillance	Catégorie	Évaluation du Conseil quant	
			à la réponse de Meta	à l'implémentation
<b>Revendication d'une cure contre le COVID</b>				
3	Dans les cas où les utilisateurs publient des informations sur les traitements soignant le COVID-19 qui contredisent les conseils spécifiques des autorités sanitaires et où Facebook identifie un potentiel préjudice physique non imminent, le Conseil recommande à l'entreprise d'adopter une série de mesures moins intrusives. Celles-ci pourraient inclure un étiquetage qui alerte les utilisateurs de la nature litigieuse du contenu de la publication et qui fournit des liens vers les vues de l'Organisation mondiale de la Santé et des autorités sanitaires nationales. Dans certaines situations, il peut s'avérer nécessaire d'introduire un point de friction supplémentaire à une publication - par exemple, en empêchant les interactions ou le partage afin de réduire l'amplification organique et influencée par un algorithme. Il peut également être envisagé de rétrograder le contenu de façon à en réduire la visibilité dans le fil d'actualité des autres utilisateurs. Toutes les mesures de mise en application, y compris l'étiquetage ou d'autres méthodes d'introduction de point de friction, devraient être clairement communiquées aux utilisateurs et devraient pouvoir faire l'objet d'un appel.	A	✘	
4	Publier un rapport de transparence sur la manière dont les Standards de la communauté ont été appliqués durant la crise sanitaire mondiale du COVID-19. Ce rapport devrait inclure : des données en termes absolus et en pourcentage du nombre de suppressions, ainsi que des données sur les autres mesures de mise en application, sur les Standards de la communauté spécifiques concernés par une mise en application, y compris sur la proportion qui était totalement automatisée ; une répartition par type de contenu concerné par une mise en application (y compris les publications et comptes individuels ainsi que les groupes) ; une répartition par source de détection (y compris l'automatisation, le signalement par un utilisateur, les partenaires de confiance ou les autorités policières) ; une répartition par région et par langue ; des indicateurs sur l'efficacité des mesures moins intrusives (p. ex. l'impact de l'étiquetage ou de la rétrogradation) ; des données sur la disponibilité des appels tout au long de la crise, y compris le nombre total de cas où l'appel a été totalement retiré ainsi que le pourcentage d'appels automatisés ; les conclusions et leçons à retenir, y compris les informations sur toute modification apportée par Facebook afin de garantir une plus grande conformité avec ses responsabilités relatives aux droits de l'homme à l'avenir.	T	✘	
<b>Protestation contre la France en Inde</b>				
1	Pour s'assurer que les utilisateurs différencient bien les contenus autorisés de ceux qui ne le sont pas, le Conseil recommande à Facebook de leur fournir davantage d'informations sur la portée de ce standard de la communauté ainsi que sur sa mise en application. Les critères de modération de contenu devraient être publics et s'aligner sur les standards d'implémentation internes de Facebook. Spécifiquement, les critères de Facebook doivent répondre au contexte, à l'audience, à l'intention et à l'identité de l'utilisateur.	CP	●	
<b>Description de Zwarte Piet</b>				
1	Facebook doit lier la règle des Standards de la communauté en matière de discours incitant à la haine interdisant le blackface au raisonnement de l'entreprise, en ce compris les préjudices qu'elle cherche à éviter.	CP	●	
2	Conformément à la recommandation du Conseil dans le cas 2020-003-FB-UA, Facebook doit « garantir que les raisons de la mise en application des Standards de la communauté à leur encontre soient toujours communiquées aux utilisateurs, y compris la règle spécifique que Facebook fait appliquer ». En l'espèce, tout avis aux utilisateurs devrait préciser la règle sur le blackface et renvoyer aux ressources mentionnées ci-dessus qui expliquent les préjudices que cette règle cherche à éviter. Facebook devrait fournir une mise à jour détaillée de son « étude de faisabilité » des recommandations antérieures du Conseil à ce sujet, y compris la nature spécifique de toute limitation technique et la manière dont elle peut être surmontée.	A	●	

**CATÉGORIE:** A– Application, T–Transparence, CP–Politique de contenu

**LA RÉPONSE DE META:** ● Complète, ● Assez complète, ✘ Incomplète

**IMPLÉMENTATION**

■ Progrès rapporté

□ Progrès rapporté

□ Meta a informé le Conseil d'une implémentation ou celle-ci a été décrite comme « Travail déjà en cours de réalisation par Meta », mais aucune information n'a été publiée comme preuve de cette application

■ Recommandation omise, refusée ou redéfinie

Identifiant de la recommandation	Recommandation du Conseil de surveillance	Catégorie	Évaluation du Conseil quant	
			à la réponse de Meta	à l'implémentation
<b>Préoccupations au Pendjab par rapport au RSS en Inde</b>				
1	Facebook doit traduire ses Standards de la communauté et ses standards d'implémentation internes en pendjabi. Facebook doit chercher à rendre ses Standards de la communauté accessibles dans toutes les langues parlées par un grand nombre de ses utilisateurs. Cela permettrait aux utilisateurs de comprendre parfaitement les règles auxquelles ils doivent se conformer lorsqu'ils utilisent les produits de Facebook. Cela simplifierait également la tâche des utilisateurs désireux de contacter Facebook au sujet de contenus susceptibles d'enfreindre leurs droits.	CP	●	■
2	Conformément à la recommandation du Conseil dans le cas 2020-004-IG-UA, l'entreprise doit restaurer dès que possible l'examen manuel et l'accès à un processus d'appel manuel aux niveaux appliqués avant la pandémie tout en protégeant pleinement la santé de son personnel et de ses sous-traitants.	A	●	■
3	Facebook doit améliorer ses rapports de transparence pour augmenter la qualité des informations qu'elle fournit au public au sujet du taux d'erreurs de modération en ajoutant ces informations par pays, par langue et par standard de la communauté. Le Conseil soutient que des rapports de transparence plus détaillés aideront le public à détecter les endroits où les erreurs sont les plus fréquentes, y compris les répercussions spécifiques susceptibles de toucher les groupes minoritaires, et à alerter Facebook pour que l'entreprise y remédie.	T	◐	■
<b>Suspension du compte de l'ancien président des États-Unis, Donal Trump</b>				
1	Le Conseil estime qu'il n'est pas toujours utile d'opérer une nette distinction entre les dirigeants politiques et les autres utilisateurs influents. Il est important de reconnaître que les autres utilisateurs disposant de vastes audiences peuvent également contribuer à des risques sérieux de préjudice. Les mêmes règles doivent être appliquées à tous les utilisateurs sur la plateforme, mais le contexte revêt une grande importance dans l'évaluation des questions de causalité ainsi que de la probabilité et de l'imminence du préjudice. C'est le degré d'influence qu'un utilisateur exerce sur les autres qui importe. [...] Facebook doit évaluer les publications des utilisateurs influents dans leur contexte, en fonction de la manière dont elles sont susceptibles d'être interprétées, même si le message incendiaire qu'elles contiennent est formulé dans un langage destiné à éviter toute responsabilité, tel qu'un encouragement superficiel à agir pacifiquement et légalement. En l'espèce, Facebook a utilisé les six facteurs contextuels du Plan d'action de Rabat. Le Conseil estime qu'il s'agit d'un moyen utile d'évaluer les risques contextuels d'un discours potentiellement préjudiciable. Le Conseil souligne que le temps est un facteur essentiel dans de telles situations. Agir avant que les utilisateurs influents ne puissent causer un préjudice majeur doit prévaloir sur l'intérêt médiatique et d'autres valeurs de la communication politique.	CP	◐	■
2	Lorsque les publications d'utilisateurs influents présentent une forte probabilité de préjudice imminent tel qu'évalué selon les normes internationales des droits de l'homme, Facebook doit prendre des mesures pour faire appliquer ses règles dans les plus brefs délais.	A	◐	■
3	Bien que tous les utilisateurs soient tenus de respecter les mêmes règles relatives au contenu, des facteurs uniques doivent être pris en considération dans l'évaluation des discours des dirigeants politiques. Les chefs d'État et les hauts représentants de gouvernement sont davantage susceptibles de causer des préjudices que d'autres personnes. Facebook doit reconnaître que les publications des chefs d'États et d'autres hauts représentants de gouvernement sont susceptibles de présenter un risque accru d'encouragement, de légitimation ou d'incitation à la violence, soit parce que leur poste de confiance élevé insufflé davantage de force et de crédibilité à leurs propos, soit parce que leurs partisans peuvent en déduire qu'ils sont autorisés à agir en toute impunité. Dans le même temps, il demeure primordial de protéger le droit des personnes à écouter des discours politiques. Toutefois, si un chef d'État ou un haut représentant de gouvernement a publié à maintes reprises des messages présentant un risque de préjudice en vertu des normes internationales des droits de l'homme, Facebook doit suspendre le compte concerné durant une période fixe suffisante à la protection contre ce préjudice imminent.	CP	●	■

**CATÉGORIE:** A– Application, T–Transparence, CP–Politique de contenu

**LA RÉPONSE DE META:** ● Complète, ◐ Assez complète, ✘ Incomplète

**IMPLÉMENTATION**

■ Progrès rapporté

□ Progrès rapporté

■ Meta a informé le Conseil d'une implémentation ou celle-ci a été décrite comme « Travail déjà en cours de réalisation par Meta », mais aucune information n'a été publiée comme preuve de cette application

■ Recommandation omise, refusée ou redéfinie

Identifiant de la recommandation	Recommandation du Conseil de surveillance	Catégorie	Évaluation du Conseil quant	
			à la réponse de Meta	à l'implémentation
<b>Suspension du compte de l'ancien président des États-Unis, Donal Trump</b>				
4	Les périodes de suspension doivent être suffisamment longues pour décourager les comportements répréhensibles et peuvent, le cas échéant, inclure la suppression du compte ou de la page.	CP	●	
5	Les limitations de la liberté d'expression sont souvent imposées par ou sur ordre d'acteurs étatiques puissants à l'encontre de voix dissidentes et des membres de l'opposition politique. Facebook doit résister aux pressions exercées par les gouvernements pour faire taire leur opposition politique. Lors de l'évaluation des risques potentiels, Facebook doit être particulièrement attentive au contexte politique.	A	◐	
6	Dans le cadre de l'évaluation du discours politique des utilisateurs extrêmement influents, Facebook doit signaler rapidement la procédure de modération du contenu à des membres du personnel spécialisés qui connaissent bien le contexte linguistique et politique et qui sont tenus à l'écart de toute interférence politique, économique et induite. Cette analyse vise à examiner le comportement des utilisateurs extrêmement influents en dehors des plateformes Facebook et Instagram pour évaluer dûment l'ensemble du contexte pertinent des discours potentiellement préjudiciables. Par ailleurs, Facebook doit s'assurer d'allouer des ressources et une expertise adéquates à l'évaluation des risques de préjudice que présentent les comptes influents à l'échelle mondiale.	A	✘	
7	Facebook doit détailler publiquement les règles utilisées lors de l'imposition de sanctions aux comptes d'utilisateurs influents. Ces règles visent à garantir que Facebook vérifie que le risque de préjudice a diminué avant l'expiration du délai lorsqu'elle impose une suspension limitée dans le temps du compte d'un utilisateur influent pour réduire le risque de préjudice majeur. Si Facebook établit que l'utilisateur présente un risque sérieux d'incitation à des actes de violence imminents, à la discrimination ou à toute autre action illégale à ce moment-là, une autre suspension, elle aussi limitée dans le temps, devra être infligée si ces mesures sont nécessaires pour favoriser la sécurité publique et proportionnées par rapport au risque.	CP	●	
8	Lorsque Facebook met en place des procédures spéciales applicables aux utilisateurs influents, ces dernières doivent être bien documentées. Il n'apparaissait pas clairement si Facebook avait appliqué des standards différents en l'espèce et le Conseil a entendu de nombreuses préoccupations quant à l'application éventuelle de l'intérêt médiatique. Il est essentiel que Facebook remédie à ce manque de transparence et à la confusion provoquée.	T	●	
9	L'entreprise doit informer davantage les utilisateurs afin de les aider à comprendre et évaluer les procédés et critères d'application de l'intérêt médiatique. Elle doit expliquer clairement comment l'intérêt médiatique est appliqué aux comptes influents, en ce compris les dirigeants politiques et autres personnalités publiques.	CP	●	
10	En ce qui concerne la vérification croisée, Facebook doit préciser expressément les motifs, les standards et les procédés d'examen, en ce compris les critères permettant de déterminer les pages et les comptes qui peuvent être sélectionnés.	CP	◐	
11	Elle doit également signaler les taux d'erreur relatifs et la cohérence thématique des décisions effectuées sur la base de la vérification croisée par rapport aux procédures de mise en application ordinaires.	T	✘	
12	Lorsque la plateforme Facebook a fait l'objet d'une utilisation abusive de la part d'utilisateurs influents de manière à entraîner de graves répercussions négatives sur les droits de l'homme, l'entreprise doit mener une enquête approfondie sur l'incident. Elle doit déterminer l'influence qu'elle a exercée ainsi que les changements qu'elle pourrait apporter pour identifier, prévenir, atténuer et justifier des répercussions négatives à l'avenir.	A	✘	

**CATÉGORIE:** A– Application, T–Transparence, CP–Politique de contenu

**LA RÉPONSE DE META:** ● Complète, ◐ Assez complète, ✘ Incomplète

**IMPLÉMENTATION**

■ Progrès rapporté

□ Progrès rapporté

□ Meta a informé le Conseil d'une implémentation ou celle-ci a été décrite comme « Travail déjà en cours de réalisation par Meta », mais aucune information n'a été publiée comme preuve de cette application

□ Recommandation omise, refusée ou redéfinie

Identifiant de la recommandation	Recommandation du Conseil de surveillance	Catégorie	Évaluation du Conseil quant	
			à la réponse de Meta	à l'implémentation
<b>Suspension du compte de l'ancien président des États-Unis, Donal Trump</b>				
13	En l'espèce, Facebook doit entreprendre un examen approfondi de son éventuelle contribution au récit de fraude électorale et aux tensions exacerbées qui ont débouché sur les scènes de violence aux États-Unis le 6 janvier 2021. Il doit s'agir d'une réflexion ouverte sur les choix de conception et de politique opérés par Facebook qui pourraient permettre l'utilisation abusive de sa plateforme. Facebook doit exercer cette diligence raisonnable, mettre en œuvre un plan d'action pour agir conformément à ses conclusions et communiquer ouvertement sur la manière dont elle traite les répercussions négatives sur les droits de l'homme auxquelles elle a été associée.	T	✘	
14	Si des utilisateurs de Facebook ou d'Instagram sont impliqués dans des crimes atroces, de graves infractions des droits de l'homme ou des incitations au sens de l'article 20 du PIDCP, la suppression du contenu et la désactivation des comptes peuvent, outre réduire potentiellement le risque de préjudice, saper les efforts de responsabilisation, notamment en supprimant les preuves. Facebook a la responsabilité de recueillir, de protéger et, le cas échéant, de partager les informations afin de faciliter les enquêtes et les poursuites éventuelles effectuées par les autorités compétentes et des mécanismes de responsabilisation concernant des infractions graves du droit pénal international, des droits de l'homme et du droit humanitaire. La politique d'entreprise de Facebook sur les droits de l'homme doit préciser les protocoles mis en place par l'entreprise à cet égard. Elle doit également indiquer clairement la manière dont des informations autrefois publiques sur la plateforme peuvent être mises à la disposition d'enquêteurs menant des investigations conformes aux normes internationales et aux lois applicables en matière de protection des données.	T	●	
15	Facebook doit expliquer sa procédure de sanction et de suspension en matière de restriction de profils, de pages, de groupes et de comptes sur Facebook et Instagram de manière claire, exhaustive et accessible dans ses Standards et ses Règles de la communauté. Ces règles doivent fournir aux utilisateurs des informations suffisantes pour comprendre à quel moment les suspensions sont imposées (en ce compris les exceptions ou autorisations applicables) et comment les sanctions sont définies.	CP	●	
16	Facebook doit également fournir aux utilisateurs des informations accessibles sur le nombre d'infractions, de suspensions et de sanctions qui leur ont été infligés, ainsi que sur les conséquences qui découleront de futures infractions.	A	●	
17	L'entreprise doit inclure dans ses rapports de transparence le nombre de limitations de profils, pages et comptes imposées, en ce compris pourquoi et comment les mesures de mise en application ont été prises, avec des informations ventilées par région et par pays.	T	✘	
18	Facebook doit développer et publier une politique régissant sa réponse face aux crises ou aux situations nouvelles lorsque ses procédures habituelles ne permettent pas de prévenir ou d'empêcher un préjudice imminent. Bien que ces situations ne puissent pas toujours être anticipées, les directives de Facebook doivent définir des paramètres appropriés pour ces actions, en ce compris l'obligation de revoir sa décision dans un délai fixé.	CP	●	
<b>« Même aux deux boutons »</b>				
1	Prendre les mesures techniques nécessaires pour s'assurer que les avis aux utilisateurs fassent référence au standard de la communauté appliqué par l'entreprise. Si Facebook détermine que (i) le contenu n'enfreint pas le standard de la communauté initialement communiqué à l'utilisateur et que (ii) le contenu enfreint un autre standard de la communauté, l'utilisateur devrait en être informé correctement et bénéficier d'une nouvelle occasion de faire appel. L'utilisateur devrait toujours avoir accès aux bonnes informations avant de contacter le Conseil.	A	●	
2	Jouter l'exception en matière de satire, qui n'est actuellement pas accessible aux utilisateurs, à la partie accessible au public de son standard de la communauté relatif au discours incitant à la haine.	CP	●	

**CATÉGORIE:** A– Application, T–Transparence, CP–Politique de contenu

**LA RÉPONSE DE META:** ● Complète, ● Assez complète, ✘ Incomplète

**IMPLÉMENTATION**

■ Progrès rapporté

□ Progrès rapporté

□ Meta a informé le Conseil d'une implémentation ou celle-ci a été décrite comme « Travail déjà en cours de réalisation par Meta », mais aucune information n'a été publiée comme preuve de cette application

□ Recommandation omise, refusée ou redéfinie

Identifiant de la recommandation	Recommandation du Conseil de surveillance	Catégorie	Évaluation du Conseil quant	
			à la réponse de Meta	à l'implémentation
<b>« Même aux deux boutons »</b>				
3	S'assurer que les procédures adéquates sont mises en place pour évaluer correctement les contenus satiriques et leur contexte respectif. Cela nécessite notamment de fournir aux modérateurs de contenu : (i) la possibilité de contacter les équipes d'exploitation locales de Facebook afin de recueillir les informations culturelles et le contexte pertinents ; et (ii) suffisamment de temps pour consulter les équipes d'exploitation locales et pour procéder à l'évaluation. Facebook doit s'assurer que ses politiques à l'attention des modérateurs de contenu encouragent ces derniers à se renseigner davantage ou à faire remonter le cas lorsqu'un modérateur n'est pas certain qu'un même est satirique ou non.	A	●	
4	Permettre aux utilisateurs d'indiquer dans leur appel que leur contenu est couvert par l'une des exceptions à la politique relative au discours incitant à la haine. Cela concerne notamment les exceptions en matière de contenu satirique ainsi que les cas où les utilisateurs partagent du contenu à caractère haineux afin de sensibiliser à la haine et de la condamner.	A	●	
5	S'assurer que l'examen manuel est accordé en priorité aux appels fondés sur des exceptions aux politiques.	A	●	
<b>Manifestations de soutien à Navalny</b>				
1	Facebook doit modifier et remanier son Standard de la communauté Facebook sur le harcèlement et l'intimidation afin de préciser la relation qui unit la justification de sa politique en matière de harcèlement et d'intimidation et la section « Ne pas » ainsi que les autres règles de limitation du contenu qui la suivent.	CP	●	
2	Différencier le harcèlement de l'intimidation et donner des définitions qui distinguent ces deux actes. Le Standard de la communauté doit également expliquer clairement aux utilisateurs en quoi le harcèlement et l'intimidation diffèrent des discours uniquement offensants qui peuvent être protégés par la législation internationale des droits de l'homme.	CP	◐	
3	Définir clairement son approche des différentes catégories d'utilisateurs cibles et donner des exemples de chaque catégorie de cibles (p. ex. qui peut être considéré comme une personnalité publique). Présenter le Standard de la communauté sur le harcèlement et l'intimidation par catégories d'utilisateurs actuellement énumérées dans la politique.	CP	●	
4	Inclure des exemples de contenus en infraction et autorisés dans le Standard de la communauté sur le harcèlement et l'intimidation en vue de clarifier les lignes d'action établies ainsi que la manière dont ces distinctions peuvent reposer sur le statut identitaire de la cible.	CP	●	
5	Lors de l'examen d'un contenu composé d'une « déclaration à caractère négatif » à l'encontre d'un adulte particulier, Facebook doit modifier son Standard de la communauté pour exiger un examen du contexte sociopolitique du contenu. L'entreprise doit reconsidérer la mise en application de cette règle dans les débats politiques ou publics où la suppression du contenu étoufferait le débat.	CP	✘	
6	Chaque fois que Facebook supprime un contenu en raison d'une déclaration à caractère négatif qui ne représente qu'un seul mot ou une seule phrase dans une publication plus importante, elle doit en informer rapidement l'utilisateur afin que ce dernier ait la possibilité de publier à nouveau le contenu sans cette déclaration à caractère négatif.	A	●	
<b>Isolement d'Öcalan</b>				
1	Rétablir immédiatement l'instruction égarée en 2017 dans les Standards d'implémentation internes et les Questions fréquemment posées (les instructions internes pour les modérateurs de contenu), en informant tous les modérateurs de son existence et en organisant immédiatement une formation sur le sujet.	A	●	

**CATÉGORIE:** A– Application, T–Transparence, CP–Politique de contenu

**LA RÉPONSE DE META:** ● Complète, ◐ Assez complète, ✘ Incomplète

**IMPLÉMENTATION**

■ Progrès rapporté

□ Progrès rapporté

□ Meta a informé le Conseil d'une implémentation ou celle-ci a été décrite comme « Travail déjà en cours de réalisation par Meta », mais aucune information n'a été publiée comme preuve de cette application

□ Recommandation omise, refusée ou redéfinie

Identifiant de la recommandation	Recommandation du Conseil de surveillance	Catégorie	Évaluation du Conseil quant	
			à la réponse de Meta	à l'implémentation
<b>Isolement d'Öcalan</b>				
2	Évaluer les procédés de modération automatisés destinés à la mise en application de la politique sur les individus et les organismes dangereux. Le cas échéant, Facebook est tenue de mettre à jour les classificateurs afin d'exclure de la formation les données provenant de précédentes erreurs d'application résultant de l'incapacité à appliquer la directive de 2017. De nouvelles données d'entraînement reflétant la restauration de cette directive doivent être ajoutées.	A	✘	
3	Publier les résultats de la procédure d'examen continu pour déterminer si d'autres politiques ont été égarées, en ce compris la description de toute politique perdue, la période durant laquelle elle a été égarée et les mesures prises pour la restaurer.	T	●	
4	Indiquer, dans la « justification de sa politique » sur les individus et les organismes dangereux, que le respect des droits de l'homme et de la liberté d'expression, en particulier des discussions ouvertes sur les infractions des droits de l'homme et les abus liés au terrorisme ainsi qu'aux efforts de lutte antiterroriste, peut faire progresser la valeur de sécurité, et qu'il est essentiel que la plateforme fournisse un espace dédié à ces discussions. Bien que les valeurs de sécurité et de liberté d'expression peuvent parfois entrer en conflit, la justification de la politique doit préciser plus en détail les « dangers réels » que la politique cherche à éviter et empêcher en supprimant la valeur « liberté d'expression ».	CP	●	
5	Ajouter à la politique sur les individus et les organismes dangereux une explication claire des exclusions que comporte la notion de « soutien ». Les utilisateurs doivent être libres de discuter des infractions et des abus présumés des droits de l'homme des membres d'organismes désignés. Ce point ne doit pas se limiter aux individus placés en détention. Facebook doit inclure les discussions sur les droits protégés par les conventions des Nations unies sur les droits de l'homme, tels que cités dans sa politique d'entreprise en matière des droits de l'homme. Les utilisateurs pourraient ainsi débattre, par exemple, des allégations de torture ou autre peine ou traitement cruel, inhumain ou dégradant, des infractions au droit à un procès équitable, des exécutions extrajudiciaires, sommaires ou arbitraires, des disparitions forcées, des restitutions extraordinaires et des révocations de citoyenneté rendant une personne apatride. Les appels à la responsabilisation pour les infractions et les abus des droits de l'homme doivent également être protégés. Les contenus qui incitent à des actes de violence ou recrutent des personnes pour qu'elles rejoignent ou apportent un soutien matériel à des organismes désignés par Facebook doivent être exclus de cette protection, même s'ils traitent également de préoccupations relatives aux droits de l'homme. L'intention de l'utilisateur, le contexte plus large dans lequel la publication s'inscrit et la manière dont les autres utilisateurs comprennent le message s'avèrent essentiels afin de déterminer la probabilité qu'un danger réel résulte de la publication.	CP	●	
6	Expliquer, dans les Standards de la communauté, comment les utilisateurs peuvent faire comprendre à Facebook l'intention sous-entendue par leurs publications. Pour ce faire, il convient de mettre en œuvre la recommandation du Conseil visant à rendre publique la liste des individus et des organismes désignés par l'entreprise (voir : cas 2020-005-FB-UA). Facebook doit également fournir des exemples illustratifs pour définir la limite entre les contenus autorisés et interdits, notamment quant à l'application de la règle précisant les exclusions que comporte la notion de « soutien ».	CP	●	
7	Assurer un engagement significatif des parties prenantes envers les modifications de politique proposées par le biais du Forum de politique des produits de Facebook, notamment par un appel public à contributions. Facebook doit mener cet engagement en plusieurs langues dans toutes les régions, afin de garantir la participation effective des personnes les plus touchées par les dangers que cette politique cherche à éviter. Cet engagement devrait également inclure des organisations de défense des droits de l'homme, de la société civile et du monde académique possédant des connaissances spécialisées sur ces dangers, ainsi que sur les dangers pouvant résulter d'une mise en application excessive de la politique existante.	CP	●	

**CATÉGORIE:** A– Application, T–Transparence, CP–Politique de contenu

**LA RÉPONSE DE META:** ● Complète, ● Assez complète, ✘ Incomplète

**IMPLÉMENTATION**

■ Progrès rapporté

□ Progrès rapporté

■ Meta a informé le Conseil d'une implémentation ou celle-ci a été décrite comme « Travail déjà en cours de réalisation par Meta », mais aucune information n'a été publiée comme preuve de cette application

■ Recommandation omise, refusée ou redéfinie

Identifiant de la recommandation	Recommandation du Conseil de surveillance	Catégorie	Évaluation du Conseil quant	
			à la réponse de Meta	à l'implémentation
<b>Isolement d'Öcalan</b>				
8	Veiller à ce que des formations et des instructions internes soient fournies aux modérateurs de contenu concernant toute nouvelle politique. Les modérateurs de contenu doivent disposer de ressources adéquates pour pouvoir comprendre la nouvelle politique et de suffisamment de temps pour prendre des décisions lors de sa mise en application.	A	●	
9	Veiller à ce que les utilisateurs soient prévenus de la suppression de leur contenu. La notification doit indiquer si la suppression est due à une demande gouvernementale ou à une infraction aux Standards de la communauté, ou encore à un gouvernement affirmant que sa législation nationale est enfreinte (et la portée juridictionnelle de toute suppression).	A	●	
10	Préciser aux utilisateurs d'Instagram que les Standards de la communauté Facebook s'appliquent à Instagram de la même manière qu'ils s'appliquent à Facebook, conformément à la recommandation du cas 2020-004-IG-UA.	CP	●	
11	Inclure des informations sur le nombre de demandes de suppression de contenu reçues par Facebook de la part des gouvernements sur la base d'infractions des Standards de la communauté (par opposition aux infractions de la législation nationale), ainsi que le résultat de ces demandes.	T	◐	
12	Inclure des informations plus complètes sur les taux d'erreur dans l'application des règles relatives aux « éloges » et au « soutien » envers les individus et les organismes dangereux, réparties par région et par langue.	T	◐	
<b>Le bot du Myanmar</b>				
1	Facebook doit s'assurer que ses standards d'implémentation internes soient disponibles dans la langue dans laquelle les modérateurs de contenu examinent le contenu. S'il est nécessaire d'établir des priorités, Facebook doit d'abord se concentrer sur les contextes où les risques pour les droits de l'homme sont plus sérieux.	A	◐	
<b>Confinements liés au COVID au Brésil</b>				
1	Facebook devrait procéder à une analyse de proportionnalité pour identifier une série de mesures moins intrusives que la suppression du contenu. Lorsque cela est nécessaire, des mesures moins intrusives devraient être utilisées lorsque le contenu lié au COVID-19 déforme les recommandations des autorités de santé internationales et lorsqu'un risque de blessures physiques est identifié, mais pas imminent. Les mesures recommandées incluent : (a) le classement de contenu pour alerter les utilisateurs de la nature contestée du contenu d'une publication et fournir des liens vers les points de vue de l'Organisation mondiale de la Santé et des autorités sanitaires nationales ; (b) l'introduction d'un point de friction dans les publications pour éviter les interactions ou le partage ; et (c) la rétrogradation afin de réduire la visibilité dans les Fils d'actualité d'autres utilisateurs. Toutes ces mesures de mise en application devraient être clairement communiquées à l'ensemble des utilisateurs et pouvoir faire l'objet d'un appel.	A	✘	
2	Étant donné le contexte de la pandémie de COVID-19, Facebook devrait mettre en place des dispositions techniques pour donner la priorité à la vérification des informations concernant la désinformation sanitaire potentielle partagée par les autorités publiques qui est portée à la connaissance de l'entreprise, en tenant compte du contexte local.	A	✘	
3	Facebook devrait être plus transparent dans le Standard de la communauté sur les fausses informations concernant les cas où le contenu est éligible à la vérification des informations, et notamment la possibilité ou non pour les comptes des institutions publiques d'être soumis à la vérification des informations.	CP	●	

**CATÉGORIE:** A– Application, T–Transparence, CP–Politique de contenu

**LA RÉPONSE DE META:** ● Complète, ◐ Assez complète, ✘ Incomplète

**IMPLÉMENTATION**

■ Progrès rapporté

□ Progrès rapporté

□ Meta a informé le Conseil d'une implémentation ou celle-ci a été décrite comme « Travail déjà en cours de réalisation par Meta », mais aucune information n'a été publiée comme preuve de cette application

■ Recommandation omise, refusée ou redéfinie

Identifiant de la recommandation	Recommandation du Conseil de surveillance	Catégorie	Évaluation du Conseil quant	
			à la réponse de Meta	à l'implémentation
<b>Partage d'un post d'Al Jazeera</b>				
1	Ajouter des critères et des exemples concrets à sa politique relative aux individus et aux organismes dangereux afin de permettre aux utilisateurs de mieux comprendre les exceptions accordées en cas de discussions neutres, de condamnation et de reportages.	CP	●	
2	Faire réaliser sous peu la traduction des mises à jour apportées aux Standards de la communauté dans toutes les langues disponibles.	CP	◐	
3	Charger un organisme indépendant, qui n'est associé à aucune des parties au conflit israélo-palestinien, d'effectuer un examen minutieux pour déterminer si les pratiques de Facebook en matière de modération des contenus publiés en arabe et en hébreu (notamment son utilisation de l'automatisation) sont appliquées en toute impartialité. Cet examen ne doit pas seulement se concentrer sur le traitement des contenus palestiniens ou pro-palestiniens, mais également sur les publications qui incitent à la violence, quelles que soient les cibles, leur nationalité, leur ethnicité, leur religion ou leurs croyances, ou leurs opinions politiques. Il doit également passer en revue les contenus publiés par les utilisateurs de Facebook situés en territoire israélien, dans les territoires palestiniens occupés et en dehors. Le rapport et ses conclusions doivent être rendus publics.	T	●	
4	Mettre en place une procédure formelle et transparente pour recevoir toutes les demandes gouvernementales de suppression de contenus et pour y répondre, et assurer que lesdites demandes soient reprises dans les rapports de transparence. Les rapports de transparence doivent distinguer les demandes gouvernementales qui ont engendré la suppression de contenus pour infraction aux Standards de la communauté et les demandes gouvernementales qui ont engendré la suppression ou le blocage géographique de contenus pour infraction aux lois locales, en plus des demandes qui n'ont engendré aucune action.	T	●	
<b>Manifestations en Colombie</b>				
1	Publier des exemples illustratifs de la liste des injures considérées comme en infraction en vertu de son Standard de la communauté sur les discours incitant à la haine, en ce compris des cas limites comprenant des termes susceptibles d'être nuisibles dans certains contextes, mais pas dans d'autres.	CP	●	
2	Établir un lien entre l'explication succincte de l'intérêt médiatique reprise dans l'introduction des Standards de la communauté et l'explication plus détaillée de l'Espace modération de Facebook sur l'application de cette politique. L'entreprise devrait compléter cette explication au moyen d'exemples illustratifs tirés de divers contextes, en ce compris le signalement de manifestations à grande échelle.	CP	●	
3	Élaborer et publier des critères clairs destinés aux examinateurs de contenu afin qu'ils puissent soumettre à un examen supplémentaire les contenus d'intérêt public susceptibles d'enfreindre les Standards de la communauté, mais qui pourraient relever de l'intérêt médiatique. Ces critères devraient couvrir les contenus représentant de grandes manifestations sur des questions politiques.	A	✘	
4	Notifier tous les utilisateurs qui ont signalé un contenu jugé en infraction, mais laissé sur la plateforme pour des raisons d'intérêt public, car l'intérêt médiatique a été appliqué à la publication. Cette notification devrait comporter un lien vers l'explication de l'intérêt médiatique disponible dans l'Espace modération.	A	●	

**CATÉGORIE:** A– Application, T–Transparence, CP–Politique de contenu

**LA RÉPONSE DE META:** ● Complète, ◐ Assez complète, ✘ Incomplète

**IMPLÉMENTATION**

■ Progrès rapporté

□ Progrès rapporté

□ Meta a informé le Conseil d'une implémentation ou celle-ci a été décrite comme « Travail déjà en cours de réalisation par Meta », mais aucune information n'a été publiée comme preuve de cette application

□ Recommandation omise, refusée ou redéfinie

Identifiant de la recommandation	Recommandation du Conseil de surveillance	Catégorie	Évaluation du Conseil quant	
			à la réponse de Meta	à l'implémentation
<b>Insultes en Afrique du Sud 1</b>				
1	Informers les utilisateurs de la règle spécifique du Standard de la communauté sur les discours incitant à la haine qui a été enfreinte dans la langue qu'ils utilisent sur Facebook, tel que commandé dans la décision sur le cas 2020-003-FB-UA (Arméniens en Azerbaïdjan) et la décision sur le cas 2021-002-FB-UA (Représentation de Zwarte Piet). Par exemple, en l'espèce, l'utilisateur aurait dû être informé qu'il avait enfreint l'interdiction de proférer des insultes. Le Conseil a pris note de la réponse de Facebook à la deuxième recommandation de la décision sur le cas 2021-002-FB-UA, laquelle décrit un nouveau classificateur qui devrait être en mesure d'informer en langue anglaise les utilisateurs Facebook que leur contenu a enfreint la règle sur les insultes. Le Conseil attend de Facebook qu'elle fournisse des informations confirmant l'implémentation pour les utilisateurs anglophones ainsi que des renseignements sur le calendrier d'implémentation pour les utilisateurs des autres langues.	A	●	
<b>Collier de wampum</b>				
1	Informers les utilisateurs en temps utile de toute mesure prise par l'entreprise au sujet du contenu sur lequel porte leur appel. Le cas échéant, y compris en cas d'erreur de mise en application comme celle-ci, la notification à l'utilisateur doit mentionner que la mesure découle du processus d'examen du Conseil de surveillance. Meta est tenue de partager les messages envoyés à l'utilisateur lorsque les mesures du Conseil ont une incidence sur les décisions qui font l'objet d'un appel de la part de l'utilisateur afin de démontrer qu'elle s'est conformée à cette recommandation. Ces mesures doivent être prises dans tous les cas rectifiés à l'étape d'éligibilité du processus du Conseil.	A	●	
2	Étudier les conséquences de la modification de l'approche du second examen sur la précision et la productivité de l'examineur. Le Conseil demande en particulier une évaluation du taux de précision lorsque les modérateurs de contenu sont informés qu'ils effectuent un second examen, si bien qu'ils savent que la décision initiale a été contestée. Cette expérience devrait idéalement inclure la possibilité pour les utilisateurs de fournir un contexte pertinent susceptible d'aider les examinateurs à évaluer leur contenu, conformément aux recommandations précédentes du Conseil. Meta doit partager les résultats de ces évaluations de précision avec le Conseil et les résumer dans son rapport trimestriel de transparence au Conseil pour démontrer qu'elle s'est conformée à cette recommandation.	A	●	
3	Procéder à des évaluations de précision en se concentrant sur les autorisations de la politique en matière de discours incitant à la haine qui couvrent l'expression artistique et les propos sur les infractions aux droits de l'homme (p. ex. les condamnations, les sensibilisations, les emplois autoréférentiels et les utilisations valorisantes). Cette évaluation devrait également analyser spécifiquement l'incidence de l'origine d'un examinateur sur la capacité des modérateurs à évaluer avec précision les discours incitant à la haine et les contre-discours provenant de la même région ou non. Le Conseil comprend que cette analyse nécessitera probablement la constitution d'échantillons appropriés et étiquetés avec précision des contenus pertinents. Meta doit partager les résultats de cette évaluation avec le Conseil, en ce compris la manière dont ces résultats permettront d'améliorer les opérations de mise en application et l'élaboration des politiques. L'entreprise doit également indiquer si elle prévoit d'effectuer régulièrement des évaluations de la précision des examens sur la base de ces autorisations, ainsi que résumer les résultats dans son rapport trimestriel de transparence au Conseil pour démontrer qu'elle s'est conformée à cette recommandation.	A	◐	

**CATÉGORIE:** A– Application, T–Transparence, CP–Politique de contenu

**LA RÉPONSE DE META:** ● Complète, ◐ Assez complète, ✘ Incomplète

**IMPLÉMENTATION**

■ Progrès rapporté

□ Progrès rapporté

□ Meta a informé le Conseil d'une implémentation ou celle-ci a été décrite comme « Travail déjà en cours de réalisation par Meta », mais aucune information n'a été publiée comme preuve de cette application

□ Recommandation omise, refusée ou redéfinie

Identifiant de la recommandation	Recommandation du Conseil de surveillance	Catégorie	Évaluation du Conseil quant	
			à la réponse de Meta	à l'implémentation
<b>Ayahuasca</b>				
1	Le Conseil répète sa recommandation de la décision sur le cas 2020-004-IG-UA et la décision sur le cas 2021-006-IG-UA selon laquelle Meta devrait expliquer aux utilisateurs qu'elle applique les Standards de la communauté Facebook sur Instagram, avec plusieurs exceptions spécifiques. Le Conseil prend note de la réponse de Meta à ces recommandations. Bien que Meta puisse prendre d'autres mesures pour se conformer aux recommandations, le Conseil recommande à Meta de mettre à jour l'introduction des Règles de la communauté d'Instagram (les Règles de la communauté « en bref ») dans un délai de 90 jours afin d'informer les utilisateurs que si un contenu est considéré en infraction sur Facebook, il est également considéré comme en infraction sur Instagram, comme indiqué dans l'Espace modération de l'entreprise, à quelques exceptions près.	A	●	
2	Le Conseil répète sa recommandation de la décision sur le cas 2021-005-FB-UA et de la décision sur le cas 2020-005-FB-UA selon laquelle Meta devrait expliquer précisément quelle règle de la politique de contenu ils ont enfreint.	A	●	
3	Afin de respecter les diverses expressions et pratiques traditionnelles et religieuses, le Conseil recommande à Meta de modifier les Règles de la communauté Instagram et les Standards de la communauté Facebook sur les marchandises réglementées afin d'autoriser une discussion positive sur les utilisations traditionnelles et religieuses des substances non médicales lorsqu'il existe des preuves historiques de cette utilisation. Le Conseil recommande également à Meta de rendre publiques toutes les autorisations de ces politiques, y compris les autorisations existantes.	CP	●	
<b>Crimes présumés à Raya Kobo</b>				
1	Meta devrait réécrire sa valeur « Sécurité » pour qu'elle reflète le fait que les discours en ligne peuvent présenter des risques pour la sécurité physique et le droit à la vie des personnes, tout comme le risque d'intimider, d'exclure et de réduire au silence.	CP	●	
2	Les Standards de la communauté Facebook devraient refléter le fait que, dans des contextes de guerre et de conflits violents, les rumeurs non vérifiées créent davantage de risques pour le droit à la vie et la sécurité des personnes. Cette réalité devrait être reflétée à tous les niveaux du processus de modération.	CP	✘	
3	Meta devrait commander une évaluation indépendante sur les droits de l'homme, avec la diligence requise, pour comprendre comment Facebook et Instagram ont été utilisés pour diffuser des discours d'incitation à la haine et des rumeurs non vérifiées qui intensifient la violence en Éthiopie. L'évaluation devra examiner la réussite des mesures prises par Meta pour éviter la mauvaise utilisation de ses produits et services en Éthiopie. L'évaluation devra aussi examiner la réussite des mesures prises par Meta pour permettre la réalisation de rapports corroborés et d'intérêt public sur les atrocités portant atteinte aux droits de l'homme commises en Éthiopie. L'évaluation devra examiner les capacités linguistiques de Meta en Éthiopie et si elles permettent de protéger de façon adéquate les droits de ses utilisateurs. L'évaluation devra porter sur la période commençant le 1er juin 2020, jusqu'à aujourd'hui. L'entreprise devra réaliser l'évaluation dans un délai de six mois à compter du moment où elle répond à ces recommandations. L'évaluation devra être publiée dans sa totalité.	T	●	

**CATÉGORIE:** A– Application, T–Transparence, CP–Politique de contenu

**LA RÉPONSE DE META:** ● Complète, ● Assez complète, ✘ Incomplète

**IMPLÉMENTATION**

■ Progrès rapporté

□ Progrès rapporté

□ Meta a informé le Conseil d'une implémentation ou celle-ci a été décrite comme « Travail déjà en cours de réalisation par Meta », mais aucune information n'a été publiée comme preuve de cette application

■ Recommandation omise, refusée ou redéfinie





