Rapport sur les droits humains

Meta

Statistiques et mesures 2024

Sommaire

À propos de ce rapport

Ce rapport annuel sur les droits humains couvre les statistiques et les mesures, du 1er janvier 2024 au 31 décembre 2024. Nous rendons compte des services et produits Meta, notamment Facebook, Messenger, Instagram, WhatsApp, Threads et Reality Labs.

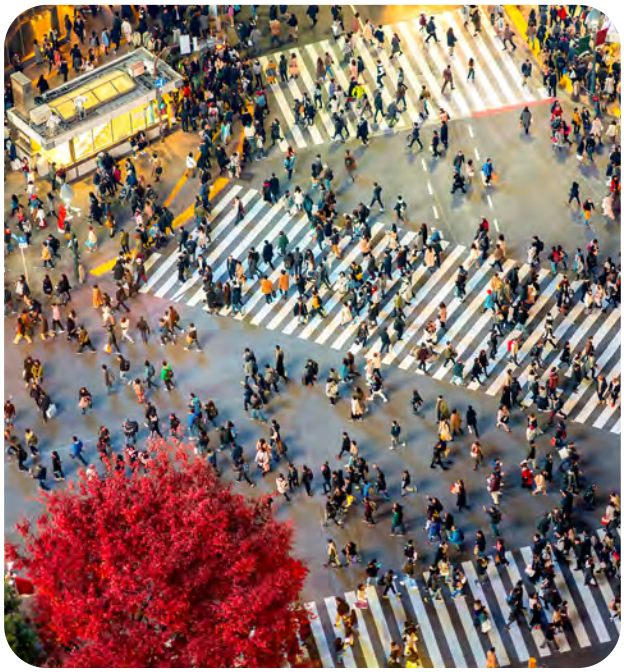
Il s'appuie sur le travail de Meta en vue de respecter les droits humains et reflète les progrès réalisés dans le cadre de nos engagements vis-à-vis des Principes directeurs des Nations unies relatifs aux entreprises et aux droits humains et de notre Politique d'entreprise en matière de droits humains. Le rapport montre comment nous avons appliqué ces principes dans l'entreprise en 2024 et fournit des conseils permettant de trouver des informations approfondies.

Le contenu de ce rapport se base sur notre évaluation globale des risques principaux en matière de droits humains, réalisée en 2022. L'objectif de cette évaluation était d'identifier et de hiérarchiser nos impacts négatifs potentiels les plus significatifs en matière de droits humains[1] sur les personnes qui utilisent nos produits et sur d'autres personnes susceptibles d'être concernées par nos mesures. Ce rapport décrit ces principaux risques potentiels ainsi que les exemples de nos mesures et de nos atténuations en 2024.[2]

Les droits humains sont toujours un sujet d'une importance cruciale pour notre entreprise et pour nos parties prenantes en 2024. Nous nous efforçons de fournir une représentation fidèle de notre travail au sein de plusieurs équipes et de l'engagement des parties prenantes à travers le monde.

Politiques et progrès

Outre ce rapport sur les droits humains, Meta rend compte chaque année de ses politiques et de ses progrès grâce aux mécanismes suivants :



- Rapport annuel
- Déclaration de procuration
- Rapport sur les pratiques professionnelles responsables
- Espace modération
- Rapport sur la durabilité
- Rapport CDP sur le changement climatique
- Pacte mondial des Nations unies

Ce rapport complète le Rapport sur les pratiques professionnelles responsables Meta le plus récent. Nous rapportons séparément nos efforts pour identifier et atténuer les risques d'esclavage moderne et de traite d'êtres humains dans nos activités commerciales et nos chaînes d'approvisionnement. De plus, nous nous conformons aux rapports obligatoires au niveau national et de l'Union européenne, disponibles dans notre Espace modération. Des liens vers d'autres publications de Meta se trouvent dans l'Annexe du présent rapport.

Accédez à l'annexe

[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 mesure supprime ou réduit la capacité d'une personne à jouir de ses droits humains.

[2] Elle ne comprend pas la politique de contenu ni d'autres modifications que nous avons annoncées en janvier 2025, lorsque nous avons mis à jour notre politique relative aux comportements haineux, anciennement connue sous le nom de politique relative au discours haineux, afin de répondre aux préoccupations concernant les excès de mise en application et de permettre une plus grande liberté d'expression.

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 qui peuvent avoir des répercussions différentes sur les droits humains. Le présent rapport fait référence à des mesures 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 et/ou en toutes circonstances.[3]

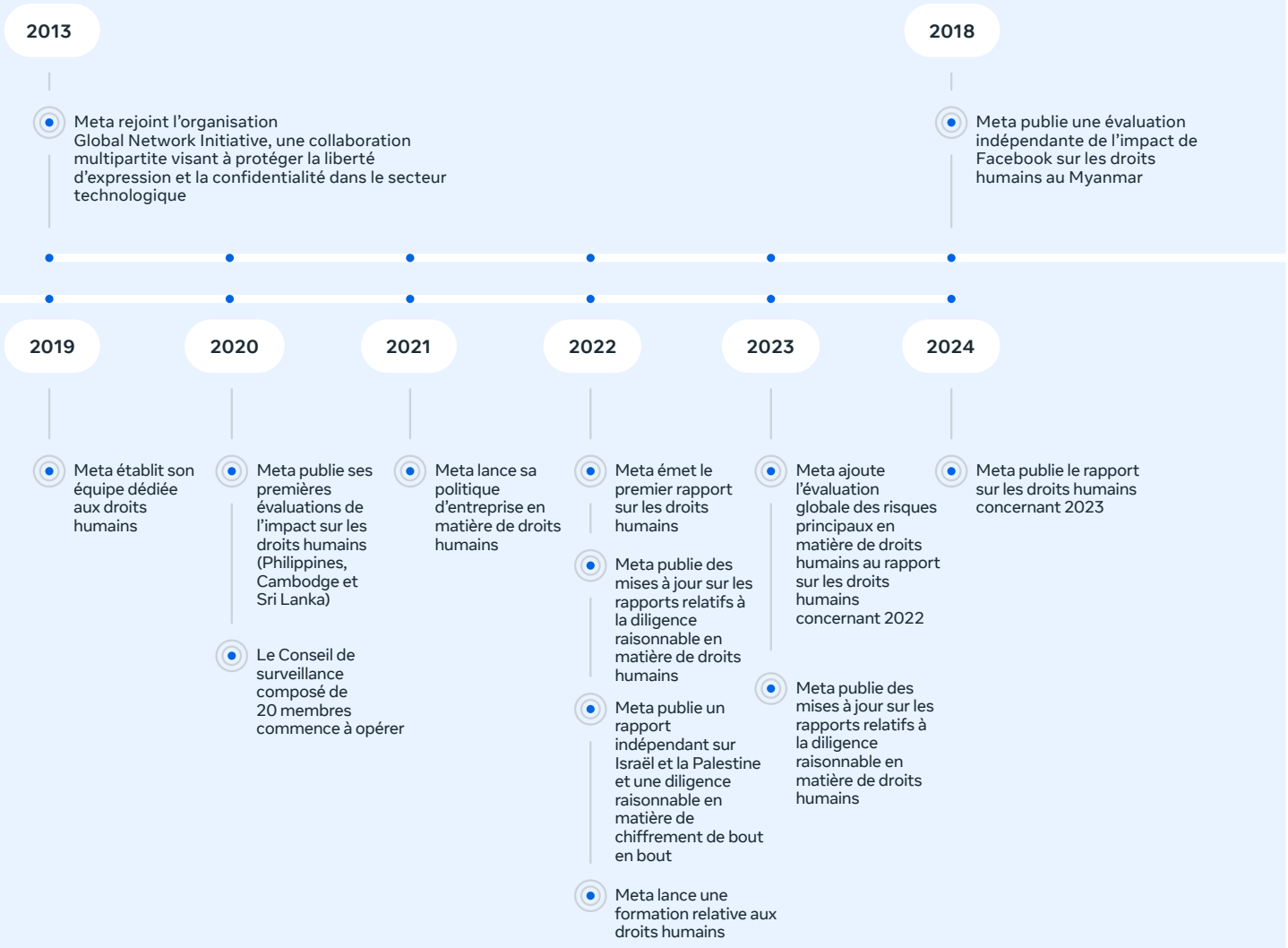
[3] L'analyse de la modération des contenus et des actions connexes sur Facebook et Instagram présentée dans ce rapport ne s'applique pas à WhatsApp et, à moins qu'une politique ou une mesure ne soit spécifiée comme s'appliquant à WhatsApp, elle ne s'applique pas à WhatsApp. En outre, bien que de nombreuses mesures 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 analytique

Il s'agit du quatrième rapport annuel sur les droits humains de Meta. Il fournit des statistiques sur le travail réalisé par Meta en 2024 pour gérer les risques en matière de droits humains à grande échelle et respecter nos engagements envers les Principes directeurs des Nations unies relatifs aux entreprises et aux droits humains (UNGP).

Calendrier relatif aux droits humains

Le tableau suivant décrit notre parcours relatif aux droits humains et la manière dont notre travail a évolué depuis l'adoption des UNGP par le Conseil des droits humains des Nations unies en 2011.



Évaluation des risques principaux

Lire la suite

En 2024, nos priorités reflétaient les risques principaux identifiés dans l'évaluation globale des risques principaux en matière de droits humains de 2022 : 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.

Accélérer l'innovation en matière d'IA

Accélération des progrès en matière d'intelligence artificielle (IA) en 2024. Notre idée est de créer une superintelligence personnelle et de la rendre accessible à grande échelle afin que tout le monde puisse en bénéficier.

Les applications d'IA générative sont de plus en plus utilisées et elles ont considérablement transformé notre manière de communiquer, d'apprendre, de créer et de travailler. Nous avons continué de promouvoir une approche ouverte de l'IA capable de renforcer les droits humains. La présente approche a contribué à faciliter l'accès à l'information et la liberté d'expression, ainsi qu'à faire progresser les droits à l'égalité et à la non-discrimination, notamment en améliorant l'accessibilité et en élargissant l'inclusivité linguistique.

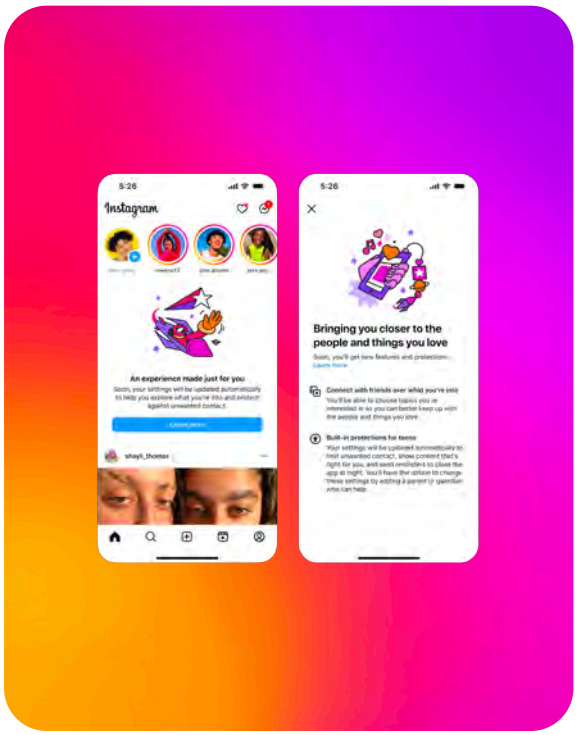
Lire la suite

2024 : l'année des élections

2024 a été la plus grande année de l'histoire sur le plan électoral. Plus de 70 pays, comptant plus de la moitié de la population mondiale, ont organisé des élections nationales. Près de 2 milliards de personnes ont pu voter. Nous nous sommes attachés à garantir les droits à la liberté d'expression, à la participation aux processus politiques et à l'accès à l'information des personnes vivant dans les pays où des élections avaient lieu.

Notre approche s'est récemment affinée au fil de centaines d'élections. Des efforts ont été nécessaires pour gérer les risques liés à l'IA, mettre en application nos politiques en matière d'ingérence électorale ou dans le recensement, perturber les réseaux conflictuels, augmenter la transparence de la publicité politique et fournir aux électeurs des informations fiables. Dans ce rapport, vous trouverez les exemples des États-Unis, du Mexique, de l'Inde et de l'Union européenne.

Lire la suite



Sécurité des enfants et des jeunes

Nous avons poursuivi notre engagement en faveur de la sécurité des enfants et des jeunes. Parmi ces initiatives, notre travail de 2024 comprenait le lancement des comptes Ado d'Instagram, une nouvelle expérience pour les ados, avec les parents aux commandes. Les comptes Ado sont dotés de protections intégrées qui limitent les personnes pouvant contacter les ados et le contenu que ces derniers peuvent voir, ainsi que de moyens pour gérer le temps que les ados passent sur l'application, tout en leur fournissant de nouvelles méthodes pour explorer leurs centres d'intérêt. Nous avons dû trouver un équilibre entre la promotion de l'autonomie des jeunes et les droits et devoirs des parents et représentants légaux. Nous les avons développés conformément aux conseils de spécialistes et au principe de l'évolution des capacités de l'enfant décrit dans la Convention des Nations unies relative aux droits de l'enfant.

Lire la suite

Services de crise

Nous avons continué d'intégrer les principes relatifs aux droits humains dans notre manière de nous préparer aux crises et d'y répondre. Notre Protocole de politique de crise sert de guide dans l'utilisation rapide des leviers à notre disposition pour atténuer les préjudices potentiels. En 2024, nous avons mentionné 19 situations à travers le monde relevant de ce protocole. Dans le présent rapport, nous fournissons des exemples de notre service de crise au Bangladesh, en Géorgie, au Moyen-Orient et au Soudan.

Lire la suite

Implication des parties prenantes

Notre Politique d'entreprise en matière de droits humains soutient notre engagement proactif avec les parties prenantes. En 2024, nous avons contacté un grand nombre de parties prenantes pour définir l'approche de l'entreprise sur les questions liées à la liberté d'expression, au contenu haineux, aux fausses informations et à la confidentialité. Les parties prenantes concernées comprenaient un large éventail de groupes dédiés aux droits humains, de communautés vulnérables, de membres de la société civile, d'universitaires, de groupes de réflexion et d'organismes de réglementation. Parmi les sujets clés, notre approche en matière d'IA responsable et d'intégrité des élections, mais aussi nos signaux de désignation pour les organisations et personnes dangereuses, et les évènements violents.

En 2024, nous avons organisé six Forums politiques, pendant lesquels des spécialistes de Meta ont partagé différents points de vue et discuté des modifications potentielles des Standards de la communauté et des Standards publicitaires. Nous avons aussi accueilli des Forums communautaires pour exploiter des données publiques sur des questions où il existait des compromis contradictoires et aucune réponse claire. Ils nous ont aidé à améliorer les produits et à anticiper les risques potentiels liés aux technologies émergentes, et ont permis à des personnes extérieures à l'entreprise d'avoir davantage leur mot à dire dans notre prise de décisions.

Nous avons continué d'impliquer nos partenaires de confiance dans le monde entier pour aider à identifier des tendances, et mieux comprendre l'impact des contenus en ligne et des comportements sur les communautés locales. Nous avons également découvert comment renforcer les canaux pertinents en ce qui concerne la remontée des problèmes. Leur expertise a été particulièrement précieuse lors des nombreuses élections qui ont eu lieu en 2024, ainsi que dans des situations de troubles accrus. Ils ont également fourni des statistiques et identifié des contenus potentiellement en infraction dans les pays et régions suivants : Bangladesh, Brésil, Côte d'Ivoire, République démocratique du Congo, France, Grèce, Inde, Indonésie, Kenya, Kurdistan irakien, Mexique, Nigeria, Pakistan, Sénégal, Afrique du Sud, Syrie et Venezuela.

Lire la suite

Conseil de surveillance

En 2024, le Conseil de surveillance a pris en considération des cas mentionnant nos efforts pour respecter les droits humains, notamment la liberté d'expression, le droit à la santé, ainsi que le droit à l'égalité et à la non-discrimination, pour ne citer que ces sujets. Le Conseil de surveillance est un organisme indépendant qui examine des cas transmis par Meta ou qui font l'objet d'un appel par des particuliers sur Facebook, Instagram ou Threads, car ils sont en désaccord avec nos décisions en matière de modération de contenu. Il rend des décisions contraignantes indiquant si le contenu doit être retiré ou conservé. En réponse à une recommandation émise par le Conseil de surveillance, Meta a évalué la rapidité et l'efficacité des réponses vis-à-vis du contenu signalé via le programme des partenaires de confiance.

Lire la suite

Gestion des demandes gouvernementales

Tout au long de l'année, nous avons continué à nous laisser guider par notre engagement en faveur de la Global Network Initiative pour respecter la liberté d'expression et la confidentialité, y compris en répondant aux demandes gouvernementales visant à restreindre le contenu. En 2024, nous avons publié des études de cas liées à des discours politiques au Brésil, en Allemagne, en Inde, en Irak, en Israël, à Singapour et en Turquie.

Voir des études de cas

Gestion des risques en matière de droits humains

Les Principes directeurs des Nations unies relatifs aux entreprises et aux droits humains mettent en lumière le fait que les entreprises doivent identifier leurs impacts négatifs sur les droits humains afin de les prévenir ou de les atténuer avec efficacité.

Étant donné l'ampleur des opérations de Meta et l'éventail des droits qu'elles peuvent impliquer, il est important d'anticiper et de gérer nos risques principaux, mais cela s'avère complexe. Nous gérons deux types de risques inhérents : ceux provenant de nos propres activités et ceux provenant des activités de tiers, notamment les personnes qui utilisent nos plateformes.

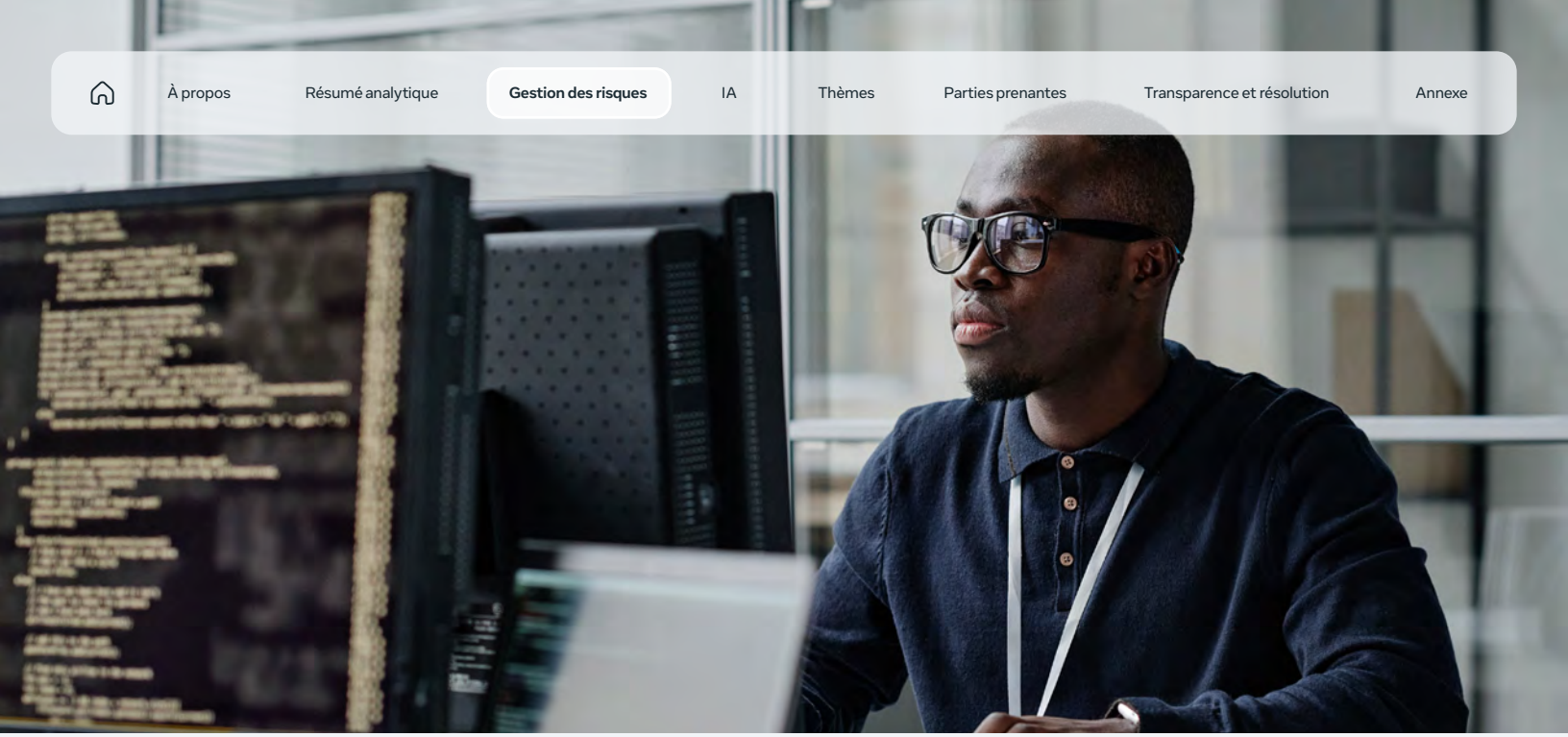
Une fois les processus implémentés pour faire face à ces risques, il restera toujours un certain degré de risque. On appelle cela le risque « résiduel ». Même si les risques résiduels existent dans tous les systèmes de gestion des risques, ceux associés aux technologies digitales et leur impact sur les droits humains persistent en raison de l'évolution rapide et dynamique de ces technologies et du degré élevé d'activités de tiers.

Le tableau qui figure dans les pages suivantes présente nos risques principaux en matière de droits humains, tels qu'ils sont définis dans notre évaluation globale des risques principaux en matière de droits humains de 2022 (EGRP), que nous avons publiée dans notre Rapport sur les droits humains 2022. Ce tableau fournit des exemples illustrant la manière dont nous avons fait face aux potentiels risques en 2024. Pus loin dans ce rapport, nous étudierons plus en détail certains des présents exemples ainsi que la manière dont nous avons géré les risques potentiels en lien avec l'intelligence artificielle (IA), les élections et les conflits.

1. Liberté d'opinion et d'expression

Le droit à la liberté d'opinion et d'expression comprend le droit à rechercher, recevoir et partager des informations et des idées en tout genre. Il s'agit d'un droit fondamental, essentiel à la protection de la dignité humaine, de l'autonomie individuelle et de la démocratie. La liberté d'expression est partie intégrante de notre mission, en accord avec notre valeur qui consiste à donner la parole à tout le monde.

Exemples de principaux risques potentiels inhérents aux droits humains identifiés lors de l'EGRP	Exemples de mesures prises par Meta pour remédier aux risques potentiels en 2024
La mise en application et les politiques relatives à la modération du contenu de Meta peuvent limiter la liberté d'expression.	Nous avons continué à élaborer nos politiques avec la liberté d'expression en ligne de mire. En 2024, nous avons organisé plusieurs Forums politiques qui ont cherché à développer une appréciation nuancée des défis en matière de liberté d'expression dans un grand nombre de domaines.
Limites gouvernementales abusives en matière de contenu	Nous nous efforçons de respecter nos engagements vis-à-vis de la Global Network Initiative (GNI). Cela comprenait les rapports sur nos réponses aux demandes gouvernementales en matière de restrictions relatives aux données ou au contenu (ici et ici). Notre approche pour répondre aux demandes gouvernementales est détaillée dans notre Rapport sur les droits humains 2023. 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. En 2024, parmi les études de cas notables en matière de transparence impliquant des discours politiques, on peut citer celles du Brésil, de l'Allemagne, de l'Inde, de l'Irak, d'Israël, de Singapour et de la Turquie.
Des perturbations sur Internet et des blocages des réseaux sociaux empêchent les citoyens d'exercer leur droit à la liberté d'expression et les privent de la possibilité de recevoir et d'envoyer des nouvelles et des informations vitales.	Afin d'empêcher les blocages des réseaux sociaux et des messages, nous pouvons nous conformer aux demandes légitimes du gouvernement tout en nous efforçant de respecter nos engagements vis-à-vis de la GNI en matière de liberté d'expression. Nous continuons également de fournir la fonctionnalité WhatsApp par proxy pour les personnes qui ne peuvent pas se connecter directement à nos applications.



2. Confidentialité

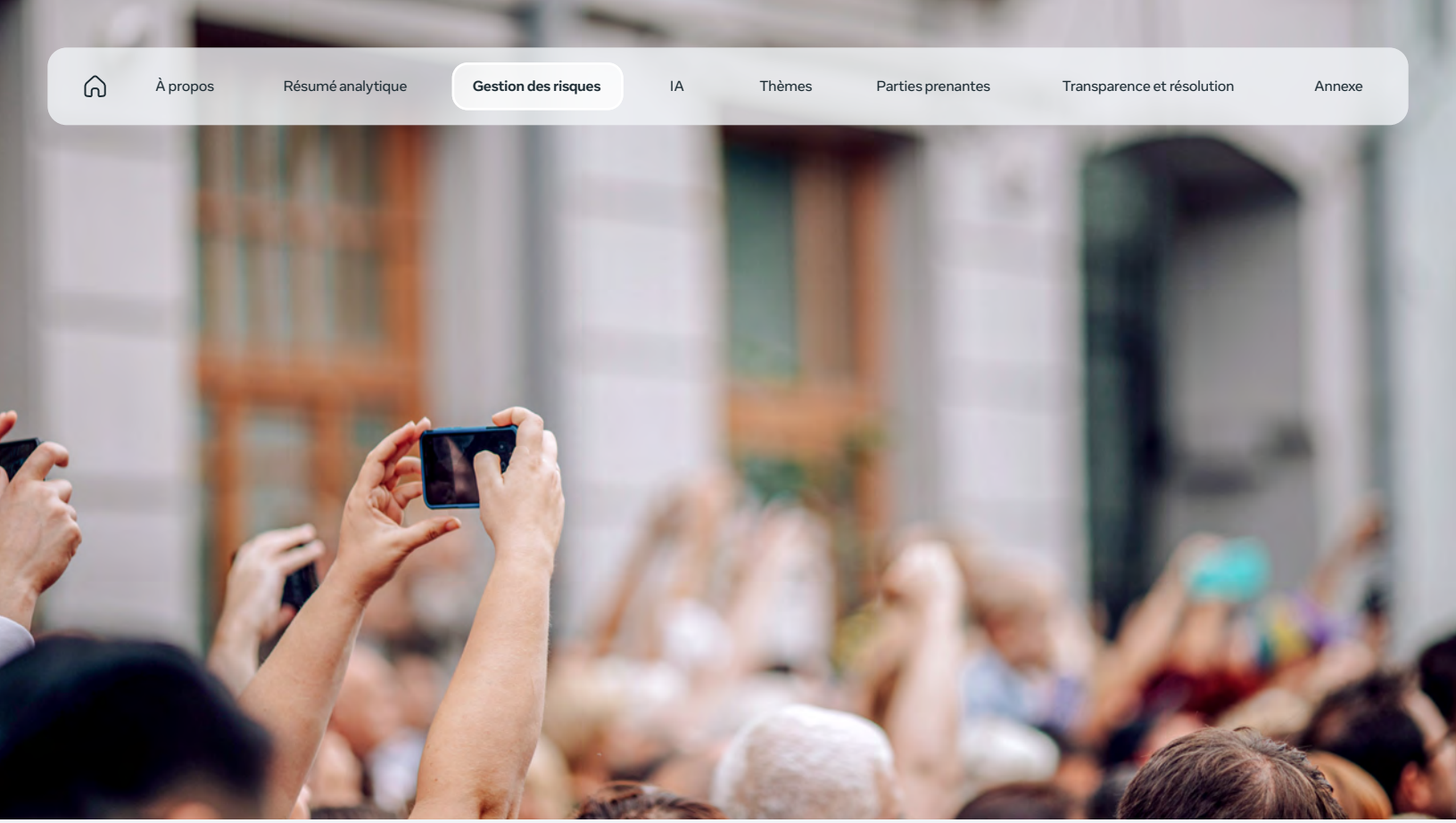
Le droit à la confidentialité constitue une condition nécessaire au respect d'autres droits humains, tels que la liberté d'expression, la liberté de réunion et d'association, ainsi que la liberté de croyance et de religion. Un des principes de base décrits dans notre Politique d'entreprise en matière de droits humains consiste à garantir la sécurité des utilisateurs et à protéger leur confidentialité.

Exemples de principaux risques potentiels inhérents aux droits humains identifiés lors de l'EGRP	Exemples de mesures prises par Meta pour remédier aux risques potentiels en 2024
Les modèles d'IA générative peuvent impliquer le traitement de données personnelles d'une manière que les personnes ne prévoient pas ou ne comprennent pas.	Nous faisons preuve de transparence concernant l'utilisation des informations par Meta pour les fonctionnalités et les modèles d'IA générative, et nous disposons d'un processus d'examen de la confidentialité en interne pour une utilisation des données responsable, y compris l'IA générative. Une mise à jour sur nos progrès réalisés en 2024 en matière de confidentialité sont disponibles ici et ici. Lire la suite
Le contenu ou le comportement sur les applications Meta peut avoir un effet négatif sur la confidentialité et les droits en matière de protection des données.	En octobre 2024, Meta a réintroduit la technologie de reconnaissance faciale sur Facebook et Instagram pour aider les utilisateurs à récupérer les comptes compromis et empêcher les arnaques impliquant de faux soutiens de la part de célébrités. Afin de trouver un équilibre entre les risques potentiels pour la confidentialité et l'intégrité, nous offrons aux personnalités publiques, dont l'image est utilisée à des fins frauduleuses pour arnaquer les utilisateurs, la possibilité de participer au programme ou de s'en retirer.

3. Égalité et non-discrimination

Le droit à l'égalité et à la non-discrimination prévoit une protection égale contre la discrimination. Dans le cadre du respect de ce droit, nous n'autorisons pas les comportements haineux sur nos plateformes, tel que défini dans notre politique.

Exemples de principaux risques potentiels inhérents aux droits humains identifiés lors de l'EGRP	Exemples de mesures prises par Meta pour remédier aux risques potentiels en 2024
Certaines langues et certains dialectes peuvent être plus difficiles à modérer que d'autres.	Nous avons conçu et déployé de nouveaux mécanismes afin de classer les contenus en langue arabe par dialecte pour garantir une modération plus efficace et plus précise, notamment au Soudan. Le nouveau système détecte et donne la priorité à l'acheminement du contenu vers les modérateurs les plus susceptibles de comprendre ce dialecte arabe particulier.
Contenu ayant un effet négatif sur l'égalité et la non-discrimination (p. ex., comportements haineux)	Grâce à des études, des engagements externes et des enquêtes menées sur nos plateformes, nous avons mis à jour notre politique relative aux comportements haineux en ce qui concerne le contenu visant à attaquer les « sionistes ». Lors de la formation de nos modèles d'IA, nous avons testé les données de formation afin de détecter tout contenu ou toute propriété susceptible d'augmenter le risque de générer du contenu potentiellement nuisible, par exemple en vérifiant si un ensemble de données était représentatif de plusieurs données démographiques.



4. Vie, liberté et sécurité de la personne

Le droit à la vie, à la liberté et à la sécurité de la personne concerne le droit de ne pas subir de blessures physiques ni d'être privé de liberté. Pour Meta, le respect de ce droit humain implique l'atténuation du risque que le contenu provoque des dommages, notamment les risques de violence et de traite d'êtres humains, les menaces en ligne soutenues par des États et les groupes non étatiques qui se livrent à la violence ou à la haine ou qui en font l'apologie.

Exemples de principaux risques potentiels inhérents aux droits humains identifiés lors de l'EGRP	Exemples de mesures prises par Meta pour remédier aux risques potentiels en 2024
Les acteurs malveillants qui : • Se servent des services et applications Meta pour coordonner des préjudices en ligne et hors ligne • Abusent des services et applications pour mettre en place des cyberattaques ou de l'hameçonnage • Menacent ou harcèlent des défenseurs des droits humains, des militants ou d'autres groupes vulnérables	La politique de Meta relative aux attaques coordonnées et à la promotion d'actions criminelles interdit de faciliter, d'organiser, de promouvoir ou d'admettre certaines activités criminelles ou préjudiciables. En 2024, nous avons fourni des conseils sur les prisonniers de guerre dans la politique, pour que les équipes d'examen de contenu puissent mieux supprimer les contenus en infraction à grande échelle, y compris au Soudan. Nous avons continué de soutenir le Fonds pour les défenseurs des droits humains et avons remodelé notre programme des partenaires de confiance afin d'améliorer les réponses d'urgence pour les défenseurs des droits humains et d'autres personnes vulnérables.

5. Intérêt supérieur de l'enfant

La Convention internationale des droits de l'enfant (CIDE) énonce que « l'intérêt supérieur de l'enfant doit être une considération principale » dans le cadre de toutes les actions concernant ces derniers. Le cadre de l'intérêt supérieur de l'enfant de Meta est conforme aux valeurs fondamentales de la CIDE. La protection de l'enfant en ligne est une priorité de Meta. Nous proposons des outils aux ados, aux parents et aux représentants légaux sous la forme de protections intégrées afin de garantir leur sécurité tout en leur offrant de l'espace pour exercer leur droit à la liberté d'expression et l'accès à l'information.

Exemples de principaux risques potentiels inhérents aux droits humains identifiés lors de l'EGRP	Exemples de mesures prises par Meta pour remédier aux risques potentiels en 2024
Les enfants peuvent être exposés à des contenus inappropriés et non sollicités ou même à des comportements de prédateurs.	Nous avons lancé les comptes Ado d'Instagram, une nouvelle expérience pour les ados dotée de protections intégrées, avec les parents aux commandes. Lire la suite

6. Participation publique, processus de vote et éligibilité

Le droit à la participation publique, au processus de vote et à l'éligibilité lors d'élections libres et équitables est un des fondements de la démocratie. Protéger l'intégrité des élections sur nos services et applications est l'une de nos priorités. Nous mettons tout en œuvre pour protéger les élections en ligne avant, pendant et après la période des élections.

Exemples de principaux risques potentiels inhérents aux droits humains identifiés lors de l'EGRP	Exemples de mesures prises par Meta pour remédier aux risques potentiels en 2024
Contenu en infraction pouvant avoir une incidence négative sur la participation publique, le vote ou la candidature à une fonction publique Cela peut venir d'activités comprenant, sans s'y limiter : • Les acteurs malveillants coordonnés qui interfèrent avec les élections • La menace de préjudice hors ligne et de violence envers les candidats • Les tentatives individuelles d'empêcher des personnes de voter, l'augmentation du nombre de spams, l'ingérence étrangère ou les signalements de contenus qui enfreignent nos politiques	En 2024, les élections ont été une priorité absolue. Nous nous sommes préparés à des élections à grande échelle, notamment à des élections à haut risque, et nous avons aidé les personnes en droit de voter à trouver des informations, entre autres mesures. Lire la suite

7. Liberté de réunion et d'association

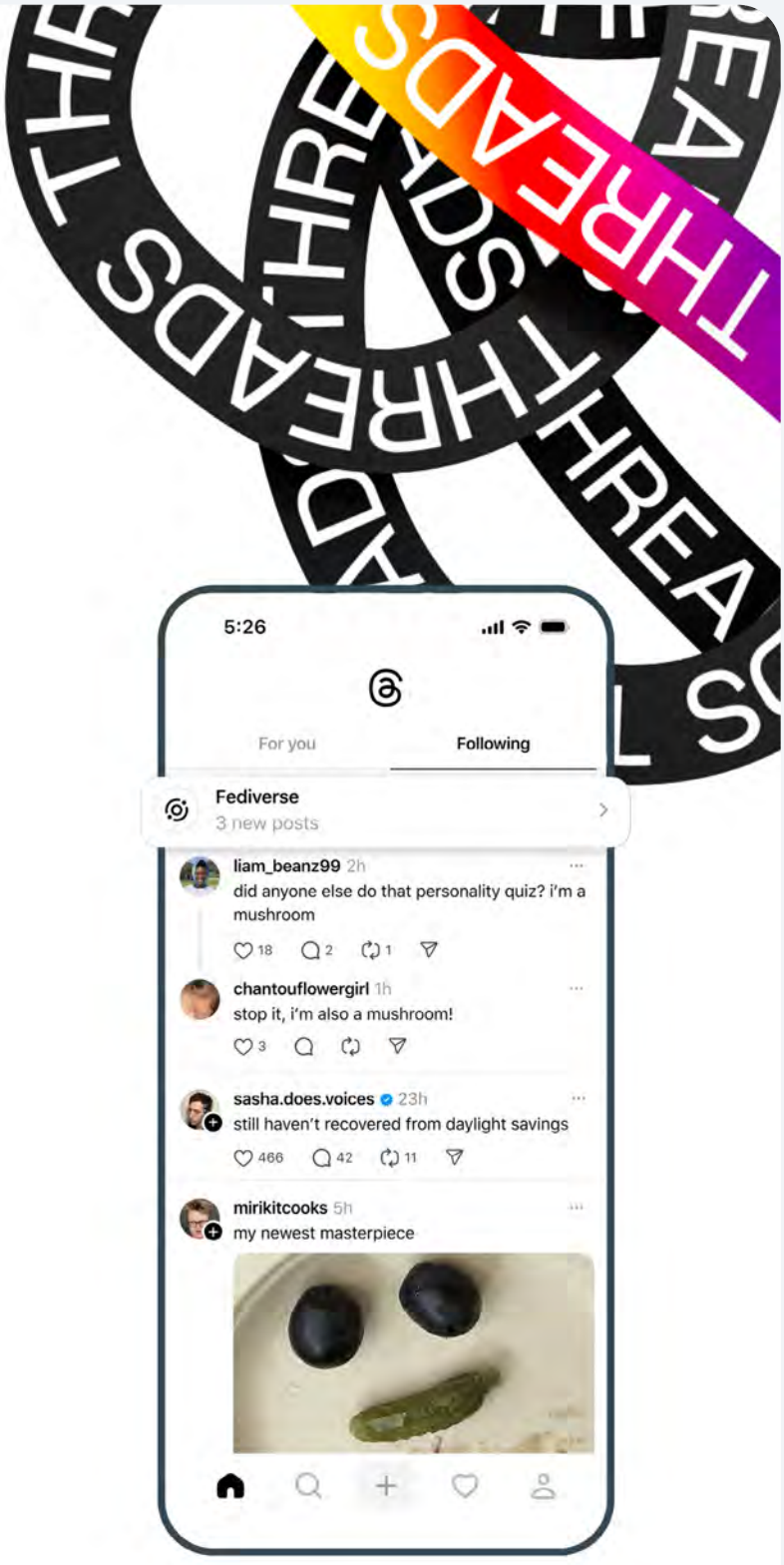
Le droit à la liberté de réunion et d'association est essentiel à la démocratie et interdépendant avec de nombreux autres droits garantis par le droit international relatif aux droits humains, notamment le droit à la liberté d'expression et le droit de participation à des affaires publiques. Pour Meta, ce droit est lié à nos valeurs fondamentales qui consistent à donner la parole à tout le monde et à créer des liens et une communauté.

Exemples de principaux risques potentiels inhérents aux droits humains identifiés lors de l'EGRP	Exemples de mesures prises par Meta pour remédier aux risques potentiels en 2024
Les contenus ou les comportements non authentiques coordonnés sur les plateformes de Meta peuvent donner l'impression à certaines personnes qu'elles ne sont pas en mesure de se rassembler librement sur les applications Meta ou hors ligne.	Nous avons déployé notre Protocole de politique de crise pour soutenir nos efforts visant à lutter contre les contenus en infraction liés aux manifestations de masse, par exemple au Bangladesh et en Géorgie. Nous nous sommes également préparés bien avant les élections pour réduire le risque de contenus en infraction qui pourraient donner l'impression à des personnes qu'elles ne sont pas en sécurité pour se rassembler pendant et après les élections. Threads a rejoint le fédivers, un réseau mondial ouvert composé de serveurs de réseaux sociaux. Ceci a permis aux utilisateurs d'élargir leurs communautés et de toucher de nouvelles audiences.

8. Droit à la santé

Le droit à la santé désigne le droit de toute personne au meilleur état de santé physique et mentale susceptible d'être atteint. Meta respecte ce droit en améliorant l'accès à des informations en matière de santé, en permettant aux personnes ayant des problèmes de santé similaires d'entrer en contact les unes avec les autres et en leur donnant les moyens de prendre des décisions éclairées concernant leur santé et leur bien-être.

Exemples de principaux risques potentiels inhérents aux droits humains identifiés lors de l'EGRP	Exemples de mesures prises par Meta pour remédier aux risques potentiels en 2024
Contenu enfreignant les politiques qui incite ou est destiné à causer des préjudices hors ligne	Nous avons lancé le programme Thrive, un programme de partage de signaux dans plusieurs secteurs conçu pour empêcher la propagation de contenus liés au suicide et aux blessures intentionnelles, en collaboration avec Snap et TikTok. Nous avons organisé un Forum politique sur le contenu commercial présentant des risques pour la santé et la sécurité, en tenant compte des réglementations en vigueur. Nous avons mis à jour nos Standards de la communauté et nos Standards publicitaires pour y inclure les produits rappelés.



Nouveaux produits et services.

Comme évoqué dans notre Rapport sur les droits humains 2023, Meta s'efforce de respecter les droits humains lors de la conception et du développement de nos produits et services.

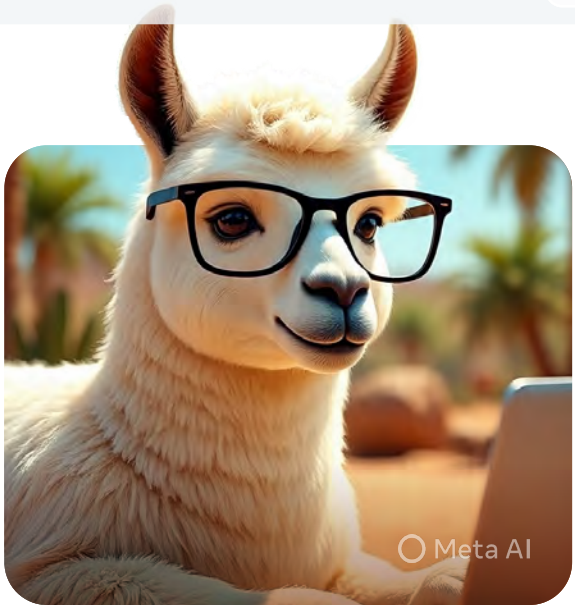
En 2024, Threads a rejoint le fédivers, un réseau mondial ouvert composé de serveurs de réseaux sociaux. Si un utilisateur décide d'activer le partage vers le fédivers, des personnes de différentes plateformes (p. ex., Mastodon ou Flipboard) peuvent suivre cet utilisateur et interagir avec son contenu Threads, même si elles n'ont pas de profil Threads. Cela permet aux personnes d'exercer leur droit à la liberté d'expression ainsi qu'à la liberté de réunion et d'association en touchant de nouvelles audiences, en élargissant leurs communautés et en rejoignant des discussions publiques sur des sujets qui les intéressent. Cela aide également à créer un écosystème d'informations plus varié.

Nous avons aussi fourni des informations aux personnes qui utilisent Threads via une section dédiée dans nos pages d'aide ainsi qu'un nouveau guide sur le fédivers dans notre Centre de confidentialité sur la manière dont la décentralisation et l'interopérabilité impactent leur confidentialité.

Accélérer l’innovation en matière d’IA tout en respectant les droits humains

Accélération des progrès en matière d’intelligence artificielle (IA) en 2024. Les applications et les outils d’IA générative sont de plus en plus utilisés et elles ont considérablement transformé la manière de communiquer, d’apprendre, de créer et de travailler. Chez Meta, nous sommes pleinement conscients que ce développement et cette adoption rapides de l’IA font apparaître des avantages et des risques importants, et souvent nouveaux, en matière de droits humains.

Notre idée à long terme est de créer une superintelligence personnelle et de la rendre accessible à grande échelle afin que tout le monde puisse en bénéficier.



En 2024, nous avons sorti nos grands modèles de langage (LLM) ouverts Llama 3, Llama 3.1, Llama 3.2 et Llama 3.3. Nous avons également lancé notre assistant Meta AI et l'avons intégré à l'ensemble de nos technologies. Meta AI Studio a vu le jour en tant que plateforme permettant de créer des personnages d'IA personnalisés, une suite d'outils d'IA générative a aidé les annonceurs à développer leurs activités, et Meta AI a été intégrée à nos lunettes Ray-Ban Meta. Nous avons continué à mener et à publier des études innovantes en matière d'IA, notamment nos modèles Movie Gen qui génèrent des vidéos et permettent des montages vidéo précis basé sur des instructions, ainsi que notre modèle Video Seal qui donne la possibilité d'apposer un filigrane durable sur les vidéos générées par l'IA, entre autres progrès.

Fin 2024, les développeurs avaient téléchargé plus de 650 millions de fois nos modèles ouverts Llama, et Meta AI comptait près de 600 millions d'utilisateurs actifs mensuels dans le monde, faisant ainsi de nos modèles les plus adoptés au niveau mondial. Cette base d'utilisateurs étendue, composée de développeurs et d'utilisateurs finaux, souligne notre responsabilité de développer l'IA dans le respect des droits humains.

Notre approche ouverte

Nous pensons que l'IA open source est un élément important pour s'assurer que tout le monde puisse profiter des avancées dans le domaine de l'IA. Comme nous l'avons décrit dans notre Rapport sur les droits humains 2023, une approche ouverte présente des avantages importantes pour les droits humains. Modèles d'IA open source :

- Ils sont intrinsèquement plus résistants à la censure et aux autres restrictions du droit à la liberté d'expression, car ils peuvent être téléchargés et utilisés hors ligne, ce qui réduit l'impact des potentielles demandes gouvernementales visant à restreindre les publications après leur sortie.
- Ils permettent mieux l'adaptation et l'ajustement pour refléter le contexte et les nuances au niveau local, conformément au droit à l'égalité, en améliorant aussi bien l'accessibilité que l'inclusion linguistique.
- Ils permettent aux développeurs de créer plus facilement des modèles plus petits et plus efficaces capables de réduire les obstacles auxquels sont confrontées les communautés traditionnellement défavorisées, et de soutenir ainsi les droits économiques, sociaux et culturels.
- Ils soutiennent la recherche essentielle sur les mesures de protection et la sécurité de l'IA en permettant à tout le monde d'examiner attentivement les modèles afin d'identifier les risques potentiels, contribuant ainsi à atténuer les éventuels impacts négatifs sur les droits humains.

Nous constatons déjà les avantages tangibles de notre approche ouverte. Lancé en 2023, notre programme des bourses Llama Impact s'est poursuivi en 2024. Conjointement avec nos Llama Impact Innovation Awards de 2024, ce programme soutient et met en lumière les cas d'utilisation de nos modèles ouverts ayant un impact social positif.

Par exemple, les développeurs ont utilisé Llama pour créer le modèle Vax-Llama, un service de chatbot conçu pour fournir des informations précises sur la vaccination, destiné à être adopté par les professionnels de santé du monde entier. Il a aussi servi pour le projet Llama-Suho, une initiative visant à ajuster Llama à l'aide de données spécifiques à la Corée afin d'améliorer les mesures de sécurité de l'IA dans le contexte coréen.

Recentrer les protections

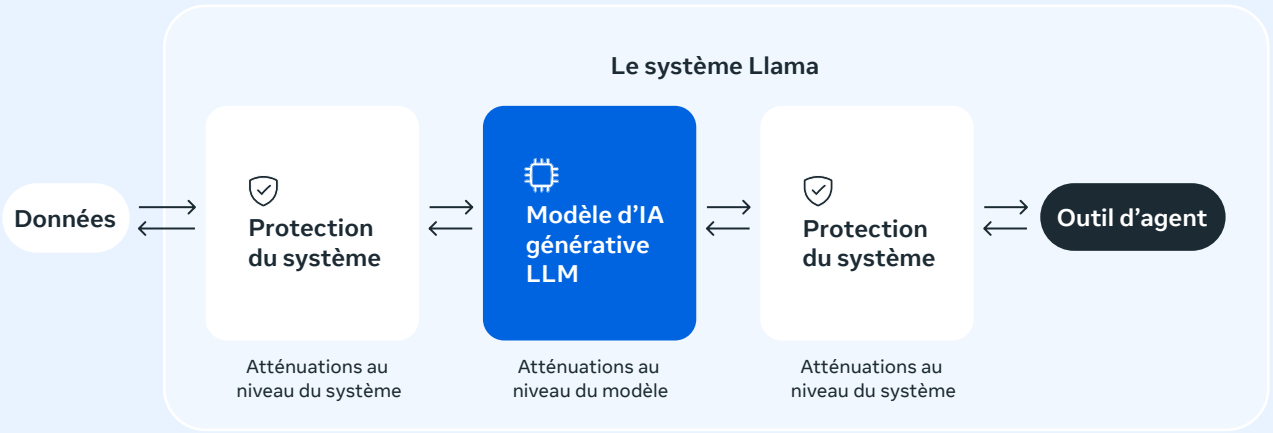
Nous faisons toujours preuve de détermination quant au développement et au déploiement des produits d'IA à la pointe de la technologie, tout en tenant compte des standards en matière de droits humains et des mesures de protection contre les abus.

Notre Politique d'entreprise en matière de droits humains fait explicitement référence à l'applicabilité de nos engagements vis-à-vis de l'IA en matière de droits humains.

Avec la sortie de Llama 3 en avril 2024, nous avons commencé à mettre en avant une approche des mesures de protection de l'IA basée sur les systèmes. Cette approche offre aux développeurs une flexibilité supplémentaire pour appliquer les couches de protection appropriées aux différents cas d'utilisation et aux différentes audiences. Par exemple, nous fournissons des protections pour certains types de discours potentiellement offensants, mais légaux, sous forme de mesures d'atténuation facultatives au niveau du système. Cela s'ajoute à l'intégration continue de protections de base contre la génération de contenus exploitant des enfants dans nos modèles de base.

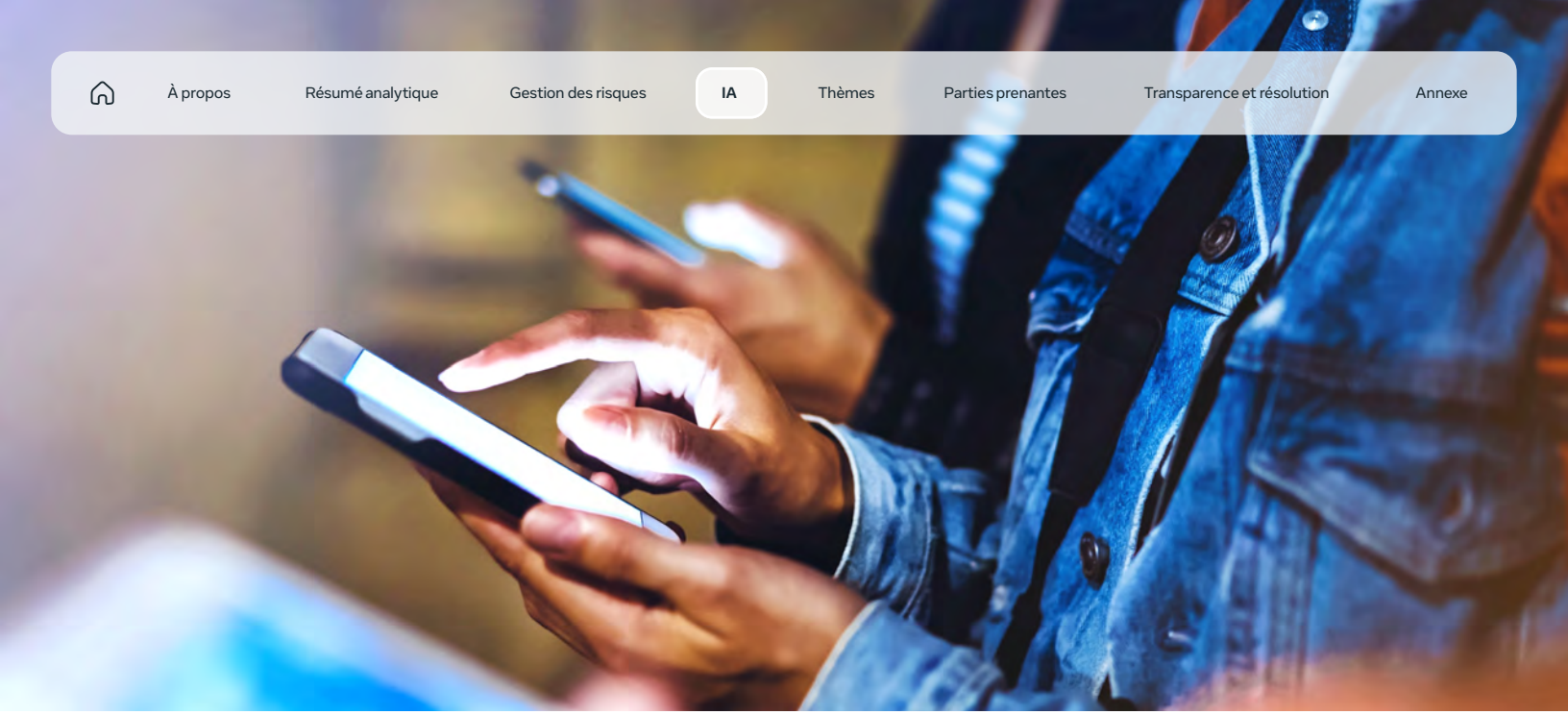
Nous pensons que cette approche basée sur les systèmes favorise un équilibre adéquat entre la liberté d'expression et d'autres droits humains.

Systèmes de sécurité relatifs à l'IA



Dans le cadre de notre approche basée sur les systèmes, nous avons mis à disposition en open source trois outils clés (Llama Guard, Protection des prompts et Code Shield) qui peuvent être personnalisés et utilisés conjointement ou indépendamment par les développeurs afin d'implémenter des protections contre les abus.

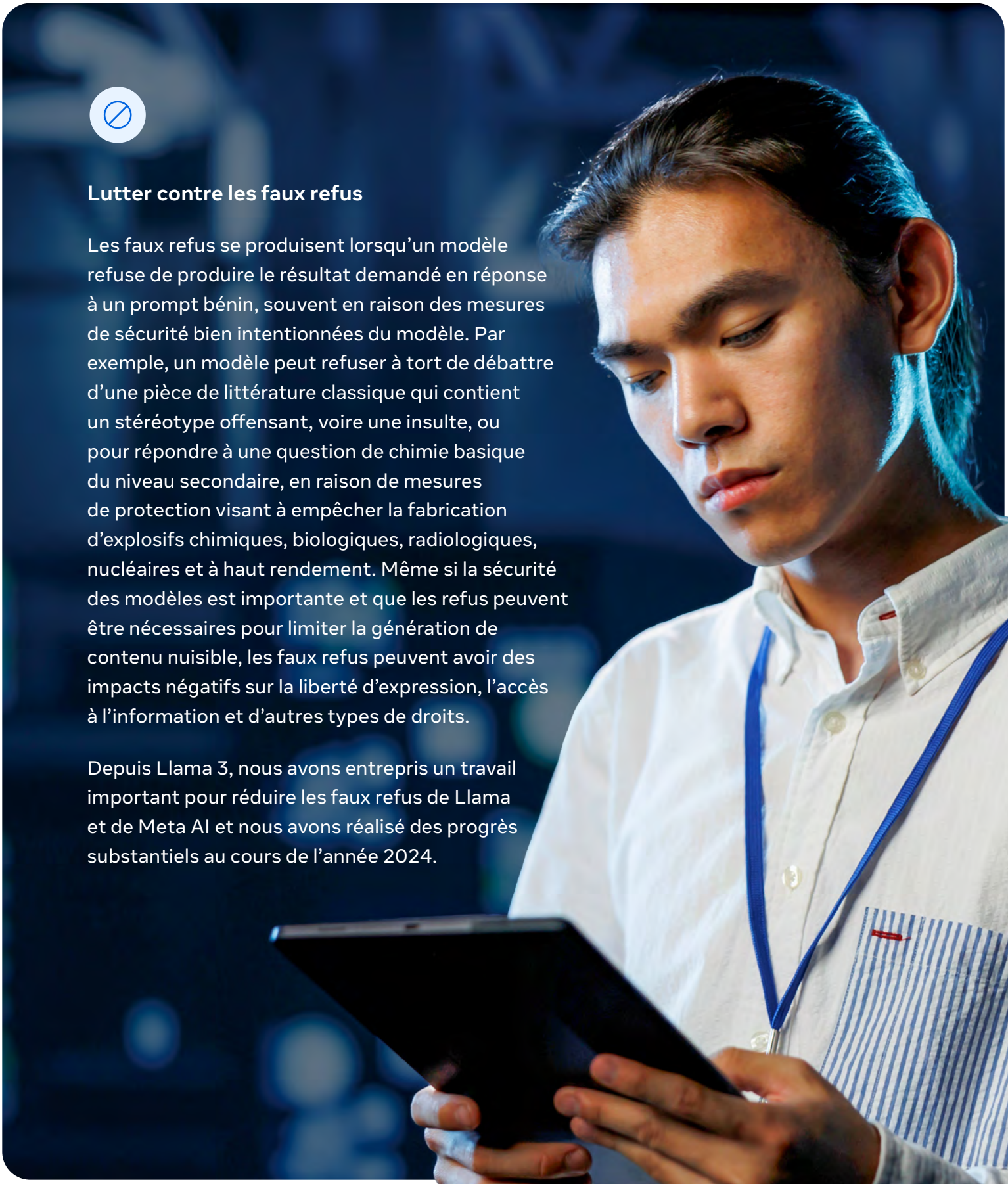
Notre Guide d'utilisation pour les développeurs fournit des conseils détaillés pour apprendre à déployer de manière responsable nos modèles de base et nos systèmes de sécurité dans divers contextes. Notre Politique d'utilisation acceptable continue à gouverner les déploiements de nos modèles ouverts.



Outre ces outils fournis en 2024, nous avons également pris des mesures importantes pour atténuer les risques associés à nos propres déploiements de l'IA générative. Exemples de nos pratiques en 2024 :

- Red-teaming approfondi de nos modèles et produits propriétaires avant leur lancement afin d'identifier et d'atténuer les risques potentiels, y compris ceux liés à d'éventuels impacts négatifs sur les droits humains.
- Mise à jour de notre manière de gérer les médias manipulés sur la base des commentaires émis par le Conseil de surveillance indépendant, notamment en ajoutant des étiquettes « AI Info » et du contexte à un plus large éventail de contenus vidéo, audio et image, et en exigeant des Creators qu'ils mentionnent leur utilisation de l'IA.
- Affinement des règles en interne et du processus que nous utilisons pour tester la production de modèles acceptables afin de mieux refléter les cas d'utilisation du monde réel et de correspondre aux standards internationaux en matière de droits humains.

Nous reconnaissons également que les protections de l'IA nécessitent une collaboration entre plusieurs parties prenantes et plusieurs secteurs. En février 2024, conjointement avec les pairs du secteur, nous avons signé l'Accord sur l'IA et les élections, dans lequel nous nous engageons à empêcher que des contenus trompeurs générés par l'IA n'interfèrent avec les élections mondiales. En mai 2024, nous avons rejoint le Frontier Model Forum, un organisme soutenu par le secteur et entièrement dédié à l'amélioration de la sécurité des modèles d'IA de pointe.



Lutter contre les faux refus

Les faux refus se produisent lorsqu'un modèle refuse de produire le résultat demandé en réponse à un prompt bénin, souvent en raison des mesures de sécurité bien intentionnées du modèle. Par exemple, un modèle peut refuser à tort de débattre d'une pièce de littérature classique qui contient un stéréotype offensant, voire une insulte, ou pour répondre à une question de chimie basique du niveau secondaire, en raison de mesures de protection visant à empêcher la fabrication d'explosifs chimiques, biologiques, radiologiques, nucléaires et à haut rendement. Même si la sécurité des modèles est importante et que les refus peuvent être nécessaires pour limiter la génération de contenu nuisible, les faux refus peuvent avoir des impacts négatifs sur la liberté d'expression, l'accès à l'information et d'autres types de droits.

Depuis Llama 3, nous avons entrepris un travail important pour réduire les faux refus de Llama et de Meta AI et nous avons réalisé des progrès substantiels au cours de l'année 2024.

Internationaliser avec précaution

En 2024, nous avons rendu Meta AI accessible dans plus de 40 pays supplémentaires avec plusieurs nouvelles langues, dont l'arabe, l'indonésien, le philippin, le français, l'allemand, l'hindi, l'italien, le portugais, l'espagnol, le thaï et le vietnamien.

Avant le lancement dans ces pays avec ces langues, nous avons évalué les risques potentiels en matière de droits humains et effectué des exercices de red-teaming spécifiques au contexte, une pratique courante visant à atténuer les comportements à risque dans les LLM.

Les pays où Meta AI est disponible ne disposent pas tous de protections renforcées pour la liberté d'expression dans leur législation nationale. Dans le cadre de notre travail d'internationalisation en 2024, nous avons élaboré une approche qui s'appuie sur les droits humains pour répondre aux demandes gouvernementales visant à restreindre ou à limiter la production de Meta AI, en se basant sur nos politiques de longue date et conformément à nos engagements en tant que membre de la Global Network Initiative et à notre Politique d'entreprise en matière de droits humains.

Interagir avec les parties prenantes

Tous les espaces où la technologie progresse aussi rapidement que l'IA constituent de nouveaux défis pour les interactions avec les parties prenantes. Tout au long de 2024, nous avons souhaité former les parties prenantes et solliciter leurs commentaires constructifs.

Quelques exemples de nos efforts :

Nous avons organisé des tables rondes sur l'IA aux États-Unis pour obtenir l'avis de groupes pluridisciplinaires sur les lancements des produits et des modèles, notamment grâce à des spécialistes provenant des États-Unis, du Brésil, de Bruxelles, de Jordanie, du Mexique ou encore de toute l'Afrique.

Nous avons partagé les conclusions des Forums communautaires qui ont eu lieu aux États-Unis, au Brésil, en Allemagne et en Espagne pour étudier les principes des chatbots de l'IA générative, en partenariat avec le Deliberative Democracy Lab de l'Université de Stanford.

Nous avons organisé une série d'ateliers Open Loop conçus pour lutter contre les complexités et exploiter les opportunités de l'IA open source. Ces ateliers ont réuni des décideurs politiques, des dirigeants du secteur, des universitaires et des représentants de la société civile du monde entier afin d'élaborer ensemble des politiques efficaces et responsables en matière d'IA.

Parallèlement au 13e Forum des Nations unies sur les entreprises et les droits humains à Genève, nous avons développé et animé une simulation interactive multipartite sur la diligence raisonnable en matière de droits humains pour les produits d'IA générative, partageant notre approche et favorisant une meilleure compréhension mutuelle des risques et des défis.

Alors que nous continuons d'innover dans le domaine de l'IA, nous restons engagés à favoriser une implication et une consultation inclusives et solides de toutes les parties prenantes à l'échelle mondiale.

Lire la suite

Thèmes à ne pas rater

2024 : l'année des élections

2024 a été la plus grande année de l'histoire sur le plan électoral. Plus de 70 pays, comptant plus de la moitié de la population mondiale, ont organisé des élections nationales ; et près de 2 milliards de personnes ont pu voter.

Meta reconnaît l'importance de garantir les droits à la liberté d'expression, au vote et à la participation aux affaires publiques. Tout au long de l'année, notre travail électoral a été au centre de nos préoccupations. Nous nous sommes préparés au développement, à la propagation et au rythme des élections, et nous nous sommes efforcés d'atténuer les risques connexes pour les utilisateurs, y compris les risques potentiels liés à l'utilisation grandissante de l'IA.

Dans les pages suivantes, nous allons revenir sur nos efforts de 2024 et fournir des résumés accompagnés d'illustrations concernant les élections dans l'Union européenne, en Inde, au Mexique et aux États-Unis.

Préparation pour le jour des élections à grande échelle

Meta a amélioré notre approche de base en matière d'élections au cours des dernières années. Nous la déployons dans tous les pays où nos services sont utilisés, en prenant soin d'adapter notre stratégie en fonction des besoins et des éventuels risques au niveau local. Dans le cadre de notre préparation aux élections de 2024, une équipe dédiée était chargée de coordonner les efforts à l'échelle de l'entreprise. Elle était composée de spécialistes issus de nos équipes chargées du renseignement, de la science des données, des produits et de l'ingénierie, de la recherche, des opérations, du contenu, des droits humains, des politiques publiques et des affaires juridiques. Tout au long de l'année, nous avons cherché à permettre aux personnes de s'exprimer, de voter et d'être élues.

Notre approche comprenait des efforts pour gérer les risques liés à l'IA, mettre en application nos politiques en matière d'ingérence électorale, perturber les réseaux conflictuels, assurer la transparence de la publicité politique et fournir aux électeurs des informations fiables. Nous avons également évalué la couverture linguistique appropriée des classificateurs et des équipes d'examen manuel pour les pays organisant des élections afin de soutenir nos efforts visant à prendre des mesures contre les contenus qui enfreignent nos règles. Certains éléments de notre travail à la une :

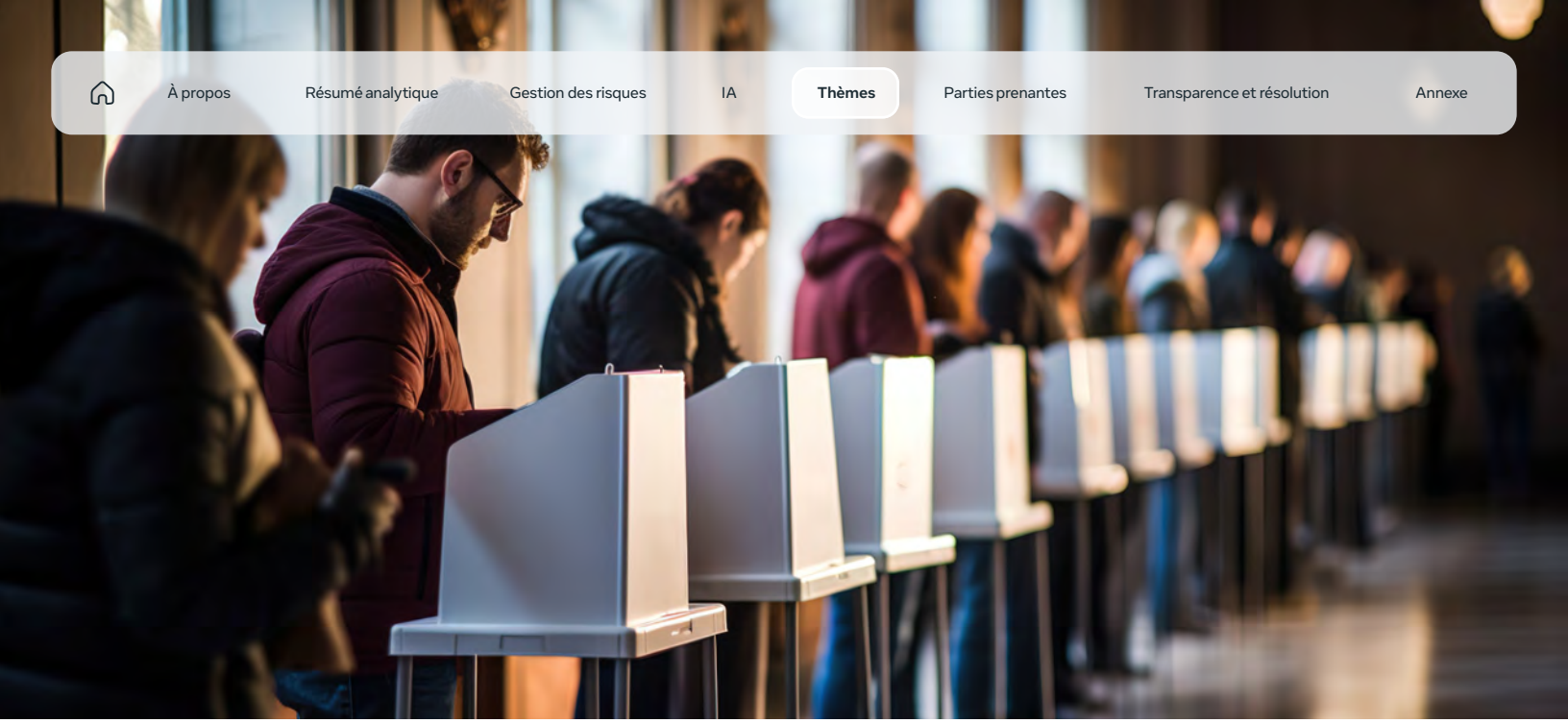
Gestion des risques liés à l'influence de l'IA

Au début de l'année, de nombreuses personnes s'inquiétaient des éventuels risques liés à l'IA générative quant au bon déroulement des élections, notamment le risque de propagation à grande échelle de deepfakes et de campagnes de désinformation basées sur l'IA. Nous nous y sommes préparés et avons surveillé de près les menaces adverses ainsi que les éventuelles perturbations des élections avec l'IA. D'après ce que nous avons observé dans nos services, il est apparu que les présents risques ne se sont pas concrétisés de manière significative et que leur impact a été modeste et limité. Par exemple, pendant la période électorale dans un groupe d'élections majeures, les évaluations sur le contenu IA lié aux sujets électoraux, politiques et sociaux représentaient moins de 1 % de toutes les fausses informations soumises à une vérification. Nos processus et nos politiques actuels ont semblé être suffisants pour réduire les risques liés au contenu basé sur l'IA générative.

Tout au long de l'année, nous avons axé nos efforts sur la lutte contre les opérations d'influence et l'exploitation des partenariats mondiaux afin de préserver l'intégrité des élections.

Nous avons observé de près la potentielle utilisation détournée de l'IA générative par des campagnes coordonnées utilisant des faux comptes. Nous avons constaté qu'elles ont uniquement obtenu des gains incrémentaux en matière de productivité et de génération de contenu grâce à l'IA générative. Ces gains incrémentaux n'ont pas entravé notre capacité à perturber ces opérations d'influence, car nous nous concentrons sur le comportement lorsque nous enquêtons et retirons ces campagnes, et non sur le contenu qu'elles publient, qu'il soit créé ou non à l'aide de l'IA.

Nous avons également coopéré avec d'autres personnes de notre secteur pour lutter contre d'éventuelles menaces en raison de l'utilisation de l'IA générative. Par exemple, en février 2024, nous avons signé l'Accord sur l'IA et les élections avec des dizaines d'autres dirigeants du secteur, nous engageant à empêcher les contenus trompeurs générés par l'IA d'interférer avec les élections mondiales de 2024. Des exemples d'initiatives en matière d'IA, adaptées selon les pays, sont décrites dans les pages suivantes.



Autres efforts relatifs à l'intégrité des élections

Outre l'atténuation des risques liés à la potentielle influence de l'IA sur les élections, nous avons également cherché à encourager les électeurs, à prévenir l'ingérence étrangère, à améliorer la sécurité des candidats, à établir des partenariats et à garantir la transparence des annonceurs.

Encourager les électeurs

L'accès à des informations fiables et l'utilisation responsable de plateformes en ligne ont une importance particulière en période d'élections. Dans de nombreux pays, nous avons fourni des informations relatives aux électeurs et des rappels le jour J grâce à des notifications dans l'application sur Facebook et Instagram. Ces fonctionnalités ont permis aux citoyens d'accéder à des informations fiables provenant des autorités électorales officielles sur la manière, le lieu et le moment de voter le jour du scrutin. Par exemple, lors d'élections locales au Brésil, on a dénombré environ 9,7 millions d'interactions avec ces notifications sur Facebook et Instagram. Plus de 63 millions d'utilisateurs sur Facebook et 118 millions sur Instagram ont vu le sticker d'inscription sur les listes électorales les redirigeant vers des informations officielles à propos des élections et du vote.

Prévenir l'ingérence étrangère

Nos équipes de sécurité ont enquêté et retiré des réseaux coordonnés de comptes, de Pages et de Groupes non authentiques. De plus, selon nos estimations, notre détection automatisée des faux comptes a empêché la création de millions de faux comptes chaque jour. Nos équipes ont retiré environ 20 opérations d'influence secrètes à travers le monde, notamment au Moyen-Orient, en Asie, en Europe et aux États-Unis. Par exemple, en Moldavie, nous avons supprimé un réseau ciblant des audiences de langue russe dans le cadre de notre enquête sur les comportements non authentiques coordonnés suspects dans la région.

Sécurité des candidats

Meta a également fourni une protection renforcée contre le piratage, l'usurpation d'identité et le harcèlement pour les comptes d'élus, de candidats et de leur personnel. Nous avons organisé plusieurs formations sur la sécurité pour les candidats, en dévoilant les conseils disponibles pour lutter contre le harcèlement sur nos plateformes, et publié du contenu éducatif afin qu'il soit largement accessible à l'ensemble des participants aux élections.

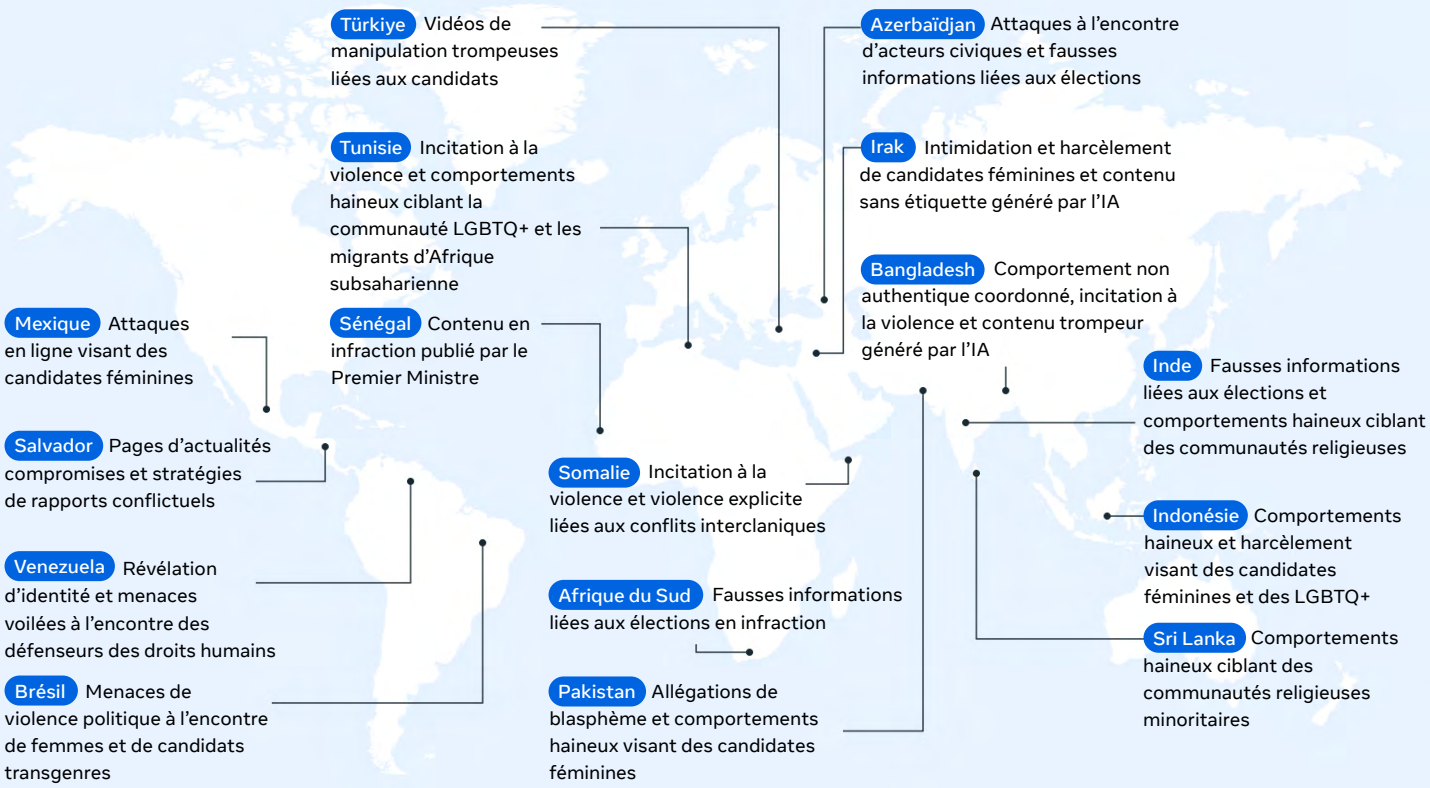
Sensibilisation et partenariats

Nous avons mené des actions de sensibilisation et établi des canaux de communication avec les autorités gouvernementales et les forces de l'ordre afin qu'elles puissent signaler tout contenu susceptible d'enfreindre nos Standards de la communauté ou la législation locale. Nous nous sommes également associés à des groupes de la société civile, à des médias de vérification et à d'autres entreprises technologiques afin de pouvoir identifier et stopper les menaces émergentes et la propagation des fausses informations.

Transparence des publicités

Nous avons continué à proposer une transparence sans équivalent pour les publicités portant sur un enjeu social, électoral et politique. Dans la plupart des marchés où nous proposons ces publicités, les annonceurs passent par un processus d'autorisation et doivent ajouter l'avertissement « Financé par » sur leurs contenus pour diffuser leurs publicités. L'avertissement pourrait inclure des informations sur l'organisation ou sur la personne responsable de la publicité, même si les critères requis peuvent varier d'un pays à l'autre. Les publicités sont ensuite stockées dans notre Bibliothèque publicitaire disponible de manière publique. En 2024, nous avons ajouté la condition requise suivante : les annonceurs doivent mentionner quand ils utilisent l'IA ou d'autres techniques digitales pour créer ou modifier une publicité portant sur un enjeu social, électoral ou politique dans certains cas.

Les partenaires de confiance ont soutenu les efforts en matière d'intégrité des élections dans 25 pays en 2024



Se préparer à des élections à haut risque

Nous avons estimé que certaines élections présentaient un risque plus élevé, nécessitant davantage de préparation, des ressources supplémentaires et un travail sur mesure. Par exemple, nous avons pris en considération le type d'élection, la taille du pays par rapport à notre base d'utilisateurs, les risques de violence politique, le ciblage des groupes vulnérables et notre capacité opérationnelle. Par efforts supplémentaires, on entend la création de mesures de surveillance spécifiques et de mesures temporaires de réponse aux risques pouvant être conçues et appliquées dans tous les pays et toutes les langues.

Nous avons mis en place plusieurs centres opérationnels électoraux à travers le monde afin de surveiller et de réagir rapidement aux problèmes qui pourraient survenir, notamment dans le cas d'élections à haut risque. Vous trouverez davantage de détails en ligne à propos des efforts fournis au Brésil, en France, en Inde, en Indonésie, au Mexique, au Pakistan, en Afrique du Sud, au Royaume-Uni, aux États-Unis et au Parlement européen.

Exemples d'élections nationales

Les quatre brefs exemples nationaux présentés dans les pages suivantes illustrent la manière dont nous avons cherché à gérer les risques liés aux élections en 2024. Dans chacun des contextes, nous avons entamé les préparatifs un an à l'avance, au minimum.

États-Unis

Dans le cadre de la préparation à l'élection américaine, nous avons concentré nos efforts pour aider les citoyens à accéder à des informations fiables concernant les électeurs, à lutter contre l'ingérence étrangère et à garantir la transparence des annonceurs.

Informations relatives aux électeurs	Lors de l'élection générale de 2024 aux États-Unis, les rappels en haut du fil sur Facebook et Instagram ont reçu plus d'un milliard d'impressions. Ces rappels comportaient des informations concernant l'inscription sur les listes électorales, le vote par correspondance, le vote anticipé en personne et le vote le jour du scrutin. Les citoyens ont cliqué sur ces rappels plus de 20 millions de fois pour consulter les sites Web officiels du gouvernement afin d'obtenir de plus amples informations.
Ingérence étrangère	Nous nous sommes préparés à une ingérence étrangère en ligne lors des élections, en élargissant la mise en application de nos mesures coercitives à l'encontre des médias contrôlés par l'État russe, et nous avons continué à perturber l'une des campagnes d'influence secrètes les plus importantes et les plus persistantes, appelée Doppelganger. La grande majorité des tentatives de Doppelganger ciblant les États-Unis en octobre et novembre ont été stoppées de manière proactive avant même que les contenus ne soient visibles.
Période de restriction publicitaire	Pendant la dernière semaine de la campagne électorale,nous avons interdit les nouvelles publicités portant sur un enjeu social, électoral ou politique, une pratique que nous avons conservée depuis 2020. La raison d'être de cette période de restriction est restée la même que les années précédentes : dans les derniers jours d'une campagne électorale, nous avons reconnu qu'il n'y aurait peut-être pas assez de temps pour contester les nouvelles réclamations faites dans les publicités.

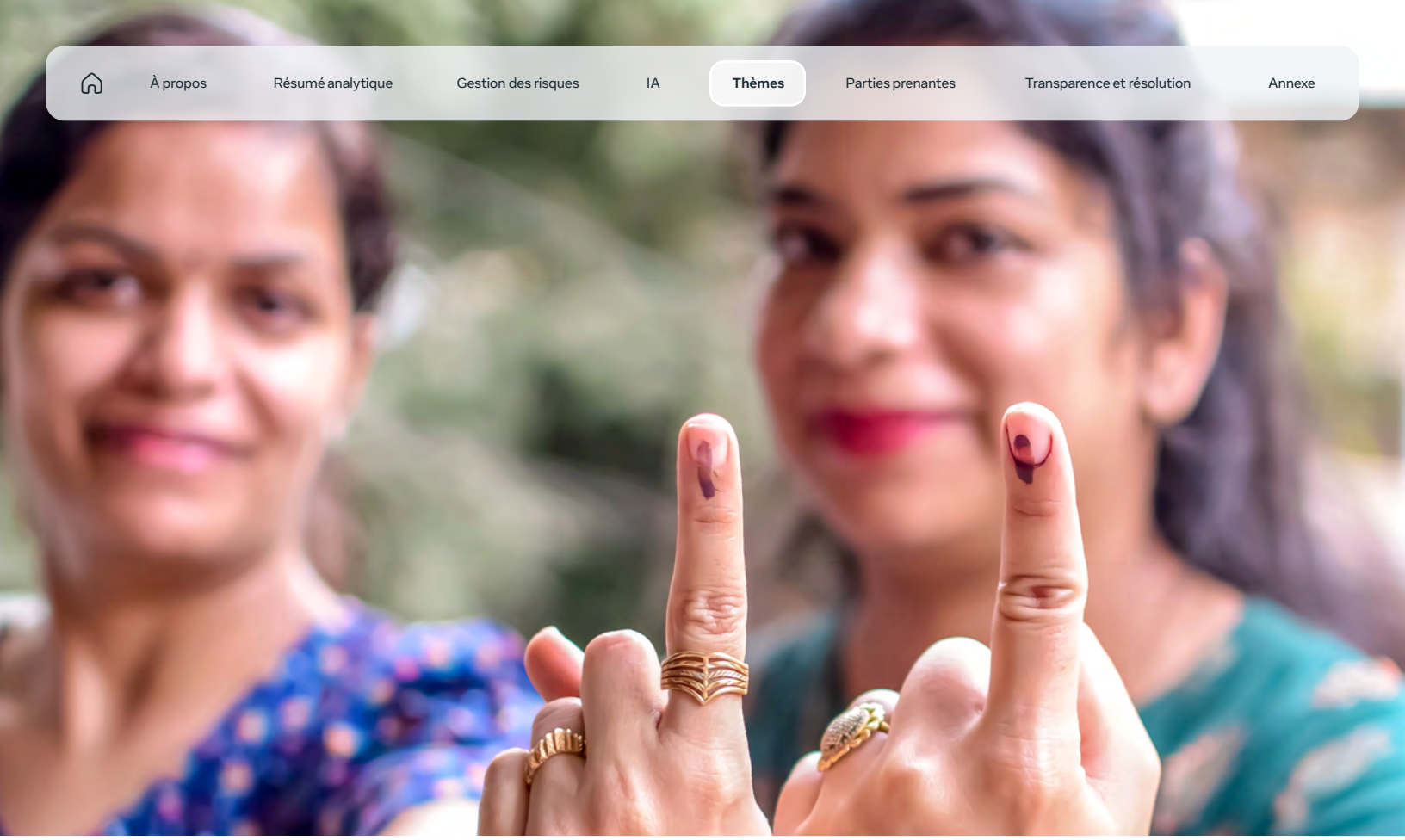
Mexique

2024 a été la plus grande année de l'histoire du Mexique en matière d'élections, avec environ 90 000 candidats en lice pour plus de 20 000 fonctions publiques. Les violences pendant les campagnes électorales ont également atteint un niveau sans précédent. Au moins 37 candidats ont été tués, et plus de 828 attaques non létales ont été enregistrées. Plus de femmes se sont présentées aux élections que lors de tout autre cycle dans l'histoire du Mexique sur le plan électoral, et les candidates ont été victimes d'un taux élevé de violences sexistes et d'assassinats.

Nous avons fourni les mêmes efforts que ceux déployés dans d'autres contextes à haut risque et ils ont bénéficié de l'expertise des spécialistes Meta sur le terrain. Nous avons supprimé plus de contenus en infraction que d'habitude avant et pendant les élections. Par exemple, l'ingérence électorale, les ventes de votes, les contenus haineux et les menaces de harcèlement sexiste et de violences à l'encontre des candidates sur Facebook et Instagram.

Afin de prévenir les perturbations et de réduire les risques de préjudice hors ligne, nous avons concentré nos efforts sur la sécurité des candidates, en fournissant des informations faciles à utiliser sur les électeurs et une éducation aux médias.

Sécurité des candidats	Nous avons inscrit plus de 3 000 candidates, y compris toutes les candidates aux élections fédérales et aux postes de gouverneures, à notre programme de vérification croisée afin d'empêcher les erreurs de mise en application et/ou avons appliqué la Protection avancée à leurs comptes. Cela comprenait la surveillance des éventuelles menaces de piratage. Nous avons élaboré « Vote Against Violence », une campagne éducative en collaboration avec des organisations à but non lucratif et des groupes de médias afin de dissuader les violences sexistes en ligne. Cette campagne a touché plus de 1,2 million de personnes sur nos plateformes et a été amplifiée sur d'autres canaux. Les autorités ont envoyé des demandes de retrait lorsqu'elles constataient des épisodes de violence ou des menaces de violence à l'encontre des candidates.
Informations relatives aux électeurs	En collaboration avec l'Institut national électoral (INE), nous avons lancé le chatbot « Inés » sur WhatsApp afin d'aider les électeurs. Le chatbot a répondu à des questions concernant le processus électoral, par exemple, où et comment voter, comment traiter les cartes d'identité des électeurs et la procédure de vote pour les Mexicains qui vivent à l'étranger. Le jour du scrutin, nous avons envoyé des rappels sur Facebook et Instagram, puis lancé des stickers sur ces deux applications dans le but d'inciter les citoyens à voter.
Éducation aux médias	Pour aider à prévenir la propagation de fausses informations, nous avons lancé la campagne « Soy Digital » (« Nous pensons digital ») en collaboration avec l'INE et Movilizatorio, une organisation de la société civile. La campagne a fourni des ressources et des modules de formation accessibles pour développer les compétences en matière de citoyenneté digitale et de maîtrise de l'information, notamment sur la manière de rester en sécurité en ligne. La campagne a touché plus de 15 millions de personnes. Nous avons également formé 300 dirigeants électoraux au niveau des circonscriptions, qui ont ensuite formé des milliers d'agents électoraux à l'éducation aux médias.



Inde

Meta a commencé à se préparer aux élections générales de 2024 en Inde, avec 18 mois d'avance. L'accent a été mis sur la promotion de l'éducation des électeurs et la garantie de l'intégrité de la plateforme. Notre approche flexible était en mesure de tenir pendant une période électorale de 60 jours, durant laquelle plus de 640 millions de personnes ont exprimé leur vote. Nos préparatifs comprenaient :

Éducation et sensibilisation des électeurs	La notification Alerte élection, lancée depuis la page Facebook de la Commission électorale indienne, a touché 145 millions de personnes. La Commission électorale indienne a également déployé l'interface de programmation d'application (API) de WhatsApp pour diffuser des campagnes de rappel pour aller voter, touchant ainsi environ 400 millions de personnes.
Garantir l'intégrité de la plateforme	Nous avons pris des mesures pour empêcher l'utilisation détournée de nos plateformes. Les équipes d'examen de contenu ont travaillé sur les contenus figurant sur Facebook, Instagram et Threads dans plus de 20 langues indiennes et en anglais. Nous avons supprimé les faux comptes et honoré nos engagements pris dans le cadre du Code d'éthique volontaire auquel nous avons adhéré en 2019, aux côtés d'autres entreprises de réseaux sociaux.
Lutter contre les fausses informations	Nous avons lancé une assistance téléphonique de vérification des informations sur WhatsApp, en collaboration avec l'alliance intersectorielle Misinformation Combat Alliance (MCA), afin de lutter contre les fausses informations générées par l'IA. Nous avons lancé une assistance téléphonique WhatsApp avec la MCA, qui a mis en place un Service d'analyse des deepfakes de classe internationale pour évaluer tous les contenus audio et vidéo que les utilisateurs pensaient être des deepfakes. Nous avons également formé des centaines d'agents des forces de l'ordre en collaboration avec la MCA pour lutter contre les deepfakes.

Élections au Parlement européen

Les préparatifs de Meta pour les élections au Parlement européen se sont appuyés sur les enseignements tirés des précédentes élections dans le monde entier, ainsi que le cadre réglementaire défini par la Législation sur les services numériques et nos engagements dans le Code de bonnes pratiques de l'Union européenne (UE) sur la désinformation.

Nos mesures électorales adaptées à l'UE se sont concentrées sur plusieurs actions :

Promouvoir les informations relatives au vote et l'engagement civique	Nous avons fourni des informations électorales fiables et orienté les utilisateurs vers des informations sur le processus électoral grâce aux rubriques « Unités d'informations pour les électeurs » et « Informations sur le jour du scrutin » de l'application. On a dénombré plus de 41 millions d'interactions avec ces notifications sur Facebook et plus de 58 millions sur Instagram.
Lutter contre les opérations d'influence	Nos efforts visant à stopper les comportements non authentiques coordonnés se sont concentrés sur des menaces spécifiquement associées aux élections au Parlement européen. Nous avons désactivé plusieurs réseaux qui ciblaient l'Union européenne, notamment en intervenant à plusieurs reprises sur le réseau d'origine russe connu sous le nom de Doppelganger.
Lutter contre les fausses informations	Nous nous sommes associés avec le Réseau européen des normes de vérification des informations afin de lutter contre les médias générés par l'IA et modifiés de manière digitale, et nous avons mené une campagne d'éducation aux médias afin de sensibiliser le public aux risques associés.
Lutter contre les risques liés à l'utilisation abusive des technologies d'IA générative	Grâce aux politiques et aux mesures que nous avons mises en place pour lutter contre les contenus d'IA générative, près de 6 000 publicités portant sur un enjeu social, électoral ou politique et plus de 5,7 millions de contenus sur Facebook et Instagram dans l'UE ont reçu des avertissements liés à l'IA à l'occasion des élections au Parlement européen, ce qui a permis de renforcer la transparence.

En savoir plus dans notre Espace modération.

Accéder à l'Espace modération

Sécurité des enfants et des jeunes

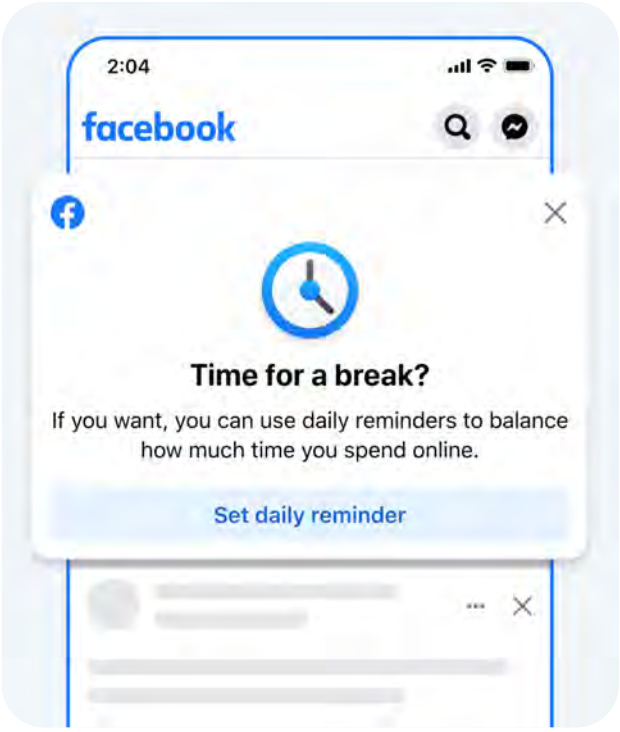
La sécurité des enfants en ligne est une priorité absolue de Meta. Nous proposons des protections intégrées, ainsi que des outils pour les ados et leurs parents, afin de garantir leur sécurité dans nos applications et nos services.

Protection intégrée pour les ados

La sécurité des ados en ligne nécessite une collaboration entre plusieurs parties prenantes dans le monde entier, notamment les parents, les spécialistes de l'enfance, les universitaires, les pairs du secteur, le gouvernement, la société civile, etc. Nous continuons à respecter notre engagement pour protéger les ados tout en leur fournissant un espace leur permettant d'exercer leur liberté d'expression et d'accéder à l'information avec leurs parents pour guides.

Nous avons élaboré plus de 50 outils et ressources sur plusieurs années pour soutenir les ados ainsi que leurs parents et représentants légaux. Nous avons passé plus de dix ans à développer des politiques et des technologies afin de répondre aux contenus et aux comportements qui enfreignent nos règles.

En 2024, nous avons mis à jour nos politiques et le design de nos produits pour créer une expérience unique et différenciée avec davantage de clarté pour les ados. Ces mises à jour continuent d'aider les ados à voir des contenus adaptés à leur âge selon notre cadre de l'intérêt supérieur de l'enfant. Nous avons renforcé les protections existantes sur Instagram en déployant les comptes Ado d'Instagram aux États-Unis, au Royaume-Uni, au Canada et en Australie, avec d'autres déploiements à suivre au niveau mondial. Les comptes Ado repensés disposent de protections intégrées pour limiter qui peut contacter les ados et le contenu qu'ils peuvent voir, ainsi que de méthodes permettant de gérer le temps qu'ils passent sur l'application. Les modifications apportent également aux ados de nouveaux moyens pour explorer leurs centres d'intérêt, avec les parents aux commandes. Cette nouvelle expérience des comptes Ado d'Instagram est conforme aux conseils de spécialistes et au principe de l'évolution des capacités de l'enfant décrit dans la Convention des Nations unies relative aux droits de l'enfant.



Nous avons développé et lancé à l'échelle mondiale un tableau de bord de supervision parentale où les parents et les représentants légaux qui utilisent nos outils de supervision peuvent voir et gérer les comptes appartenant à leurs enfants ; et tout ceci en un seul et même endroit. Ceci permet aux parents et aux représentants légaux de définir des contrôles via leurs propres comptes pour voir et gérer tout contact indésirable ou tout contenu inapproprié, mais aussi de définir des limites de temps d'écran.

Nous avons également mené une série d'ateliers aux États-Unis dans le cadre de notre programme Utiliser les écrans intelligemment (Screen Smart) pour aider les parents à aborder avec leur famille la question de l'utilisation sécurisée des appareils et à en savoir plus sur les outils de supervision parentale proposés par Meta, en utilisant les limites et les protections qui leur conviennent le mieux.

« Les nouveaux comptes Ado d'Instagram lancés par Meta contribuent grandement à donner aux parents la possibilité de guider leurs ados sans pour autant priver les plus âgés de leur autonomie. Les nouveaux paramètres, avec des conseils et des outils améliorés en matière de sécurité et de confidentialité, constituent une avancée majeure. »

— Larry Magid, PDG, ConnectSafely

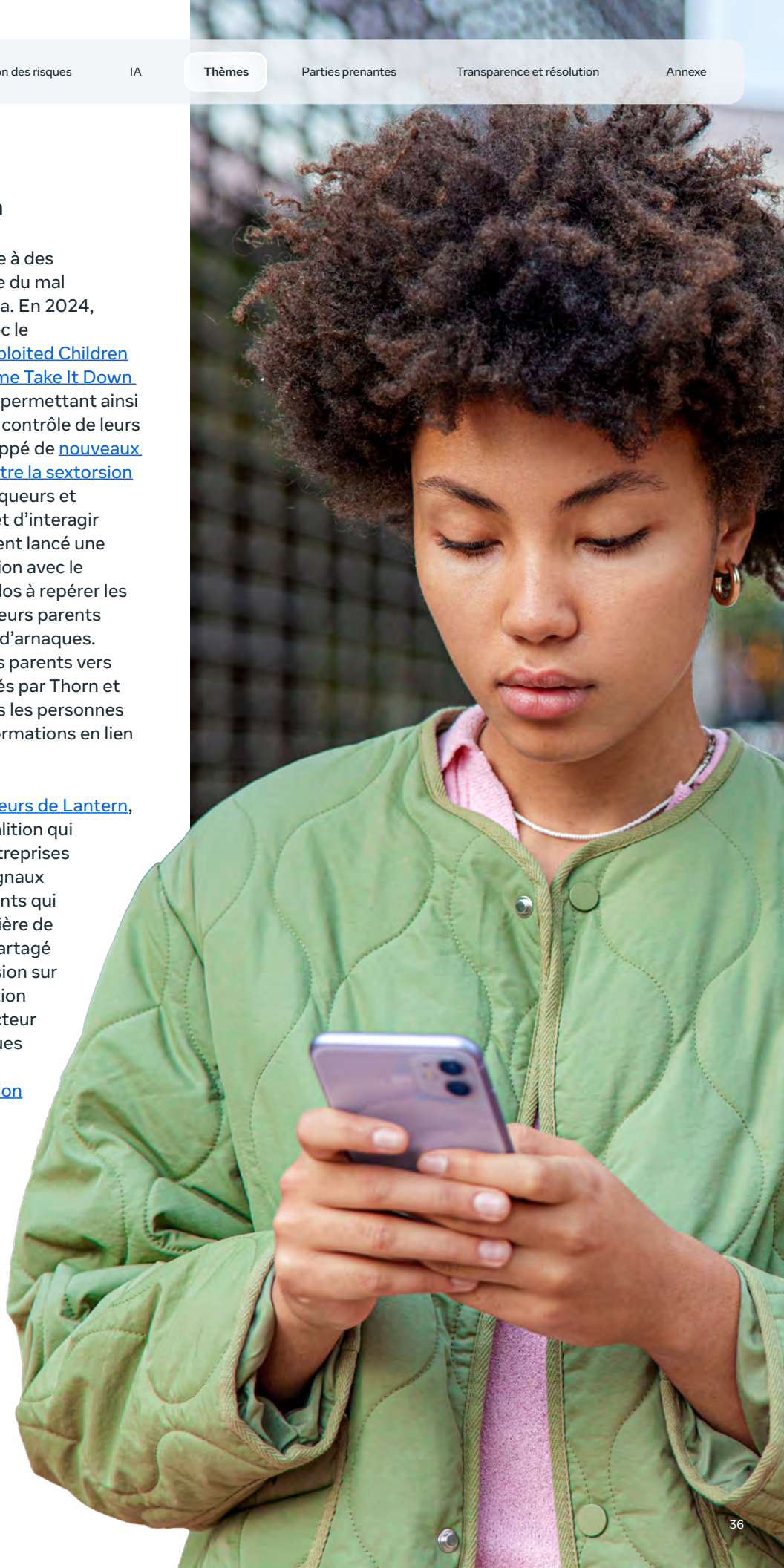
Lutter contre la sextorsion

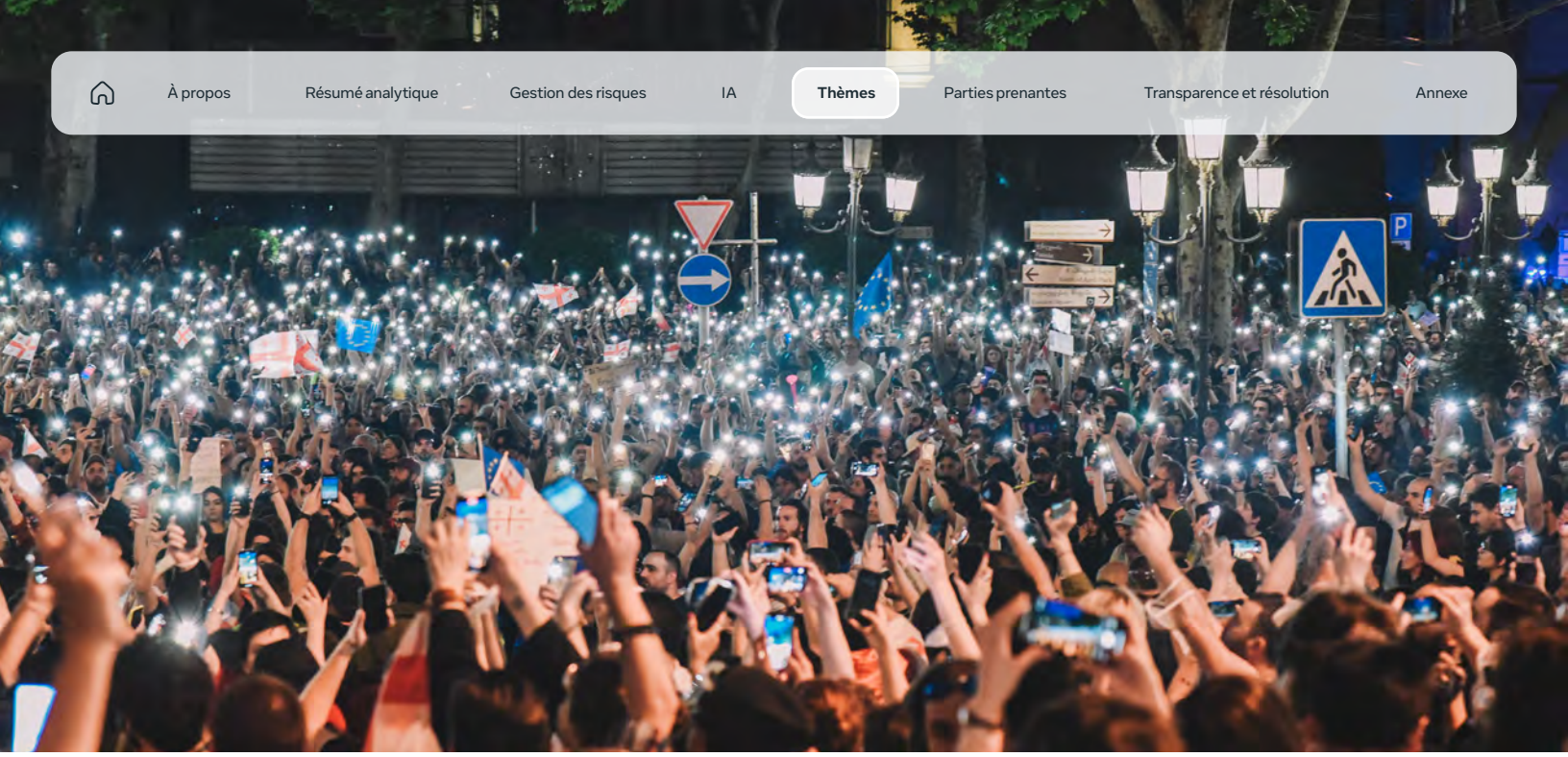
Garantir la sécurité des enfants face à des utilisateurs qui essaient de leur faire du mal reste une priorité absolue pour Meta. En 2024, nous avons continué à travailler avec le National Center for Missing and Exploited Children (NCMEC) pour étendre le programme Take It Down à davantage de pays et de langues, permettant ainsi à davantage d'ados de reprendre le contrôle de leurs images intimes. Nous avons développé de nouveaux outils pour aider à la protection contre la sextorsion et rendre plus difficile pour les arnaqueurs et les criminels potentiels de trouver et d'interagir avec des ados. Nous avons également lancé une campagne éducative, en collaboration avec le NCMEC et Thorn, afin d'aider les ados à repérer les arnaques par sextorsion et d'aider leurs parents à les soutenir pour éviter ces types d'arnaques. La campagne renvoie les ados et les parents vers des conseils de spécialistes, élaborés par Thorn et adaptés par Meta, destinés à toutes les personnes qui recherchent de l'aide et des informations en lien avec la sextorsion.

Nous sommes les membres fondateurs de Lantern, un programme géré par la Tech Coalition qui nous permet, ainsi qu'à d'autres entreprises technologiques, de partager des signaux sur les comptes et les comportements qui enfreignent leurs politiques en matière de sécurité des enfants. Nous avons partagé des signaux spécifiques à la sextorsion sur Lantern pour établir cette coopération importante entre les acteurs du secteur afin de tenter de stopper les arnaques en matière de sextorsion sur les plateformes. En 2024, la participation au programme a doublé, portant à 26 le nombre total d'entreprises ayant adopté Lantern.

Cliquez ici pour obtenir la liste complète de nos outils, fonctionnalités et ressources afin d'aider les ados et leurs parents.

Lire la suite





Comment se préparer aux crises et y répondre

Nous nous préparons à intervenir dans de nombreuses situations de crise à travers le monde, notamment les conflits, les violences intercommunautaires, les troubles civils, les manifestations de masse, les catastrophes environnementales, mais aussi les attentats terroristes et les fusillades. En 2024, nous avons initié et coordonné des services de crise au Bangladesh, en Géorgie, au Kenya, en Nouvelle-Calédonie, au Nigeria, en Corée du Sud, au Royaume-Uni et au Venezuela, pour ne citer que ces pays et territoires. Nous avons poursuivi nos efforts pour les crises mentionnées dans le Protocole de politique de crise pour les conflits qui ont lieu en Ukraine, au Soudan et au Moyen-Orient.

Le Protocole de politique de crise est un outil essentiel que nous utilisons en période de crise. Le protocole sert de guide dans l'utilisation rapide des leviers à notre disposition pour atténuer les préjudices potentiels dans les domaines suivants :

- Politique, comme la publication de conseils supplémentaires pour les équipes d'examen. Par exemple, fournir des conseils sur la suspension des sanctions pour certaines violations de nos politiques relatives au contenu violent et explicite afin d'éviter de pénaliser ou de restreindre de manière excessive les utilisateurs qui tentent de sensibiliser aux conséquences d'un conflit.
- Produits, comme la modification de l'expérience produits. Par exemple, modifier les paramètres pour que seuls les amis et la famille puissent commenter les publications.
- Personnes, y compris le déplacement des ressources pour se concentrer sur des questions spécifiques.

Le protocole de politique de crise nous aide à évaluer hors ligne les situations susceptibles d'entraîner des risques sur la plateforme. Une fois désignés, nous réalisons une évaluation pour identifier les risques sur la plateforme et déterminer si des mesures supplémentaires sont nécessaires. Les types spécifiques de réponses déployés sont cohérents avec les risques observés. Ils s'appuient sur les interventions menées lors de crises passées, les principes des droits humains et le droit des conflits armés.

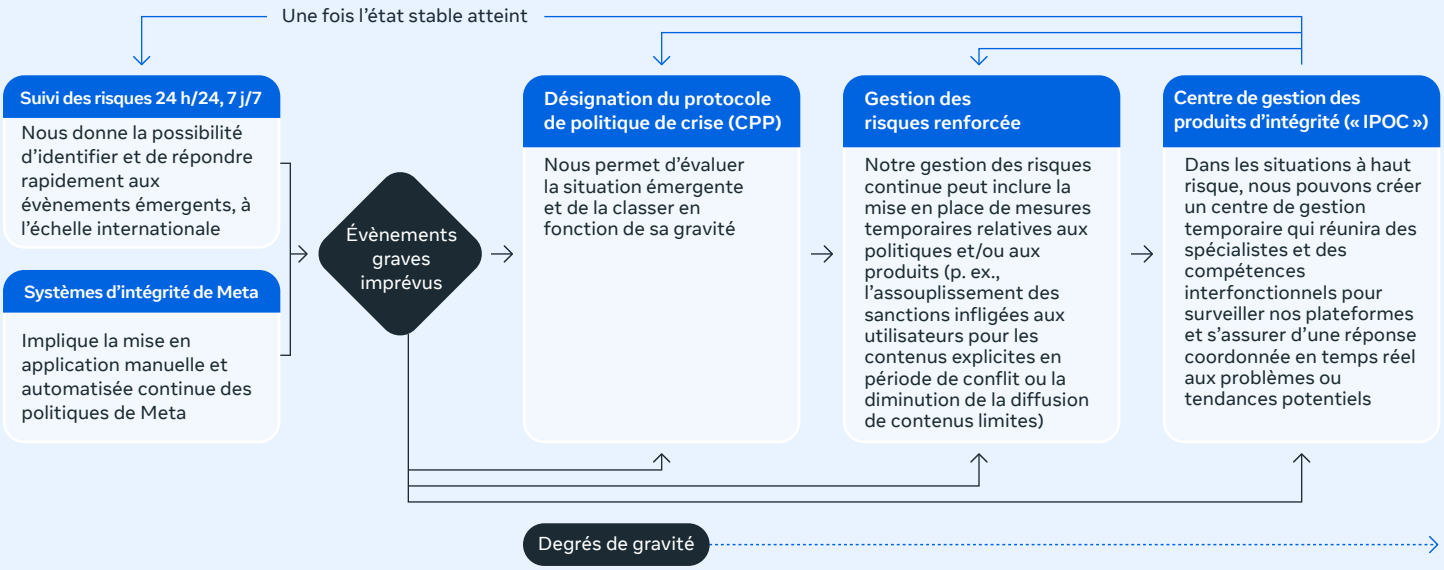
La page suivante illustre notre manière de nous préparer et de répondre aux crises et aux conflits. Nous donnons également quelques exemples pour montrer comment nous utilisons notre Protocole de politique de crise et la diversité de nos efforts sur le plan géographique.

Se préparer et répondre aux crises et aux conflits[4]

Notre Protocole de politique de crise et notre travail sur les pays à risque sont des outils essentiels que nous utilisons pour prévenir, détecter et atténuer les risques. Nos équipes en charge des produits, des politiques et des opérations évaluent l'évolution de la situation sur le terrain afin d'orienter des réponses efficaces et proportionnées.

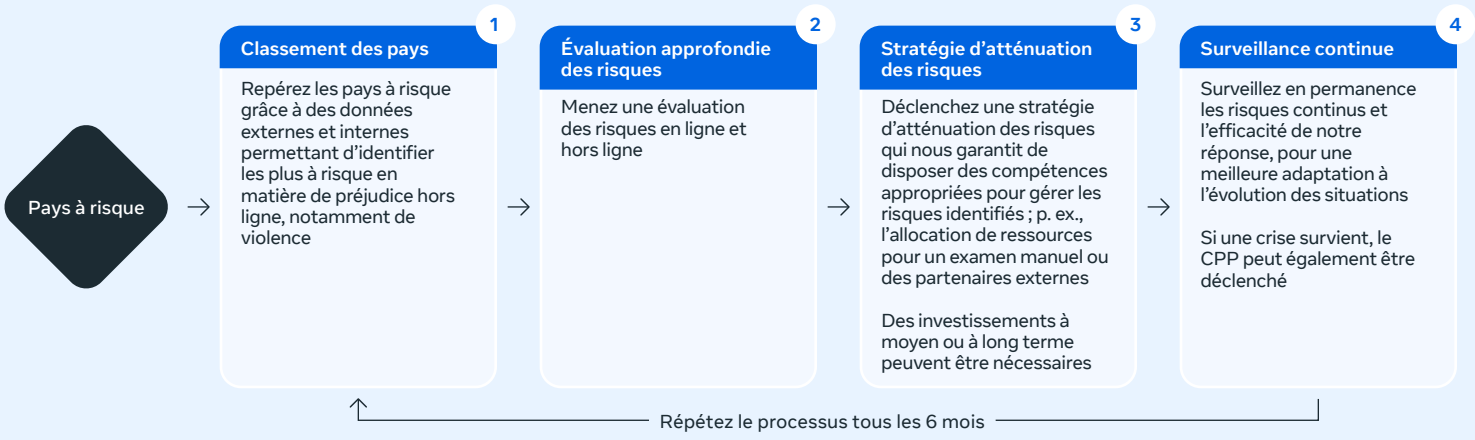
Réponse réactive

Comment réagir rapidement à des évènements graves imprévus ?



Mesures à long terme

Comment prendre des mesures à long terme pour atténuer les risques de conflit ?



[4] Notre intervention en cas de crise couvre de nombreuses situations dans le monde, notamment les conflits, les violences intercommunautaires, les troubles civils, les manifestations de masse et les catastrophes environnementales, mais aussi les attentats terroristes ou autres attaques criminelles.

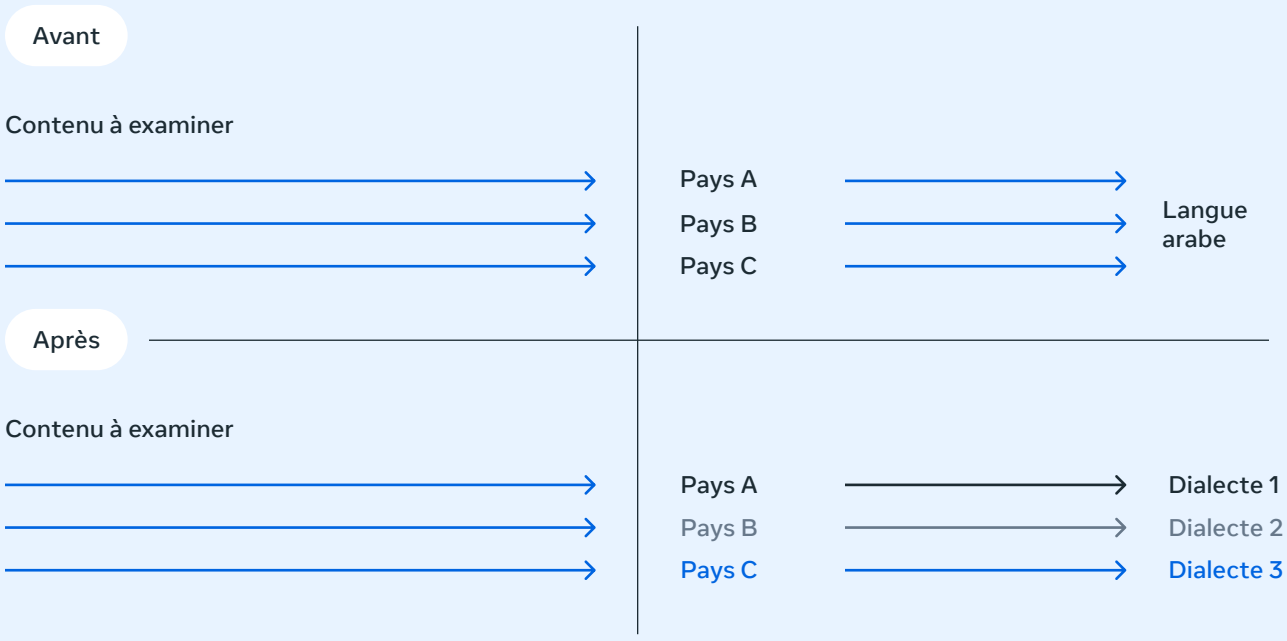
Soudan

En 2024, le conflit au Soudan entre les Forces armées soudanaises (FAS) et les Forces de soutien rapide (RSF) s'est encore intensifié, exacerbant l'instabilité et la crise humanitaire dans le pays. Le volume des contenus en infraction, notamment la violence et l'incitation, le préjudice coordonné, l'exploitation d'êtres humains ainsi que les organisations et personnes dangereuses, a augmenté par rapport aux niveaux précédant le conflit et est resté élevé toute l'année.

Pour réduire la prévalence des contenus qui enfreignent nos politiques, nous nous sommes servi des mesures prises en 2023, en nous appuyant sur notre protocole de politique de crise. Alors que le conflit se poursuivait, nous avons déployé des mesures temporaires et élaboré des mesures d'atténuation à long terme pour faire face aux risques liés au volume élevé et constant de contenus en infraction.

L'une de ces atténuations à long terme consistait à concevoir, élaborer et lancer un système qui identifie les dialectes arabes spécifiques et hiérarchise le contenu pour les équipes d'examen qui sont plus susceptibles de comprendre les nuances linguistiques et le contexte local. Le système précédent identifiait l'arabe comme une seule langue et le renvoyait à des modérateurs capables d'examiner le contenu. Le nouveau système, quant à lui, peut identifier le dialecte exact utilisé, et renvoie le contenu à l'équipe d'examen la plus susceptible de le comprendre. Pour le Soudan, cette modification a entraîné l'examen plus précis de davantage de contenus, réduisant ainsi les erreurs au niveau de la mise en application des règles. Ce travail s'est appuyé sur les résultats de la Diligence raisonnable en matière de droits humains en Israël et en Palestine.

Acheminement en fonction des nuances linguistiques



En 2024, les deux parties du conflit ont de plus en plus souvent révélé l'identité des prisonniers de guerre (PG) en ligne. La révélation de leurs identités a augmenté le risque de préjudices réels et a porté atteinte à la protection de la dignité et de la sécurité des PG en vertu de la Convention de Genève relative au traitement des prisonniers de guerre. Guidés par une recommandation du Conseil de surveillance formulée en 2023, puis à nouveau en 2024 dans le cas de la vidéo captive des Forces de soutien rapide soudanaises, nous avons également reconnu que certains contenus relatifs aux prisonniers de guerre pouvaient présenter un intérêt public, par exemple en sensibilisant aux éventuels abus en matière de droits humains ou en aidant à localiser les prisonniers de guerre disparus. Meta a donc fourni des conseils aux équipes d'examen de contenu sur les PG dans la politique relative aux attaques coordonnées et à la promotion d'actions criminelles afin qu'elles soient plus à-même de répondre aux contenus potentiellement en infraction dans la région à grande échelle.

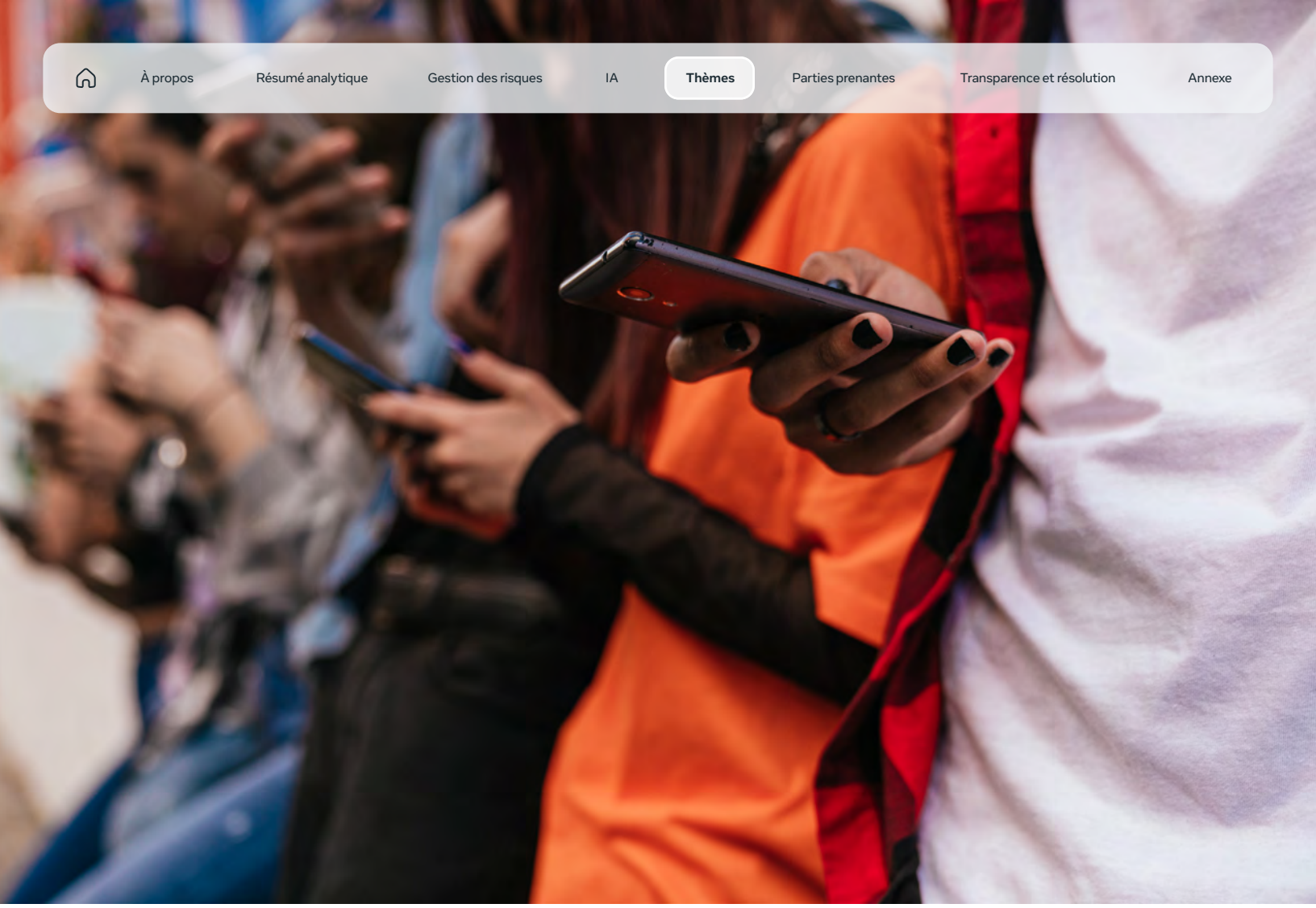
Lire le rapport du Conseil de surveillance pour le premier semestre 2024

Voir le cas du Conseil de surveillance

Les partenaires de confiance ont joué un rôle primordial en fournissant des statistiques essentielles sur les développements au niveau local et sur les contenus potentiellement en infraction liés au conflit. Ces statistiques ont aidé à la mise en application des politiques Meta pertinentes, notamment concernant les comportements haineux, l'intimidation et le harcèlement, l'exploitation d'êtres humains et la désignation des réclamations potentiellement préjudiciables pré-examinées dans le cadre de la politique sur la désinformation et les préjudices, et ont finalement contribué à rendre l'environnement en ligne plus sûr. En vertu de notre politique relative à l'exploitation d'êtres humains, nous avons pu identifier les risques potentiels liés aux images et au recrutement d'enfants soldats, supprimer ce contenu et réduire sa prévalence.

Nous avons dirigé des sessions de formation avec des défenseurs des droits humains, des journalistes ainsi que des organisations nationales et de la diaspora pour les aider à former les utilisateurs soudanais, y compris les migrants et les réfugiés. Ces sessions étaient axées sur les politiques relatives au contenu, la sécurité digitale et le renforcement de leur présence sur les plateformes de Meta.

Les conflits armés déclenchent le déplacement de personnes en masse, c'est la raison pour laquelle, au Soudan, nous avons choisi de nous concentrer sur l'identification d'une éventuelle exploitation d'êtres humains, notamment le trafic d'êtres humains, toute forme de trafic, l'exploitation sexuelle des femmes et des filles ainsi que le mariage forcé. Nous avons supprimé pas moins de 19 100 publications de groupes proposant des services de trafic d'êtres humains et supprimé des contenus faisant l'apologie du mariage forcé. Les interactions avec la diaspora sont restées une stratégie importante pour mettre en évidence les tendances en matière de contenu et clarifier la surveillance au niveau national. Cela a donné lieu à plus de 30 interactions uniques visant à protéger les utilisateurs sur nos plateformes, notamment des efforts pour identifier les mots et expressions nouveaux et émergents liés aux discours haineux et au trafic d'êtres humains. Ces statistiques nous ont permis de détecter plus efficacement les contenus qui ont enfreint nos politiques et d'y répondre.



Moyen-Orient

Le conflit au Moyen-Orient est toujours une priorité pour Meta. En 2024, nous nous sommes concentrés sur les risques découlant de la violence persistante en Israël et à Gaza, alors que la guerre s'étendait à l'ensemble de la région et que d'autres acteurs régionaux s'impliquaient davantage et aggravaient le conflit. Nous avons mis tout en œuvre pour garantir que nos plateformes pouvaient être utilisées pour la liberté d'expression, tout en cherchant à empêcher la propagation des contenus qui incitent à des actes de terrorisme ou de violence et à d'autres préjudices réels.

Notre approche fondamentale n'a pas changé depuis 2023. Cette approche comprenait le maintien de la désignation de l'attaque perpétrée par le Hamas le 7 octobre 2023 en tant qu'attaque terroriste, en vertu de notre politique relative aux organisations et personnes dangereuses (DOI) et le traitement des contenus en infraction dans le cadre de nos politiques. Nous avons cessé les modifications de produits temporaires présentées en 2023.

Immédiatement après les attaques terroristes du 7 octobre, Meta a classé ces violences et le conflit qui s'en est suivi au niveau le plus élevé de notre protocole de politique de crise et a implémenté des mesures immédiates en réponse à la crise. Les mesures prises comprenaient la mise en place d'une équipe interfonctionnelle dédiée disponible 24 h/24 et 7 j/7, ainsi que des mesures temporaires concernant les produits et les politiques. Nous avons choisi de nous tourner vers les Principes directeurs des Nations unies relatifs aux entreprises et aux droits humains, qui constituent la base de notre Politique d'entreprise en matière de droits humains, et sur notre travail de diligence raisonnable en 2022 pour éclairer notre approche. Vous trouverez nos réponses détaillées dans notre Rapport sur les droits humains 2023 et nos publications Newsroom.

Tout au long de l'année 2024, nous avons continué à interagir avec divers acteurs issus du gouvernement, de la société civile et d'autres secteurs en Israël et dans les pays arabes du Moyen-Orient, mais aussi à l'échelle mondiale, afin de faire preuve de transparence et de réactivité. Nous avons également répondu à plusieurs cas du Conseil de surveillance.

De plus, nous avons continué à implémenter les recommandations issues du rapport de 2022 relatif à la Diligence raisonnable en matière de droits humains. Nous avons rapporté les progrès réalisés par Meta pour la période allant du 30 juin 2023 au 30 juin 2024, notamment l'augmentation de nos ressources pour la modération de contenu en hébreu et la mise à jour de notre politique relative aux organisations et personnes dangereuses (DOI) afin de permettre davantage de discours sociaux et politiques dans certaines situations. Ceci faisait suite à des commentaires selon lesquels notre politique relative aux organisations et personnes dangereuses (DOI) capturait trop souvent des contenus tels que des reportages d'actualité, des discussions neutres sur des évènements actuels ou même des condamnations de groupes terroristes et haineux. Les contenus qui font l'éloge ou qui soutiennent des organismes ou des individus dangereux, ou leurs actions ou missions violentes, restent interdits.

Entre le 30 juin 2024 et le 30 juin 2025, nous avons lancé un système qui détecte et hiérarchise le contenu vers les modérateurs les plus susceptibles de comprendre ce dialecte arabe particulier. Nous avons également retravaillé notre canal de remontée des problèmes des partenaires de confiance, ce qui a permis l'amélioration de la rapidité de réponse à ces remontées. Vous trouverez les détails de nos avancées durant cette période dans notre Dernière mise à jour de décembre 2025 : diligence raisonnable en matière de droits humains en Israël et en Palestine.

↗ Progrès en 2023 ↗ Progrès en 2024

Bangladesh

Nos préparatifs en vue des élections de janvier 2024 nous ont permis d'anticiper et de faire face aux nombreux risques qui pourraient survenir pendant cette période. Nous avions pour objectif d'aider à la protection des utilisateurs tout en soutenant leur capacité à voter et à s'exprimer. Ces préparations aux élections nous ont permis de répondre en milieu d'année aux manifestations étudiantes, à la répression violente et au changement de gouvernement qui s'en est suivi.

Compte tenu de la gravité des troubles, nous avons déployé notre Protocole de politique de crise. Nous avons identifié les risques de manière proactive, notamment les discours haineux, l'incitation à cibler les communautés religieuses minoritaires, les fausses informations et le comportement non authentique coordonné. Nous avons mis en place des atténuations, notamment l'utilisation de notre Politique relative aux lieux temporairement à haut risque, les interactions avec nos partenaires de confiance et notre réseau de vérification tierce, ainsi que l'application de protections renforcées aux comptes des défenseurs des droits humains.

Voici nos autres mesures :

- Établir des signaux de détection précis pour identifier les pics de contenu liés aux violations qui pourraient être mises en application en temps réel, comme la violence explicite et les comportements haineux.
- Appliquer des outils et des techniques, par exemple la détection de l'IA pour identifier et mettre en application les contenus en infraction et les recherches par mots-clés.
- Désigner des réclamations potentiellement préjudiciables pré-examinées supplémentaires dans le cadre de la politique sur la désinformation et les préjudices.

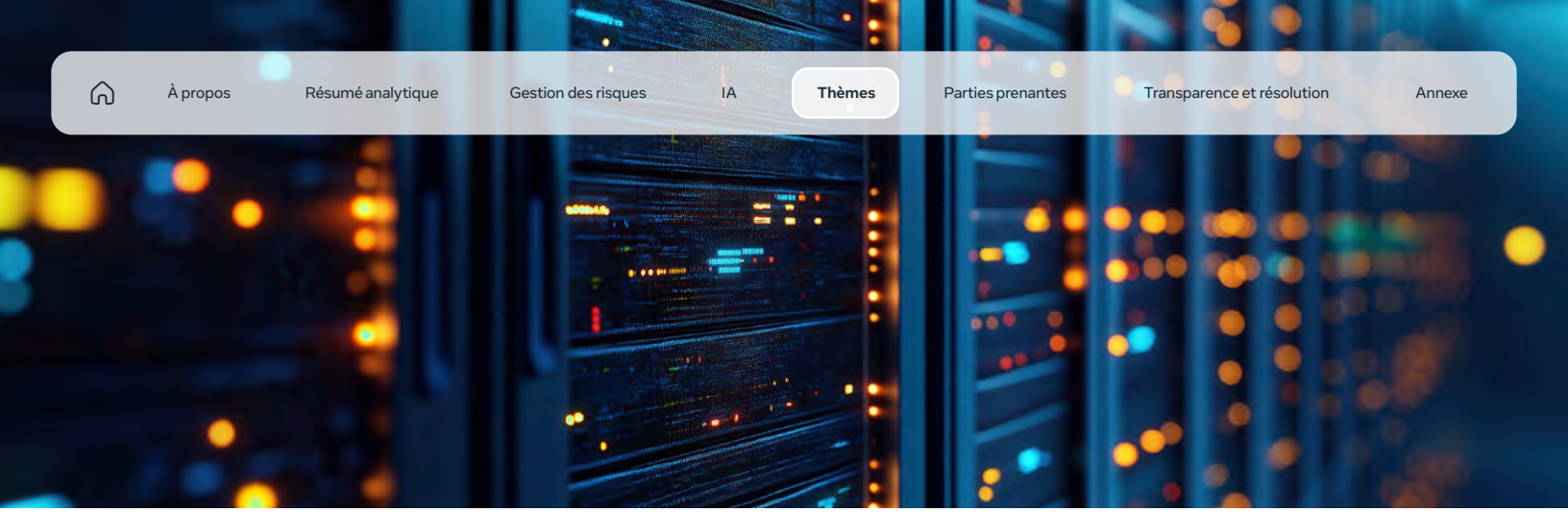
Nous n'avons pas donné suite aux demandes de retrait du gouvernement liées au contenu concernant les manifestations qui étaient contraires aux standards internationaux en matière de droits humains. Nous avons pris cette décision conformément à nos engagements en tant que membre de la Global Network Initiative et à notre Politique d'entreprise en matière de droits humains.

Géorgie

Nous avons déployé le Protocole de politique de crise à deux reprises pour la Géorgie en 2024. En mars 2024, nous l'avons implémenté une première fois après une série de manifestations de masse contre le projet de loi sur la transparence de l'influence étrangère. Nous l'avons réactivé en décembre 2024 à la suite des élections nationales, lors de la survenue de manifestations de masse, accompagnées d'une escalade de la violence de la part de la police et des autres forces de sécurité.

Le déploiement du Protocole de politique de crise a permis à notre équipe de renforcer ses efforts en matière d'atténuation des risques, de faire face aux pics de contenu en infraction et aux risques accrus de violence physique, mais aussi d'aider à la protection des défenseurs des droits humains. Nous avons réalisé un audit sur notre liste d'insultes (des mots qui ont été historiquement utilisés pour attaquer certains groupes) pour identifier et gérer les contenus haineux sur l'ensemble de nos plateformes. Nous avons retiré les faux comptes qui étaient conçus pour manipuler l'opinion publique ou pour diffuser des contenus potentiellement nuisibles. Nous avons également perturbé un réseau de comportements non authentiques coordonnés (CIB) qui visait la Géorgie, ainsi que d'autres comptes non authentiques.

Tout au long des périodes de crise, nous avons collaboré avec des organisations de la société civile, des médias de vérification et des partenaires de confiance, qui nous ont permis de comprendre la tournure que prenait la situation et ont facilité le partage d'informations avec la société civile au sens large et l'opposition en Géorgie. Les partenaires de confiance ont fourni des statistiques et des signaux essentiels sur les contenus en infraction qui ciblaient des groupes d'opposition. Nous avons également interagi avec des partenaires de confiance pour les aider à mieux comprendre ce qui constitue un contenu en infraction selon nos Standards de la communauté. De plus, nous avons contacté des partenaires de la société civile pour identifier les défenseurs des droits humains à risque afin de garantir une protection renforcée de leurs comptes.



Cybersécurité

Nos politiques en matière de sécurité sont importantes pour les droits des utilisateurs à la liberté d'expression, l'accès à l'information et la confidentialité, pour ne citer que ces droits. Nous avons continué à mettre tout en œuvre dans l'entreprise pour identifier et lutter contre les menaces liées aux plateformes adverses, notamment les opérations d'influence, le cyberespionnage, la surveillance, la fraude et les arnaques. Un élément important de notre intervention en matière de sécurité consiste à perturber les réseaux conflictuels qui se livrent à des activités malveillantes.

En 2024, nous avons retiré 20 réseaux de comportements non authentiques coordonnés (CIB) au Moyen-Orient, en Asie, en Europe et aux États-Unis pour avoir enfreint notre politique relative aux comportements non authentiques coordonnés. Ce sont des réseaux qui ont cherché à manipuler le débat public à des fins stratégiques en utilisant de faux comptes ou des stratégies trompeuses. Nous surveillons et supprimons les tentatives de reconstitution sur nos plateformes des réseaux que nous avons déjà supprimés, partageons publiquement des informations via nos rapports sur les menaces et nous efforçons d'intégrer les statistiques issues de nos enquêtes dans nos systèmes de détection et la conception de nos produits afin de les rendre plus résilients.

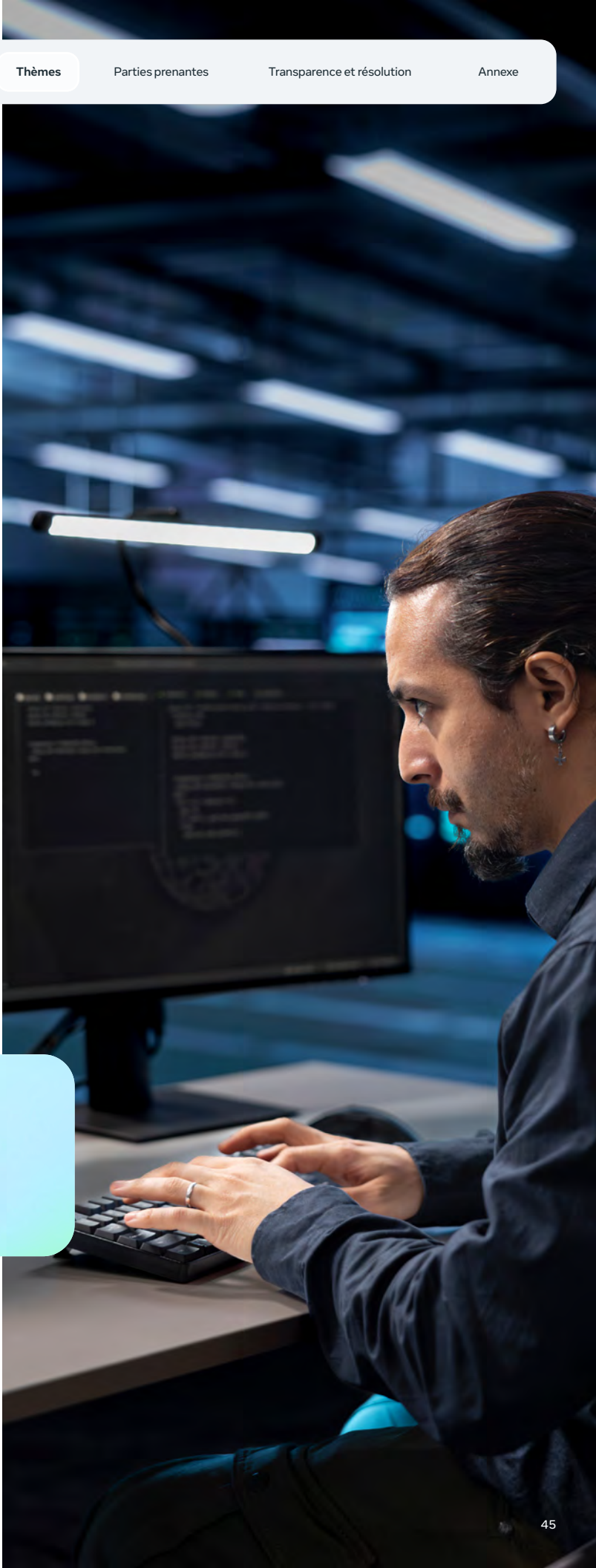
Nous avons continué à détecter et à retirer des réseaux de comportements non authentiques coordonnés (CIB) qui ciblaient et/ou se faisaient passer pour des groupes ethniques ou religieux spécifiques. En 2024, parmi les nombreux exemples, un réseau provenant du Bangladesh a été retiré pour comportement non authentique coordonné. Il ciblait des audiences nationales à l'aide de faux comptes pour publier du contenu et gérer des Pages. Le réseau se faisait passer pour des organes de presse fictifs et utilisait les noms d'organismes de presse actuels pour propager des contenus hostiles au Parti nationaliste du Bangladesh et soutenir le parti au pouvoir. L'opération était en lien avec des particuliers associés au parti de la ligue Awami et à une organisation à but non lucratif au Bangladesh.

Prenons un autre exemple avec un réseau provenant de Chine qui ciblait la communauté sikhe dans le monde, avec des faux comptes et des comptes piratés se faisant passer pour des sikhs, afin de promouvoir un mouvement activiste fictif, Operation K, qui appelait à des manifestations pro-sikhs, notamment en Nouvelle-Zélande et en Australie. L'opération a utilisé des publications et des images générées par l'IA, en anglais et en hindi, concernant les inondations dans la région du Pendjab, la communauté sikhe à travers le monde, le mouvement indépendantiste du Khalistan, l'assassinat de Hardeep Singh Nijjar et les critiques vis-à-vis du gouvernement indien.

Lire la suite

Dans le cadre de nos efforts en matière de lutte contre les entreprises spécialisées dans les logiciels espions, nous avons perturbé et mis fin aux activités de Paragon Solutions, un éditeur de logiciels espions qui ciblait un certain nombre d'utilisateurs de WhatsApp, notamment des journalistes et des membres de la société civile. Nous avons contacté les personnes impactées utilisant WhatsApp et leur avons fourni des ressources pour apprendre à se protéger. Nous leur avons aussi transmis des informations concernant The Citizen Lab à l'Université de Toronto, qui fournit des ressources supplémentaires aux membres de la société civile. En 2024, nous faisions partie des signataires fondateurs du Mémorandum de Pall Mall, une initiative multinationale visant à limiter les abus liés aux logiciels espions.

En décembre 2024, un juge fédéral américain a déclaré le NSO Group responsable d'avoir enfreint les lois fédérales et celles des États, ainsi que les conditions de service de WhatsApp. C'est la première fois qu'une entreprise spécialisée dans les logiciels espions est jugée responsable en vertu de la législation américaine. Meta et WhatsApp ont intenté cette action en justice en 2019 contre le NSO Group, qui avait accédé sans autorisation aux serveurs de WhatsApp afin d'installer le logiciel espion Pegasus sur les appareils mobiles de plus de 1 400 utilisateurs de WhatsApp, notamment des journalistes, des militants des droits humains, des dissidents politiques et d'autres personnes.



20

réseaux de comportements non authentiques coordonnés (CIB) retirés

Implication des parties prenantes

Les interactions structurées et proactives avec notre communauté mondiale d'utilisateurs aident à façonner les politiques de Meta ; elles sont essentielles à notre gestion des risques en matière de droits humains.

Nous avons interagi avec un vaste éventail de parties prenantes en 2024, notamment des membres de la société civile, des universitaires, des groupes de réflexion, des spécialistes des droits humains et des organismes de réglementation. Parmi les questions clés, notre approche en matière d'intelligence artificielle (IA) responsable et d'intégrité des élections, mais aussi nos signaux de désignation pour les organisations et personnes dangereuses, et les évènements violents.

Par exemple, pour évaluer si notre politique relative au terme « sioniste » était appropriée, nous avons réalisé des consultations auprès de 145 parties prenantes issues de la société civile et du secteur universitaire à l'échelle internationale. Parmi ces participants, se trouvaient des politologues, des historiens, des juristes, des groupes de défense des droits numériques et civils, des défenseurs de la liberté d'expression et des spécialistes en droits humains. Nous avons également interagi avec des parties prenantes, notamment des organisations non-gouvernementales issues de notre programme des partenaires de confiance ainsi qu'un large éventail de communautés de la diaspora représentant différents points de vue.

En 2024, nous avons créé des groupes de travail sur la voix et l'expression avec des organisations de la société civile au niveau local pour les régions d'Afrique subsaharienne, du Moyen-Orient et d'Afrique du Nord afin de comprendre leurs préoccupations concernant les propositions législatives au Royaume d'Arabie saoudite, en Jordanie, au Nigeria et au Sénégal, pour ne citer que quelques pays. Lors de ces sessions,nous avons étudié comment protéger l'accès à nos plateformes tout en respectant les restrictions de contenu imposées par la législation locale et nos engagements vis-à-vis de la Global Network Initiative visant à défendre la liberté d'expression et la confidentialité des utilisateurs.

Nous avons aussi piloté un programme de Commissions relatives aux droits humains, qui comprenait des institutions nationales de droits humains provenant d'Éthiopie, du Ghana, du Kenya, du Nigeria et d'Afrique du Sud. Il était axé sur la manière dont Meta traite les contenus potentiellement nuisibles et la réglementation des contenus en ligne.

De plus, nous avons dirigé des ateliers pour répondre aux conflits en Éthiopie, en Palestine, en Somalie, au Soudan et en Tunisie. Nous avons formé des défenseurs des droits humains et des journalistes dans les pays organisateurs d'élections, et leur avons fourni des outils leur permettant de protéger leur présence digitale.

Dans le cadre de notre programme Open Loop India et de notre initiative intitulée Open Loop Sprint, nous avons collaboré avec des entreprises, des législateurs et des spécialistes de l'IA pour fournir des informations sur le rôle de l'implication des parties prenantes dans la chaîne de valeur et le cycle de vie de l'IA.

Notre approche en matière d'implication des parties prenantes

- **Intégrer un large éventail de points de vue et une certaine expertise** : dévoilez des statistiques importantes et impliquez des spécialistes en la matière dans toutes les régions afin d'obtenir des points de vue variés ainsi que des nuances sur le plan local.
- **Faire preuve de transparence** : échangez avec des parties prenantes externes grâce à des défis et des améliorations.
- **Créer une boucle de retours** : montrez comment nos politiques évoluent au fil du temps.
- **Instaurer la confiance** : renforcez la légitimité de nos politiques et de leur mise en application.

464

parties prenantes dans **34** pays ont contribué aux flux de travail de **6** Forums politiques.

121

parties prenantes ont contribué à d'autres initiatives de développement de Meta.

Plus de 100

parties prenantes ont participé à des réunions d'information sur les élections, et **7** newsletters relatives aux élections ont été publiées.

Plus de 290

journalistes, défenseurs des droits humains et activistes ont été formés.

Cycle de développement des politiques de Meta

Examen constant

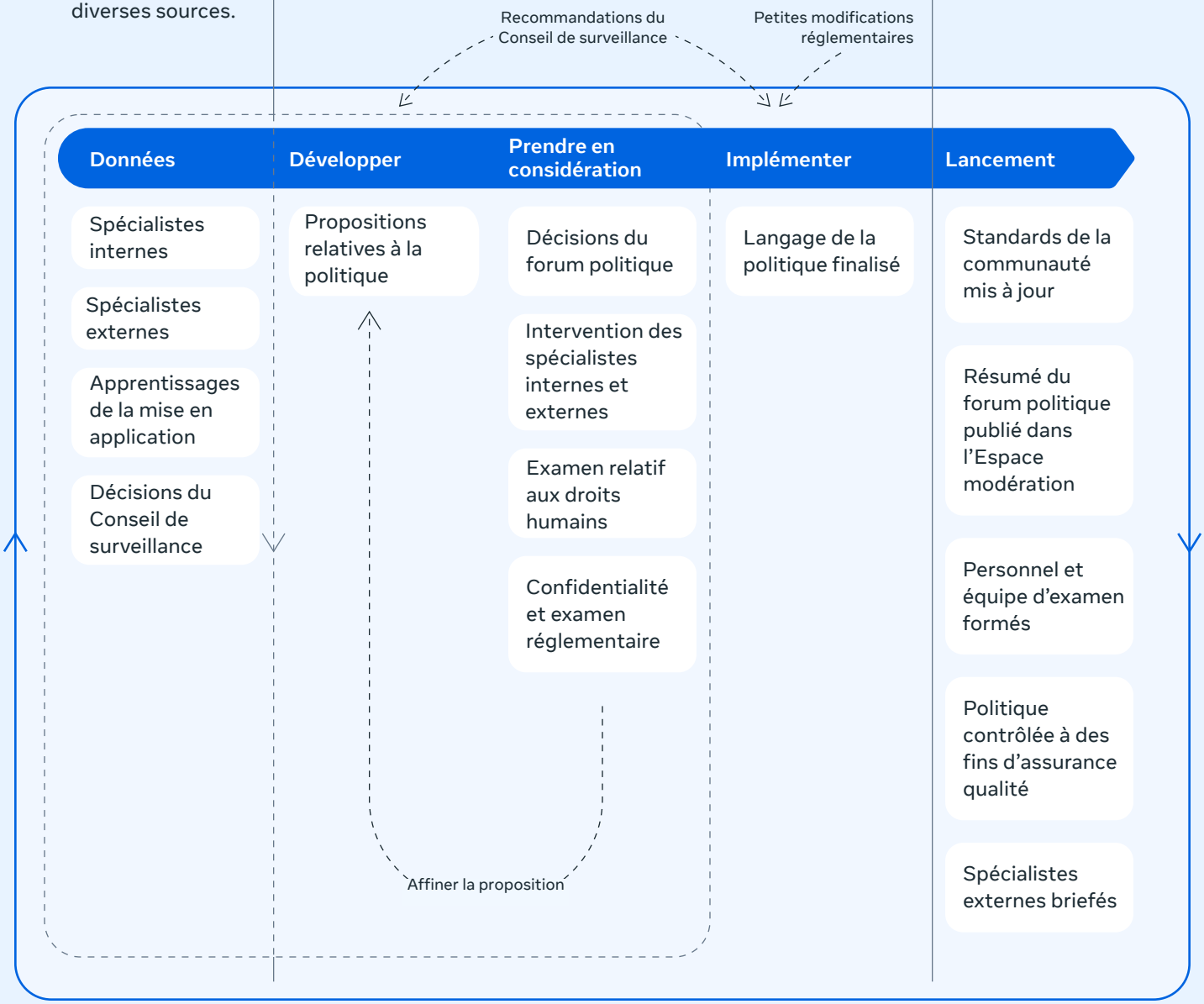
Nous passons constamment en revue nos politiques en fonction des données provenant de diverses sources.

Développement

Les propositions passent par un processus de développement rigoureux pour s'assurer qu'elles sont fondées sur des principes, réalisables et faciles à expliquer.

Lancement

Les systèmes de mise en application sont mis à jour et les politiques sont « diffusées en direct » sur nos services.



Forums politiques

Nous cherchons à élaborer des politiques qui respectent les droits humains et adoptent différents points de vue, où les opinions et les croyances multiples peuvent être entendues et prises en considération. Le Forum politique de Meta est une réunion standard pendant laquelle des spécialistes discutent des modifications potentielles des Standards de la communauté et des Standards publicitaires. Ces réunions permettent de proposer de nouvelles politiques ou d'en modifier des actuelles, de suivre le processus d'élaboration d'une politique qui comprend des interactions approfondies avec des parties prenantes à l'échelle internationale et un examen des études en interne comme en externe.

Nous avons organisé six Forums politiques en 2024 :

1. Le terme « sioniste » comme synonyme de comportement haineux
2. Évènements violents en infraction
3. Contenu commercial avec des risques potentiels en matière de santé et de sécurité
4. Suppression d'images sensibles
5. Contenu lié aux troubles alimentaires
6. Condoléances pour des personnes considérées comme dangereuses

Forums communautaires

Fondés sur une gouvernance délibérative, les Forums communautaires de Meta sont conçus pour exploiter les données publiques sur des sujets où il existe des compromis contradictoires et aucune réponse claire. Grâce à notre approche, les voix en dehors de l'entreprise ont davantage leur mot à dire dans notre prise de décisions, cela nous permet de voir comment l'opinion publique peut évoluer à l'avenir.

En 2024, Meta a organisé un Forum communautaire, en partenariat avec le Deliberative Democracy Lab de Stanford, axé sur les principes que les utilisateurs souhaitent voir sous-tendre le développement des agents IA. Le forum a impliqué pas moins de 1 000 personnes en Inde, au Nigeria, en Arabie saoudite, en Afrique du Sud et en Turquie Un rapport détaillé est disponible ici.

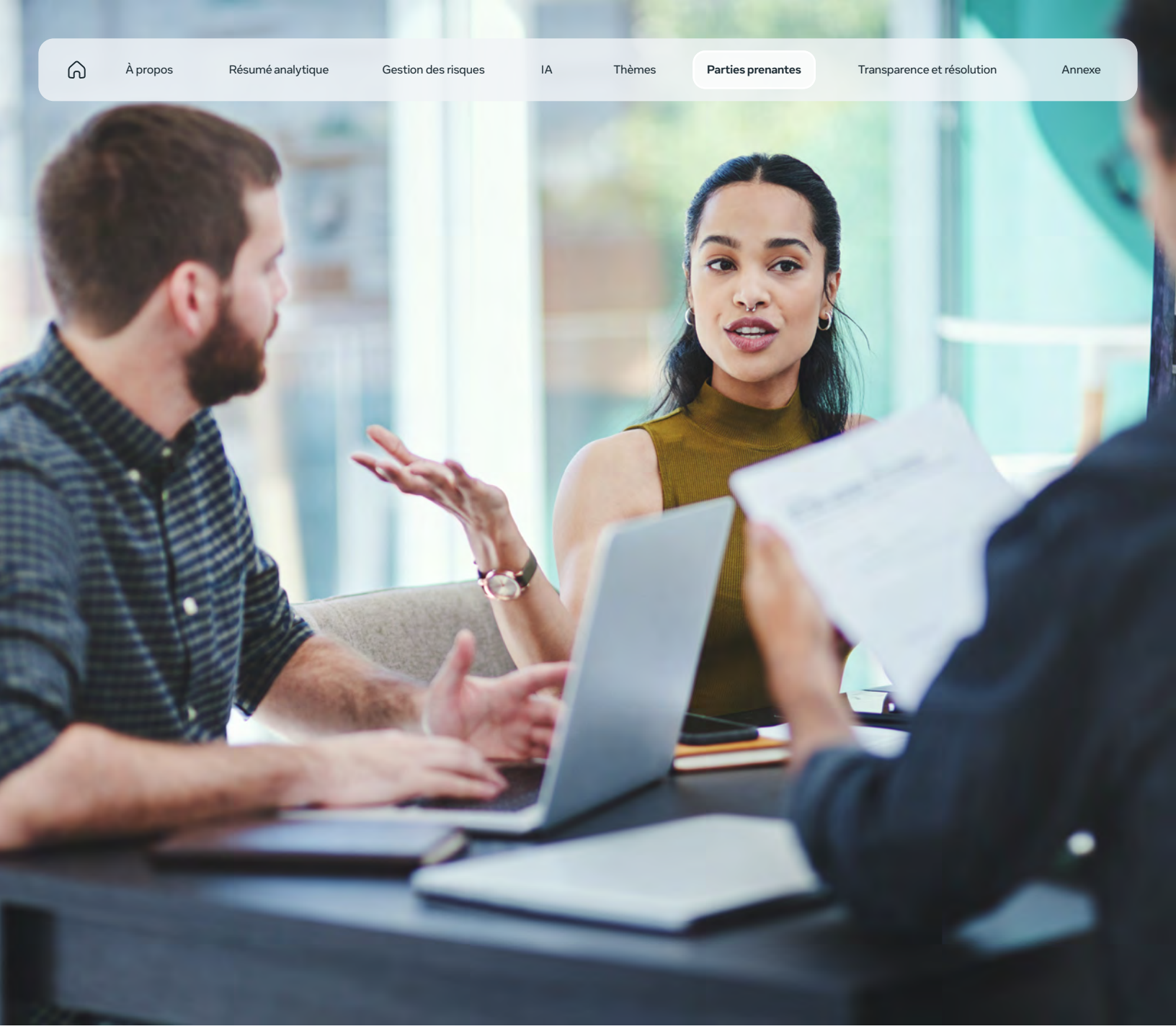
Dans le cadre du forum, les participants ont pu entendre directement des spécialistes en la matière, délibérer entre eux et fournir des avis précieux à Meta. La méthode de délibération a permis aux participants d'aborder les tensions inhérentes à la fourniture d'expériences personnalisées, en évaluant la valeur de la personnalisation par rapport à des compromis tels que la collecte ou encore le stockage des données.

Notre approche des contrôles des utilisateurs et des expériences personnalisées

Les conclusions ont éclairé notre approche des contrôles des utilisateurs et des expériences personnalisées avec des agents IA. Cela comprenait :

- Les participants ont soutenu les agents IA en se remémorant leurs précédentes conversations pour personnaliser leur expérience, en particulier si la transparence et les contrôles des utilisateurs étaient en place.
- Les participants se sont montrés plus favorables aux agents IA adaptés à la culture ou à la région qu'aux agents IA standardisés.
- Les participants étaient en faveur d'agents IA de genre humain qui peuvent réagir à des signaux émotionnels

De plus, nous avons lancé un programme pilote pour faire interagir le public sur ce qui contribue, selon lui, à un modèle d'IA pertinent sur le plan culturel, à développer des ensembles de données sur les préférences à partir de ces avis et à mettre ces données à la disposition des développeurs sous forme open source. Une collection d'ensembles de données facilement accessible verrait alors le jour pour renforcer la pertinence de notre grand modèle de langage Llama ainsi que son utilité dans différents contextes culturels.



Partenaires de confiance

Nous avons continué d'impliquer nos partenaires de confiance pour identifier des tendances, mieux comprendre l'impact des contenus en ligne et des comportements sur les communautés locales, mais aussi découvrir comment renforcer nos canaux de remontée des problèmes vers la société civile.

Les partenaires de confiance sont des alliés importants pour l'identification des violations graves de nos Standards de la communauté et ils ont été particulièrement utiles pendant l'année électorale de 2024. Ils ont fourni des statistiques et identifié des contenus nuisibles dans les pays qui ont connu une augmentation des troubles. Voici quelques pays et régions concernés : Bangladesh, Brésil, Côte d'Ivoire, République démocratique du Congo, France, Grèce, Inde, Indonésie, Kenya, Kurdistan irakien, Mexique, Nigeria, Pakistan, Sénégal, Afrique du Sud, Syrie et Venezuela.

En 2024, nous avons supprimé pas moins de 100 000 contenus qui enfreignaient nos politiques dans le cadre de notre programme des partenaires de confiance.

Les partenaires de confiance ont fourni des informations sur les tendances des contenus liés aux élections pour éclairer les efforts en matière d'intégrité, nous aider à détecter et à supprimer les contenus en infraction, et identifier les utilisateurs à haut risque de nos plateformes pour renforcer les mesures de protection. Les partenaires de confiance ont fait preuve d'efficacité pour identifier les pics de discours hostiles visant les communautés marginalisées et les attaques à l'encontre de journalistes et de défenseurs des droits humains, ainsi qu'une utilisation détournée des contenus générés par l'IA.

Pour gérer les risques de comportements haineux, nous avons supprimé les insultes désignées. Nous avons collaboré avec nos partenaires de confiance pour mieux comprendre le contexte dans lequel les insultes étaient utilisées, afin de pouvoir mettre en application nos politiques avec plus de précision.

Nous avons consulté plus de 40 partenaires de confiance dans 20 pays différents pour éclairer les processus de développement des produits et des politiques en rapport avec la suppression d'images sensibles, l'exploitation d'êtres humains, les signaux de désignation d'organisations et de personnes dangereuses, l'utilisation du terme « sioniste » comme synonyme de comportement haineux, les chatbots d'IA, etc.

En réponse à la recommandation émise par le Conseil de surveillance, Meta a évalué la rapidité et l'efficacité des réponses vis-à-vis du contenu signalé via le programme des partenaires de confiance. Sur une période de deux ans, entre le 2e trimestre 2022 et le 4e trimestre 2024, Meta a réalisé des améliorations considérables en matière de délai de réponse vis-à-vis du contenu signalé via le programme des partenaires de confiance.

Son investissement dans la formation, les systèmes simplifiés de mise en application et les nouveaux outils ont permis d'améliorer le volume des signalements et l'efficacité des examens en 2024.

Résultats au niveau mondial

À l'échelle internationale, le programme des partenaires de confiance a reçu plus de **11 800** contenus signalés au 2e trimestre 2022, chiffre qui a atteint plus de **49 200** contenus signalés au 2e trimestre 2024, soit **quatre fois plus** d'augmentation.

Croissance mondiale du canal des partenaires de confiance sur 2 ans

T2 2022 - T4 2024

+4	+12 points	+15 %	+15
Contenus signalés via le programme des partenaires de confiance	Pourcentage de cas résolus dans les 5 jours qui suivent la remontée d'un problème	Gain d'efficacité sur le délai moyen de traitement en jours	Contenu signalé pour un autre examen de la politique

Vous trouverez des exemples de l'impact des partenaires de confiance au Pakistan, en Syrie et au Venezuela dans les pages suivantes.

ÉTUDE DE CAS

Rapports sur les statistiques provenant de Syrie

Au lendemain de la chute du régime Assad en décembre 2024, les partenaires de confiance ont joué un rôle essentiel dans le signalement et l'analyse des développements au niveau local, en fournissant des informations sur les tendances en matière de contenu local et en faisant remonter les violations graves.

Les rapports établis par les partenaires de confiance, forts de leur expertise locale, ont éclairé les efforts déployés par Meta pour répondre à la crise et nous ont permis de mettre en application nos politiques, mais aussi d'atténuer les risques plus rapidement et avec plus d'efficacité. Le programme des partenaires de confiance a soulevé des préoccupations concernant les risques de révélation d'identité et les allégations d'affiliation au régime déchu visant les minorités ethniques et religieuses, notamment les groupes alaouites, chrétiens et kurdes, et a signalé la montée en puissance de différentes factions extrémistes au sein de l'ancienne armée syrienne. Ces statistiques ont soutenu nos efforts pour atténuer le risque de retrouver des organisations et personnes dangereuses sur nos plateformes et le risque d'attaques physiques en fonction des caractéristiques personnelles.

ÉTUDE DE CAS

Atténuer les risques pour les acteurs civiques au Venezuela

Lors des préparatifs des élections du 28 juillet 2024 au Venezuela, nous avons collaboré avec nos partenaires de confiance et tissé de nouveaux liens avec des organisations de la société civile pour se préparer aux risques liés aux élections et pour augmenter les signalements des contenus potentiellement en infraction.

Après les élections, des manifestations ont éclaté, suivies d'un épisode de répressions de la part du gouvernement. Il s'agissait de détentions en masse et d'arrestations ciblées d'opposants politiques. Nos partenaires de confiance ont fourni des informations essentielles sur les avancées sur le terrain. Ils ont signalé du contenu nuisible, notamment des menaces voilées et la révélation de l'identité de manifestants et de partisans de l'opposition, qui les exposaient à un risque de détention arbitraire et de blessures physiques. Les partenaires de confiance ont également signalé des attaques visant les comptes d'acteurs civiques, tels que les journalistes, les membres de l'opposition et les défenseurs des droits humains, pour ne citer qu'eux.

Ces statistiques ont éclairé une détection proactive et nous ont permis d'activer la Protection avancée pour ces comptes et de prévenir les erreurs de mise en application grâce à des vérifications croisées. Ces mesures ont contribué à soutenir le journalisme et l'engagement civique dans un environnement répressif.

ÉTUDE DE CAS

Les partenaires de confiance s’attaquent aux allégations de blasphème et aux discours hostiles au Pakistan

Au Pakistan, les partenaires de confiance ont joué un rôle primordial pour nous alerter sur les contenus potentiellement nuisibles qui ciblent les communautés marginalisées, notamment les minorités religieuses et de genre.

Pendant la période électorale de février 2024, les partenaires de confiance ont signalé des contenus liés aux élections, axés sur les discours hostiles visant les candidats politiques et les allégations de blasphème qui peuvent constituer une incitation à la haine. Au Pakistan, les allégations de blasphème peuvent déclencher des poursuites judiciaires et des épisodes de violence physique.

Suite à ces signalements, nous avons pu supprimer les contenus liés aux allégations de blasphème dans le cadre de notre politique relative aux attaques coordonnées et à la promotion d’actions criminelles.

Les partenaires de confiance ont également fourni des signaux et des informations pendant d’autres moments importants, comme les épisodes de violence sectaire. Leur travail nous a permis de réagir rapidement, de supprimer les contenus en infraction sur nos plateformes et de renforcer notre détection et notre mise en application.

Implication des parties prenantes au Pakistan

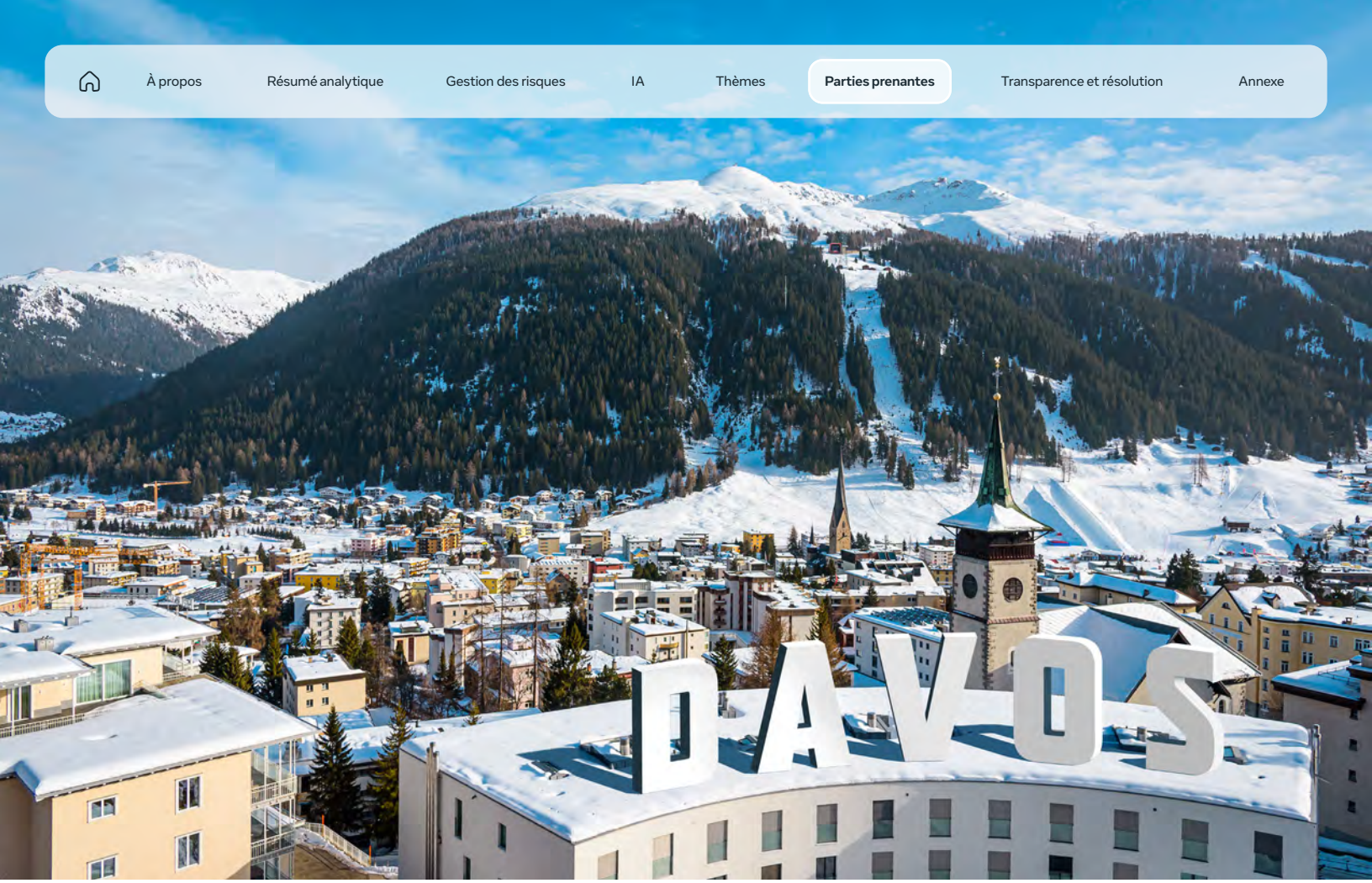
Meta a pris une série d'engagements au Pakistan avec plusieurs parties prenantes gouvernementales ou non dans le cadre de notre diligence raisonnable en matière de droits humains, en se concentrant sur l'équilibre entre la sécurité et la liberté d'expression. Voici quelques exemples :

Une discussion sous forme de table ronde à propos de la sécurité des jeunes en ligne, organisée conjointement avec le Ministère des droits de l'homme, la Commission nationale sur les droits des enfants, la Commission nationale sur les droits humains et la Digital Rights Foundation. Nous avons discuté des comptes Ado et du lancement du portail Take It Down en ourdou pour les utilisateurs pakistanais.

Les interactions avec un groupe varié de défenseurs des droits humains pour rassembler des informations sur les perturbations relatives à Internet et découvrir les possibilités de collaboration en matière de plaidoyer. Ils ont fourni des informations précieuses sur l'impact du « pare-feu » et de l'autorisation des réseaux privés virtuels (VPN) mis en place par le gouvernement.

Une discussion sous forme de table ronde avec des organisations de la société civile pour approfondir les engagements clés de Meta en matière de droits humains ainsi que l'initiative de notre équipe dédiée aux droits humains. Cela comprenait une discussion sur les moyens de répondre dans certaines situations sans fermer Internet ni limiter la bande passante des plateformes de réseaux sociaux, y compris notre famille d'applications.

Lors de chaque évènement, l'ensemble des interlocuteurs a évoqué l'impact des accusations conflictuelles et frivoles de blasphème ciblant les utilisateurs. Ils ont été rassurés en apprenant que Meta avait élaboré une politique relative aux risques liés à la révélation de l'identité et déployé des efforts continus pour assurer la sécurité des personnes visées.



Organisations internationales

En 2024, les États membres des Nations unies ont entamé des négociations, puis adopté le Pacte numérique mondial (GDC), un cadre complet pour la gouvernance des technologies digitales et de l'IA au niveau mondial. Nous avons collaboré avec les États membres des Nations unies, des agences des NU et des coalitions du secteur pour finaliser le texte du GDC. Notre travail visait à soutenir la liberté d'expression tout en créant un avenir digital ouvert, plus sûr et plus inclusif pour tout le monde.

Meta a également interagi sur l'ensemble du système des NU d'autres manières tout au long de l'année. Notre travail comprenait la contribution au projet Technologies digitales, droits et bien-être des enfants de l'UNICEF pour la diligence raisonnable dans le secteur technologique, et avec l'UNESCO sur la gouvernance des plateformes digitales à propos de la désinformation. Nous avons également soutenu l'UNESCO grâce à une interface de traduction qui s'appuie sur le modèle IA Meta No Language Left Behind (NLLB) pour aider à réaliser des traductions de qualité dans 200 langues. Vous trouverez, parmi elles, des langues marginalisées comme l'asturien, le luganda, le maori, le swahili et l'ourdou, contribuant ainsi à promouvoir la diversité linguistique et l'accès à l'information.

Meta a continué à travailler en étroite collaboration avec le Haut-Commissariat aux droits de l'homme (HCDH). Vous avez rencontré régulièrement le personnel du HCDH et avons interagi de manière active dans le projet B-Tech, qui fournit des ressources et des conseils faisant autorité pour l'implémentation des Principes directeurs des Nations unies relatifs aux entreprises et aux droits humains dans le secteur des technologies, ainsi qu'à sa communauté de pratique, un espace de dialogue confidentiel avec d'autres entreprises technologiques. Nous avons participé de manière active à des discussions continues autour de l'IA et des standards en matière de droits humains. De plus, nous avons participé au Forum des Nations unies sur l'entreprise et les droits humains de 2024 et avons fait des présentations sur les thèmes « Discours haineux en ligne » et « Protéger la liberté de la presse ».

Meta a participé aux discussions politiques en marge du Sommet de l'avenir et de la 79e Assemblée générale des Nations unies. Exemples de sujets évoqués : le rôle de l'IA dans la gouvernance mondiale, l'autonomisation des créateurs digitaux, l'innovation économique menée par la diaspora, ou encore l'impact de la cybercriminalité et des lois relatives au contenu sur la liberté d'expression. Nous avons également participé à des discussions concernant la protection des défenseurs des droits humains et l'utilisation des réseaux sociaux pour fournir des informations capitales en période de crises humanitaires.

De plus, nous avons consulté les procédures spéciales des NU en matière de droits humains (des spécialistes indépendants en droits humains), notamment les rapporteurs spéciaux sur la liberté d'opinion et d'expression et sur les défenseurs des droits humains, entre autres.

Tout au long de l'année, Meta a collaboré avec le G7, le G20, l'UNESCO et l'OCDE sur des flux de travail liés à la gouvernance et à l'inclusion de l'IA. Nous avons aussi participé à des conversations avec les autorités gouvernementales à propos de l'importance de l'intégrité des informations. Nous avons poursuivi notre participation à la Coalition mondiale pour la sécurité digitale du Forum économique mondial, qui a abouti à la publication de The Intervention Journey: A Roadmap to Effective Digital Safety Measures (Le parcours de l'intervention : une feuille de route pour des mesures efficaces en matière de sécurité digitale).

De plus, nous avons participé de manière active et interagi avec des parties prenantes dans plusieurs forums multipartites, notamment le Sommet pour l'éradication de la haine, le Forum sur la liberté d'Internet en Afrique (FIFAfrica), le Forum mondial sur Internet pour lutter contre le terrorisme, le Forum sur la gouvernance de l'Internet (IGF), RightsCon, la Conférence Tech Against Trafficking et la Commission des Nations unies sur la condition de la femme.

Par le biais de notre adhésion à la Global Network Initiative et de notre participation au Digital Trust and Safety Partnership, Meta a assisté à la conférence « European Rights and Risks: Stakeholder Engagement Forum », qui a permis d'éclairer nos évaluations des risques systémiques dans le cadre de la Législation sur les services numériques.

Transparence et résolution

Le Conseil de surveillance, en tant qu'organisme indépendant, nous aide à résoudre certaines des questions les plus complexes qui soient en matière de liberté d'expression en ligne : quel contenu retirer, quel contenu conserver, et pourquoi. Il examine des cas transmis par Meta ou qui font l'objet d'un appel par des particuliers sur Facebook, Instagram ou Threads, car ils sont en désaccord avec nos décisions en matière de modération de contenu. Il rend aussi des décisions contraignantes indiquant si le contenu doit être retiré ou conservé. Le Conseil de surveillance fournit également des recommandations pour renforcer nos pratiques relatives à la modération de contenu et offre des avis consultatifs en matière de politique sur simple demande.

Où trouver des informations sur l’impact du Conseil de surveillance

En 2024, nous sommes passés d’un rythme trimestriel à un rythme semestriel pour les rapports sur les cas que Meta a transmis au Conseil de surveillance et les mises à jour sur nos progrès dans l’implémentation de ses recommandations. De plus, nous avons lancé une page Espace modération qui suit l’impact des recommandations du Conseil de surveillance. Elle vient s’ajouter à notre page des recommandations du Conseil de surveillance, où nous décrivons les recommandations liées à un cas transmis par le Conseil de surveillance, notre niveau d’engagement ainsi que le statut de l’implémentation.

Actions associées aux recommandations du Conseil de surveillance en 2024

Recommandations émises par le Conseil de surveillance

48

(66 en 2023)

Évaluation et/ou implémentation par Meta en cours[5]

70

(69 en 2023)

Recommandations implémentées[5]

41

(61 en 2023)

En 2024, le Conseil de surveillance a pris en considération des cas mentionnant la mise en application de nos règles en matière de contenu à la lumière du cadre international des droits humains, notamment la liberté d'expression, le droit à la santé, ainsi que le droit à l'égalité et à la non-discrimination, pour ne citer que ces sujets. Voici quelques exemples illustrés des mesures que nous avons prises en réponse à des décisions du Conseil de surveillance en 2024. Vous trouverez plus de détails dans les Rapports semestriels de Meta sur le Conseil de surveillance.

Voici quelques exemples de décisions prises par le Conseil de surveillance en 2024 :

Le Conseil de surveillance a annulé les décisions de Meta de supprimer trois publications Facebook montrant des images de l'attentat terroriste de Moscou en mars 2024, exigeant que le contenu soit rétabli avec des écrans d'avertissement « Marquer comme dérangeant ». Le Conseil de surveillance a constaté que même si les publications enfreignaient la politique de Meta relative à la diffusion d'attaques ciblées contre des victimes visibles, leur suppression n'était pas conforme aux responsabilités de l'entreprise en matière de droits humains.

Le Conseil de surveillance a confirmé la décision de Meta de supprimer la vidéo d'une personnalité politique pakistanaise livrant un discours dont le texte affirme que celle-ci « dépassait toutes les limites de la loyauté » et utilise le mot « kufr » pour suggérer un blasphème, en raison du risque de préjudice hors ligne.

[5] Certaines évaluations et/ou implémentations en cours ou recommandations entièrement mises en œuvre incluent des recommandations des années précédentes (voir notre Rapport sur les droits humains 2023 pour en savoir plus).

Exemples de mesures prises à la suite de recommandations du Conseil de surveillance :

Suite à une série de recommandations (p. ex., ici) sur l'IA, nous avons apporté des modifications à notre manière de traiter le contenu généré par l'IA, notamment la mise à jour des étiquettes et politiques, telles que notre politique relative à la désinformation.

Suite à des recommandations provenant du Conseil de surveillance sur la politique relative au contenu, Meta a modifié la politique relative aux organisations et personnes dangereuses pour autoriser les contenus utilisant le terme « shaheed » dans toutes les langues qui utilisent ce terme, sauf lorsque ces contenus sont accompagnés de signaux de violence ou enfreignent nos politiques (p. ex., en glorifiant des personnes désignées comme dangereuses).

En 2024, le Conseil de surveillance a également expérimenté les délais plus courts dans des cas urgents. Par exemple, suite à l'élection présidentielle de juillet au Venezuela qui a donné lieu à une flambée de la violence, nous avons transmis deux contenus concernant les « Colectivos » en vue d'un examen rapide. « Colectivos » est un terme générique utilisé pour désigner des gangs armés irréguliers ou des groupes de type paramilitaire étroitement liés au gouvernement. Les cas présents ont été décidés dans un délai accéléré de 14 jours.

Nous avons également collaboré avec le Conseil de surveillance afin de mobiliser des régulateurs et des organisations de la société civile, notamment en Afrique, en Amérique latine, au Moyen-Orient et en Turquie, afin de mieux faire connaître le mandat du Conseil et son processus de sélection des cas.

Annexe

Comment les droits humains sont régis et gérés chez Meta

Nos spécialistes en droits humains guident l'implémentation de notre Politique d'entreprise en matière de droits humains, qui est supervisée par le président des affaires mondiales (maintenant chef 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. En 2024, le directeur des droits humains a présenté un rapport au comité d'audit et de surveillance des risques du conseil d'administration.

En 2024, Meta a lancé le secteur des risques liés aux droits humains du programme de gestion des risques tiers de l'entreprise. Ce contrôle témoigne de notre engagement visant à améliorer en permanence notre gestion des risques en matière de droits humains et à nous efforcer d'interagir avec des tiers qui soient responsables et respectueux des 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. Notre formation aux droits humains 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 toutes les personnes qui utilisent nos services.

Nous avons lancé notre formation *Bigger than Meta: Human Rights* (Au-delà de Meta : les droits humains) en 2022, et elle s'est poursuivie tout au long de l'année 2024. Notre formation relative à la confidentialité soutient également nos objectifs de la formation en matière de droits humains. Il s'agit de développer notre capacité collective à protéger les personnes, y compris, en particulier, les catégories de personnes marginalisées, contre les préjudices découlant du traitement de leurs données.

Liens vers les rapports référencés

Rapport sur les pratiques professionnelles responsables 2025

Rapport sur la durabilité 2025

Rapport sur les droits humains 2023, Rapport sur les droits humains 2022, Rapport sur les droits humains 2021

Rapport 2024 sur la lutte contre l'esclavage et la traite d'êtres humains

Rapport sur les minerais de conflit 2024

Rapports de transparence de Meta

Rapports réglementaires et autres rapports de transparence

Précédentes publications des Évaluations de l'impact sur les droits humains : Chiffrement de bout en bout, Philippines, Myanmar, Indonésie, Cambodge, Inde, Sri Lanka, et Israël et Palestine


Meta