



∞ Meta

Rapport sur les droits humains

Statistiques et mesures
2023



Table des matières

À propos de ce rapport	3
Résumé	5
Avant-propos	8
L'IA dans le contexte des droits humains	9
Adopter une approche ouverte	10
Prendre des mesures appropriées en fonction des résultats potentiellement nuisibles de l'IA générative	11
Gestion des risques en matière de droits humains	13
Diligence raisonnable au quotidien	13
Politique de contenu	16
Développement de produits	17
Comment nous nous préparons et réagissons aux événements critiques et aux crises mondiales	19
Comment Facebook se prépare aux élections	19
Brésil	20
Pakistan	22
Réponse aux crises mondiales	24
Soudan	24
Haut-Karabakh	26
Migration en Amérique centrale et du Sud	28
Évaluations de l'impact sur les droits humains	29
Israël et Palestine : Mise à jour de nos actions	29
Mesures politiques temporaires	30
Mesures temporaires sur les produits	31
Chiffrement de bout en bout Mise à jour de nos actions	32
Problématiques clés	33
Combattre les menaces contre la sécurité	33
Sécurité des enfants et des jeunes	34
Exploitation d'enfants	34
Sécurité et bien-être des jeunes	36
Interactions avec les parties prenantes	37
Groupes marginalisés et défenseurs des droits humains	37
Partenaires de confiance	39
Étude de cas : Apparition de nouveaux préjudices en Haïti	40
Étude de cas : Renforcement de l'information en Éthiopie	40
Étude de cas : Atténuer les conflits intra-religieux au Bangladesh	40
Organisations internationales et multipartites	41
Transparence et recours	44
Répondre aux demandes gouvernementales	45
Conseil de surveillance	50
Qu'en est-il de l'avenir ?	53
Annexe	55
Comment les droits humains sont régis et gérés chez Meta	55
Formation du personnel de Meta aux droits humains	55
Liens vers les rapports référencés	56



À propos de ce rapport

Il s'agit de notre troisième rapport annuel sur les droits humains. Il s'appuie sur le travail de Meta et reflète les progrès réalisés dans le cadre de nos engagements vis-à-vis des [Principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains](#) et de notre [Politique d'entreprise en matière de droits humains](#). Il couvre nos actions du 1er janvier 2023 au 31 décembre 2023. Nous rendons compte des services et produits Meta, notamment Facebook, Messenger, Instagram, WhatsApp, Threads et Reality Labs.

Les droits humains sont un sujet d'une importance cruciale pour nos parties prenantes et pour notre entreprise. Le contenu de ce rapport reflète les commentaires reçus de nos parties prenantes à la suite de la publication de notre [Rapport 2022 sur les droits humains](#) et s'appuie sur notre [Évaluation globale des risques liés aux droits humains](#), qui est résumée dans le rapport 2022. L'objectif de cette évaluation était d'identifier et de hiérarchiser nos impacts négatifs potentiels les plus significatifs en matière de droits humains¹ sur les personnes, y compris les utilisateurs et autres personnes susceptibles d'être concernées par nos actions dans l'ensemble de l'entreprise au niveau mondial. Le présent rapport met en évidence la manière dont Meta a géré, à travers ses opérations mondiales, les 8 droits humains les plus importants suivants :

Les 8 principaux droits humains potentiellement importants selon notre évaluation complète des risques liés aux droits humains

Droits humains importants

Section du rapport

Liberté d'opinion et d'expression

Politique de contenu, Comment Meta se prépare aux élections, Réponse aux crises mondiales, Israël et Palestine : Mise à jour de nos actions, Sécurité des enfants et des jeunes, Réponse aux demandes gouvernementales, Conseil de surveillance

Confidentialité

Politique de contenu, chiffrage de bout en bout : Mise à jour de nos actions, Lutte contre les menaces à la sécurité, Sécurité des enfants et des jeunes, Réponse aux demandes gouvernementales

Égalité et non-discrimination

L'IA dans le contexte des droits humains, Développement des produits, Israël et Palestine : Mise à jour de nos actions

Vie, liberté et sécurité de la personne

Comment Meta se prépare aux élections, Réponse aux crises mondiales

Intérêt supérieur de l'enfant

Politique de contenu, Sécurité des enfants et des jeunes, Engagement des parties prenantes

Participation publique, processus de vote et éligibilité

Comment Facebook se prépare aux élections : Brésil, Pakistan

Liberté de réunion et d'association

Chiffrage de bout en bout : Mise à jour de nos actions, Réponse aux crises mondiales : Soudan

Droit à la santé

Réponse aux crises mondiales : Soudan

¹ L'expression « impact négatif sur les droits humains » est conforme aux principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains et désigne un impact qui se produit lorsqu'une action supprime ou réduit la capacité d'une personne à jouir de ses droits humains.



Nos principaux risques en matière de droits humains figurent en tête des sections du présent rapport.

Nous nous sommes engagés avec de nombreuses parties prenantes internes et externes et nous nous efforçons de donner une image représentative de notre travail au sein de plusieurs équipes dans le monde entier. Le rapport a pour but de fournir une vue d'ensemble des questions relatives aux droits humains que nous traitons en tant qu'entreprise, tout en guidant le lecteur vers un contenu plus approfondi si nécessaire.

Ce rapport complète le plus récent [Rapport sur les pratiques commerciales responsables de Meta](#). Nous [rapportons](#) séparément nos efforts pour identifier et atténuer les risques d'esclavage moderne et de traite des êtres humains dans nos activités commerciales et nos chaînes d'approvisionnement. Des liens vers d'autres publications de Meta se trouvent dans l'[Annexe](#) du présent rapport.

Notre politique en matière de droits humains s'applique à l'ensemble de l'entreprise. Chaque service et entité de Meta dispose de ses propres politiques et procédures avec, parfois, des répercussions différentes sur les droits humains. Le présent rapport fait référence à des actions prises par Meta en tant qu'entreprise et concernant une ou plusieurs entités de Meta. Les déclarations n'ont pas pour objectif de suggérer que Meta a pris les mêmes mesures pour toutes les entités.²

² Par exemple, WhatsApp, une application d'appels et de messages chiffrés de bout en bout, dispose d'un point de contact distinct pour les questions de droits humains. L'analyse de la modération des contenus et des actions connexes sur Facebook et Instagram présentée dans ce rapport ne s'applique pas à WhatsApp et, à moins qu'une politique ou une action ne soit spécifiée comme s'appliquant à WhatsApp, elle ne s'applique pas à WhatsApp. En outre, bien que de nombreuses actions décrites dans ce rapport s'appliquent à Facebook et Instagram, il existe des distinctions intentionnelles dans les politiques et les procédures entre les services. Si une politique est qualifiée de politique « Facebook », elle peut ne pas s'appliquer à Instagram. Aucune déclaration du présent rapport n'est destinée à créer, ou ne doit être interprétée comme créant, de nouvelles obligations (juridiques ou autres) concernant l'application d'une politique ou d'une procédure à d'autres services ou entités.



Résumé



Avec ce rapport sur les droits humains, nous cherchons à donner un aperçu des politiques, des processus et des standards que Meta utilise pour gérer les risques liés aux droits humains pour des milliards de personnes, à grande échelle et en temps réel. Notre objectif est de montrer comment nous nous efforçons de respecter les droits des utilisateurs et des autres communautés dans le cadre de notre travail et des systèmes que nous utilisons chaque jour, qu'il s'agisse de la manière dont nous réagissons aux crises, dont nous nous préparons aux élections ou dont nous répondons aux demandes des gouvernements.

Le rapport de cette année couvre nos activités de l'année civile 2023 et reflète les commentaires des parties prenantes concernant le [Rapport sur les droits humains de l'année dernière](#). Le rapport de cette année met en lumière la manière dont Meta

a géré les huit principaux droits humains suivants : liberté d'opinion et d'expression ; confidentialité ; égalité et non-discrimination ; vie, liberté et sécurité de la personne ; intérêt supérieur de l'enfant ; participation publique, droit de vote et droit d'être élu ; liberté de réunion et d'association ; et droit à la santé.

Les services et les applications de Meta ont continué à être utilisés pour faire le bien, aidant ainsi à construire une communauté, à encourager l'innovation et à mobiliser des ressources pour soutenir l'aide humanitaire, entre autres efforts. Au fur et à mesure de l'émergence des défis, notre [Politique sur les droits humains](#) a guidé nos actions pour résoudre des problèmes tels que l'utilisation de nos services dans des pays en conflit, la cybersécurité et les demandes excessives de données de la part des gouvernements.



Nous poursuivons notre travail pour garder les droits humains à l'esprit dans la manière dont nous comprenons et développons l'intelligence artificielle (IA). Nos produits d'IA générative permettent aux personnes de faire progresser leurs droits humains de manière novatrice, mais nous reconnaissons que ces systèmes sont imparfaits, même s'ils présentent des possibilités d'amélioration continue. Pour limiter les risques potentiels, nous avons élaboré une [Politique d'utilisation acceptable](#), intégrée au [Contrat de licence relatif à la communauté](#) pour notre modèle ouvert Llama 2, et fourni des outils de sécurité pour aider les équipes de développement à réaliser leur travail de manière responsable. Nous pensons qu'une approche ouverte peut favoriser l'innovation par l'itération et conduire à des produits meilleurs et plus sûrs grâce à des commentaires, un examen, un développement et des mesures d'atténuation continus.

L'engagement proactif avec les parties prenantes est au cœur de notre approche des droits humains et est intégré à notre politique d'entreprise relative aux droits humains. Le développement de nos services, nos politiques de contenu et de modération, ainsi que les Standards de la communauté sont [fondés](#) sur des informations issues de nombreuses organisations de la société civile, de défenseurs des droits humains, de groupes marginalisés, d'organisations internationales, d'investisseurs, d'annonceurs, d'utilisateurs et de [partenaires de confiance](#). Ces groupes couvrent l'ensemble du spectre politique et idéologique afin d'apporter des contributions diverses. En janvier 2023, notre réseau de Partenaires de confiance comprenait plus de 400 organisations non gouvernementales, agences humanitaires, défenseurs des droits humains et chercheurs de 113 pays du monde entier.

L'année 2023 a marqué le dixième anniversaire de l'adhésion de Meta à la [Global Network Initiative](#) (GNI). Nous appliquons les principes de la GNI, et nous en sommes responsables, sur la manière dont les entreprises technologiques doivent respecter la liberté d'expression et la vie privée de leurs utilisateurs lorsqu'elles répondent à des demandes, notamment de restrictions, émanant de gouvernements. Dans ce rapport, nous expliquons comment la GNI a guidé notre réponse face aux demandes gouvernementales, y compris les restrictions de contenu et d'accès aux données des utilisateurs.

Nous avons également travaillé à l'intégration des principes des droits humains afin de nous préparer et de répondre aux événements et problèmes critiques en 2023. Le rapport traite de notre préparation à des défis accrus dans les pays à haut risque, de la manière dont nous nous préparons à des événements critiques planifiés, tels que les élections, et de notre réponse coordonnée aux situations de crise. En 2023, Meta a permis la préparation de plus de 130 élections dans le monde. Nous rendons compte dans le présent rapport de notre réponse aux conflits au Soudan et dans le Haut-Karabakh, ainsi qu'aux questions de droits humains liées aux migrations en Amérique centrale et en Amérique du Sud. Nos efforts de réponse à la crise ont impliqué de nombreux spécialistes de la politique de contenu, des global operations (opérations globales), des droits humains et des produits, ainsi que d'autres spécialistes internes qui ont évalué et traité ces situations, en s'appuyant sur notre [Protocole de politique de crise](#). Nous décrivons également la manière dont nous avons donné suite aux recommandations d'une précédente diligence raisonnable en matière de droits humains concernant [Israël](#).

et la [Palestine](#) et nos actions dans le cadre du conflit le plus récent, ainsi qu'une mise à jour sur le chiffrage de bout en bout.

Nous avons renforcé notre engagement en faveur de la transparence et du recours en tant que principes fondamentaux de la gestion des risques liés aux droits humains. Nous avons continué à faire participer le [Conseil de surveillance](#), une initiative de premier plan créée pour aider Meta à répondre à certaines des questions les plus difficiles concernant la liberté d'expression et la sécurité en ligne. L'année 2023 a été une année charnière pour le Conseil de surveillance, qui a dépassé son objectif de 50 décisions par an, triplant ainsi le nombre de décisions prises en 2022.

En outre, nous avons amélioré notre expérience utilisateur sur Facebook et Instagram pour le contenu restreint par Meta sur la base de demandes légales émanant de gouvernements et de tribunaux.

Dans la majorité des cas, nos notifications informaient les utilisateurs de l'autorité de l'État qui avait envoyé la demande à l'origine de la restriction et du pays dans lequel le contenu était restreint.

En 2023, nous avons soumis à la base de données [Lumen](#) notre première série de demandes de retrait de l'Autriche, du Vietnam, de Singapour, de l'Inde et du Mexique. Ce projet de recherche indépendant, hébergé par le Berkman Klein Center for Internet and Society de l'université de Harvard, permet aux chercheurs d'étudier les demandes de retrait de contenu en ligne émanant de gouvernements et d'acteurs privés. Il s'agit d'une étape supplémentaire pour permettre à la communauté mondiale d'analyser, de signaler et de défendre les droits numériques des utilisateurs de l'internet.





Avant-propos

Nous continuons à faire face à des défis nouveaux et complexes à mesure que le monde évolue. Notre réponse à ces défis est guidée par notre [Politique d'entreprise relative aux droits humains](#) et notre engagement envers les [Principes directeurs des Nations Unies sur les entreprises et les droits humains](#) et le [Pacte mondial des Nations Unies](#). Nous disposons en outre d'un système de contrôle et de régulation. Le [Conseil de surveillance](#) offre un contrôle expert et indépendant afin de garantir un processus décisionnel équitable et équilibré, notamment en ce qui concerne les questions de liberté d'expression.

Nos politiques et procédures évoluent également en fonction des changements technologiques et sociaux. Nous prenons en compte les contributions d'un grand nombre de parties prenantes en mettant l'accent sur les multiples dimensions de la diversité, y compris la diversité des points de vue. Lors de la révision de ces politiques et procédures, nous sollicitons l'avis du personnel, des spécialistes et des défenseurs des droits humains, des groupes marginalisés et des organisations internationales, ainsi que d'un large éventail de groupes de la société civile.

Nous utilisons divers outils et protocoles d'évaluation dans l'ensemble de l'entreprise afin d'identifier, de prévenir et d'atténuer les risques en matière de droits humains susceptibles d'être liés aux produits, aux politiques et aux opérations de Meta à l'échelle mondiale. Nous rendons compte des évaluations précédentes afin de partager nos progrès et d'ouvrir des canaux de commentaires.

En 2023, nous avons appliqué nos outils de gestion des risques pour répondre à des situations de crise, notamment des conflits et des guerres en Afrique et au Moyen-Orient, ainsi que pour se préparer aux événements mondiaux prévisibles, tels que les

nombreuses élections. Nos politiques sont donc dynamiques afin de garantir une réponse appropriée et adaptée à chaque situation, tout en équilibrant les intérêts concurrents et les considérations relatives aux droits humains.

Meta est pionnière dans le développement de l'intelligence artificielle (IA) depuis plus de dix ans. Nous savons que le progrès et la responsabilité doivent aller de pair. Les outils d'IA générative offrent de vastes possibilités, et nous pensons qu'il est à la fois possible et nécessaire que ces technologies soient développées dans le respect des droits humains.

Un marché numérique fort et ouvert est d'une importance vitale pour aider les personnes, la société et les pays à prospérer et à créer une communauté. Aujourd'hui, plus que jamais, nous avons besoin de la technologie pour nous aider à rapprocher le monde. Nous nous efforçons de jouer notre rôle en tenant compte des droits humains.

Signé par :



Nick Clegg,
Président des affaires mondiales



Jennifer Newstead,
Responsable des affaires juridiques



L'IA dans le contexte des droits humains

Risques majeurs pertinents :

- liberté d'opinion et d'expression
- égalité et non-discrimination
- confidentialité

Les rapides avancées de l'intelligence artificielle (IA), particulièrement l'[l'IA générative](#), ont attiré l'attention du monde entier en 2023. Compte tenu de l'attention accrue portée à l'IA et des opportunités et défis potentiels qu'elle présente, nous présentons une analyse approfondie de nos travaux en matière d'IA du point de vue des droits humains.

En 2023, nous avons lancé notre grand modèle de langage (LLM) ouvert [Llama 2](#), l'[assistant Meta AI](#), le modèle de génération d'images [Emu](#) et les outils de création [dans l'application](#).

À mesure que les chatbots d'IA sophistiqués et les outils de génération d'images font l'objet d'une utilisation généralisée, ils ont également suscité la surveillance du public et de la réglementation. Par exemple, l'Union européenne a fait progresser sa loi historique sur l'IA, [Règlement sur l'IA](#), tandis que les États-Unis ont lancé les [White House Voluntary Commitments on AI](#) avec le soutien de Meta et d'autres équipes de développement en la matière.

Nous nous engageons à développer et à déployer l'IA de manière responsable tout en atténuant les effets négatifs potentiels sur les droits humains. L'IA est incluse dans notre [Politique d'entreprise relative aux droits humains](#), qui reconnaît l'importance

Nous pensons que l'IA responsable peut être un outil puissant pour faire progresser les droits humains.

des [Principes de l'OCDE sur l'intelligence artificielle](#), largement adoptés et approuvés par le G20.

Nous pensons que l'IA responsable peut être un outil puissant pour faire progresser les droits humains. Nos produits d'IA générative permettent aux personnes d'exercer leur liberté d'expression de manière innovante, d'améliorer l'accès à l'information et à l'éducation, et d'améliorer l'accessibilité. Par exemple, les lunettes intelligentes Ray-Ban Meta, dotées d'un système de sous-titrage et de reconnaissance d'images alimenté par l'IA, améliorent l'accessibilité pour les personnes souffrant d'un handicap.

Nous utilisons l'IA pour tenter de détecter et de réagir rapidement à des contenus potentiellement en infraction aux politiques et à des interactions indésirables dans un nombre croissant de langues, afin d'éviter que la situation ne s'aggrave. Nombre de nos efforts [Data for Good](#) s'appuient sur l'IA pour soutenir les réponses humanitaires aux crises et faciliter le travail de santé publique.

Adopter une approche ouverte

Meta s'efforce d'adopter une approche responsable dans le développement et le déploiement de produits et de modèles d'IA générative.

Nous pensons qu'une approche ouverte peut favoriser l'innovation par l'itération et conduire à des produits meilleurs et plus sûrs grâce à des commentaires, un examen, un développement et des mesures d'atténuation continus de la part de la communauté. Cela peut également :

- favoriser l'équité et soutenir la diversité des points de vue en permettant aux détenteurs de droits concernés de mieux identifier les préjugés potentiels.
- réduire les obstacles à l'innovation, ce qui se traduit par des avantages économiques.
- permettre plus facilement aux communautés du monde entier d'adapter les progrès de l'IA à des langues et des contextes spécifiques.

Cependant, nous reconnaissons qu'une approche ouverte peut également présenter des risques potentiels pour les droits humains, notamment car les équipes de développement ignorent les instructions d'utilisation responsable de Meta et déploient les modèles sans se soucier de la sécurité. Pour tenter d'atténuer ces risques, nous effectuons des tests de sécurité et avons élaboré une [Politique d'utilisation acceptable](#) détaillée, intégrée au [Contrat de licence relatif à la communauté](#) pour nos modèles Llama. Nous avons également partagé un [Guide d'utilisation responsable](#) détaillé et fourni des outils de sécurité, par le biais de notre projet [Purple Llama](#), y compris [Llama Guard](#), afin d'aider les équipes de développement à faire leur travail de manière responsable en s'appuyant sur ces modèles de base librement disponibles et à les déployer en toute sécurité pour leurs propres cas d'utilisation.

Nous pensons qu'une approche ouverte peut favoriser l'innovation par l'itération et conduire à des produits meilleurs et plus sûrs grâce à des commentaires, un examen, un développement et des mesures d'atténuation continus de la part de la communauté.

Le déploiement de l'IA en toute sécurité est une responsabilité partagée par l'ensemble de l'écosystème. C'est pourquoi nous collaborons depuis de nombreuses années avec des organisations qui travaillent à la mise en place d'une IA sûre et fiable. Par exemple :

Nous avons travaillé avec [MLCommons](#) et un ensemble de partenaires pour créer des [critères de responsabilité](#) au bénéfice de la communauté open source.

- Nous avons participé à des initiatives multipartites telles que le [Partenariat sur l'AI](#) afin d'aborder des questions allant du déploiement responsable de modèles de base aux moyens les plus appropriés d'identifier le contenu synthétique.
- Nous avons cofondé l'Alliance [AI](#), une coalition d'entreprises, d'universitaires, de défenseurs et de gouvernements qui travaillent à l'élaboration d'outils permettant de créer un écosystème d'IA ouvert et sûr.

Nous avons consulté les parties prenantes en matière de droits humains pour les aider à comprendre notre utilisation de l'IA et à nous faire part de leurs commentaires à ce sujet. Dans le cadre de ces efforts, nous avons collaboré avec le

Projet B-Tech du Bureau des droits de l'homme des Nations Unies sur son travail sur l'IA générative, en s'appuyant sur trois documents fondamentaux fondés sur les principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains. Nous avons également fait appel à un conseiller du bureau du rapporteur spécial des Nations Unies pour la liberté d'opinion et d'expression et de l'UNESCO pour nous aider à mieux connaître les concepts fondamentaux de l'IA et les risques, et nous avons offert un cadre propice à une discussion éclairée pour comprendre les préoccupations des différentes parties prenantes. En outre, nous avons organisé un forum communautaire consultatif, mené en partenariat avec l'université de Stanford, qui a intégré des questions relatives aux droits humains pour aider Meta et d'autres entreprises, chercheurs et gouvernements à prendre des décisions concernant les chatbots d'IA générative. Les constatations ont souligné l'importance du cadre des droits humains et de l'intégration d'informations provenant de sources relatives aux droits humains.

Prendre des mesures appropriées en fonction des résultats potentiellement nuisibles de l'IA générative

Nous savons que la technologie d'IA générative peut produire des résultats susceptibles, par exemple, de générer un contenu potentiellement haineux, offensant ou discriminatoire, de renforcer les préjugés, de présenter des informations inexactes et/ou de soulever des questions relatives à la protection de la vie privée. Nous savons également que des acteurs malveillants peuvent essayer d'abuser des modèles et des outils que nous diffusons, de manière délibérément nuisible. Nous avons pris une série de mesures pour atténuer ces risques dans nos modèles de base, notamment le Llama et Emu. Nous pensons que l'adoption d'une approche ouverte et itérative nous aidera à nous améliorer en temps réel si ces problèmes surviennent.

Le projet B-Tech de l'ONU met en évidence un certain nombre de domaines de risque pertinents dans plusieurs droits internationalement reconnus. Nous nous appuyons sur ces orientations, ainsi que sur les principaux risques de Meta en matière de droits humains, pour éclairer notre travail.

Pour nos modèles de base Llama 2 et Emu, ces mesures d'atténuation visent en partie à traiter les risques potentiels importants en matière de droits humains, en abordant les risques dans le cadre de la formation, en traitant les associations sensibles au contexte et en procédant à des évaluations et à des réglages de la sécurité.

Alors que nous continuons à développer nos modèles de base, nous avons partagé publiquement des informations détaillées sur notre approche responsable.

En plus de nos efforts proactifs pour atténuer les risques associés aux résultats potentiellement dangereux générés par l'IA, nous nous efforçons également de réagir aux résultats potentiellement dangereux dès que nous en avons connaissance.



Image générée par Meta AI



Dans les jours qui ont suivi les attentats terroristes du 7 octobre en Israël, les parties prenantes nous ont signalé plusieurs problèmes avec nos outils, que nous avons tous rapidement reconnus et corrigés. Les principaux problèmes concernaient des traductions automatiques dans lesquelles le mot « terroriste » était ajouté aux biographies traduites du profil Instagram de certains utilisateurs, ainsi que des stickers générés par l'IA qui représentaient les Palestiniens comme des personnes violentes.

Dès que nous avons eu connaissance de ces problèmes, nos équipes d'ingénieurs ont entamé une enquête pour déterminer les causes profondes et mettre en œuvre des correctifs. Nos équipes ont identifié que ces problèmes semblaient être liés aux hallucinations du modèle et aux données d'entraînement utilisées. Ces deux questions constituent des défis bien documentés pour les produits alimentés par l'IA (voir [ici](#) et [ici](#)).

Nous avons apporté une correction d'urgence à notre produit de traduction automatique dans les 90 minutes qui ont suivi la découverte du problème. Nous avons également travaillé aussi rapidement que possible pour atténuer les associations potentiellement problématiques dans les images générées et les stickers produits à l'occasion du 7 octobre. Par exemple, nous avons ajouté des noms de famille connus au modèle pour éviter les hallucinations. En outre, nous avons affiné nos modèles de base pour la génération d'images afin de mieux prendre en compte un large éventail d'associations potentiellement problématiques liées aux conflits, et nous les avons déployés dans l'ensemble de nos produits.

Voir [Israël et Palestine : Pour en savoir plus, consultez notre section actions.](#)

Nous sommes impatients de poursuivre notre engagement auprès d'une série de détenteurs de droits aux opinions diverses, car nous nous efforçons de placer la sécurité, la responsabilité et les droits humains au cœur de notre travail de développement et de déploiement de produits d'IA générative. Nous prendrons également les mesures appropriées pour nous conformer aux obligations légales dans ce domaine.

Gestion des risques en matière de droits humains

Chaque jour, nous cherchons activement à traduire les principes des droits humains en actions concrètes. Notre travail repose sur la [Politique de Meta en matière de droits humains](#). Notre engagement et notre approche s'appuient sur les [Principes directeurs des Nations Unies sur les entreprises et les droits humains](#), ainsi que sur les normes internationales et régionales en matière de droits humains énumérées dans notre Politique d'entreprise relative aux droits humains.

Notre travail comprend, entre autres, nos efforts de diligence raisonnable, [l'engagement des parties prenantes](#), [le conseil en matière de produits](#), [la formation aux droits humains](#) et la fourniture de conseils fondés sur des données dans tous les domaines de l'entreprise. Dans ce rapport, nous nous concentrons sur les différents types de diligence raisonnable en matière de droits humains que nous appliquons chez Meta. Dans le cadre de la diligence raisonnable, nous appliquons une série d'outils et de protocoles d'évaluation afin d'identifier, de prévenir et d'atténuer les risques susceptibles d'être liés à nos produits, à nos politiques et à nos opérations dans le monde entier.

Ce travail alimente notre réflexion sur la réglementation, parallèlement à nos engagements dans le cadre de la [Global Network Initiative \(GNI\)](#).



L'environnement réglementaire évolue et a des répercussions sur les droits humains, comme le montrent les tendances mondiales récentes. Par exemple, en 2023, la loi britannique sur la sécurité en ligne a été promulguée et la loi sur les services numériques de l'Union européenne est entrée en vigueur, tout comme d'autres lois sur la sécurité en ligne, notamment à Singapour, à Taïwan et en Uruguay.

Le tableau de la page 15 présente certaines des mesures que nous prenons pour gérer les risques liés aux droits humains. De plus amples

informations sont également disponibles dans nos rapports sur les droits humains [2021](#) et [2022](#).

Diligence raisonnable au quotidien

Dans le cadre de notre travail quotidien de diligence raisonnable, de nombreuses équipes différentes s'associent pour tenir compte des droits humains dans nos activités, et ce en permanence. Par exemple, les spécialistes en droits humains et en droits civils de Meta fournissent des conseils permanents pour

évaluer et atténuer les risques liés aux droits humains en matière de politique de contenu, de développement de produits, de réponse aux crises et aux conflits, et de préparation aux élections. Ce travail porte souvent sur des questions liées au discours haineux, aux fausses informations, au contenu explicite susceptible de documenter les violations des droits humains, les mesures visant à protéger les défenseurs des droits humains, ainsi que d'autres questions liées aux droits humains.

Avec 3,19 milliards de personnes en moyenne qui utiliseront quotidiennement au moins une de nos applications d'ici décembre 2023, les impacts potentiels de nos produits sur les droits humains varient considérablement en fonction du temps, du lieu, du contenu et des audiences concernées. C'est pourquoi nous établissons des priorités dans notre travail sur la base de notre [évaluation globale des risques liés aux droits humains](#) de 2022, qui a analysé les incidences sur tous les droits humains reconnus au niveau international. L'analyse a permis de hiérarchiser les huit risques les plus importants sur la base des critères du [guide des principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains](#), présentés dans le graphique ci-dessous.

La question relative à la hiérarchisation de nos actions face aux risques potentiels pour les droits humains revient souvent. Cette tension est inhérente au cadre connexe et est mise en évidence lorsque nous cherchons à concilier plusieurs droits humains parfois, concurrents. Nous cherchons à donner la priorité aux risques les plus importants en matière de droits humains, c'est-à-dire ceux qui sont susceptibles d'avoir les conséquences négatives les plus graves pour les parties prenantes.

Nous cherchons à donner la priorité aux risques les plus importants en matière de droits humains, c'est-à-dire ceux qui sont susceptibles d'avoir les conséquences négatives les plus graves pour les parties prenantes.

Nos priorités sont fondées sur les principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains et sur notre Politique d'entreprise en matière de droits humains, et enrichies par l'engagement des parties prenantes organisé autour de multiples dimensions de la diversité. Par exemple, l'établissement de priorités est nécessaire pour permettre l'utilisation de nos produits et services par des catégories de personnes vulnérables, telles que les jeunes, au sein d'une population d'utilisateurs large et diversifiée. Meta s'appuie sur la [Convention des Nations Unies relative aux droits de l'enfant](#), en particulier sur le principe de « l'intérêt supérieur de l'enfant », pour définir sa manière de servir les jeunes. Cela conduit parfois à des décisions complexes entre les impacts potentiels sur les droits qui sont prioritaires : par exemple, reconnaître la sécurité et le bien-être des enfants tout en respectant leur droit à rechercher des informations en privé sur des sujets tels que la politique, la santé ou l'identité de genre. Le [cadre de l'intérêt supérieur de l'enfant](#) a guidé l'adoption et la conception d'éléments tels que le [Centre familial](#) et la [supervision parentale](#).

Notre approche de la gestion des risques en matière de droits humains

Engagements en faveur des droits humains

Nous mettons en œuvre notre engagement en faveur des droits humains en suivant les approches définies dans les principes directeurs des Nations Unies, notamment :

Application des politiques en matière de droits humains

Transparence et diligence raisonnable en matière de droits humains

Accès aux voies de recours

Maintien de la surveillance, de la gouvernance et de la responsabilité

Protection des défenseurs des droits humains

Hiérarchisation des droits

L'évaluation globale des risques liés aux droits humains a identifié huit droits comme prioritaires :

Liberté d'opinion et d'expression

Confidentialité

Égalité et non-discrimination

Vie, liberté et sécurité de la personne

Intérêt supérieur de l'enfant

Participation publique, processus de vote et éligibilité

Liberté de réunion et d'association

Droit à la santé

Outils et exemples clés

Nous déployons les outils et processus suivants pour atténuer les risques, avec quelques exemples illustratifs abordés dans le présent rapport :

Développement de la politique de contenu



Mise à jour de la politique sur les personnes et organisations dangereuses

Avis sur les produits



IA générative ; Threads

Droits humains intégrés à la préparation des élections



Élections au Pakistan

Droits humains intégrés aux efforts de réponse coordonnés



Priorité à la réponse active Statut des pays

Réponse à la crise



Conflits dans le Haut-Karabakh et au Soudan

Cadre GNI pour les demandes gouvernementales



Restrictions de contenu conformément à la législation locale et accès aux données utilisateur

Conseil de surveillance



Contribution aux cas de référence, par exemple les otages en droit international

Interactions avec les parties prenantes



Sessions de co-conception pour les parents et les ados

Protection des défenseurs des droits humains



Fonds pour les défenseurs des droits humains



Risques majeurs pertinents :

- liberté d'opinion et d'expression
- confidentialité
- égalité et non-discrimination
- vie, liberté et sécurité de la personne

Politique de contenu

Dans le cadre de notre travail de diligence raisonnable, nos spécialistes en droits humains soutiennent le processus d'élaboration de la politique de contenu. Cela signifie qu'ils examinent les propositions de mise à jour de nos politiques de contenu sous l'angle de la législation relative aux droits humains, afin de tenir compte de la liberté d'expression et de la protection contre la discrimination, ainsi que d'autres droits humains. En 2023, il s'agissait notamment d'affiner notre politique relative aux personnes et organisations dangereuses. Ces travaux ont été menés en différentes phases en 2023, et nous avons publié une mise à jour de nos travaux en janvier 2024.

En vertu de la politique relative aux personnes et organisations dangereuses, nous n'autorisons pas les personnes ou les organisations qui proclament une mission violente ou qui sont engagées dans des actes de violence à être présentes sur les services de Meta. Notre approche globale repose sur des définitions complètes des types et niveaux d'organisations dangereuses, telles que publiées dans notre Espace modération. Nous disposons d'un processus indépendant pour identifier les personnes et les organisations dangereuses, chaque cas étant examiné par des spécialistes compétents sur la base de preuves corroborées. Nos définitions et nos seuils ne dépendent pas de la région ou de l'idéologie et ont été élaborés en consultation avec des spécialistes et des universitaires du monde entier. En outre, nous appliquons la loi aux organisations terroristes étrangères désignées par les États-Unis, aux plaques

tournantes du trafic de stupéfiants spécialement désignées et aux terroristes internationaux spécialement désignés.

Conformément aux recommandations du Conseil de surveillance et de Business for Social Responsibility en matière de diligence raisonnable à l'égard d'Israël et de la Palestine, nous avons mis à jour notre politique relative aux personnes et organisations dangereuses de la manière suivante :

- Nous avons mis à jour notre politique afin de permettre un discours plus social et politique, notamment sur les élections, la résolution des conflits, les catastrophes et l'aide humanitaire, tout en continuant à respecter nos obligations légales.
- Nous avons reçu des commentaires selon lesquels notre précédente définition de « l'éloge d'une personne ou d'une organisation dangereuse » était trop large, et nous avons donc mis à jour notre politique pour qu'elle soit plus précise et plus adaptée à la situation. Nous interdisons désormais la « glorification » de la violence et de la haine de personnes et d'organisations dangereuses, ce qui constitue une définition plus précise que le mot « éloge ».
- Nous avons simplifié la manière dont nous évaluons et classons les acteurs dangereux en fonction de leurs liens avec les préjudices hors ligne et la violence.
- Nous avons mis à jour notre procédure de retrait de la liste afin de fournir des critères plus détaillés et complets à remplir pour qu'une personne ou une organisation dangereuse soit considérée comme pouvant être supprimée de la liste. Cette mise à jour nous permet de nous assurer que nos désignations peuvent refléter l'évolution des circonstances.



En septembre 2023, nous avons également modifié notre politique en matière de [violence et d'incitation](#) afin d'affiner la formulation relative à la violence de haute gravité. En mars 2023, nous avons modifié notre [Politique en matière d'intimidation et de harcèlement](#) afin d'y inclure une définition des « personnalités publiques », comme l'a recommandé le Conseil de surveillance. En juin 2023, nous avons organisé un forum politique sur les contenus violents et explicites dans les contextes d'abus des droits humains, en réponse à une recommandation du Conseil de surveillance dans le cas [Vidéo explicite au Soudan](#). Nous avons examiné si nos politiques en matière de contenu violent et explicite trouvaient un juste équilibre entre le respect de la vie privée et de la dignité des victimes, la liberté d'expression et le bien-être de nos communautés. Notre analyse a démontré que l'équilibre était approprié, nous n'avons donc pas modifié nos politiques.

Dans un souci de transparence, les modifications apportées aux politiques sont reflétées dans nos [Standards de la communauté](#) et peuvent être consultées dans le journal des modifications correspondant et dans les [procès-verbaux publics de nos forums de politiques](#).

En outre, nous avons mis à jour notre [système de sanctions](#) en 2023. Ce nouveau système a réduit le nombre de périodes de restriction et vise à garantir une réponse proportionnée aux manquements aux politiques. Nous continuerons à appliquer des restrictions de compte aux contrevenants récidivistes, généralement à partir du septième manquement, après avoir transmis suffisamment d'avertissements et d'explications pour aider la personne à comprendre pourquoi nous avons supprimé son contenu. Pour les manquements plus graves, tels que la publication de contenu évoquant le terrorisme, l'exploitation des enfants, la traite des êtres humains, la

promotion du suicide, l'exploitation sexuelle, la vente de drogues non médicales ou la promotion de personnes et d'organisations dangereuses, nous continuerons à appliquer des conséquences immédiates.

Développement de produits

Nos équipes d'ingénieurs s'efforcent de trouver des solutions créatives aux problèmes du monde réel. Les droits humains guident notre travail de développement de pratiques d'innovation responsable, et nous nous engageons à faire preuve de diligence raisonnable pour les nouveaux produits.

En 2023, nous avons lancé un certain nombre de nouveaux produits passionnants, tels que [Threads](#) en juillet, les [lunettes intelligentes Ray-Ban Meta](#) et les [canaux WhatsApp](#) en septembre, et [Quest 3](#), notre casque de réalité mixte, en octobre. Pour chacun de ces produits, les considérations relatives aux droits humains ont été intégrées au processus de développement.

En outre, nous avons continué à développer le métavers, notamment pour contribuer à améliorer l'éducation et à lutter contre les préjugés et la haine. Par exemple, nous avons lancé plusieurs expériences de réalité virtuelle pour renforcer la communauté, notamment [MLK : Now is the Time](#), [Inside the Mosque](#) (une expérience [en deux parties](#)), et [ABLE](#).

Dans un environnement rapide et itératif, les équipes construisent, testent, affinent et déploient des produits et des services - y compris ceux rendus possibles par l'IA. Nous construisons [en pensant à la confidentialité](#), et nous avons conçu des approches d'atténuation des risques liés aux produits qui correspondent aux besoins particuliers de notre

Risques majeurs pertinents :

- liberté d'opinion et d'expression
- confidentialité
- égalité et non-discrimination

processus dynamique de développement de produits. Dès le début du processus de développement, les équipes chargées des produits appliquent les conseils en matière d'innovation responsable, conformément aux principes des droits humains. Cela permet aux équipes d'anticiper et d'atténuer les préjudices potentiels pour les individus, les communautés et la société. Nos spécialistes en droits civils et en droits humains peuvent fournir des conseils ou mener des actions rapides, par exemple pour identifier les pays les plus à risque pour le lancement de produits.

Par exemple, au cours du processus de développement du produit Threads, nos spécialistes en droits humains se sont appuyés sur les processus existants d'examen des produits et des politiques pour évaluer

l'impact potentiel sur les droits humains.

L'équipe a travaillé avec celle de Threads pour mettre en œuvre des mesures d'atténuation, notamment en ce qui concerne les demandes potentielles du gouvernement de censurer le contenu.

Nous continuons également à développer le projet Height, un cadre permettant aux équipes de produits d'évaluer les problèmes de droits civils présentés lors du lancement de produits. Cela permet aux équipes de développement de produits de travailler en tenant compte des droits civils au cours du processus de développement du produit. Il complète les autres processus de gestion des risques que nous avons mis en place pour évaluer la sécurité de nos produits.



MLK: Now is the Time par Meta Quest



Risques majeurs pertinents :

- vie, liberté et sécurité de la personne
- participation publique, processus de vote et éligibilité
- liberté d'opinion et d'expression
- égalité et non-discrimination
- confidentialité
- intérêt supérieur de l'enfant

Comment nous nous préparons et réagissons aux événements critiques et aux crises mondiales

Nous travaillons dans toute l'entreprise pour intégrer les principes des droits humains dans les efforts d'intégrité de Meta.³ Cela comprend la façon dont nous nous préparons à relever les défis dans les pays à haut risque, ainsi que les événements critiques prévus, tels que les élections, et notre réponse coordonnée aux situations de crise. Ce travail coordonné entre des équipes spécialisées nous permet d'anticiper et de répondre efficacement à l'escalade des tensions, où qu'elles se produisent.

Comment Facebook se prépare aux élections

En 2023, de nombreux pays se sont rendus aux urnes, et d'autres encore se préparent à un cycle électoral chargé en 2024. En 2023, Meta a permis de préparer plus de 130 élections dans le monde, notamment au Pakistan, en Argentine, en Turquie et au Nigeria. Nous partageons ci-dessous des informations et nos actions issues de deux élections pertinentes, au Brésil et au Pakistan.

Nos équipes spécialisées dirigent les préparatifs à l'échelle de Meta afin de

mettre en place des mesures de protection avant les élections internationales. Il peut s'agir d'opérations de sécurité avancées pour lutter contre les menaces, de politiques et de processus actualisés pour supprimer les contenus potentiellement nuisibles de nos plateformes, de notre réseau mondial de vérification des informations, le plus performant du secteur, et de la transparence autour des publicités portant sur un enjeu social et politique.

Nous évaluons également en permanence les risques de préjudices imminents lors d'événements critiques, tels que les élections, afin de pouvoir réagir par des actions ciblées, limitées dans le temps, au niveau des politiques et des produits, qui contribueront à assurer la sécurité des personnes.

Bien que la plupart de nos travaux soient proactifs et entrepris bien avant les dates des élections, nous nous préparons également à réagir à des événements à haut risque. Dans ces cas-là, Meta propose plusieurs approches pour répondre à une série de scénarios, par exemple, modifier ou limiter les caractéristiques du produit, introduire des limites de taux de messagerie et limiter la diffusion du contenu.

Nous identifions les tendances potentiellement nuisibles en recueillant des informations dans des rapports publics,

³ « Intégrité » fait partie de la terminologie de Meta. La confiance et la sûreté, la sécurité des ordinateurs et des comptes, la réduction des mauvaises expériences, les questions liées à la confidentialité et bien d'autres questions encore sont gérées par un réseau d'équipes au sein de Meta, dont le nom fait souvent référence à « Intégrité ». Ces équipes construisent notamment des outils pour prévenir les préjudices, modérer nos plateformes et faire respecter nos politiques.

en examinant les recommandations de nos partenaires de confiance, en observant en permanence les tendances en matière de contenu, en faisant preuve de diligence raisonnable en matière de droits humains et en examinant les évaluations de nos spécialistes en matière de renseignement. Ces informations, ainsi que les Standards de la communauté existants et les systèmes d'application, permettent de déterminer le type de produit et les mesures d'atténuation des politiques que nous pouvons utiliser pour prévenir les abus lors des élections à haut risque. Cela inclut notre [Protocole de politique de crise](#) et les mesures distinctes que nous pouvons prendre pour adapter nos systèmes en réponse à [un risque accru](#). Nous fournissons également des outils de transparence, tels que notre [Bibliothèque publicitaire](#).

Brésil

Nous avons commencé à nous préparer aux [élections présidentielles brésiliennes de 2022](#) un an avant le scrutin. Nous avons poursuivi nos efforts pendant la période post-électorale, ce qui nous a permis de réagir rapidement aux [attentats du 8 janvier 2023](#) contre le Congrès brésilien, la Cour suprême et d'autres bâtiments publics. Meta a également qualifié les troubles post-électorales de crise dans le cadre du Protocole de politique de crise afin de pouvoir évaluer comment atténuer les risques liés au contenu.

Nos préparatifs pour l'élection de 2022 comprenaient un établissement des scénarios de risque pendant et après l'élection. Nous avons pris en compte les normes internationales relatives à la liberté d'expression et aux autres droits humains. Nous avons également modifié nos produits et adapté nos politiques afin de protéger l'intégrité des élections brésiliennes. Il s'agissait notamment d'établir un partenariat avec la Cour supérieure électorale afin de permettre aux citoyens d'accéder à des informations fiables sur le vote.

Nous avons déployé une série d'outils et de méthodes pour lutter contre les discours haineux, l'incitation à la violence et aux fausses informations avant, pendant et après les attentats du 8 janvier 2023. Peu avant l'élection, nous avons désigné le Brésil comme un pays temporairement à haut risque, afin de supprimer les appels à apporter des armements ou à pénétrer de force dans les bâtiments gouvernementaux, comme indiqué dans le cas du [Conseil de surveillance « Discours du général brésilien »](#).

Depuis le début de la campagne électorale, le 16 août 2022, jusqu'à la fin du mois de janvier 2023, [nous avons supprimé](#) plus d'un million de contenus sur Facebook et plus de 960 000 contenus sur Instagram pour infraction à nos politiques en matière de violence et d'incitation au Brésil. Ce contenu comprenait des publications appelant à une intervention militaire.

Au cours de cette période de cinq mois au Brésil, nous avons supprimé des contenus qui enfreignaient nos politiques en matière de discours haineux :



Plus de 570 000



Plus de 520 000

Nous avons également supprimé des contenus au Brésil pour manquement à nos politiques en matière d'intimidation et de harcèlement :



Plus de 380 000

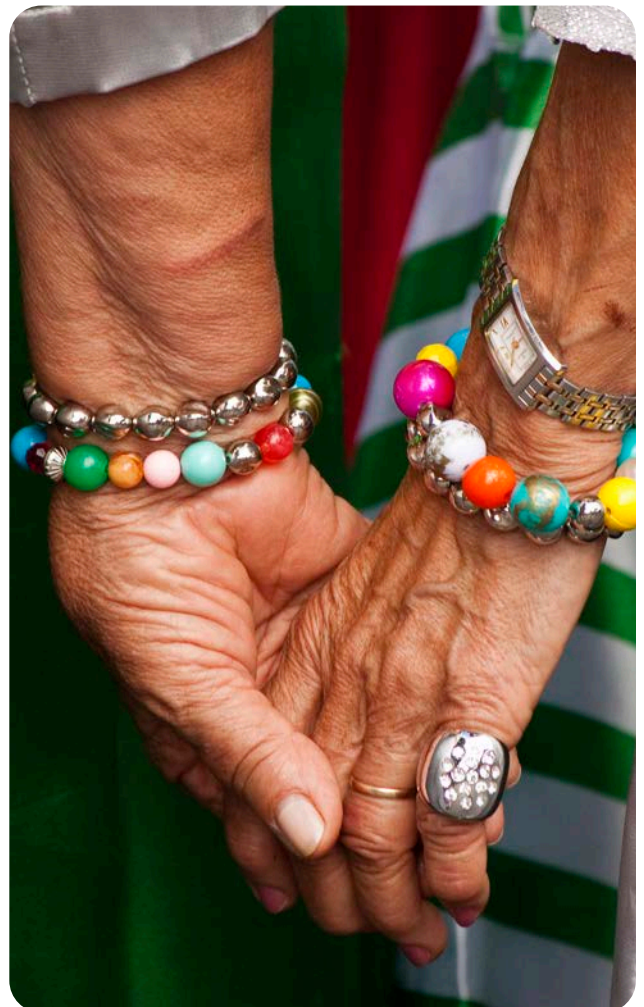


Plus de 630 000

Violence fondée sur le genre

La menace de harcèlement basée sur le genre et de violence à l'encontre des candidates, des journalistes et des défenseuses des droits humains s'intensifie souvent pendant les élections. Pour atténuer ces risques et dans le cadre de notre approche plus large visant à renforcer la protection des femmes sur les plateformes et à les mettre en relation avec les services publics, nos actions au Brésil ont notamment été les suivantes :

- collaborer avec le ministère des Femmes pour lancer un canal officiel sur WhatsApp au Brésil permettant aux femmes de transmettre des plaintes ou de demander des informations au médiateur sur la violence à l'égard des femmes, en tant que moyen de communication supplémentaire avec le médiateur. Les utilisateurs peuvent obtenir des informations sur les lois applicables, trouver les adresses des services d'aide aux femmes et s'adresser directement aux préposés.
- publier un guide en ligne en portugais pour protéger les femmes contre la violence en ligne.
- s'engager avec des groupes industriels, des organisations de la société civile et des régulateurs, y compris le ministère des Femmes, dans des initiatives conjointes visant à lutter contre la violence et le harcèlement fondés sur le genre.



Pakistan

Bien que les élections pakistanaises aient eu lieu en 2024, nos préparatifs électoraux ont commencé en 2022. Il s'agit notamment d'efforts déployés à l'échelle de l'entreprise, au niveau des produits, des politiques et des opérations, afin d'éviter et d'atténuer les risques pour les droits humains liés à l'utilisation de nos plateformes.

Nous avons notamment mis au point des signaux permettant d'identifier les contenus posant des problèmes en matière de droits et les cas à traiter en priorité. Conformément aux [principes de la GNI](#), nous avons publié une [étude de cas](#) en temps réel dans notre Espace modération sur les demandes de retrait reçues en décembre 2023 de la part du gouvernement pakistanais qui n'enfreignaient pas nos Standards de la communauté ni la loi locale.

Les interactions avec les parties prenantes étaient essentielles. Avant les élections, nous avons organisé des réunions d'information avec les acteurs régionaux des droits humains sur les efforts en matière d'intégrité des élections. Nous nous sommes également engagés auprès de la Commission électorale pour expliquer notre approche des demandes d'action gouvernementale et des droits humains. L'Internet a été régulé pendant la période précédant l'élection, et Meta a partagé des [informations](#) avec les parties prenantes sur la manière de mettre en place un serveur proxy pour se connecter à WhatsApp.

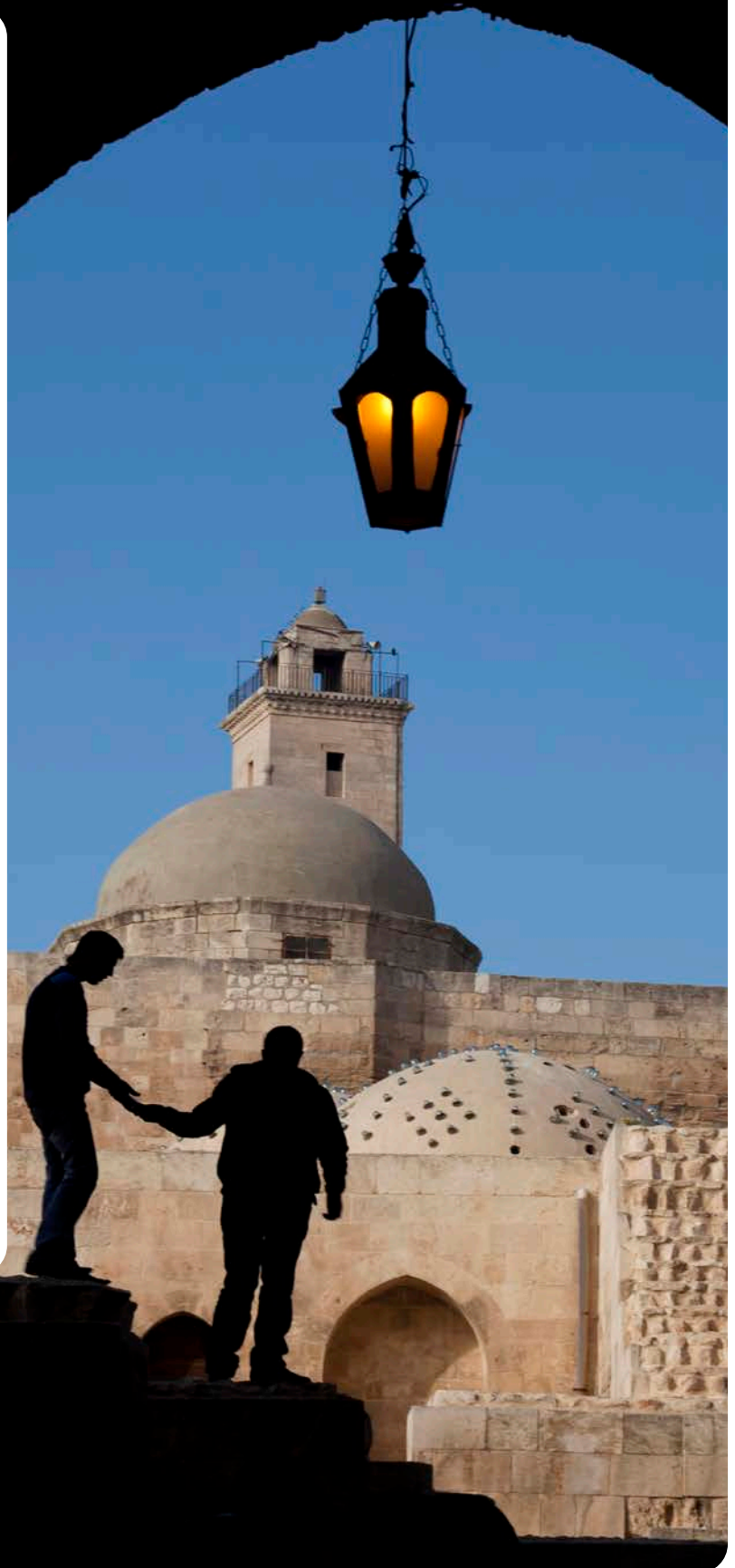
En tant que membre de groupes multipartites, dont la [GNI](#) et la [Asia Internet Coalition](#), Meta a contribué à des déclarations d'inquiétude concernant l'impact sur les droits des perturbations de l'Internet et de la législation portant atteinte aux droits avant les élections.

Plusieurs mois avant les élections, nous avons également pris des mesures pour fournir aux défenseurs des droits humains et aux autres personnes vulnérables des mesures de protection renforcées au niveau de la plateforme. Il s'agit notamment d'assurer une protection contre les suppressions de contenu fondées sur de fausses accusations de manquement aux Standards de la communauté. Notre approche était inclusive, conformément à notre [Politique d'entreprise en matière de droits humains](#), qui adopte la définition large des défenseurs des droits humains dans la [Déclaration sur les défenseurs des droits humains des Nations Unies](#).



Préparation aux défis dans les pays à réponse active

La classification des pays dans le cadre de la réponse active est déterminée par un processus fondé sur des preuves, piloté par des équipes spécialisées et nos spécialistes en droits humains, et s'appuie sur notre processus de diligence raisonnable en matière de droits humains pour examiner et classer par ordre de priorité les pays présentant le risque le plus élevé de préjudices et de violences hors ligne. Cela déclenche une stratégie approfondie d'atténuation des risques. Il peut s'agir d'une intensification des efforts de surveillance et d'atténuation des risques et d'investissements supplémentaires, par exemple en ce qui concerne l'examen du contenu dans les langues en rapport avec le conflit afin de supprimer le contenu contraire à la politique et de fournir un soutien sur mesure aux produits. En raison de la complexité des risques, nous faisons également appel à des partenaires de confiance et à des médias de vérification des informations tiers pour soutenir nos efforts dans les pays à réponse active.



Réponse aux crises mondiales

Nous répondons à des crises inattendues dans le monde entier, y compris des conflits interétatiques, des attaques terroristes et des crises non violentes telles que des catastrophes environnementales. Nos efforts de réponse aux crises impliquent de nombreux spécialistes en matière de politique de contenu, de global operations (opérations globales), de droits humains et de produits, ainsi que d'autres spécialistes internes qui évaluent la situation et y répondent, en appliquant notre [Protocole de politique de crise](#). Comme décrit dans notre [Rapport sur les droits humains 2022](#), le protocole guide notre utilisation accélérée de leviers politiques ciblés ou exceptionnels pour atténuer les préjudices potentiels en fonction des risques observés et en s'appuyant sur les interventions de crise passées et les principes des droits humains.

Notre évaluation globale des risques liés aux droits humains nous aide à établir des priorités entre les différents risques potentiels liés aux droits humains dans les situations de conflit, par exemple, entre le droit à la liberté et à la sécurité de la personne, le droit à la liberté d'expression, et la sécurité et le bien-être de nos communautés. Dans le cadre de notre boîte à outils, nous pouvons adopter un certain nombre de mesures temporaires afin d'assurer la sécurité des personnes et d'atténuer les risques que nos plateformes soient utilisées pour exacerber les tensions en ligne et hors ligne. Nous pouvons prendre en considération les développements pertinents du droit international public, y compris le droit international humanitaire, le droit des conflits armés, conformément aux

principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains, qui préconisent une diligence accrue en matière de droits humains en période de conflit.

Nous présentons ci-dessous des exemples illustratifs de notre travail. Nous fournissons



une [mise à jour](#) de nos actions sur la diligence raisonnable en matière de droits humains en Israël et en Palestine et sur les conflits connexes dans la sous-section suivante.

Soudan

En avril 2023, des combats ont éclaté entre les Forces armées soudanaises (SAF) et les Forces de soutien rapide (RSF), provoquant des déplacements massifs, l'insécurité alimentaire, le manque d'accès aux soins de santé et d'autres problèmes nécessitant une aide humanitaire. Sur la base du Protocole de politique de crise de Meta, le Soudan a été classé au niveau de risque le plus élevé.

Risques majeurs pertinents :

- vie, liberté et sécurité de la personne
- liberté d'opinion et d'expression
- liberté de réunion et d'association
- confidentialité
- droit à la santé

Guidés par notre Protocole de politique de crise, nous avons pris des mesures pour respecter les droits humains et minimiser les risques de préjudice, ainsi que pour soutenir la fourniture d'une aide humanitaire.

En août 2023, nous avons désigné les RSF dans le cadre de notre [politique relative aux organisations et personnes dangereuses](#), qui n'autorise pas les personnes ni les organisations qui proclament une mission violente ou qui ont perpétré des actes de violence à avoir une présence sur nos plateformes. Cela a permis d'endiguer la diffusion de contenus potentiellement nuisibles. La décision a été prise après concertation avec les agences des Nations Unies et sur la base des recherches et des rapports de diligence raisonnable en matière de droits humains établis par les principales organisations non gouvernementales et faisant état de graves violations des droits humains. Il a été examiné et approuvé par le Conseil de surveillance. Nous avons également désigné le Soudan comme un pays temporairement à haut risque, ce qui a permis de supprimer les contenus qui appelaient les civils à prendre les armes ou à les apporter dans un endroit précis du pays. Nous avons travaillé avec des médias de vérification des informations tiers pour déchiffrer, étiqueter et réduire la visibilité des fausses informations liées au conflit sur la plateforme. Nous avons également supprimé les fausses informations qui, selon les partenaires de confiance, pouvaient contribuer au risque de blessures physiques ou de violence imminentes.

En raison des déplacements massifs de la capitale, Khartoum, nombre de nos parties prenantes, y compris les fonctionnaires, la société civile, les médias et les défenseurs des droits humains, étaient difficiles, voire impossibles à atteindre. Pour ce faire, il a fallu identifier de nouveaux groupes et s'engager avec eux pour soutenir les efforts humanitaires, ainsi que pour aider à identifier et à arrêter la diffusion de contenus potentiellement nuisibles et de fausses informations. Nous nous sommes engagés avec les groupes les plus actifs de la société civile, notamment les médecins, les médias, les journalistes citoyens et les groupes de

femmes, à la fois au Soudan et principalement au sein de la diaspora. Nous les avons informés sur la politique de contenu et leur avons fourni une formation à la sécurité numérique et des outils de crise Meta. Il s'agissait notamment d'organiser des ateliers sur l'utilisation des communautés WhatsApp, qui ont ensuite été utilisées pour coordonner les évacuations et l'aide humanitaire.

Malgré la détérioration de la situation sécuritaire, nos partenaires de confiance ont continué à signaler des contenus liés à la violence et à l'incitation, aux fausses informations, au discours haineux, à l'intimidation et au harcèlement. Les évaluations des partenaires de confiance ont contribué à l'application de notre politique relative aux fausses informations et au danger sur les contenus utilisant des images hors contexte pour attiser les tensions, ainsi que sur les affirmations faites à l'encontre des agences humanitaires opérant au Soudan.

Comme pour les [autres conflits](#), nous avons supprimé certains contenus explicites violents sans appliquer de [pénalités](#), des sanctions pour les manquements qui entraînent des [restrictions de compte](#) au fur et à mesure qu'elles s'accumulent, et ce, afin d'éviter de trop pénaliser ou de restreindre les utilisateurs qui tentaient de sensibiliser le public aux conséquences du conflit. Cette approche s'inspire d'une recommandation du Conseil de surveillance dans le [cas de la vidéo explicite au Soudan](#) en 2022. Ce cas a donné lieu à un examen en 2023 de nos politiques sur le partage de contenus violents et explicites pour sensibiliser aux violations des droits humains, ainsi qu'à une explication plus détaillée de la manière dont nous appliquons notre [marge de tolérance d'intérêt médiatique](#).

Le pays subissant une importante pénurie de médicaments disponibles en raison d'un manque d'infrastructures médicales, nous avons modifié temporairement notre politique afin de permettre aux utilisateurs de demander ou de donner des médicaments, normalement interdits par notre

Renforcer la résilience : sécurité numérique et formation à la politique de contenu

2023 :

6

sessions de formation
organisées

300

bénéficiaires
concernés

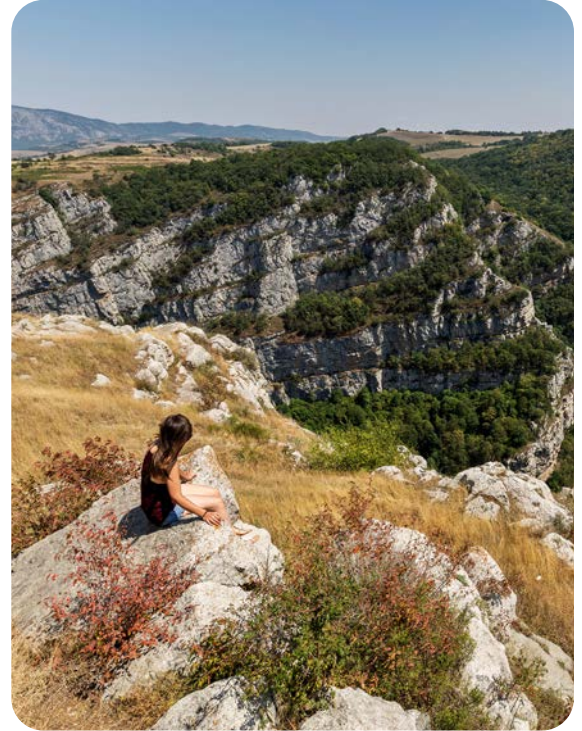
Après le conflit :

2

sessions de formation
organisées

70

bénéficiaires
concernés



politique sur les produits soumis à restriction. En outre, nous avons travaillé avec le Sudan Medical Specialization Board pour lancer la [SMSB Sudan Clinic](#), une application de télémédecine utilisant l'API de WhatsApp. Cela a été possible grâce à notre programme d'exonération des frais d'impact social de WhatsApp, un programme destiné aux partenaires gouvernementaux et à but non lucratif dans le cadre duquel nous renonçons à faire payer le coût des messages.

Nous devons également équilibrer les messages sur la mobilité et la sécurité, la communication d'informations qui aident les personnes à voyager tout en contribuant à assurer leur sécurité, sans soutenir la migration irrégulière vers les pays voisins, à savoir la République centrafricaine, le Tchad, l'Égypte, l'Éthiopie, la Libye et le Sud-Soudan. En effet, le risque d'exploitation humaine augmente à mesure que les besoins des populations déplacées s'accroissent, ce qui entraîne l'apparition de personnes malveillantes dans ces

zones. Nous nous sommes efforcés d'identifier les personnes malveillantes et de supprimer les contenus susceptibles d'exposer les personnes à un danger, y compris ceux relatifs à la traite des êtres humains.

Haut-Karabakh

Les 19 et 20 septembre 2023, l'Azerbaïdjan a lancé une offensive militaire dans la République autoproclamée d'Artsakh, qui fait partie de la région contestée du Haut-Karabakh. L'offensive a provoqué la fuite des Arméniens de souche de la région, déclenchant une crise des réfugiés.

Une équipe interfonctionnelle, comprenant des spécialistes en droits humains, a suivi la situation et a rapidement mobilisé des ressources et appliqué notre [Protocole de politique de crise](#). Des spécialistes internes possédant des compétences linguistiques et une connaissance du contexte ont traité la prévalence accrue des discours haineux, du harcèlement, des contenus explicites et des

Risques majeurs pertinents :

- liberté d'opinion et d'expression
- vie, liberté et sécurité de la personne
- confidentialité

références à des personnes et des organisations dangereuses. Nous nous sommes également engagés avec les agences des Nations Unies sur le terrain pour mieux comprendre la dynamique de la crise. Meta a fourni des mesures de protection supplémentaires aux acteurs civiques des deux côtés afin de réduire les risques d'usurpation d'identité, de harcèlement et de piratage.

Comme dans d'autres conflits, nous avons dû décider comment trouver un équilibre entre la préservation de la vie privée et de la dignité des prisonniers de guerre et ce qu'apporte la documentation des violations des droits humains. Les Standards de la communauté de Meta interdisent la publication d'informations susceptibles de révéler l'identité ou la localisation de prisonniers de guerre. Cette démarche est conforme aux responsabilités de l'entreprise en matière de droits humains en vertu des principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains. Ces responsabilités sont accrues en période de conflit armé et doivent s'appuyer sur les règles du droit international humanitaire. La révélation de l'identité et de la localisation des prisonniers de guerre peut mettre en péril leur sécurité, leur vie privée et leur dignité, ainsi que la sécurité de leur famille.

En même temps, cette exposition peut éclairer le débat public et sensibiliser aux mauvais traitements potentiels, y compris les violations des droits humains et du droit humanitaire international. Elle peut également créer une dynamique en faveur d'actions qui protègent les droits et garantissent la responsabilité. L'ampleur et la rapidité avec lesquelles les images de prisonniers de guerre peuvent être partagées via les réseaux sociaux compliquent la tâche consistant à résoudre ces intérêts divergents.

En 2023, le [Conseil de surveillance](#) a confirmé notre décision de laisser en ligne une publication Facebook contenant une vidéo montrant des prisonniers de guerre identifiables, avec l'écran d'avertissement « marquer comme dérangent ». Le Conseil s'est rangé à l'avis de Meta, estimant que dans cette situation, les publications pouvaient éclairer le débat public et sensibiliser l'opinion à d'éventuels mauvais traitements, y compris des violations des droits humains et du droit humanitaire international, ce qui « l'emportait sur le risque pour la sécurité et la dignité des prisonniers de guerre ».

Conservation des preuves

Nous soutenons la justice dans le cadre de tous les crimes internationaux. Nous travaillons depuis plusieurs années à l'étude d'initiatives respectueuses des droits en matière de conservation et de divulgation des preuves, en consultant largement la société civile, le monde universitaire et les spécialistes et organismes internationaux spécialisés dans les poursuites judiciaires.

Comme nous l'[avons indiqué](#) dans notre réponse aux recommandations du Conseil de surveillance, nous avons travaillé à l'élaboration d'une approche permettant aux tribunaux internationaux et aux mécanismes de responsabilisation autorisés par les Nations Unies, tels que les missions d'établissement des faits et les commissions d'enquête, de nous demander une conservation prolongée des données pertinentes pour leurs enquêtes en cours. Ce travail est en grande partie achevé et est distinct de nos [politiques de longue date pour répondre aux demandes de conservation des forces de l'ordre](#).

Nous avons informé une série de mécanismes autorisés par les Nations Unies et de rapporteurs spéciaux de notre approche et nous avons décrit la procédure à suivre pour présenter des demandes d'extension de la durée de conservation à Meta. Nous examinerons attentivement toutes les demandes reçues pour vérifier qu'elles sont conformes à nos politiques et à la législation applicable.

Il s'agit d'un nouveau domaine sans recommandations établies ou testées, et il subsiste d'importantes considérations juridiques, politiques et de protection de la vie privée inhérentes à ce travail. Nous prévoyons de faire le point sur nos travaux dans ce domaine dans nos [mises à jour semestrielles sur le Conseil de surveillance](#) et dans notre rapport annuel sur les droits humains.



Risques majeurs pertinents :

- vie, liberté et sécurité de la personne
- confidentialité

Migration en Amérique centrale et du Sud

Le trafic d'êtres humains peut entraîner l'exploitation de personnes vulnérables qui tentent de quitter leur pays d'origine, souvent à la recherche d'une vie meilleure. Nous nous sommes efforcés de supprimer les contenus qui facilitent ou coordonnent l'exploitation des êtres humains, y compris la traite des êtres humains et le trafic de migrants.

En 2023, nous avons identifié des risques potentiels pour les droits humains concernant le trafic de migrants à travers l'Amérique du Nord, l'Amérique centrale et l'Amérique du Sud, y compris le long du Darién Gap, une bande de jungle dangereuse et densément boisée qui traverse le nord de la Colombie et le sud du Panama. Ces zones présentent un risque accru d'activités commerciales illégales, de présence en ligne d'organisations dangereuses et de fausses informations pour les migrants et les réfugiés, en particulier lorsque certains d'entre eux tentent de se frayer un chemin jusqu'à la frontière entre les États-Unis et le Mexique.

En réponse à ces risques, nous avons appliqué nos [Standards de la communauté](#) et nos politiques commerciales sur le [commerce](#) qui interdisent l'exploitation humaine, y compris le trafic d'êtres humains. Nous avons collaboré avec la société civile et les autorités pour lutter contre les fausses informations, et nous avons continué à contrôler et à appliquer nos politiques en matière de contenus organiques, payants et commerciaux illicites. Nous avons également facilité la tâche des médias de vérification des informations tiers pour trouver et évaluer le contenu relatif à la frontière entre les États-Unis et le Mexique, car nous reconnaissons qu'il convient d'être rapide lors d'événements critiques. Nous avons utilisé la détection de mots-clés pour regrouper les contenus connexes en un seul endroit pour nos partenaires de vérification des informations, qui ont évalué le contenu sur notre plateforme et publié des articles en anglais et en espagnol pour lutter contre les fausses informations.



Nous soutenons également les communautés de migrants et de réfugiés ainsi que les organisations non gouvernementales par le biais de programmes d'intégrité et d'éducation tels que We Think Digital, qui met l'accent sur les connaissances dans le digital pour identifier les fausses informations, les fraudes et les arnaques.

Évaluations de l'impact de Meta sur les droits humains

Les méthodes de diligence raisonnable en matière de droits humains évoluent. Nous examinons les risques potentiels et tirons les leçons des efforts passés afin de déterminer comment nous pourrions améliorer notre approche et renforcer les efforts qui ont bien fonctionné.

Nous réalisons des études d'impact sur les droits humains pour des pays, des applications, des services, du matériel ou des initiatives stratégiques spécifiques, sur la base de notre évaluation globale des risques liés aux droits humains. Nous effectuons ces évaluations pour nous aider à anticiper les impacts potentiels, en particulier en préparation de nouveaux produits et de nouvelles fonctionnalités, tels que le [chiffrement de bout en bout](#). Les études d'impact sur les droits humains impliquent l'engagement des [parties prenantes](#), un élément fondamental de la diligence raisonnable en matière de droits humains. Lorsque nous entreprenons une évaluation, chaque recommandation requiert souvent plusieurs équipes et groupes de travail pour la mettre en œuvre dans l'ensemble de l'entreprise.

Nous avons partagé les résumés et les recommandations des études d'impact sur les droits humains dans le passé, et ils

sont énumérés dans l'[Annexe](#). Nous faisons également le point sur les mesures prises en réponse aux évaluations menées sur les risques potentiels pour les droits humains liés aux plateformes Meta aux [Philippines](#), en [Inde](#), en [Israël](#) et en [Palestine](#), et liés au [chiffrement de bout en bout](#).

Israël et Palestine : Mise à jour de nos actions

Peu de temps après la publication de notre mise à jour de [septembre 2023 sur la diligence raisonnable en matière de droits humains en Israël et en Palestine](#), les attaques terroristes du 7 octobre du Hamas contre Israël ont eu lieu, suivies de la réponse militaire d'Israël à Gaza et de l'implication et de l'incitation à la violence d'autres acteurs régionaux. Nous savons que les habitants de la région et du monde entier ont été profondément touchés par notre réponse à la situation de conflit qui perdure dans la région.

Immédiatement après l'attaque, Meta a classé la violence au niveau le plus élevé de notre protocole de politique de crise et a mis en œuvre des mesures immédiates de réponse à la crise, y compris une équipe interfonctionnelle dédiée, disponible 24 heures sur 24 et 7 jours sur 7. Les principes fondamentaux des droits humains ont guidé nos actions. Nous nous sommes appuyés sur les [Principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains](#), qui sont à la base de notre [Politique d'entreprise connexe](#), afin de donner la priorité et d'atténuer les risques les plus importants en la matière. Nous avons également utilisé le droit international humanitaire comme référence importante.

Nous avons initialement communiqué les détails de notre réponse dans une publication de blog en [anglais](#), [arabe](#) et [hébreu](#) le

Risques majeurs pertinents :

- vie, liberté et sécurité de la personne
- liberté d'opinion et d'expression
- égalité et non-discrimination
- confidentialité



13 octobre 2023, et avons fourni d'autres mises à jour le 18 octobre, le 5 décembre et le 8 décembre. Nous avons continué à affiner notre approche pour tenir compte de l'évolution de la dynamique, notamment de la crise humanitaire en cours à Gaza et des otages toujours retenus en captivité par le Hamas.

L'équilibre entre la sécurité et le bien-être de nos communautés et la liberté d'expression sur nos plateformes est particulièrement difficile dans les situations de conflit, et encore plus difficile lorsqu'il s'agit d'entités sanctionnées par les États-Unis, telles que le Hamas, le Hezbollah et le Jihad islamique palestinien, qui sont également désignées comme des organisations dangereuses dans le cadre de nos politiques.

Le travail que nous avons déjà effectué en réponse à la [diligence raisonnable en matière de droits humains en Israël et en Palestine](#) a contribué à éclairer notre approche. Par exemple, nous avons été en mesure d'acheminer le contenu arabe de manière plus appropriée à travers nos systèmes pour offrir une plus grande précision au public. Nous avons également augmenté nos ressources en matière de modération de contenu sur le marché hébraïque. Pour une mise à jour plus détaillée, voir [Mise à jour Meta 2024 : Diligence raisonnable en matière de droits humains en Israël et en Palestine](#).

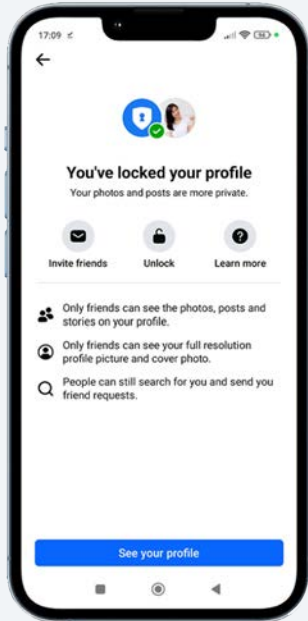
Notre réponse au conflit est allée au-delà des mesures décrites dans la [diligence raisonnable](#) et a également été guidée par l'utilisation de notre [Protocole de politique de crise](#). Nous avons adopté un certain nombre de [mesures temporaires liées aux produits et aux politiques](#) pour atténuer les principaux risques en matière de droits humains. Nous savons que dans certains cas, ces mesures, par exemple, l'abaissement des seuils d'application automatisée, peuvent involontairement limiter la discussion sur les événements mondiaux critiques, ce qui a un impact sur la liberté d'expression. C'est pourquoi nous cherchons à prendre des mesures limitées dans le temps et proportionnées aux risques tels que nous les connaissons.

Mesures politiques temporaires

Nous nous sommes d'abord appuyés sur notre politique relative aux personnes et organisations dangereuses et avons élargi notre politique sur la violence et l'incitation afin de supprimer les contenus montrant des otages identifiables enlevés ou retenus en captivité, même si cela a été fait pour condamner leur situation ou sensibiliser l'opinion à celle-ci. Nous avons agi ainsi pour protéger la vie privée et la dignité des victimes, conformément aux normes du droit humanitaire international, et pour empêcher la propagande du Hamas d'apparaître sur nos plateformes, conformément à notre politique relative aux personnes et organisations dangereuses.

Au fur et à mesure que la guerre se poursuivait, nous avons constaté que certaines personnes commençaient à partager des contenus sur les otages afin de réfuter les récits émergents niant l'existence des attaques terroristes du 7 octobre. Nous avons donc adopté une approche plus équilibrée et autorisé le partage de contenus montrant des otages en train d'être enlevés afin de sensibiliser le public ou de condamner les attaques. Nous avons continué à supprimer des contenus et à privilégier la sécurité si l'intention de l'utilisateur à l'origine du partage n'était pas claire.

Nous avons apporté d'autres modifications visant à protéger la liberté d'expression. Pour certains domaines d'actions, comme certains types de [contenu violent et explicite](#), nous avons supprimé le contenu en infraction sans appliquer de [pénalités](#), des sanctions pour les manquements qui entraînent des [restrictions de compte](#) au fur et à mesure qu'elles s'accumulent, et ce, afin d'éviter de trop pénaliser ou de restreindre les utilisateurs qui tentaient de sensibiliser le public aux conséquences du conflit.



Mesures temporaires sur les produits

Nous avons adopté un certain nombre de mesures temporaires sur les produits afin d'assurer la sécurité des personnes et d'atténuer les principaux risques en matière de droits humains, comme indiqué dans notre [publication de blog décrivant notre réponse au conflit](#). Il s'agit notamment d'ajustements des seuils de confiance pour l'action automatique sur le contenu, le blocage de certains hashtags dans les recherches et les modifications apportées aux produits pour traiter les commentaires indésirables et problématiques, ainsi que des mesures de sécurité telle que l'outil [verrouiller votre profil](#).

Nous avons également réduit temporairement le seuil à partir duquel les contenus limites ou potentiellement violents, tels que les images ou les vidéos présentant une violence explicite, peuvent être rendus inéligibles à la recommandation. Cette mesure s'applique au contenu non connecté, c'est-à-dire au contenu provenant de personnes qu'un utilisateur n'a pas encore commencé à suivre et qui peut apparaître sur des surfaces telles que Fil, Rechercher, Explorer et Reels.

Comme pour nos mesures politiques temporaires, nous avons apporté d'autres modifications visant à protéger la liberté d'expression. En réponse à une forte augmentation de l'utilisation de nos produits, nous avons temporairement ajusté certaines limites de taux automatisées conçues pour empêcher les comportements définis comme [du spam dans nos politiques](#), tels que la publication à haute fréquence, pour les rendre plus permissives, réduisant ainsi le risque de restrictions pour les utilisateurs légitimes.

Par ailleurs, nos politiques plus générales sur la [recommandabilité](#) et la [rétrogradation](#), qui ne font pas partie de nos actions de réponse à la crise, peuvent avoir un impact sur la visibilité des contenus liés au conflit, notamment s'ils s'avèrent explicites ou violents.

Meta a fréquemment mené des enquêtes pour évaluer les performances de nos systèmes de modération de contenu et déterminer s'ils appliquaient trop ou pas assez les règles en matière de contenu. Pour l'avenir, nous avons identifié des domaines dans lesquels des améliorations doivent être apportées, notamment en donnant la priorité à l'examen du contenu arabe par dialecte et en renforçant le contrôle et la surveillance de nos banques d'images et de vidéos afin d'atténuer les risques potentiels d'une application excessive de la législation et les conséquences qui en découlent pour la liberté d'expression. Ce travail est en cours.

Pour des informations plus détaillées sur notre réponse, veuillez vous référer à nos lettres en réponse aux questions que nous avons reçues des organisations de la société civile en cliquant [ici](#) et [ici](#). Pour plus de détails sur notre mise à jour 2024 sur la diligence raisonnable en matière de droits humains en Israël et en Palestine, voir [ici](#).



Risques majeurs pertinents :

- confidentialité
- liberté d'opinion et d'expression
- liberté de réunion et d'association
- égalité et non-discrimination

Chiffrement de bout en bout :**Mise à jour de nos actions**

En décembre 2023, nous avons commencé à déployer le chiffrement de bout en bout par défaut pour toutes les discussions et tous les appels personnels sur Facebook et Messenger. Cela a marqué une étape importante dans notre parcours, lancé en 2019, visant à étendre le chiffrement de bout en bout par défaut à toutes nos applications de messagerie. Nous continuons également à travailler sur le chiffrement de bout en bout pour Instagram Direct. WhatsApp propose un chiffrement de bout en bout par défaut depuis 2016.

Le chiffrement de bout en bout renforce la sécurité et la confidentialité en garantissant que personne ne voit votre message, à l'exception de vous et de la personne à qui vous envoyez un message. Il s'agit d'une couche de sécurité supplémentaire qui protège vos messages et vos appels avec votre famille et vos amis et signifie que personne, y compris Meta, ne peut voir ni

écouter ce qui est envoyé ou dit. Lorsque vous signalez une conversation chiffrée de bout en bout, les messages récents de cette conversation seront envoyés en toute sécurité depuis votre appareil à nos équipes d'aide.

Le cadre des droits humains a joué un rôle clé dans l'orientation de nos travaux visant à étendre le chiffrement de bout en bout. Le chiffrement de bout en bout favorise directement le droit à la vie privée, qui à son tour permet la réalisation de toute une série d'autres droits humains, notamment la liberté d'expression, d'association, d'opinion, de mouvement et la sécurité corporelle. En parallèle, il existe toujours un risque que des personnes abusent d'un service de messagerie, y compris des fonctionnalités chiffrées, pour nuire à autrui.

En 2022, nous avons communiqué l'intégralité des résultats d'une évaluation indépendante de l'impact sur les droits humains, réalisée par BSR, que nous avons chargée d'examiner les avantages et les risques potentiels pour les droits humains de nos projets visant à étendre le chiffrement de bout en bout par défaut. L'évaluation a permis de formuler 45 recommandations à l'intention de Meta afin de maximiser les avantages potentiels et d'atténuer les effets négatifs éventuels. Depuis notre réponse initiale à ces recommandations, nous avons continué à travailler pour mettre en œuvre nos engagements déclarés et avons fourni des mises à jour publiques approfondies, notamment en janvier et en août 2023, ainsi qu'un livre blanc détaillé sur la sécurité.

Problématiques clés

Risques majeurs pertinents :

- confidentialité
- vie, liberté et sécurité de la personne

Combattre les menaces contre la sécurité

Au cours des six dernières années, Meta a publié des rapports publics sur les efforts déployés pour détecter et contrer les menaces à la sécurité sur nos plateformes. Notre analyse et nos enquêtes postérieures à la perturbation ont servi de base au Citizen Lab et à 45 groupes de la société civile pour présenter un dossier au Sommet pour la démocratie afin d'interdire la vente de logiciels espions jusqu'à ce que des garanties soient mises en place pour prévenir les violations des droits humains. Par la suite, le président des États-Unis, Joe Biden, a édicté un décret interdisant aux entités fédérales américaines d'utiliser des logiciels espions pour commettre des violations des droits humains.

Certains exemples de menaces que nous avons atténuées en 2023 concernent le secteur de la surveillance pour compte d'autrui. Les sociétés de surveillance pour compte d'autrui sont des entreprises qui proposent des services de surveillance et de suivi de personnes ou d'organisations à des fins diverses. La protection de la vie privée et le recours abusif à la surveillance par des gouvernements autoritaires, des organisations criminelles et d'autres acteurs malveillants qui tentent d'espionner les détracteurs du régime, les personnalités de l'opposition, les journalistes et les défenseurs des

droits humains afin de censurer la liberté d'expression suscitent des inquiétudes.

En 2023, nous avons identifié et supprimé six réseaux distincts de comptes sur Facebook et Instagram, liés à huit sociétés de surveillance en Italie, en Espagne et dans les Émirats arabes unis. Parmi leurs tactiques figurent l'ingénierie sociale et l'hameçonnage sur l'Internet, ainsi que l'utilisation de personnages se faisant passer pour des manifestants, des journalistes et des jeunes femmes afin d'inciter les personnes à communiquer leurs adresses e-mail et leurs numéros de téléphone et à cliquer sur des liens malveillants.

« L'essor de l'industrie des logiciels espions privés démocratise l'accès à des mesures de surveillance sophistiquées tout en occultant les clients qui achètent et déploient ces services. Ce ciblage aveugle et opaque constitue une menace importante pour les défenseurs des droits humains. C'est pourquoi nous nous engageons à travailler avec nos partenaires de la société civile et avec l'ensemble de l'industrie pour que les fabricants de logiciels espions abusifs soient tenus pour responsables ».

David Agranovich, directeur, Threat Disruption, Meta

Risques majeurs pertinents :

- intérêt supérieur de l'enfant
- confidentialité
- liberté d'opinion et d'expression

Sécurité des enfants et des jeunes

La sécurité en ligne des enfants et des jeunes⁴ est un défi à l'échelle de l'Internet qui requiert des efforts et une collaboration entre le secteur, le gouvernement et la société civile, car nous cherchons à assurer la sécurité des utilisateurs tout en leur offrant un espace pour exercer divers droits humains, la liberté d'expression et l'accès à l'information, entre autres.

Nous avons passé une décennie à travailler sur cette question et nous continuons à utiliser une approche fondée sur les droits pour la conception de nos services et de notre politique de contenu. Notre cadre relatif à [l'intérêt supérieur de l'enfant](#) s'appuie sur [la Convention des Nations Unies relative aux droits de l'enfant](#), ainsi que sur des réglementations et des orientations telles que le code britannique de conception adaptée à l'âge, les principes fondamentaux de l'enfance de la Commission irlandaise de protection des données et la recommandation de la Commission nationale de l'informatique et des libertés (CNIL) sur les mineurs.

Exploitation d'enfants

La sécurité des enfants est toujours l'une de nos priorités. Nous cherchons à prévenir les dommages en appliquant nos politiques et en développant des outils de prévention. Nous avons travaillé avec diligence pour rendre les outils de signalement de

l'exploitation des enfants plus faciles à trouver, ce qui a conduit à une augmentation de 75 % des signalements qui nous ont été envoyés par les ados au premier trimestre 2023 par rapport au premier trimestre 2022 sur Messenger et Instagram. Nous avons également introduit de nouvelles méthodes innovantes pour trouver, supprimer ou restreindre de manière proactive les comptes susceptibles d'enfreindre nos politiques en matière de sécurité des enfants. Par exemple, nous avons lancé une technologie proactive pour détecter et désactiver automatiquement les comptes (et tous les comptes qui leur sont liés) s'ils interagissent dans des activités et des interactions sur la plateforme qui démontrent un intérêt sexuel pour les enfants. Plus de 90 000 comptes ont été identifiés et automatiquement [supprimés](#) entre le 1er août 2023 et le 31 décembre 2023, grâce à cette méthode. Nous nous efforçons de détecter de manière



⁴ Il n'existe pas de définition universelle de la « jeunesse ». La [Convention des Nations Unies relative aux droits de l'enfant](#) définit un « enfant » comme tout être humain âgé de moins de 18 ans (article 1), ce qui englobe également les ados de plus de 13 ans. Dans ce rapport, nous utilisons indifféremment les termes « jeunes » et « ados ».



proactive et de restreindre les comptes appartenant à des adultes dès les premiers signes d'interactions indésirables avec des ados ou sur des ados, ou de tentatives de contacts potentiellement dangereux.

En tant que membre fondateur de [Take It Down](#), un outil conçu pour empêcher de manière proactive la diffusion en ligne d'images intimes de jeunes personnes, nous avons travaillé avec le [National Center for Missing and Exploited Children](#) pour l'étendre à d'autres langues et avons encouragé d'autres acteurs du secteur à y adhérer, ce qui a permis à des millions d'ados supplémentaires d'y avoir accès.

Nous sommes également un membre fondateur de [Lantern](#), un nouveau programme de la Tech Coalition lancé en novembre 2023 qui permet aux entreprises technologiques de partager des signaux sur les comptes et les comportements prédateurs sur les plateformes en ligne. Business for Social Responsibility a réalisé une [évaluation de l'impact de ce programme sur les droits humains](#).

« La prévention de l'exploitation des enfants est l'un des défis les plus importants auxquels notre secteur est confronté aujourd'hui. Les prédateurs en ligne sont des criminels déterminés qui utilisent plusieurs applications et sites Web pour cibler les jeunes. Ils testent également les défenses de chaque plateforme et apprennent à s'adapter rapidement. C'est pourquoi, aujourd'hui plus que jamais, nous travaillons dur pour rester à la pointe de la technologie dans ce domaine. Outre le développement d'une technologie permettant d'éradiquer les prédateurs, nous recrutons des spécialistes de la sécurité des enfants en ligne et nous partageons des informations avec nos homologues du secteur et les forces de l'ordre. »

Antigone Davis, responsable mondiale de la sécurité, Meta



Sécurité et bien-être des jeunes

Nous voulons que les jeunes puissent vivre des expériences en ligne sûres et adaptées à leur âge, tout en respectant leurs droits numériques. Notre approche est axée sur la prévention des préjudices, sur la mise à disposition de moyens de contrôle de l'expérience vécue et sur une réaction rapide en cas de manquements à nos politiques. Nous avons développé de nombreux [outils, fonctionnalités et ressources](#) pour soutenir les ados et leurs parents. Ce travail est en cours.

En 2023, nous avons continué à développer notre travail autour de nos [outils de supervision parentale](#) sur l'ensemble de nos services. Nous avons organisé des [séances de co-conception](#) avec des parents, des ados, des représentants légaux et des spécialistes dans le cadre des [Labos de confiance, de transparence et de contrôle \(TTC Labs\)](#), un effort interprofessionnel visant à permettre aux personnes de contrôler leur vie privée, et avons travaillé avec des conseillers pour la jeunesse afin de comprendre les comportements des ados et ainsi d'éclairer la conception des produits. Nous avons également lancé des [comptes gérés par les parents](#) pour les préados âgés de 10 à 12 ans sur Meta Quest 2 et 3.

Nous consultons régulièrement des spécialistes en développement des ados, en psychologie et en santé mentale pour que nos plateformes soient sûres et adaptées à l'âge des jeunes, notamment en développant continuellement notre compréhension des types de contenu qui conviennent le mieux aux ados. Des consultations ont eu lieu avec des [partenaires de sécurité](#), des groupes de jeunes et des conseillers du monde entier, ainsi qu'avec l'UNICEF, afin de les informer et de les impliquer dans notre travail.

En tenant compte des avis des spécialistes, nous avons développé de [nouvelles mesures de protection](#) pour offrir des expériences plus adaptées à l'âge sur nos applications, par exemple en masquant davantage de résultats de recherche, en invitant les ados à mettre à jour leurs paramètres de confidentialité sur Instagram et en plaçant automatiquement les ados dans le paramètre de contrôle du contenu le plus restrictif sur Facebook et Instagram.

Interactions avec les parties prenantes

Nous avons organisé des [sessions de co-conception](#) dans le cadre des [TTC Labs](#) afin d'impliquer les utilisateurs, notre communauté et des spécialistes dans la création d'expériences en ligne positives pour les jeunes, en écoutant leurs besoins et leurs préoccupations afin d'éclairer la conception de nos produits.

87 ados

71 représentants légaux

221 spécialistes

13 pays

Interactions avec les parties prenantes



L'engagement proactif avec un large éventail de parties prenantes est au cœur de notre approche des droits humains chez Meta. Il apporte des connaissances externes et divers points de vue et commentaires dans nos processus d'élaboration de produits et de politiques de contenu, et contribue à assurer la responsabilité et la transparence. Les groupes avec lesquels nous nous engageons vont des organisations internationales aux communautés marginalisées et sous-représentées, en passant par divers groupes de la société civile et des universitaires, ainsi que des utilisateurs tels que des [ados, des parents et des soignants](#). Nous recueillons les points de vue de toutes les parties d'une question et de toutes les régions du monde afin de garantir l'équilibre.

En outre, nous avons continué à impliquer nos [partenaires de confiance](#) afin d'identifier les tendances, de mieux comprendre les défis posés par nos canaux d'escalation et d'explorer comment nous pouvons rendre ces mécanismes plus efficaces pour les partenaires de la société civile.

L'engagement proactif avec un large éventail de parties prenantes est au cœur de notre approche des droits humains chez Meta.

Groupes marginalisés et défenseurs des droits humains

Nous reconnaissons l'importance d'un engagement significatif avec les parties prenantes des communautés marginalisées, en tenant compte des variations au sein d'une région et même d'un pays, et nous cherchons à nous informer sur ces différences.

Nous nous efforçons d'écouter et de demander conseil aux spécialistes et militants des droits humains, aux universitaires et à d'autres personnes sur un large éventail de questions et nous les informons des développements pertinents au sein de Meta. Leurs avis éclairent des domaines tels que l'élaboration de nos politiques de contenu et leur mise en application afin de promouvoir la liberté d'expression et la sécurité.

Dans le monde entier, les défenseurs des droits humains qui utilisent les plateformes numériques comme instruments de leur militantisme sont souvent victimes d'intimidation, d'oppression et de poursuites judiciaires en raison de leurs activités en ligne visant à faire progresser et à protéger les droits humains. Ce phénomène est particulièrement marqué au sein des groupes marginalisés, tels que les femmes et la communauté LGBTQIA+.

Guidés par notre cadre [d'inclusion](#), nous utilisons des tables rondes, des ateliers et des réunions individuelles pour comprendre les questions politiques pertinentes en matière de contenu, identifier les lacunes politiques et rechercher des possibilités d'engagement dans l'élaboration des politiques. Par exemple, nous avons collaboré avec le [Minority Rights Group](#) pour organiser huit réunions mondiales dans des pays considérés comme exposés à un risque de conflit armé ou de violence sociétale.

Nous nous sommes également engagés auprès de plus de 250 parties prenantes issues de milieux divers, notamment des femmes, des membres de la communauté LGBTQIA+, des minorités religieuses, des minorités ethniques et des

groupes autochtones, afin de recueillir des informations précieuses sur les défis politiques auxquels ces communautés sont confrontées. Sur la base des commentaires recueillis lors de plus de 19 consultations et ateliers avec des organisations LGBTQIA+ et des défenseurs des droits humains au Moyen-Orient et en Afrique du Nord, nous avons collaboré avec la Jordan Open Source Association (JOSA) pour publier le [Digital Security Toolkit pour les défenseurs des droits humains](#). Ce guide, disponible à la fois en [arabe](#) et en [anglais](#), identifie les meilleures pratiques en matière de sécurité numérique et propose des caractéristiques de sécurité, des conseils et des mesures réactives pour améliorer la sécurité en ligne des activistes et des défenseurs des droits humains au Moyen-Orient et en Afrique du Nord. Nous avons également lancé une campagne d'influence LGBTQIA+ en Afrique subsaharienne afin de sensibiliser les militants de cette cause aux ressources en matière de sécurité et de souligner notre engagement à protéger les utilisateurs à haut risque dans la région.

Après un projet pilote réussi de deux ans axé sur l'Asie-Pacifique, nous avons élargi en 2023 notre

[Fonds pour les défenseurs des droits humains](#), géré en partenariat avec [Civil Rights Defenders](#), afin de soutenir les défenseurs des droits humains à l'échelle mondiale et avons renouvelé le financement pour un montant supplémentaire de 500 000 dollars américains à appliquer en 2024.

AfricanDefenders

En nous appuyant sur les enseignements et les principes de notre Fonds pour les défenseurs des droits humains en [Asie-Pacifique](#) et en partenariat avec [AfricanDefenders](#), nous avons lancé le [Fonds pour les défenseurs des droits humains en Afrique](#), conçu pour soutenir les défenseurs qui sont la cible de harcèlement, de persécution et/ou de poursuites en raison de leurs activités de lutte en ligne. Il s'agit notamment des membres de groupes marginalisés qui défendent leurs droits, tels que les militants et les journalistes citoyens, les militants politiques non violents, les défenseuses des droits humains et la communauté LGBTQIA+. Les micro-subventions serviront à soutenir les situations d'urgence et à financer l'achat de nouveaux appareils et de technologies de sécurité, le relogement temporaire et l'assistance juridique et de sécurité d'urgence.

Partenaires de confiance

En janvier 2023, notre réseau de partenaires de confiance comprenait plus de 400 organisations non gouvernementales, agences humanitaires, défenseurs des droits humains et chercheurs de 113 pays du monde entier. En plus de notre engagement continu, étendu et diversifié auprès des parties prenantes, ces partenaires apportent de multiples connaissances et expériences qui contribuent à la modération des contenus et à l'élaboration de politiques efficaces et transparentes. Ils ont également accès à un canal d'escalation pour signaler des contenus potentiellement nuisibles et des problèmes de sécurité de compte, auxquels nous donnons suite le cas échéant pour assurer la sécurité des utilisateurs.

En 2023, les rapports de nos partenaires de confiance nous ont aidés à supprimer 49 600 contenus qui enfreignaient nos politiques. Les partenaires de confiance nous ont également aidé à découvrir et à démanteler six réseaux trompeurs qui se livrent à un comportement non authentique coordonné aux États-Unis, en Géorgie, au Burkina Faso et au Togo. En outre, nous avons pris des mesures contre trois opérations distinctes de cyber espionnage en Asie du Sud. Ces réseaux ont été conçus pour manipuler le débat public à des fins stratégiques, ce qui peut constituer une menace pour les droits humains.

En 2023, nous avons entrepris un examen complet de nos canaux de réception des contenus d'urgence et des escalations relatives à la sécurité des comptes en provenance des partenaires de la société civile. L'organisation d'un sommet mondial auquel ont participé des spécialistes politiques, opérationnels et juridiques de Meta ainsi que 15 partenaires de la société civile de 13 pays d'Asie-Pacifique, du Moyen-Orient, d'Europe, d'Amérique latine et d'Afrique a constitué un élément clé de ce travail. Grâce à cet engagement, nous avons pu évaluer de manière globale nos canaux d'escalation de la société civile et formuler une vision commune pour leur avenir, ce qui a amélioré la hiérarchisation des rapports et renforcé les processus de réponse aux situations d'urgence et de collecte d'informations.





Étude de cas : Apparition de nouveaux préjudices en Haïti

Le 7 juillet 2023 a marqué le deuxième anniversaire de l'assassinat du président haïtien Jovenel Moïse. Les tensions restent vives dans le pays, exacerbées par la résurgence de la violence des gangs. Nos partenaires de confiance ont signalé des contenus liés à de multiples domaines d'infraction, notamment la violence et l'incitation, les personnes et organisations dangereuses, l'intimidation et le harcèlement, et la violence explicite. Ils nous ont également alertés de la présence d'un nouveau mouvement d'autodéfense, Bwa Kale, sur nos plateformes. Ces informations ont permis à nos équipes opérationnelles et d'enquête d'exploiter les mots-clés et d'identifier de manière proactive les contenus potentiellement en infraction, ce qui a permis d'atténuer efficacement les préjudices hors ligne.

Étude de cas : Renforcement de l'information en Éthiopie

Compte tenu de l'aggravation des tensions sociopolitiques et de la poursuite des violences en Éthiopie, nous avons rencontré nos partenaires de confiance tous les mois pour discuter des

tendances en matière de contenu et renforcer les pratiques de signalement des contenus à haute gravité par l'intermédiaire du canal des partenaires de confiance. Nous avons également organisé une série de dialogues bimensuels sur l'Éthiopie avec la société civile et la diaspora éthiopiennes. Nos partenaires en Éthiopie ont continué à signaler les contenus et à fournir des évaluations du préjudice pour les fausses informations et les rapports de menaces voilées. En travaillant avec ces partenaires, nous avons pu mieux identifier les principaux problèmes, notamment les images hors contexte, les faux comptes qui diffusent des fausses informations, les comptes vérifiés qui tirent parti de leur influence pour diffuser des fausses informations, les discours haineux, le harcèlement coordonné et le « doxxing », les termes à connotation politique initialement inventés par des personnalités publiques, ainsi que les personnes et les organisations dangereuses.

Étude de cas : Atténuer les conflits intra-religieux au Bangladesh

En mars 2023, des émeutes intra-religieuses ont éclaté entre musulmans sunnites et musulmans

ahmadis au Bangladesh, une communauté musulmane minoritaire. Nos partenaires de confiance au Bangladesh ont fait état de contenus comprenant des fausses informations nuisibles, des discours haineux, ainsi que des actes de violence et d'incitation visant la communauté ahmadie. En tirant parti d'allégations nuisibles préalablement désignées et examinées, nous avons été en mesure d'appliquer rapidement les rapports de fausses informations. Les partenaires de confiance ont également fourni des signaux locaux essentiels qui nous ont permis de supprimer des contenus dans le cadre de politiques d'escalation uniquement, y compris des menaces voilées, ainsi que des informations sur les termes locaux relatifs aux discours haineux.

Organisations internationales et multipartites

Au fur et à mesure que la technologie progresse et que son impact sur les droits humains s'accroît, nous continuons à renforcer notre engagement auprès d'un large éventail d'organismes gouvernementaux internationaux, notamment :

- Le [Bureau du Haut-Commissariat aux droits humains](#)
- L'[UNESCO](#)
- Le [Bureau de la prévention des génocides et de la responsabilité de protéger](#)
- Le [Bureau de l'Envoyé du Secrétaire général pour la technologie](#)
- L'[UNICEF](#)
- Le [HCR](#)
- Le [Forum sur la gouvernance de l'Internet](#)
- La [Fondation des Nations Unies](#)
- L'[OCDE](#)
- Le [Forum économique mondial](#)
- Diverses organisations régionales



Nous sommes membres de diverses initiatives multipartites, telles que le [Global Internet Forum to Counter Terrorism](#), la [Global Network Initiative](#) et le [Freedom Online Coalition Advisory Network](#). Nous nous engageons également de manière proactive dans des collaborations à l'échelle du secteur afin d'y contribuer et de soutenir la politique gouvernementale.

Nous avons participé au deuxième [Sommet pour la démocratie](#) et partagé publiquement nos engagements en réponse à l'appel lancé par le sommet au secteur privé en faveur de la démocratie. Nous avons été reconnus pour notre rôle de co-leader dans l'élaboration de principes industriels visant à « contrer les menaces croissantes posées par les cyber mercenaires » et nous avons contribué à l'appel lancé au secteur privé pour qu'il soutienne les [technologies anti-censure affirmant la démocratie pour lutter contre les régimes autoritaires](#). En outre, nous avons été récompensés pour notre travail visant à permettre des connexions sécurisées et privées, et à contrer l'utilisation abusive de la technologie numérique par les gouvernements, notamment WhatsApp via un serveur proxy et la [surveillance pour le compte d'autrui](#).

Nous avons participé aux consultations sur l'élaboration du [Global Digital Compact](#) avec le [Bureau de l'Envoyé du Secrétaire général pour la Technologie](#) afin de soutenir la nécessité d'une approche harmonisée au niveau mondial de la gouvernance de l'Internet et de la protection des droits humains en ligne, y compris les droits à la vie privée et à la liberté d'expression. Nous avons également participé aux consultations sur le Code de conduite des [Nations Unies pour l'intégrité de l'information](#) et avons présenté un rapport détaillé.

Nous avons continué à rencontrer régulièrement des fonctionnaires du Haut-Commissariat des Nations Unies aux droits humains, des spécialistes indépendants et des rapporteurs spéciaux pour discuter de questions internationales et spécifiques à certains pays, et nous avons participé activement au [projet B-Tech](#). Nous nous sommes engagés à respecter les droits humains à l'occasion du 75e anniversaire de la Déclaration universelle des droits de l'homme. Il s'agit notamment d'engagements renouvelés en matière de diligence raisonnable et de divulgation des droits humains, ainsi que d'engagement auprès de la société civile et des communautés concernées, entre autres.

Meta s'est engagée dans l'[effort de collaboration](#) sur un [Internet de confiance](#) de l'UNESCO en participant au sommet de février à Paris et au processus multipartite visant à élaborer des orientations axées sur les droits humains pour la réglementation des plateformes numériques.

En collaboration avec le [Forum économique mondial](#), nous avons participé à un vaste projet visant à dégager un consensus sur la politique technologique émergente. Nous avons contribué à l'élaboration et à l'adoption des [Principes mondiaux sur la sécurité numérique : traduire les droits humains internationaux pour le contexte numérique](#), qui établit un cadre destiné à aider les gouvernements et les fournisseurs de services en ligne à faire progresser la sécurité numérique par le biais d'une approche multipartite. Nous faisons également partie de l'[AI Governance Alliance](#) et de l'initiative [Defining](#)

« Nous ne pouvons pas le faire seuls. Nous avons besoin de ces engagements et de ces partenariats stratégiques pour que ce projet ait un impact ».

Lene Wendland, Responsable, Unité Entreprises et droits humains, Droits humains des Nations Unies

[and Building the Metaverse](#), qui vise à guider le développement d'un métavers éthique et responsable. L'année dernière, nous avons rejoint l'[AI Governance Alliance](#), qui tend à établir des normes mondiales pour la gouvernance de l'IA, en contribuant de manière positive aux avancées technologiques pour le bien-être de la société, guidée par des engagements en matière d'éthique et de diversité.

Nous avons commencé à travailler avec la [Commission des stupéfiants des Nations Unies](#) pour mettre en place The Prevent Alliance, une initiative publique-privée entre Meta, Snap, le gouvernement américain et l'Office des Nations Unies contre la drogue et le crime visant à prévenir l'utilisation abusive des plateformes numériques pour des activités illicites et nuisibles liées à l'usage non médical de drogues de synthèse.

Nous avons pris un [engagement](#) auprès du [Global Refugee Forum](#) en faveur des droits des réfugiés, des demandeurs d'asile et des personnes déplacées à l'intérieur de leur propre pays.

Data for Good chez Meta

Notre programme Data for Good continue de travailler avec des agences humanitaires, des organisations à but non lucratif, des chercheurs et des gouvernements du monde entier pour soutenir les populations marginalisées. Par exemple, ce programme continue de soutenir le travail d'organisations à but non lucratif telles qu'IMPACT Initiatives, qui aident à réinstaller les personnes déplacées par la guerre en Ukraine. Il soutient également les agences internationales, notamment l'Organisation internationale pour les migrations, dans l'amélioration des mesures globales pour soutenir les populations migrantes et les personnes déplacées à l'échelle mondiale.

par des catastrophes naturelles. En outre, Data for Good chez Meta contribue de manière significative au travail des institutions de développement, y compris la Banque mondiale, grâce à des collaborations à grande échelle telles que le Development Data Partnership, qui permet aux grandes institutions multilatérales de mieux cibler les programmes qui servent les communautés dans le besoin. Par exemple, les outils pilotés par l'IA construits par Data for Good, tels que l'index de prospérité relative, ont aidé la Banque mondiale à mieux comprendre comment la pollution de l'air affecte les populations les plus pauvres dans les pays à revenu faible et intermédiaire.





Transparence et recours



Nous nous engageons à ce que la transparence et le recours soient des principes fondamentaux de la gestion des risques liés aux droits humains. Ces deux éléments sont au cœur des [Principes directeurs des Nations Unies sur les entreprises et les droits humains](#). Nos [Rapports de transparence](#), publiés régulièrement, et notre rapport annuel sur les droits humains donnent une visibilité sur la manière dont nous appliquons nos [Standards de la communauté](#) et nos [Règles de la communauté](#), dont nous répondons aux demandes des gouvernements et dont nous protégeons la propriété intellectuelle. Nous nous efforçons également d'améliorer en permanence l'accessibilité à nos Standards de la communauté, qui étaient disponibles en 90 langues à la fin de 2023. Nous pensons que le Conseil de surveillance continue d'offrir un accès aux recours qui est unique dans le secteur des réseaux sociaux.

Rrépondre aux demandes gouvernementales

Risques majeurs pertinents :

- liberté d'opinion et d'expression
- confidentialité
- droit à un procès équitable
- droit à un recours effectif

Nous n'agissons pas sur le contenu et ne divulguons pas les données utilisateur simplement parce qu'une entité gouvernementale le demande. En tant que membres de la [Global Network Initiative](#) (GNI), nous nous engageons à respecter les principes de la GNI sur la liberté d'expression et la protection de la vie privée (principes de la GNI), qui sont fondés sur des normes internationalement reconnues en matière de droits humains. Il s'agit d'un point important, car les principes nous donnent des indications sur la manière de répondre aux demandes des gouvernements et de minimiser les effets négatifs sur la liberté d'expression et la vie privée de nos utilisateurs.

Au fil des ans, les demandes des gouvernements ont évolué. Les acteurs gouvernementaux nous demandent de prendre un large éventail de mesures allant au-delà de la restriction de

10 ans au service de la Global Network Initiative

« Les principes de la GNI sont notre référence en matière de gestion des demandes de retrait des gouvernements. Les demandes de restriction de contenu fondées sur la législation locale soulèvent souvent des dilemmes bien réels et de grande portée en matière de droits humains. Lorsque nous sommes confrontés à des conflits entre les lois locales et nos engagements en matière de droits humains, nous nous appuyons sur la GNI, les normes internationalement reconnues en matière de droits humains et notre politique d'entreprise connexe pour nous guider ».

Siobhán Cummiskey, directrice de la politique de contenu, Meta

La [GNI](#) est une initiative multipartite dont les membres s'engagent à faire respecter les droits à la liberté d'expression et au respect de la vie privée tels qu'ils sont définis dans un traité essentiel en matière de droits humains, le Pacte international relatif aux droits civils et politiques (PIDCP). La GNI établit une norme mondiale pour la prise de décision responsable des entreprises afin de promouvoir et de faire progresser la liberté d'expression et le respect de la vie privée dans l'ensemble de l'écosystème technologique. Dans le contexte des principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains, les principes de la GNI servent de lien entre les obligations de protection des droits humains incombant aux

gouvernements et la responsabilité des entreprises de respecter ces droits.

2023 marque le dixième anniversaire de l'adhésion de Meta à la GNI et de notre engagement en faveur des droits humains. Nous appliquons les principes de la GNI, et nous en sommes responsables, sur la manière dont les entreprises technologiques doivent respecter la liberté d'expression et la vie privée de leurs utilisateurs lorsqu'elles répondent à des demandes, notamment de restrictions, émanant de gouvernements. Notre implémentation fait régulièrement l'objet d'une évaluation indépendante.

l'accès au contenu dans un pays. Depuis 2020, le nombre de demandes gouvernementales a augmenté, de même que le volume de contenus restreints par Meta en vertu de la législation locale.

Les demandes des gouvernements portent à la fois sur des restrictions de contenu et sur l'accès aux données utilisateur. Ces demandes mettent en jeu les droits à la liberté d'expression et à la vie privée, respectivement. Lorsque nous recevons un signalement de restriction de contenu ou une demande d'accès aux données utilisateur de la part d'une entité gouvernementale ou d'un tribunal, nous l'examinons en suivant une procédure globale cohérente pour les deux situations ([ici](#) et [ici](#)).

- Dans le cadre d'enquêtes officielles, les autorités gouvernementales demandent parfois des données sur les personnes qui utilisent nos plateformes. Nous continuons à appliquer les [principes de la GNI](#) dans nos réponses à toutes les formes de demandes émanant des gouvernements. La grande majorité de ces demandes concerne des affaires pénales. La plupart du temps, les demandes gouvernementales portent sur les informations de base des abonnés, comme leur nom, leur date d'inscription et leur ancienneté. Certaines demandes peuvent également concerner l'adresse IP ou le contenu du compte. Nous avons mis en place des [règles](#) très strictes pour gérer les demandes gouvernementales de données : Nous indiquons le nombre de demandes reçues, le nombre d'utilisateurs/comptes demandés et le pourcentage de demandes pour lesquelles nous avons produit des données. Nous [publions](#) ces données ainsi que nos politiques de réponse aux demandes depuis 2016.
- Les autorités gouvernementales demandent parfois la suppression d'un contenu qui enfreindrait la législation locale. Nous pouvons également recevoir une ordonnance d'un tribunal exigeant la suppression d'un contenu. Si nous déterminons que le contenu ne va pas à l'encontre de nos politiques, nous pouvons restreindre l'accès au contenu dans le pays où il est considéré comme illégal après avoir mené une

évaluation juridique et des droits humains minutieuse conformément à notre [Politique d'entreprise en matière de droits humains](#) et à nos engagements en tant que membre de la GNI. Nous informons également l'utilisateur concerné, si la loi ne l'interdit pas. Nous examinons s'il y a lieu de prendre des mesures d'atténuation, telles que des mesures visant à limiter l'impact de la demande, y compris l'exécution sous protestation (qui, dans certains cas, a conduit à la révocation de ces demandes), ou un recours juridique par Meta de l'ordonnance de retrait, la transparence en temps réel via une étude de cas dans notre [Espace modération](#) et/ou la soumission d'une copie de la demande à la [base de données Lumen](#). Si nous estimons que les demandes du gouvernement ou les ordonnances des tribunaux ne sont pas valides d'un point de vue juridique, sont trop larges ou sont incompatibles avec les normes internationales en matière de droits humains, nous pouvons demander des éclaircissements, faire appel ou ne prendre aucune mesure.

- Nous recevons occasionnellement des requêtes ou des ordonnances de la part de gouvernements et de tribunaux nous demandant de prendre de nouvelles mesures fondées sur les exigences de la législation locale, y compris l'application d'un avis de correction sur le contenu (par exemple, en indiquant que le contenu est considéré par les autorités locales comme de fausses informations), la limitation de l'accès d'un utilisateur à des fonctionnalités spécifiques, la restauration d'un contenu précédemment supprimé pour infraction à nos politiques, l'obligation de restreindre automatiquement le contenu à grande échelle et dans le pays, et la restriction du contenu à l'échelle mondiale bien qu'il n'enfreigne que les lois locales d'un seul pays. Bien que nous respections la loi dans les pays où nous opérons, en cas de conflit entre les obligations légales nationales et nos engagements en matière de droits humains et de transparence, nous nous efforcerons d'honorer les principes des droits humains internationalement reconnus dans toute la mesure du possible. Nous pouvons demander des éclaircissements, contester la demande ou ne rien faire.

- Dans certains cas, la publication de demandes de retrait ou de certaines informations contenues dans ces demandes peut être interdite. Dans ce cas, nous nous efforcerons de publier autant d'informations que possible sur la demande et son existence sans enfreindre nos obligations légales.

Comment lire les rapports sur les restrictions de contenu

Nous recevons des signalements de gouvernements, d'autorités de régulation et de tribunaux, ainsi que d'entités non gouvernementales et de membres du public, concernant des contenus supposés enfreindre les lois locales. Le [Rapport sur les restrictions de contenu](#), publié deux fois par an depuis 2013, détaille les cas où nous avons restreint l'accès au contenu sur Facebook et Instagram conformément à la législation locale.

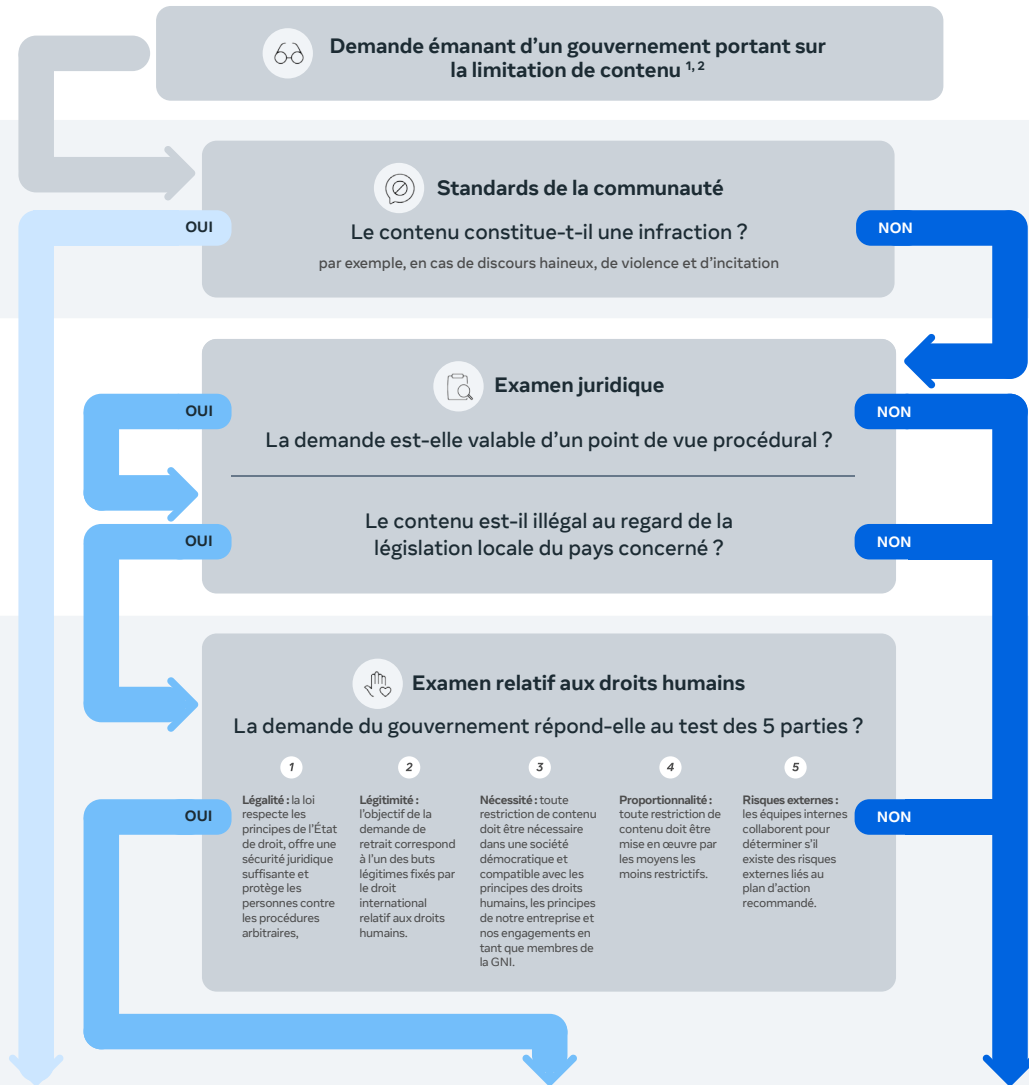
- La section sur les [données spécifiques au pays](#) fournit des détails sur le volume, la nature et le type de contenu que nous restreignons tous les six mois conformément aux lois locales. Depuis 2023, nos rapports comprennent également des informations lorsque, dans certains pays, nous sommes obligés, conformément à la législation locale, de restreindre automatiquement le contenu, à grande échelle et dans le pays, ou d'appliquer un avis de correction sur le contenu dans le pays.
- La section sur les [restrictions internationales](#) détaille les cas occasionnels où nous avons été légalement contraints de restreindre la disponibilité du contenu

au niveau mondial en raison d'exigences juridiques extraterritoriales, bien que le contenu en question n'aille à l'encontre que de la loi locale d'un pays donné.

- La section sur les [études de cas](#) fournit des détails sur certaines des demandes de retrait de contenu que nous avons reçues de la part de gouvernements et de tribunaux. À partir de 2023, nous nous attacherons à ne publier que des études de cas présentant un grand intérêt pour le public.
- La section sur [la manière dont nous évaluons les rapports de contenu enfreignant la législation locale](#) contient des informations sur notre processus d'examen des demandes de retrait des gouvernements et sur les mesures d'atténuation que nous prenons pour limiter l'impact des restrictions conformément aux normes internationales en matière de droits humains.

Nous avons publié des données sur les restrictions de contenu basées sur la législation locale pour le second semestre 2023 en mai 2024 (voir [ici](#)).

Quelles sont les étapes que traverse une demande émise par un gouvernement ?



Résultats

Nous pouvons déployer des mesures d'atténuation telles que des appels ou le partage d'informations sur la demande du gouvernement de restreindre le contenu à l'extérieur afin de limiter l'impact de la demande.



Actions dans le cadre de nos politiques

Lorsque nous identifions un contenu qui va à l'encontre des Standards de la communauté Facebook ou des Règles de la communauté Instagram, les mesures appropriées seront prises (par exemple, le contenu sera supprimé ou soumis à un contrôle de l'âge, etc.)



Restreindre l'accès dans les juridictions où il est illégal

Nous informons les personnes lorsque nous limitons un contenu qu'elles ont publié, en nous fondant sur un signalement d'une infraction du contenu aux lois locales. Nous les avertissons également lorsqu'elles essaient de consulter un contenu restreint dans leur pays suite à une demande gouvernementale de restriction du contenu. Nous informons également l'autorité gouvernementale compétente que des mesures ont été prises en réponse à leur demande.



Pas d'action OU demande d'informations complémentaires

Nous pouvons prendre contact avec l'autorité pour obtenir des informations complémentaires sur la demande.

¹ Cette illustration reflète notre processus d'examen et de réponse aux demandes des gouvernements de restreindre le contenu sur Facebook et Instagram, tel qu'il est décrit dans notre Espace modération. Pour une description complète, veuillez consulter l'Espace modération (lien ici). Certaines différences peuvent s'appliquer aux publicités et aux autres produits Meta.

² Nous pouvons être amenés à nous écarter du cycle de vie décrit ci-dessus dans des situations d'urgence. Bien que cela arrive très peu, la législation d'un pays nous oblige parfois à restreindre automatiquement l'accès au contenu, à l'échelle de certains pays, et ce, sur la base des exigences de la législation locale. Dans de tels cas, nous continuerons à nous appuyer sur nos engagements dans le cadre de la Global Network Initiative et sur notre politique d'entreprise en matière de droits humains pour guider notre approche.

Mise à jour de notre participation au projet Lumen

En mars 2022, nous nous sommes engagés à [participer à Lumen](#), un projet de recherche indépendant hébergé par le Berkman Klein Center for Internet and Society de l'université de Harvard. Le projet permet aux chercheurs d'étudier les demandes de retrait de contenu émanant de gouvernements et d'acteurs privés concernant le contenu en ligne. En novembre 2023, nous avons soumis à [Lumen](#) notre première série de demandes de retrait de l'Autriche, du Vietnam, de Singapour, de l'Inde et du Mexique, qui sont accessibles sur leur base de données. Cela permettra à la communauté mondiale d'analyser, de rédiger des rapports et de défendre les droits numériques des utilisateurs de l'Internet. Les principes qui guident notre approche en matière de partage public d'informations sur les demandes de retrait des gouvernements figurent dans notre [Rapport sur les restrictions de contenu](#).

Amélioration de l'avertissement aux utilisateurs

Afin de renforcer notre transparence et conformément aux normes relatives au droit à un recours effectif et au droit à un procès équitable, nous avons amélioré notre avertissement aux utilisateurs en 2023 sur Facebook et Instagram pour les contenus restreints par Meta, sur la base de demandes juridiques émanant de gouvernements et de tribunaux. Dans la plupart des cas, nos notifications informent les utilisateurs de l'autorité nationale qui a envoyé la demande à l'origine de la restriction et du pays dans lequel le contenu est restreint. Elles ne s'appliquent pas au contenu que nous avons supprimé pour manquement à nos politiques.



Risques majeurs pertinents :

- droit à un recours effectif
- droit à un procès équitable
- liberté d'opinion et d'expression
- vie, liberté et sécurité de la personne
- participation publique, processus de vote et éligibilité
- égalité et non-discrimination
- intérêt supérieur de l'enfant
- confidentialité

Conseil de surveillance

Le Conseil de surveillance est une initiative de premier plan créée pour aider Meta à répondre à certaines des questions les plus difficiles concernant la liberté d'expression et la sécurité en ligne, ce qu'il faut supprimer, ce qu'il faut laisser en ligne et pourquoi. Il s'agit d'un organisme indépendant qui peut examiner des cas émanant soit de Meta directement, soit de personnes sur Facebook, Instagram ou Threads qui ne sont pas d'accord avec nos décisions de modération du contenu. Outre les décisions contraignantes sur le maintien ou le retrait d'un contenu, le Conseil émet également des recommandations visant à améliorer la modération des contenus, ainsi que des avis consultatifs sur les politiques, émis à notre demande. Le Conseil de surveillance applique les normes mondiales en matière de droits humains afin de fournir à Meta des avis et des recommandations de spécialistes compétents pour garantir que les décisions concernant nos politiques et nos produits sont dans le meilleur intérêt des utilisateurs.

L'année 2023 a été une année charnière pour le Conseil de surveillance, qui a dépassé son objectif de 50 décisions par an, triplant ainsi le nombre de décisions prises en 2022.

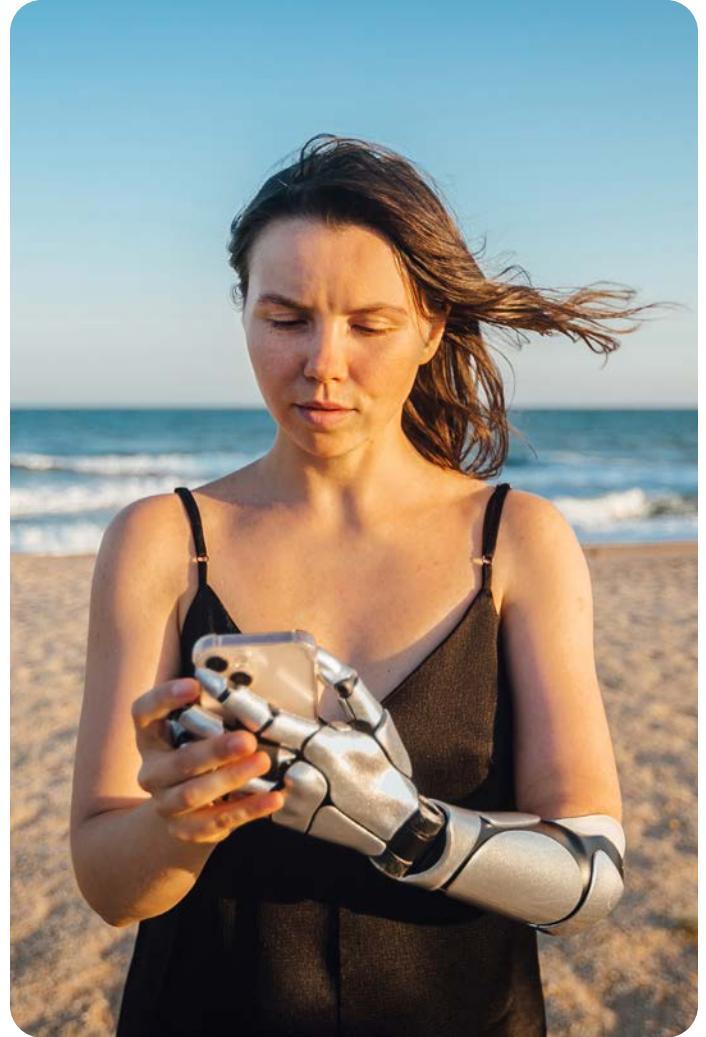
Le Conseil de surveillance a apporté des modifications importantes à son règlement afin de permettre des examens accélérés dans des circonstances exceptionnelles, notamment lorsque le contenu pourrait avoir des conséquences graves dans le monde réel. Cette procédure permet de prendre des décisions accélérées en matière de contenu dans un délai de 30 jours. Le Conseil a eu recours pour la première fois à la procédure d'examen accéléré pour deux cas liés au conflit Israël-Hamas.

Les décisions du Conseil ont porté sur de nombreux sujets différents dans un large éventail de pays, notamment le slogan de protestation iranien « Mort à Khamenei », le cas du premier ministre cambodgien Hun Sen, le genre et la nudité, les prisonniers de guerre arméniens, les élections brésiliennes et les manifestations de femmes à Cuba, parmi d'autres. Outre ses décisions, le Conseil a formulé 60 recommandations à l'intention de Meta en 2023. Il a également publié un avis consultatif sur la suppression des fausses informations sur la COVID-19. Pour plus de détails, voir les rapports de transparence du Conseil de surveillance, [ici](#), [ici](#) et [ici](#).

Marge de tolérance d'intérêt médiatique

En réponse aux recommandations du Conseil de surveillance, nous avons apporté davantage de transparence sur la manière et le moment où nous appliquons notre marge de tolérance d'intérêt médiatique pour les contenus qui pourraient enfreindre nos Standards de la communauté Facebook ou nos Règles de la communauté Instagram, mais qu'il est dans l'intérêt public de garder visibles.

- Du 1er juin 2022 au 1er juin 2023, nous avons documenté 69 marges de tolérance d'intérêt médiatique.
- Neuf (environ 13 %) de ces marges de tolérance documentées ont été accordées pour des publications faites par des personnalités politiques.
- Sur les 69 marges de tolérance, nous avons documenté un total de 17 marges de tolérance à grande échelle, ce qui signifie que chaque marge de tolérance s'applique à plus d'un seul contenu.



Notre programme de vérification croisée

L'année 2023 a également marqué un tournant important dans notre programme de vérification croisée, puisqu'elle a débuté par notre réponse initiale aux 33 recommandations du Conseil de surveillance portant sur la sophistication technique, la rigueur opérationnelle, l'équité, la gouvernance et la transparence du programme. Nous avons pris des mesures pour accroître la participation des partenaires de confiance et de la société civile au développement de la gouvernance en matière de vérifications croisées en organisant des sessions de commentaires avec des partenaires de neuf régions différentes.



En 2023, nous avons achevé notre travail sur un total de 61 recommandations sur 122 et poursuivons nos efforts sur de nombreuses autres (pour plus de détails, consultez les [mises à jour semestrielles de notre Conseil de surveillance](#)). Les recommandations que nous avons formulées couvrent nos activités, nos politiques et nos produits, contribuant ainsi à des améliorations importantes et significatives dans l'ensemble de l'entreprise et de la communauté mondiale.

Recommandations du Conseil de surveillance 2023

Recommandations
du Conseil de
surveillance

66

(91 en 2022)

Évaluation et/ou mise
en œuvre par Meta
en cours*

69

(75 en 2022)

Recommandations
totalement mises
en œuvre*

61

(14 en 2022)

* Certaines évaluations et/ou implémentations en cours ou recommandations entièrement suivies incluent des recommandations des années précédentes (voir notre [Rapport de 2022 sur les droits humains](#) pour plus de détails).

Qu'en est-il de l'avenir ?

En 2023, nous avons assisté à de nombreux développements rapides, tels que décrits dans ce rapport, et à de nombreuses opportunités d'inclure les droits humains dans la prise de décision. Nous avons cherché à le faire. Cependant, les contextes sociopolitiques plus larges sont instables et de nombreuses sources de données indiquent une érosion de la liberté d'expression et des normes démocratiques. Les technologies d'intelligence artificielle (IA) existent depuis des décennies, mais la sensibilisation du public aux opportunités et aux risques qu'elles peuvent créer a explosé en 2023.

L'évolution rapide du paysage réglementaire est à la fois prometteuse et difficile pour les droits humains : elle peut conduire à des innovations significatives dans la manière dont les entreprises appliquent la gestion des risques liés aux droits humains, en accord avec les principes fondamentaux des droits humains, notamment les principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains, et dans l'étendue de cette gestion. Mais les exigences en matière de conformité pourraient aussi étouffer la volonté et l'espace d'innovation. Il est de plus en plus évident qu'une gestion durable des risques liés aux droits humains à grande échelle implique l'utilisation permanente de données et d'outils conformes aux principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains, qui sont intégrés aux processus plus larges de l'entreprise.

« Il existe de nombreuses possibilités d'intégrer les droits humains dans les décisions quotidiennes des entreprises et des gouvernements. »

John Ruggie, auteur des principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains

Meta cherche à innover dans ses méthodes de gestion des risques liés aux droits humains, y compris dans son engagement avec les parties prenantes. L'apprentissage empirique, les exercices de simulation et les études de cas semblent promettre de bonnes pratiques, mais ne se sont pas encore généralisés. C'est pourquoi nous nous efforçons de renforcer notre travail dans ces domaines. Nous reconnaissons également l'importance des efforts concertés et de l'innovation pour faire face aux risques potentiels pour les droits humains liés à l'IA générative, comme nous l'avons évoqué dans ce rapport.

Nous savons qu'il nous reste encore beaucoup de chemin à parcourir. Nous voulons nous appuyer sur notre travail et les enseignements tirés pour contribuer à l'évolution des connaissances et aux bonnes pratiques pour le secteur technologique et les communautés des entreprises et des droits humains.

Nous nous trouvons à un moment difficile mais prometteur pour les droits humains dans le monde de la technologie. Nous tous,



y compris les entreprises, la société civile, les gouvernements, le système des Nations Unies, les investisseurs et les régulateurs, devons travailler pour développer de nouvelles compétences et connaissances, et de nouveaux modes de collaboration, pour façonner, partager et apprendre ce à quoi la gestion des risques liés aux droits humains peut ressembler à grande échelle.

Des vagues futures d'immense créativité se préparent : nous aurons besoin d'un grand nombre de personnes courageuses, d'expériences et d'interventions entrepreneuriales pour les saisir. Heureusement, le mouvement des droits humains n'a jamais manqué de courage pour innover. À présent, il est temps d'œuvrer.

Annexe

Comment les droits humains sont régis et gérés chez Meta

Des structures de gestion et de gouvernance claires nous permettent de faire progresser notre mission de respect des droits humains dans nos programmes, nos services et nos politiques. Nos spécialistes en droits humains guident la mise en œuvre de notre [politique d'entreprise en matière de droits humains](#), qui est supervisée par le président des affaires mondiales et la responsable des affaires juridiques.

Les tâches des spécialistes en droits humains consistent notamment à promouvoir l'intégration de la politique dans les politiques, programmes et services existants et en cours d'élaboration, à faire preuve de diligence raisonnable et à soutenir la formation du personnel sur la politique. La politique donne des orientations pour créer des produits respectueux des droits, répondre aux crises émergentes et travailler avec rapidité et agilité pour intégrer les droits humains à grande échelle.

Notre politique d'entreprise en matière de droits humains nous engage à régulièrement présenter des rapports au Conseil d'administration sur les principales questions relatives aux droits humains. Le Comité d'audit et de surveillance des risques du Conseil d'administration supervise les différents risques auxquels l'entreprise est exposée, y compris les risques liés aux droits humains, ainsi que les mesures prises par la direction pour contrôler ou atténuer ces risques. Le comité est régulièrement informé des travaux existants et en cours menés par les spécialistes en droits humains.

Formation du personnel de Meta aux droits humains

Chez Meta, la manière dont nous agissons est aussi importante que ce que nous faisons. Nos formations en matière de droits humains permettent à nos employés de mieux comprendre leurs responsabilités et les connaissances et compétences nécessaires pour les remplir.

Notre *Au-delà de Meta : La formation aux droits humains* a été lancée en 2022 et s'est poursuivie tout au long de l'année 2023. La formation met en évidence les impacts potentiels et réels de nos services, politiques et décisions commerciales sur les droits humains. Elle vise à promouvoir un état d'esprit axé sur les droits humains dans notre travail quotidien, en encourageant le respect des droits humains dans l'intérêt de tous ceux qui utilisent nos services. Cette formation complète notre formation sur les [droits civils](#), qui est axée sur les principes de non-discrimination, de justice et d'équité.

Nos objectifs de formation en matière de droits humains sont également soutenus par notre formation annuelle obligatoire en matière de respect de la vie privée. Il s'agit de développer notre capacité collective à protéger les individus, y compris, en particulier, les catégories d'individus marginalisés, contre les préjudices découlant du traitement de leurs données. Cette formation soutient la protection des droits des personnes à la vie privée et à la protection des données.



Liens vers les rapports référencés

- [Rapport sur la durabilité 2024](#)
- [Rapport 2024 sur les pratiques responsables des entreprises](#)
- [Rapport 2022 sur les droits humains, Rapport 2021 sur les droits humains](#)
- [Rapport 2023 sur la lutte contre l'esclavage et la traite d'êtres humains](#)
- [Rapport 2024 sur la lutte contre l'esclavage et la traite d'êtres humains](#)
- [Rapport 2023 sur les minerais de conflit](#)
- [Mise à jour 2023 sur les progrès en matière de droits civils](#)
- [Rapports de transparence de Meta](#)
- [Rapports réglementaires et autres rapports de transparence](#)
- Évaluations de l'impact sur les droits humains : [Chiffrement de bout en bout](#), [Philippines](#), [Myanmar](#), [Indonésie](#), [Cambodge](#), [Inde](#), [Sri Lanka](#), et [Israël et Palestine](#)

