تقرير حقوق الإنسان

الرؤى والإجراءات لعام 2024

Meta

جدول المحتويات

حول هذا التقرير 3

الملخص التنفيذي 6

إدارة مخاطر حقوق الإنسان 11

1. حرية الرأي والتعبير 12
2. الخصوصية 13
3. المساواة وعدم التمييز 14
4. الحياة والحرية والأمن الشخصي 15
5. مصالح الطفل الفضلى 16
6. المشاركة العامة والحق في التصويت والحق في الترشح للانتخابات 16
7. حرية تكوين الجمعيات والتجمّع 17
8. الحق في الصحة 17

تسريع الابتكار في الذكاء الاصطناعي مع احترام حقوق الإنسان 19

أبرز القضايا 25

2024: عام الانتخابات 25
الاستعداد للانتخابات على نطاق واسع 26
إدارة مخاطر تأثير الذكاء الاصطناعي 26
جهود أخرى لضمان نزاهة الانتخابات 27
الاستعداد للانتخابات عالية المخاطر 29
أمثلة من الانتخابات الوطنية 29
الولايات المتحدة 29
المكسيك 30
الهند 31
انتخابات البرلمان الأوروبي 32
سلامة الأطفال والشباب 33
وسائل حماية مدمجة للمراهقين 33
مكافحة الابتزاز الجنسي 36
كيفية الاستعداد للأزمات والاستجابة لها 37
السودان 39
الشرق الأوسط 41
بنجلاديش 42
جورجيا 43
الأمن السيبراني 44

التفاعل مع أصحاب المصلحة 46

منتديات المجتمع 51
الشركاء الموثوق بهم 52
دراسة حالة: إعداد التقارير المستندة إلى رؤى من سوريا 54
دراسة حالة: تخفيف حدة المخاطر التي تواجه الفاعلين المدنيين في فنزويلا 55
دراسة حالة: تصدي الشركاء الموثوق بهم لاتهامات الكفر والخطاب العدائي في باكستان 56
المنظمات الدولية 58

الشفافية وسبل الانتصاف 60

الملحق 64

كيفية حوكمة حقوق الإنسان والتعامل معها في Meta 65
تدريب موظفي Meta في مجال حقوق الإنسان 65
روابط إلى تقارير مرجعية 65

حول هذا التقرير

يغطي هذا التقرير السنوي لحقوق الإنسان الرؤى والإجراءات من 1 يناير 2024 حتى 31 ديسمبر 2024. كما أننا نتناول في التقرير خدمات Meta ومنتجاتها، بما في ذلك فيسبوك وMessenger وInstagram وواتساب وThreads وReality Labs.

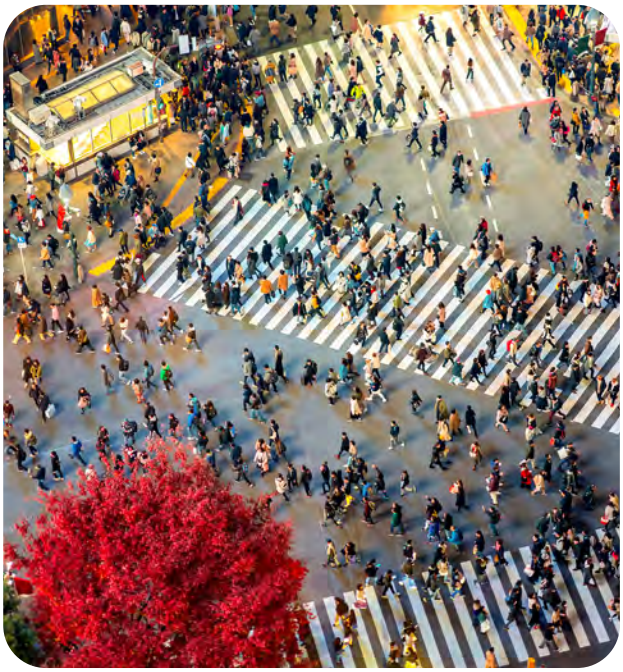
ويأتي التقرير بالبناء على جهود Meta لاحترام حقوق الإنسان، كما يعكس التقدم المحرز في تنفيذ التزاماتنا بمبادئ الأمم المتحدة التوجيهية بشأن الأنشطة التجارية وحقوق الإنسان وسياسة حقوق الإنسان الخاصة بشركتنا. يُظهر التقرير كيفية تطبيق هذه المبادئ في جميع أنحاء الشركة خلال عام 2024، ويوفر إرشادات حول مكان العثور على معلومات أكثر تفصيلاً.

يعتمد محتوى هذا التقرير على التقييم الشامل للمخاطر الجوهرية المتعلقة بحقوق الإنسان، الذي أُجري في عام 2022. وقد كان الغرض من ذلك التقييم تحديد وترتيب أولويات أبرز التأثيرات السلبية المحتملة في مجال حقوق الإنسان[1] والواقعة على الأشخاص الذين يستخدمون منتجاتنا وغيرهم ممن قد يتأثرون بإجراءاتنا. يستعرض هذا التقرير هذه المخاطر الجوهرية المحتملة وأمثلة على إجراءاتنا وتدابير التخفيف المتخذة في عام 2024[2].

ظل موضوع حقوق الإنسان محل أهمية حاسمة لشركتنا ولأصحاب المصلحة في عام 2024. نسعى لتقديم صورة تمثيلية لعملنا عبر فرق متعددة ومشاركات أصحاب المصلحة حول العالم.

السياسات والتقدم

بالإضافة إلى تقرير حقوق الإنسان هذا، تقوم Meta بالإبلاغ سنويًا عن السياسات والتقدم باستخدام الآليات التالية:



- التقرير السنوي
- بيان الوكيل
- تقرير الممارسات التجارية المسؤولة
- مركز الشفافية
- تقرير الاستدامة
- تقرير CDP حول تغير المناخ
- الميثاق العالمي للأمم المتحدة

يأتي هذا التقرير كتكملة لأحدث تقرير لشركة Meta عن الممارسات التجارية المسؤولة. ونُعدّ تقارير بصورة منفصلة عن جهودنا لتحديد والتخفيف من حدة مخاطر العبودية الحديثة والاتجار بالبشر في عملياتنا التجارية وسلاسل التوريد. بالإضافة إلى ذلك، نمتثل لمتطلبات التقارير الوطنية والأوروبية الإلزامية، والمتاحة في مركز الشفافية. وتتوفر روابط إلى عمليات إفصاح أخرى من Meta في الملحق المرفق بهذا التقرير.

الانتقال إلى الملحق

[1] يتماشى مصطلح "التأثير السلبي في مجال حقوق الإنسان" مع مبادئ الأمم المتحدة التوجيهية بشأن الأعمال التجارية وحقوق الإنسان، ويُقصد به التأثير الذي ينشأ عند اتخاذ إجراء من شأنه تقويض أو الحد من قدرة الفرد على التمتع بحقوق الإنسان المكفولة له.

[2] ولا يشمل ذلك تغييرات سياسة المحتوى وغيرها التي أعلنّا عنها في يناير 2025، عندما حدّثنا سياسة السلوك الذي يحض على الكراهية، المعروفة سابقًا باسم سياسة خطاب يحض على الكراهية، لمعالجة المخاوف المتعلقة بالمبالغة في الإنفاذ والسعي للسماح بمزيد من حرية التعبير.

تُطبَّق سياسة حقوق الإنسان الخاصة بالشركة على مستوى المؤسسة. وتوجد سياسات وإجراءات خاصة لكل خدمة وكيان في Meta، والتي قد يكون لها آثار مختلفة على حقوق الإنسان. ويشير هذا التقرير إلى الإجراءات التي اتخذتها Meta كشركة تجاه كيان أو أكثر من كيانات Meta. ولا يُقصد من البيانات الإشارة ضمنًا إلى أن Meta قد اتخذت نفس الإجراء فيما يتعلق بجميع الكيانات و/أو في جميع الظروف.[3]

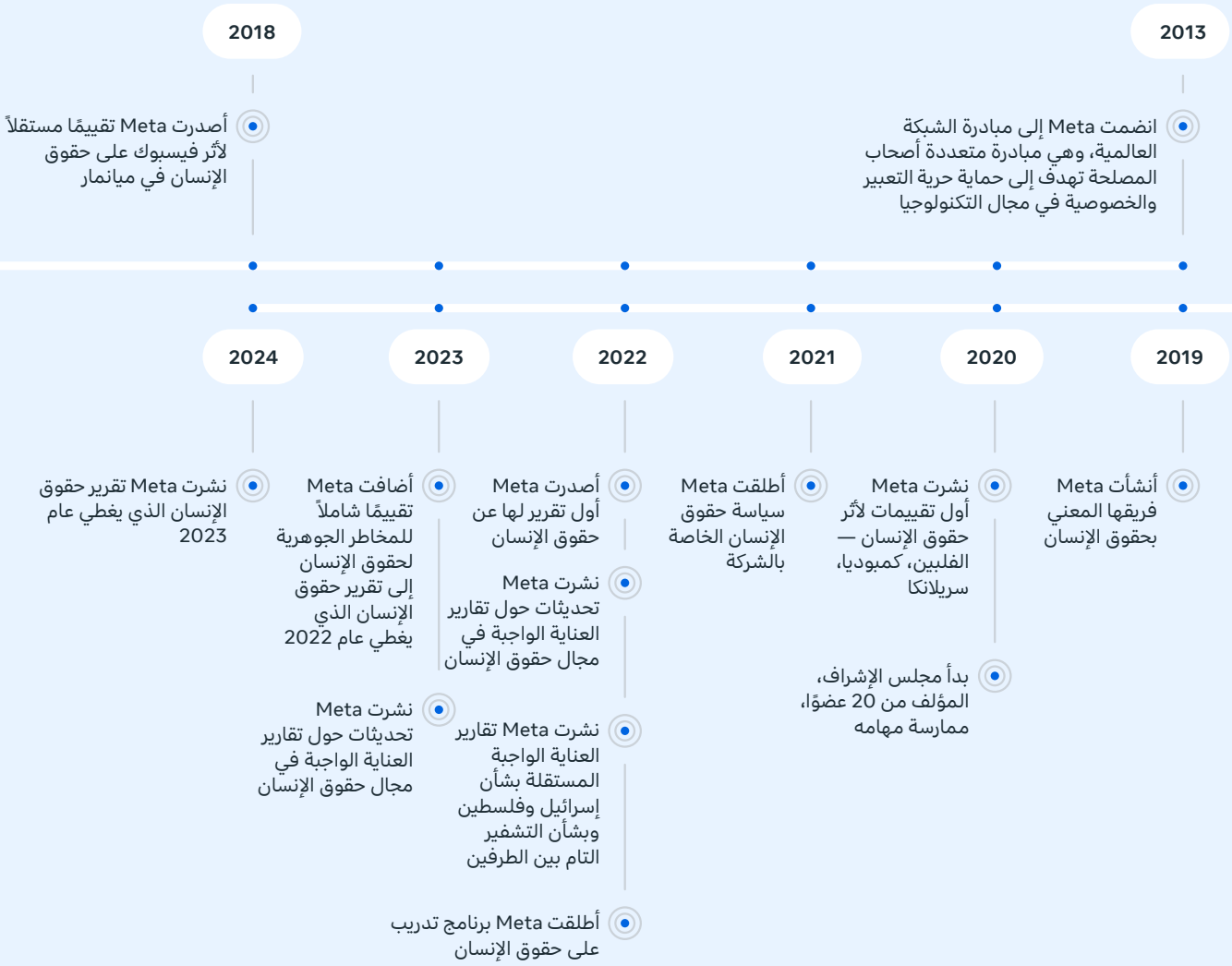
[3] إن المناقشة الواردة في هذا التقرير بخصوص الإشراف على المحتوى والإجراءات ذات الصلة على فيسبوك وInstagram لا تنطبق على واتساب، كما لا تنطبق على واتساب أي سياسات أو إجراءات معينة ما لم يُذكر أنها تنطبق على واتساب بالفعل. علاوةً على ذلك، وفي حين أن العديد من الإجراءات الموضحة في هذا التقرير تنطبق على فيسبوك وInstagram، إلا أن هناك اختلافات مقصودة في السياسات والإجراءات بين الخدمات. فإذا جرى تصنيف إحدى السياسات على أنها سياسة "فيسبوك"، فقد لا تنطبق على Instagram. ولا يهدف أي بيان في هذا التقرير إلى فرض - أو ينبغي ألا يُفسَّر على أنه يفرض - التزامات جديدة (سواء قانونية أو غير ذلك) تتعلق بتطبيق سياسة أو إجراء على أي خدمات أو كيانات أخرى.

الملخص التنفيذي

هذا هو التقرير السنوي الرابع لحقوق الإنسان من Meta. يوفر التقرير رؤى حول الجهود التي بذلتها Meta في عام 2024 لإدارة مخاطر حقوق الإنسان على نطاق واسع والوفاء بالتزاماتنا تجاه مبادئ الأمم المتحدة التوجيهية بشأن الأعمال التجارية وحقوق الإنسان (UNGPs).

الجدول الزمني لحقوق الإنسان

يوضح الرسم البياني التالي رحلتنا في مجال حقوق الإنسان وكيف تطور عملنا منذ اعتماد المبادئ التوجيهية للأمم المتحدة من قبل مجلس حقوق الإنسان التابع للأمم المتحدة في عام 2011.



تقييم المخاطر الجوهرية

عكست أولوياتنا في عام 2024 المخاطر الجوهرية التي تم تحديدها في التقييم الشامل للمخاطر الجوهرية المتعلقة بحقوق الإنسان لعام 2022، وهي: حرية الرأي والتعبير، والخصوصية، والمساواة وعدم التمييز، والحياة والحرية والأمن الشخصي، ومصالح الطفل الفضلى، والمشاركة العامة، والحق في التصويت وفي الترشح للانتخابات، وحرية تكوين الجمعيات والتجمُّع، والحق في الصحة.

قراءة المزيد

تسريع الابتكار في مجال الذكاء الاصطناعي

شهدت التطورات في مجال الذكاء الاصطناعي (AI) تسارعًا في عام 2024. رؤيتنا هي بناء ذكاء فائق شخصي وجعله متاحًا على نطاق واسع بحيث يستفيد الجميع منه.

أصبحت تطبيقات الذكاء الاصطناعي التوليدي أكثر استخدامًا، وبدأت بشكل متزايد في تحويل طرق تواصلنا وتعلمنا وإبداعنا وعملنا. واصلنا تعزيز نهج مفتوح للذكاء الاصطناعي يمكن أن يدعم حقوق الإنسان. ساهم هذا النهج في تمكين الأشخاص من الوصول إلى المعلومات وحرية التعبير، وكذلك تعزيز حقوق المساواة وعدم التمييز، بما في ذلك من خلال تحسين إمكانية الوصول وتوسيع شمولية اللغات.

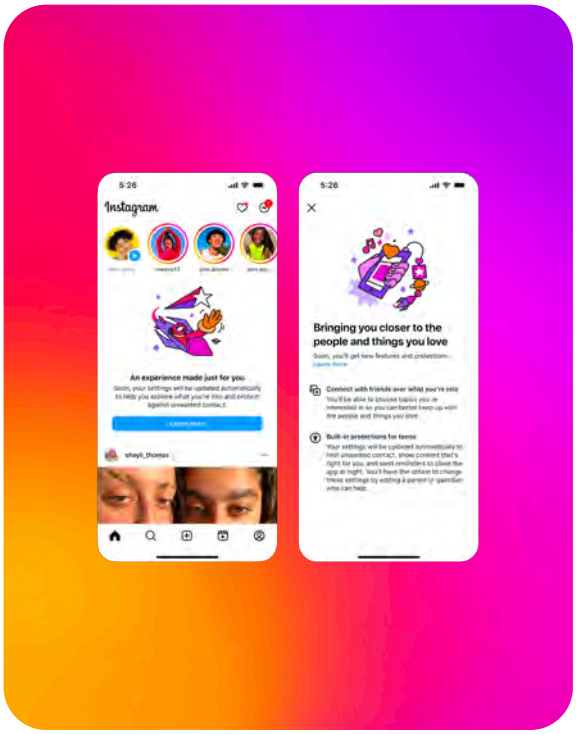
قراءة المزيد

2024: عام الانتخابات

كان عام 2024 هو أكبر عام انتخابي في التاريخ. أجرت أكثر من 70 دولة، تمثل أكثر من نصف سكان العالم، انتخابات وطنية. وكان نحو 2 مليار شخص مؤهلين للتصويت. ركزنا على تمكين حقوق حرية التعبير والمشاركة في العمليات السياسية والوصول إلى المعلومات للأشخاص في الدول التي تجري فيها الانتخابات.

نضج نهجنا على مدى مئات الانتخابات في السنوات الأخيرة. وشمل هذا الجهود المتعلقة بإدارة مخاطر الذكاء الاصطناعي وتطبيق سياساتنا المتعلقة بالتدخل في الإدلاء بالأصوات أو التعداد السكاني، وتعطيل الشبكات العدائية، وزيادة شفافية الإعلانات السياسية، وربط الناخبين بالمعلومات الموثوقة. يحتوي هذا التقرير على أمثلة من الولايات المتحدة والمكسيك والهند والاتحاد الأوروبي.

قراءة المزيد



سلامة الأطفال والشباب

استمر التزامنا بسلامة الأطفال والشباب. من بين المبادرات الأخرى، شمل عملنا في عام 2024 إطلاق حسابات المراهقين على Instagram، وهي تجربة جديدة للمراهقين بإشراف الوالدين. تتمتع حسابات المراهقين بحماية مدمجة تحد من الأشخاص الذين يمكنهم التواصل مع المراهقين والمحتوى الذي يمكنهم الاطلاع عليه، بالإضافة إلى أدوات تساعد في إدارة الوقت الذي يقضيه المراهقون على التطبيق، مع توفير طرق جديدة لاستكشاف اهتماماتهم. وازنّت جهودنا بين دعم تعزيز استقلالية الشباب وبين حقوق وواجبات الوالدين أو الأوصياء. تم تطوير هذه المبادرات وفقًا لتوجيهات الخبراء ومبدأ القدرات المتطورة للطفل المنصوص عليه في اتفاقية الأمم المتحدة لحقوق الطفل.

قراءة المزيد

الاستجابة للأزمات

واصلنا دمج مبادئ حقوق الإنسان في كيفية استعدادنا واستجابتنا للأزمات. يوجهنا بروتوكول سياسة الأزمات في استخدام أدواتنا بسرعة لتخفيف الأضرار المحتملة. في عام 2024، قمنا بتعيين 19 حالة عالمية تحت بروتوكول سياسة الأزمات. في هذا التقرير، نوفر أمثلة على استجابتنا للأزمات في بنجلاديش، وجورجيا، والشرق الأوسط والسودان.

قراءة المزيد

التفاعل مع أصحاب المصلحة

تدعم سياسة حقوق الإنسان الخاصة بالشركة تفاعلنا الاستباقي مع أصحاب المصلحة. في عام 2024، تواصلنا مع مجموعة واسعة من أصحاب المصلحة لإثراء نهج الشركة تجاه القضايا المتعلقة بالتعبير، والمحتوى المفعم بالكراهية، والمعلومات المضللة، والخصوصية. وشملت هذه المجموعة نطاقًا متنوعًا من منظمات حقوق الإنسان، والمجتمعات الضعيفة، وأعضاء المجتمع المدني، والأكاديميين، ومراكز الفكر، والهيئات التنظيمية. وشملت الموضوعات الرئيسية نهجنا تجاه الذكاء الاصطناعي المسؤول ونزاهة الانتخابات، بالإضافة إلى إشارات التعيين للمنظمات والأفراد الخطرين، والأحداث العنيفة.

في عام 2024، أجرينا ستة منتديات للسياسات، حيث يشارك خبراء الموضوع من Meta وجهات نظر مختلفة ويناقشون التغييرات المحتملة في معايير المجتمع ومعايير الإعلانات. كما عقدنا منتديات المجتمع للاستفادة من آراء الجمهور حول القضايا التي تتضمن توازنات متضاربة ولا توجد لها إجابات واضحة. ساهمت هذه المنتديات في تحسين المنتجات وتوقع المخاطر المحتملة للتقنيات الناشئة، وتمكين الأصوات خارج الشركة من المشاركة بشكل أكبر في عملية اتخاذ القرار.

واصلنا التواصل مع الشركاء الموثوق بهم حول العالم للمساعدة في تحديد المواضيع الرائجة وفهم أفضل لتأثير المحتوى والسلوك عبر الإنترنت على المجتمعات المحلية. كما بحثنا طرق تعزيز قنوات التصعيد ذات الصلة. كانت خبرتهم ذات قيمة خاصة خلال دورة الانتخابات المكثفة لعام 2024، وفي حالات الاضطرابات الشديدة. كما قدموا رؤى وحددوا المحتوى المحتمل مخالفته في بنجلاديش والبرازيل وكوت ديفوار وجمهورية الكونغو الديمقراطية وفرنسا واليونان والهند وإندونيسيا وكينيا وكردستان-العراق والمكسيك ونيجيريا وباكستان والسنغال وجنوب أفريقيا وسوريا وفنزويلا، من بين دول ومناطق أخرى.

قراءة المزيد

مجلس الإشراف

في عام 2024، نظر مجلس الإشراف في قضايا تتعلق بجهودنا لاحترام حقوق الإنسان، بما في ذلك حرية التعبير، والحق في الصحة، والحق في المساواة وعدم التمييز، من بين مواضيع أخرى. يعد مجلس الإشراف هيئة مستقلة تقوم بمراجعة القضايا التي تُحال إليها من قبل Meta أو التي يطعن عليها الأفراد على فيسبوك أو Instagram أو Threads الذين يختلفون مع قراراتنا المتعلقة بالإشراف على المحتوى. ويصدر المجلس قرارات ملزمة بشأن ما إذا كان يجب إزالة المحتوى أو تركه. استجابةً لتوصية من مجلس الإشراف، قامت Meta بتقييم مدى سرعة وفعالية الاستجابات للمحتوى المبلغ عنه من خلال برنامج الشركاء الموثوق بهم.

قراءة المزيد

إدارة الطلبات الحكومية

طوال العام، واصلنا التوجيه بناءً على التزامنا تجاه المبادرة العالمية للشبكات باحترام حرية التعبير والخصوصية، بما في ذلك عند الاستجابة لطلبات الحكومات لتقييد المحتوى. في عام 2024، نشرنا دراسات حالات تتعلق بالخطاب السياسي في البرازيل وألمانيا والهند والعراق وإسرائيل وسنغافورة وتركيا.

عرض دراسات الحالة

إدارة مخاطر حقوق الإنسان

توضح مبادئ الأمم المتحدة التوجيهية بشأن الأعمال التجارية وحقوق الإنسان أن على الشركات تحديد الآثار السلبية المحتملة على حقوق الإنسان من أجل منعها أو التخفيف منها بفعالية.

نظرًا لحجم عمليات Meta ومدى الحقوق التي قد تتأثر، فإن توقع وإدارة المخاطر الجوهرية أمر مهم ولكنه معقد. ندير نوعين من المخاطر الكامنة: تلك الناتجة عن أنشطتنا الخاصة وتلك الناتجة عن أنشطة جهات خارجية، بما في ذلك الأشخاص الذين يستخدمون منصاتنا.

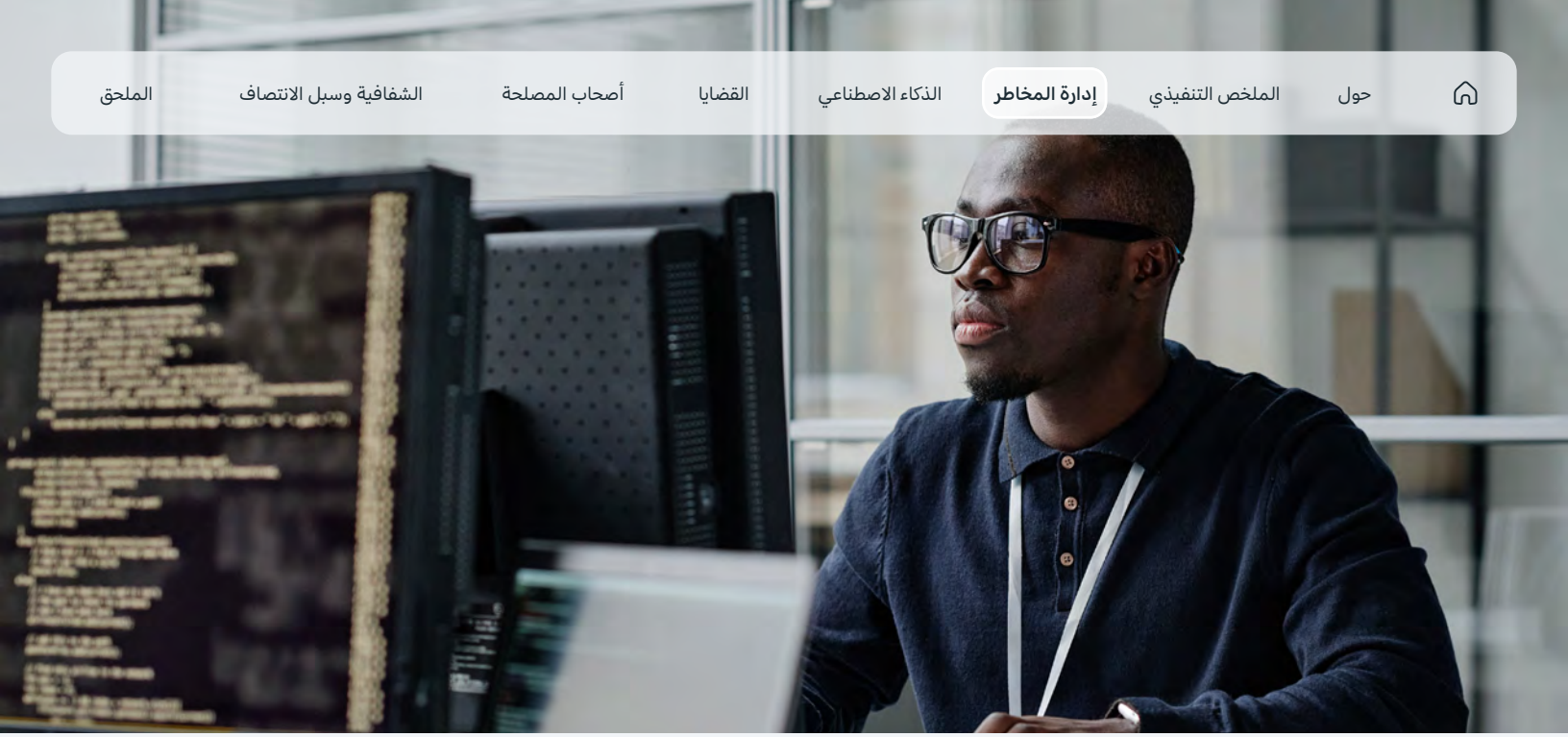
بمجرد تنفيذ العمليات لمعالجة هذه المخاطر، سيظل مستوى معين من المخاطر قائمًا دائمًا. ويُعرف هذا باسم المخاطر "المتبقية". على الرغم من وجود المخاطر المتبقية في جميع أنظمة إدارة المخاطر، فإن المخاطر المرتبطة بالتقنيات الرقمية وتأثيرها على حقوق الإنسان تستمر بسبب الطبيعة الديناميكية والمتطورة بسرعة للتقنيات الرقمية وارتفاع مستوى نشاط الجهات الخارجية.

يوضح الجدول في الصفحات التالية المخاطر الجوهرية المتعلقة بحقوق الإنسان كما حُددت في التقييم الشامل للمخاطر الجوهرية المتعلقة بحقوق الإنسان (CSRA) لعام 2022، الذي كشفنا عنه في تقرير حقوق الإنسان لعام 2022. يوفر هذا الجدول أمثلة توضيحية على كيفية معالجتنا للمخاطر المحتملة في عام 2024. ولاحقًا في هذا التقرير، نتعمق أكثر في بعض هذه الأمثلة وكيفية إدارتنا للمخاطر المحتملة المتعلقة بالذكاء الاصطناعي (AI) والانتخابات والصراعات.

1. حرية الرأي والتعبير

يشمل الحق في حرية الرأي والتعبير الحق في البحث عن المعلومات وتلقيها ومشاركتها من جميع الأنواع. إنه حق أساسي، ضروري لحماية كرامة الإنسان واستقلالية الفرد والديمقراطية. تُعد حرية التعبير جزءًا جوهريًا من مهمتنا، بما يتوافق مع قيمتنا في منح الجميع صوتًا.

أمثلة على المخاطر البارزة المتأصلة المحتملة في مجال حقوق الإنسان التي تم تحديدها في تقييم CSRA	أمثلة على كيفية معالجة Meta للمخاطر المحتملة في 2024
قد يؤدي الإنفاذ وسياسات الإشراف على المحتوى في Meta إلى تقييد حرية التعبير.	واصلنا تطوير سياساتنا مع وضع حرية التعبير كمرشد رئيسي لنا. في عام 2024، أجرينا عدة منتديات للسياسات سعت إلى تطوير فهم دقيق لتحديات حرية التعبير في مجموعة من المجالات.
القيود الحكومية المفرطة على المحتوى	نسعى إلى الالتزام بتعهداتنا في إطار مبادرة الشبكة العالمية (GNI). وشمل ذلك الإبلاغ عن استجاباتنا لطلبات الحكومات المتعلقة بالحصول على البيانات أو تقييد المحتوى (انظر هنا وهنا). ويَرِد شرح نهجنا في التعامل مع الطلبات الحكومية بالتفصيل في تقرير حقوق الإنسان لعام 2023. وبالنسبة إلى الحالات التي نعتقد فيها عدم صحة الطلبات الحكومية أو أوامر المحكمة، من الناحية القانونية أو إذا كانت فضفاضة للغاية أو غير متسقة مع المعايير الدولية لحقوق الإنسان، فقد نطلب توضيحًا أو نقدم طعنًا أو نتوقف عن اتخاذ أي إجراء. وفي عام 2024، شملت دراسات حالة بارزة في مجال الشفافية تتعلق بالخطاب السياسي حالات في البرازيل وألمانيا والهند والعراق وإسرائيل وسنغافورة وتركيا.
تمنع حالات انقطاع الإنترنت وحظر وسائل التواصل الاجتماعي الأشخاص من ممارسة حقهم في حرية التعبير وتمنعهم من تلقي وإرسال الأخبار والمعلومات الحيوية.	لمنع حظر وسائل التواصل الاجتماعي وخدمات المراسلة، قد نلتزم بالطلبات الحكومية المشروعة قانونًا، مع السعي في الوقت ذاته إلى احترام التزاماتنا ضمن مبادرة الشبكة العالمية المتعلقة باحترام حرية التعبير. كما نواصل إتاحة ميزة الاتصال عبر خادم وسيط للأشخاص الذين لا يستطيعون الاتصال بتطبيقاتنا بشكل مباشر.



2. الخصوصية

يُشكل الحق في الخصوصية شرطًا ضروريًا لإعمال حقوق الإنسان الأخرى، مثل حرية التعبير وحرية التجمّع وتكوين الجمعيات وحرية المعتقد والدين. أحد المبادى الأساسية المنصوص عليها في سياسة حقوق الإنسان الخاصة بالشركة لدينا هو الحفاظ على سلامة الأفراد وحماية الخصوصية.

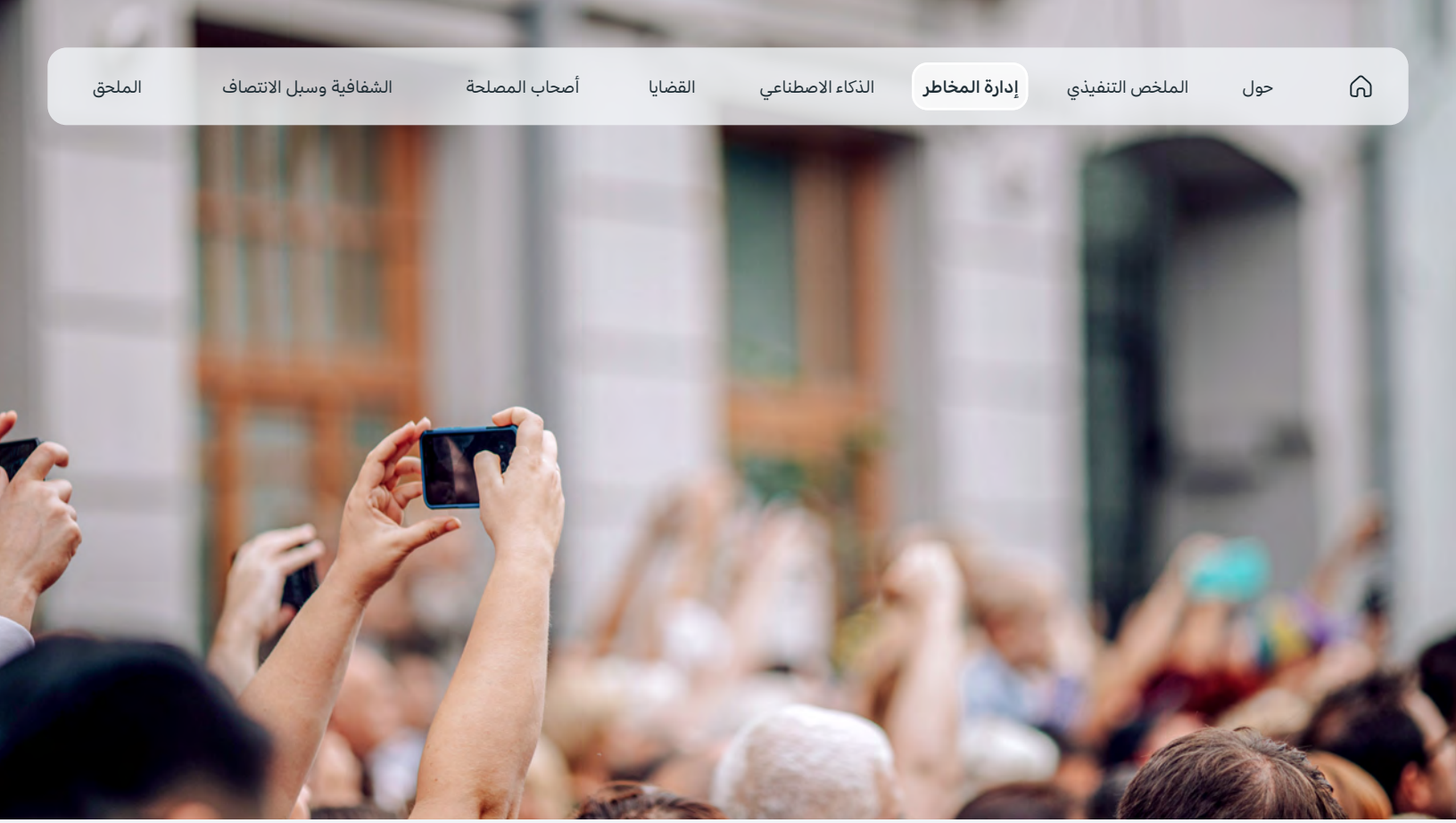
أمثلة على المخاطر البارزة المتأصلة المحتملة في مجال حقوق الإنسان التي تم تحديدها في تقييم CSRA	أمثلة على كيفية معالجة Meta للمخاطر المحتملة في 2024
قد تنطوي نماذج الذكاء الاصطناعي التوليدي على معالجة بيانات شخصية بطرق قد لا يتوقعها الأشخاص أو لا يفهمونها.	نحن نتحلى بالشفافية بشأن كيفية استخدام Meta للمعلومات في نماذج وميزات الذكاء الاصطناعي التوليدي، ولدينا عملية داخلية لمراجعة الخصوصية لضمان الاستخدام المسؤول للبيانات، بما في ذلك الذكاء الاصطناعي التوليدي. يتوفر تحديث حول تقدمنا في مجال الخصوصية خلال عام 2024 هنا وهنا.
المحتوى أو السلوك الموجود في تطبيقات Meta الذي قد يؤثر سلبًا في حقوق الخصوصية وحماية البيانات.	وفي أكتوبر 2024، أعادت Meta إدخال تقنية التعرّف على الوجوه في فيسبوك وInstagram لمساعدة الأشخاص على استعادة الحسابات التي تم اختراقها ومنع عمليات الخداع التي تتضمن تأييدات مزيفة من مشاهير. ولتحقيق التوازن بين مخاطر الخصوصية المحتملة ومخاطر النزاهة، نتيح للشخصيات العامة، التي يُساء استخدام صورها أو هيئاتها في عمليات الخداع، خيار المشاركة في هذا البرنامج أو إلغاء الاشتراك فيه.

قراءة المزيد

3. المساواة وعدم التمييز

ينص الحق في المساواة وعدم التمييز على توفير الحماية المتساوية ضد التمييز. كجزء من احترام هذا الحق، لا نسمح بالسلوك الذي يحض على الكراهية على منصاتنا، وذلك على النحو المعرّف في سياستنا.

أمثلة على المخاطر البارزة المتأصلة المحتملة في مجال حقوق الإنسان التي تم تحديدها في تقييم CSRA	أمثلة على كيفية معالجة Meta للمخاطر المحتملة في 2024
بعض اللغات واللهجات التي قد يكون الإشراف على المحتوى فيها أكثر صعوبة من غيرها.	قمنا بتصميم ونشر آليات جديدة لتوجيه المحتوى العربي بحسب اللهجات، بما يتيح إشرافًا أكثر كفاءة ودقة، بما في ذلك في السودان. ويقوم النظام الجديد باكتشاف المحتوى وتحديد أولويته وتوجيهه إلى المشرفين الأكثر قدرة على فهم تلك اللهجة العربية المحددة.
المحتوى الذي يؤثر سلبًا في المساواة وعدم التمييز (مثل سلوك يحض على الكراهية)	استنادًا إلى الأبحاث، والتفاعلات الخارجية، والتحقيقات التي أُجريت على منصاتنا، قمنا بتحديث سياسة السلوك الذي يحض على الكراهية لدينا فيما يتعلق بالمحتوى الذي يهاجم "الصهاينة". وخلال تدريب نماذج الذكاء الاصطناعي لدينا، اختبرنا بيانات التدريب للكشف عن أي محتوى أو خصائص قد تزيد من مخاطر توليد محتوى ضار محتمل، مثل مدى تمثيل مجموعة البيانات لمختلف المعلومات الديموغرافية.



4. الحياة والحرية والأمن الشخصي

يتعلق الحق في الحياة والحرية والأمن الشخصي بالتحرر من الضرر البدني والاحتجاز. بالنسبة إلى Meta، فإن احترام هذا الحق من حقوق الإنسان يشمل التخفيف من خطر تسبُّب المحتوى في أي ضرر، بما في ذلك مخاطر العنف والاتجار بالبشر، والتهديدات عبر الإنترنت برعاية الدولة، والجماعات غير الحكومية المنخرطة في أعمال عنف أو كراهية أو تدعو إليها.

أمثلة على المخاطر البارزة المتأصلة المحتملة في مجال حقوق الإنسان التي تم تحديدها في تقييم CSRA

المحتالون الذين يقومون بما يلي:

- استغلال خدمات وتطبيقات Meta لتنسيق ضرر على الإنترنت أو على أرض الواقع
- إساءة استخدام الخدمات والتطبيقات لإجراء هجمات إلكترونية أو عمليات تصيد احتيالي
- تهديد ومضايقة المدافعين عن حقوق الإنسان والناشطين وغيرهم من الفئات المستضعفة

أمثلة على كيفية معالجة Meta للمخاطر المحتملة في 2024

تحظر سياسة Meta بشأن تنسيق أنشطة ضارة وترويج ارتكاب الجرائم تسهيل أو تنظيم أو الترويج أو الإقرار بممارسة أنشطة إجرامية أو ضارة معينة. في عام 2024، قدّمنا إرشادات ضمن هذه السياسة بشأن أسرى الحرب، بما يمكّن مراجعي المحتوى من إزالة المحتوى المخالف على نطاق واسع بصورة أفضل، بما في ذلك في السودان.

كما واصلنا دعم صندوق المدافعين عن حقوق الإنسان، وأعدنا تصميم برنامج الشركاء الموثوق بهم لتحسين الاستجابة الطارئة للمدافعين عن حقوق الإنسان وغيرهم من الأشخاص المعرضين للخطر.

5. مصالح الطفل الفضلى

تنص اتفاقية الأمم المتحدة لحقوق الطفل (UNCRC) على أنه في جميع الإجراءات المتعلقة بالأطفال، يجب أن تكون "المصلحة الفضلى للطفل اعتبارًا أساسيًا". يتوافق إطار "مصالح الطفل الفضلى" الخاص بشركة Meta مع القيم الأساسية لاتفاقية الأمم المتحدة لحقوق الطفل. وتُعد حماية الأطفال على الإنترنت أولوية قصوى لدى Meta. نوفّر أدوات للمراهقين والوالدين والأوصياء تتضمن حمايات مدمجة للمساعدة في الحفاظ على سلامتهم، مع إتاحة مساحة لهم لممارسة حقهم في حرية التعبير والوصول إلى المعلومات.

أمثلة على المخاطر البارزة المتأصلة المحتملة في مجال حقوق الإنسان التي تم تحديدها في تقييم CSRA	أمثلة على كيفية معالجة Meta للمخاطر المحتملة في 2024
يمكن أن يتعرض الأطفال لمحتوى غير مرغوب فيه أو غير مناسب أو لسلوكيات استغلالية.	قمنا بإطلاق حسابات Instagram للمراهقين، وهي تجربة جديدة للمراهقين تتضمن حمايات مدمجة، تحت إشراف الوالدين. قراءة المزيد

6. المشاركة العامة والحق في التصويت والحق في الترشح للانتخابات

إن الحق في المشاركة العامة، والتصويت، والترشح للانتخابات في انتخابات حرة ونزيهة يُمثل عماد الديمقراطية. تُعد حماية نزاهة الانتخابات على خدماتنا وتطبيقاتنا من أهم أولوياتنا. نعمل جاهدين على حماية الانتخابات على الإنترنت قبل فترات الانتخابات وأثناءها وبعدها.

أمثلة على المخاطر البارزة المتأصلة المحتملة في مجال حقوق الإنسان التي تم تحديدها في تقييم CSRA	أمثلة على كيفية معالجة Meta للمخاطر المحتملة في 2024
يمكن أن يؤثر المحتوى المخالف سلبًا في المشاركة العامة أو التصويت أو الترشح للمناصب. قد ينجم ذلك عن أنشطة تشمل، على سبيل المثال لا الحصر: • المحتالون المنسقون الذين يتدخلون في الانتخابات • التهديد بإلحاق ضرر على أرض الواقع وعنف للمرشحين • الجهود الفردية المبذولة لمنع الأشخاص من التصويت، أو زيادة المحتوى غير المهم أو الاحتيالي، أو التدخل الأجنبي، أو البلاغات عن المحتوى الذي ينتهك سياساتنا	كانت الانتخابات أولوية في عام 2024. قمنا بالاستعداد للانتخابات على نطاق واسع، بما في ذلك الانتخابات الأعلى خطرًا، وساعدنا الناخبين في العثور على المعلومات، من بين إجراءات أخرى. قراءة المزيد

7. حرية تكوين الجمعيات والتجمّع

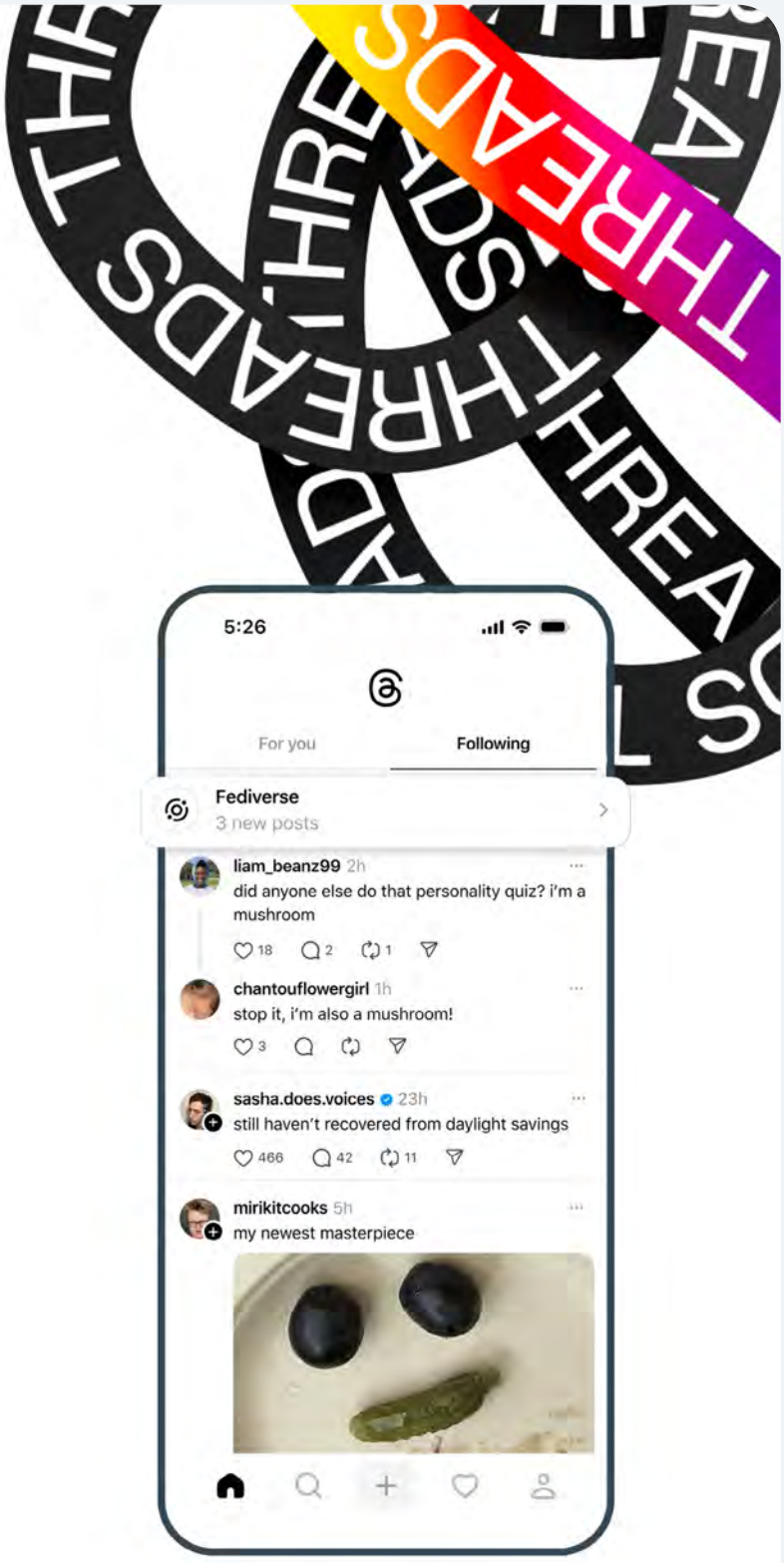
إن الحق في حرية تكوين الجمعيات والتجمّع عامل ضروري لخلق مناخ ديموقراطي سليم ومترابط مع العديد من الحقوق الأخرى التي يكفلها القانون الدولي لحقوق الإنسان، بما في ذلك الحق في حرية التعبير والمشاركة في إدارة الشؤون العامة. بالنسبة إلى Meta، يرتبط هذا الحق بقيمنا الجوهرية المتمثلة في منح الأشخاص صوتًا وبناء الروابط والمجتمعات.

أمثلة على المخاطر البارزة المتأصلة المحتملة في مجال حقوق الإنسان التي تم تحديدها في تقييم CSRA	أمثلة على كيفية معالجة Meta للمخاطر المحتملة في 2024
قد يؤدي المحتوى أو السلوك الزائف المنظم على منصات Meta إلى شعور بعض الأشخاص بعدم القدرة على التجمع بحرية على تطبيقات Meta أو خارجها.	قمنا بتطبيق بروتوكول سياسة الأزمات لدعم جهودنا في التعامل مع المحتوى المخالف المتعلق بالمظاهرات الجماهيرية، على سبيل المثال في بنجلاديش وجورجيا. كما قمنا بالاستعداد جيدًا قبل الانتخابات لتقليل مخاطر المحتوى المخالف الذي قد يجعل الأشخاص يشعرون بعدم الأمان للتجمع أثناء الانتخابات وبعدها. انضم Threads إلى fediverse، وهي شبكة مفتوحة وعالمية لخوادم وسائل التواصل الاجتماعي. وقد أتاح ذلك للأشخاص توسيع مجتمعاتهم والوصول إلى جماهير جديدة.

8. الحق في الصحة

يُعد الحق في الصحة حق الجميع في التمتع بأعلى مستوى ممكن من الصحة البدنية والعقلية. تحترم Meta هذا الحق من خلال زيادة إمكانية الوصول إلى المعلومات الصحية، وتمكين الأشخاص الذين يعانون من مشكلات صحية مماثلة من التواصل مع بعضهم بعضًا، وتمكينهم من اتخاذ قرارات مستنيرة بشأن صحتهم وسلامتهم.

أمثلة على المخاطر البارزة المتأصلة المحتملة في مجال حقوق الإنسان التي تم تحديدها في تقييم CSRA	أمثلة على كيفية معالجة Meta للمخاطر المحتملة في 2024
المحتوى الذي ينتهك السياسات ويهدف إلى التحريض على الضرر أو التسبب فيه على أرض الواقع	أطلقنا برنامج Thrive، وهو برنامج لتبادل الإشارات عبر المجالات صُمم لمنع انتشار محتوى الانتحار وإلحاق الضرر بالذات، بالشراكة مع Snap وTikTok. عقدنا منتدى السياسة حول المحتوى التجاري المرتبط بمخاطر الصحة والسلامة المستندة إلى التنظيمات. قمنا بتحديث معايير المجتمع ومعايير الإعلانات لتشمل الإشارة إلى السلع المسترجعة.



منتجات وخدمات جديدة

كما ذُكر في تقرير حقوق الإنسان لعام 2023، تسعى Meta إلى احترام حقوق الإنسان في تصميم وتطوير منتجاتنا وخدماتنا.

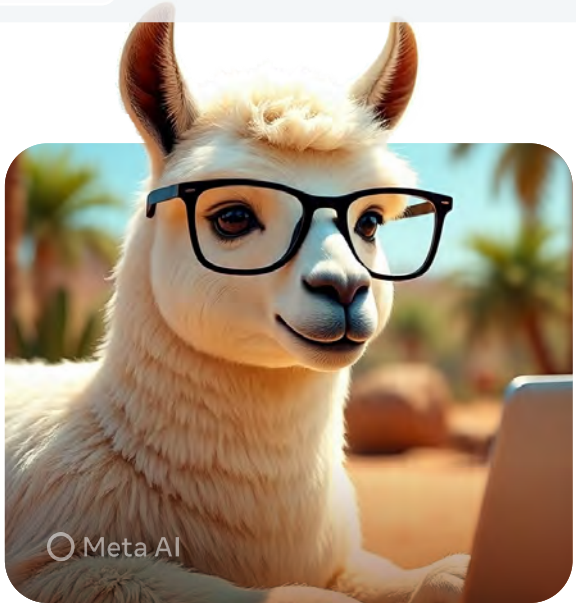
في 2024، انضم Threads إلى fediverse، وهي شبكة مفتوحة وعالمية لخوادم وسائل التواصل الاجتماعي. إذا قرر المستخدم تفعيل المشاركة مع fediverse، يمكن للأشخاص من منصات مختلفة (مثل Mastodon أو Flipboard) متابعة محتوى Threads الخاص بالمستخدم والتفاعل معه، حتى لو لم يكن لديهم ملف شخصي على Threads. يتيح هذا للأشخاص ممارسة حرية التعبير وحرية التجمع والانتماء من خلال الوصول إلى جماهير جديدة، وتوسيع مجتمعاتهم، والانضمام إلى النقاشات العامة حول المواضيع التي تهمهم. كما يساعد ذلك في خلق نظام معلومات أكثر تنوعًا.

كما قدمنا للمستخدمين موارد تعليمية حول استخدام Threads عبر قسم مخصص في مركز المساعدة ودليل fediverse جديد في مركز الخصوصية يوضح كيف تؤثر اللامركزية وقابلية التشغيل البيني على خصوصيتهم.

تسريع الابتكار في الذكاء الاصطناعي مع احترام حقوق الإنسان

شهدت التطورات في مجال الذكاء الاصطناعي (AI) تسارعًا في عام 2024. أصبحت أدوات وتطبيقات الذكاء الاصطناعي التوليدي مستخدمة على نطاق واسع وغيّرت بشكل متزايد طرق تواصلنا وتعلمنا وإبداعنا وعملنا. في Meta، ندرك أن هذا التطور السريع واعتماد الذكاء الاصطناعي يطرح فوائد ومخاطر جديدة ومهمة تتعلق بحقوق الإنسان.

تتمثل رؤيتنا طويلة المدى في بناء ذكاء فائق شخصي وجعله متاحًا على نطاق واسع بحيث يستفيد الجميع منه.



في عام 2024، أطلقنا نماذج اللغة الكبيرة (LLMs) المفتوحة لدينا Llama 3، وLlama 3.1، وLlama 3.2 وLlama 3.3. كما أطلقنا مساعد Meta AI ودمجناه في جميع تقنياتنا. ظهر Meta AI Studio كمنصة لإنشاء شخصيات ذكاء اصطناعي مخصصة، وساعدت مجموعة أدوات الذكاء الاصطناعي التوليدي المعلنين في بناء أنشطتهم التجارية، وتم دمج Meta AI مع نظارات Ray-Ban Meta الخاصة بنا. واصلنا إجراء ونشر أبحاث الذكاء الاصطناعي المتقدمة، بما في ذلك نماذج Movie Gen التي تولّد مقاطع الفيديو وتمكّن من التعديل الدقيق المستند إلى التعليمات، ونموذج Video Seal لوضع علامات مائية دائمة على مقاطع الفيديو التي تم إنشاؤها بواسطة الذكاء الاصطناعي، من بين تطورات أخرى.

بحلول نهاية عام 2024، كان المطورون قد حمّلوا نماذج Llama المفتوحة لدينا أكثر من 650 مليون مرة، وكان لدى Meta AI ما يقرب من 600 مليون مستخدم نشط شهريًا على مستوى العالم، مما يجعل نماذجنا الأكثر اعتمادًا وانتشارًا في العالم. ويؤكد هذا العدد الكبير من المستخدمين، سواء من المطورين أو المستخدمين النهائيين، مسؤوليتنا في بناء الذكاء الاصطناعي مع احترام حقوق الإنسان.

نهجنا المفتوح

نعتقد أن الذكاء الاصطناعي مفتوح المصدر هو جزء مهم لضمان استفادة الجميع من تقدمات الذكاء الاصطناعي. كما أوضحنا في تقرير حقوق الإنسان لعام 2023، فإن النهج المفتوح يحمل فوائد مهمة لحقوق الإنسان. نماذج الذكاء الاصطناعي مفتوحة المصدر:

- أكثر قدرة بطبيعتها على مقاومة الرقابة والقيود الأخرى على حق حرية التعبير، لأنها يمكن تنزيلها وتشغيلها دون اتصال بالإنترنت، مما يقلل من تأثيرات المطالبات الحكومية المحتملة بعد الإطلاق لتقييد المخرجات.
- تمكّن بشكل أفضل من التكيف والتحسين لتعكس السياق المحلي والدقة، بما يتماشى مع الحق في المساواة، وتحسين إمكانية الوصول وشمول اللغة.
- تسهّل على المطورين بناء نماذج أصغر وأكثر كفاءة، ما يمكن أن يقلل الحواجز أمام المجتمعات التي عادة لا تحظى بالاهتمام الكافي، وتدعم الحقوق الاقتصادية والاجتماعية والثقافية.
- تدعم البحث النقدي في تدابير حماية وأمان الذكاء الاصطناعي من خلال السماح لأي شخص بفحص النماذج للكشف عن المخاطر المحتملة، مما يساعد في التخفيف من الآثار السلبية المحتملة على حقوق الإنسان.

نحن نشهد بالفعل فوائد ملموسة لنهجنا المفتوح. استمر برنامج منح التطبيقات المؤثرة لنموذج Llama، الذي أُطلق في 2023، في العمل خلال 2024. يدعم هذا البرنامج، جنبًا إلى جنب مع جوائز الابتكارات المؤثرة من Llama لعام 2024، ويسلط الضوء على حالات الاستخدام ذات الأثر الاجتماعي الإيجابي لنماذجنا المفتوحة.

على سبيل المثال، استخدم المطورون Llama لبناء نموذج Vax-Llama، وهي خدمة برنامج تلقائي للدردشة صممت لتقديم معلومات دقيقة حول التطعيم، معنية بأن يعتمدها موفر خدمات الرعاية الصحية حول العالم. كما استُخدم أيضًا في مشروع Llama-Suho، وهي مبادرة لتحسين Llama بالبيانات الخاصة بكوريا لتعزيز تدابير حماية الذكاء الاصطناعي في السياق الكوري.

إرساء الحماية في صميم النهج

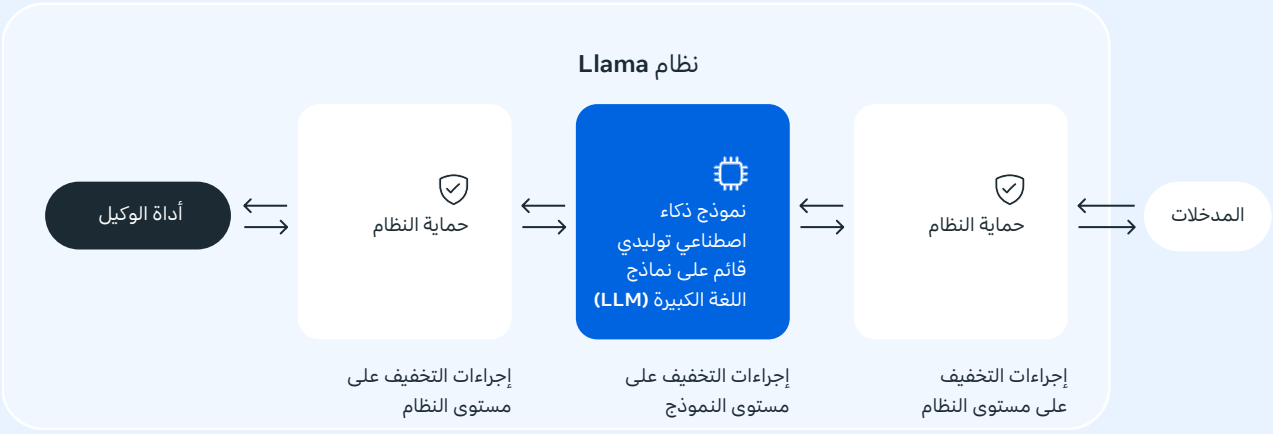
نواصل التزامنا بتطوير ونشر منتجات ذكاء اصطناعي متقدمة وفق أحدث ما توصلت إليه التكنولوجيا، مع مراعاة معايير حقوق الإنسان وتوفير الضمانات اللازمة للوقاية من إساءة الاستخدام.

وتشير سياسة حقوق الإنسان الخاصة بشركتنا صراحةً إلى انطباق التزاماتنا في مجال حقوق الإنسان على تقنيات الذكاء الاصطناعي.

ومع إطلاق Llama 3 في أبريل 2024، بدأنا في التركيز على نهج قائم على الأنظمة في ضمانات الذكاء الاصطناعي. ويتيح النهج القائم على الأنظمة للمطورين مرونة إضافية لتطبيق طبقات الحماية المناسبة بحسب اختلاف حالات الاستخدام والجمهور المستهدف. فعلى سبيل المثال، نوفر وسائل حماية لأنواع معينة من الخطاب الذي من المحتمل أن يكون مسيئًا، لكنه مشروع قانونًا، وذلك على هيئة تدابير تخفيف اختيارية على مستوى النظام. ويأتي ذلك إلى جانب استمرارنا في تضمين وسائل حماية أساسية داخل نماذجنا التأسيسية لمنع توليد محتوى استغلال الأطفال.

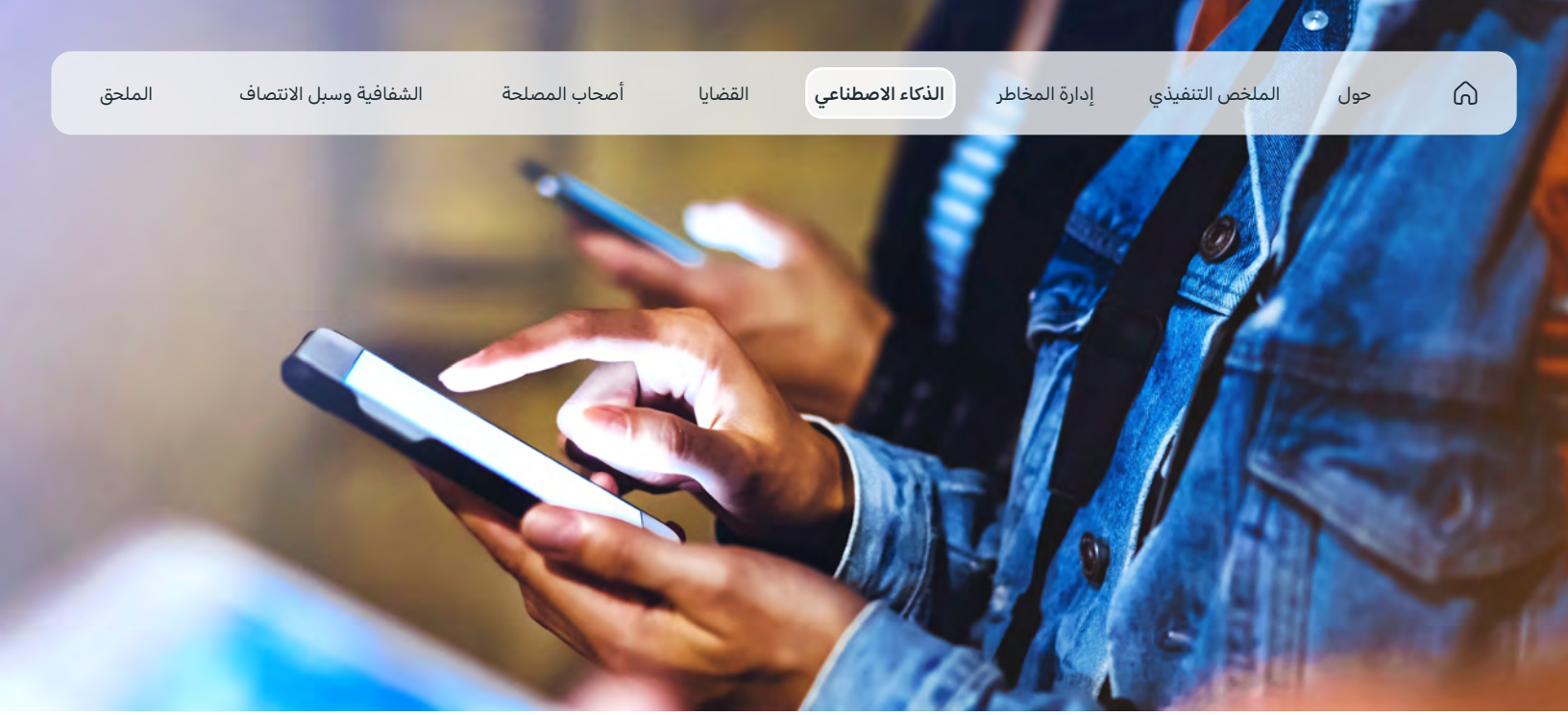
ونعتقد أن هذا النهج القائم على الأنظمة يدعم تحقيق توازن مناسب بين حرية التعبير وغيرها من حقوق الإنسان.

أنظمة سلامة الذكاء الاصطناعي



في إطار نهجنا القائم على الأنظمة، قمنا بإتاحة ثلاث أدوات رئيسية كمصادر مفتوحة؛ وهي (Llama Guard والحماية من الأوامر و Code Shield) يمكن للمطورين تخصيصها واستخدامها معًا أو بشكل مستقل لتنفيذ وسائل حماية ضد إساءة الاستخدام.

ويقدّم دليل استخدام المطورين الخاص بنا إرشادات تفصيلية حول كيفية النشر المسؤول لنماذجنا التأسيسية وأنظمة السلامة عبر مجموعة متنوعة من السياقات. ولا تزال سياسة الاستخدام المقبول تحكم عمليات نشر نماذجنا المفتوحة.



وبالإضافة إلى توفير هذه الأدوات خلال عام 2024، اتخذنا أيضًا خطوات مهمة للتخفيف من المخاطر المرتبطة بعمليات النشر الخاصة بنا للذكاء الاصطناعي التوليدي. وشملت ممارساتنا في عام 2024 ما يلي:

- إخضاع نماذجنا ومنتجاتنا المطوَّرة داخليًا لعمليات اختبار هجومي مكثّفة قبل الطرح، بهدف تحديد المخاطر المحتملة والتخفيف منها، بما في ذلك المخاطر المرتبطة بالآثار السلبية المحتملة على حقوق الإنسان.
- تحديث كيفية تعاملنا مع الوسائط المُتلاعَب بها استنادًا إلى الملاحظات الواردة من مجلس الإشراف المستقل، بما في ذلك إضافة تسميات "معلومات الذكاء الاصطناعي" وتوفير سياق توضيحي لنطاق أوسع من محتوى الفيديو والصوت والصور، إلى جانب إلزام منشئي المحتوى بالإفصاح عن استخدام الذكاء الاصطناعي.
- تحسين الإرشادات والإجراءات الداخلية التي نعتمدها لاختبار مخرجات النماذج المقبولة، بما يعكس بشكل أفضل حالات الاستخدام في العالم الواقعي ويتوافق مع المعايير الدولية لحقوق الإنسان.

كما ندرك أن ضمانات الذكاء الاصطناعي تتطلب تعاونًا عابرًا للقطاعات ومشاركة أصحاب المصلحة المتعددين. وفي فبراير 2024، وقّعنا، إلى جانب نظرائنا في القطاع، على اتفاق الذكاء الاصطناعي للانتخابات، ملتزمين بالمساعدة في منع المحتوى الخادع الذي تم إنشاؤه بواسطة الذكاء الاصطناعي من التدخل في الانتخابات العالمية. وفي مايو 2024، انضممنا إلى منتدى النماذج الرائدة، وهو كيان مدعوم من القطاع يكرّس جهوده لتعزيز أمن نماذج الذكاء الاصطناعي المتقدمة.

معالجة حالات الرفض غير الصحيح

تحدث حالات الرفض غير الصحيح عندما يرفض النموذج إنتاج المخرجات المطلوبة استجابةً لأمر سليم وغير ضار، وغالبًا ما يكون ذلك نتيجةً لضمانات السلامة الحسنة النية المدمجة في النموذج. فعلى سبيل المثال، قد يرفض النموذج على نحو غير صحيح مناقشة عمل من أعمال الأدب الكلاسيكي يتضمن صورة نمطية أو إهانات مسيئة، أو يمتنع عن الإجابة عن سؤال أساسي في الكيمياء على مستوى المرحلة الثانوية، وذلك بسبب ضمانات صُممت لمنع المساعدة في إنشاء مواد كيميائية أو بيولوجية أو إشعاعية أو نووية أو متفجرات عالية القدرة التدميرية. ورغم أن سلامة النماذج أمر مهم، وأن الرفض قد يكون ضروريًا في بعض الحالات للحد من توليد محتوى ضار، فإن حالات الرفض غير الصحيح يمكن أن تخلّف آثارًا سلبية على حرية التعبير، وإمكانية الوصول إلى المعلومات، وغيرها من الحقوق.

واعتبارًا من إطلاق Llama 3، باشرنا عملاً مكثفًا للحد من حالات الرفض غير الصحيح في Llama وMeta AI، وحققنا تقدمًا ملموسًا في هذا المجال خلال عام 2024.

إنشاء محتوى قابل للتطويع المحلي بعناية

في عام 2024، أتحنا Meta AI في أكثر من 40 دولة إضافية وعدة لغات جديدة، من بينها العربية، والبهاسا الإندونيسية، والفلبينية، والفرنسية، والألمانية، والهندية، والإيطالية، والبرتغالية، والإسبانية، والتايلاندية، والفيتنامية.

وقبل الإطلاق في كل دولة وكل لغة، قيّمنا المخاطر المحتملة على حقوق الإنسان، وأجرينا عمليات اختبار هجومي مكثّفة مخصّصة للسياق، وهي ممارسة شائعة للتخفيف من السلوكيات غير الآمنة في نماذج اللغة الكبيرة (LLM).

ولا تتمتع جميع الدول التي يتوافر فيها Meta AI بحماية قوية لحرية التعبير في تشريعاتها الوطنية. وفي إطار جهود إنشاء محتوى قابل للتطويع المحلي خلال عام 2024، طوّرنا نهجًا قائمًا على حقوق الإنسان للاستجابة لطلبات الحكومات الرامية إلى تقييد أو الحد من مخرجات Meta AI، وذلك بالاستناد إلى سياساتنا المعتمدة منذ فترة طويلة، وبما يتماشى مع التزاماتنا كعضو في مبادرة الشبكة العالمية ومع سياسة حقوق الإنسان الخاصة بشركتنا.

التفاعل مع أصحاب المصلحة

إن أي مجال تتقدم فيه التكنولوجيا بالسرعة التي يتقدم بها الذكاء الاصطناعي يطرح تحديات جديدة وغير مسبوقة تتعلق بتفاعل أصحاب المصلحة. وخلال عام 2024، ركّزنا على تثقيف أصحاب المصلحة واستطلاع ملاحظات جوهرية وذات مغزى.

ومن بين جهودنا:

عقدنا موائد مستديرة حول الذكاء الاصطناعي في الولايات المتحدة للحصول على ملاحظات من مجموعات متعددة التخصصات بشأن إطلاق المنتجات والنماذج، بما في ذلك خبراء من الولايات المتحدة والبرازيل وبروكسل والأردن والمكسيك ومن مختلف أنحاء أفريقيا.

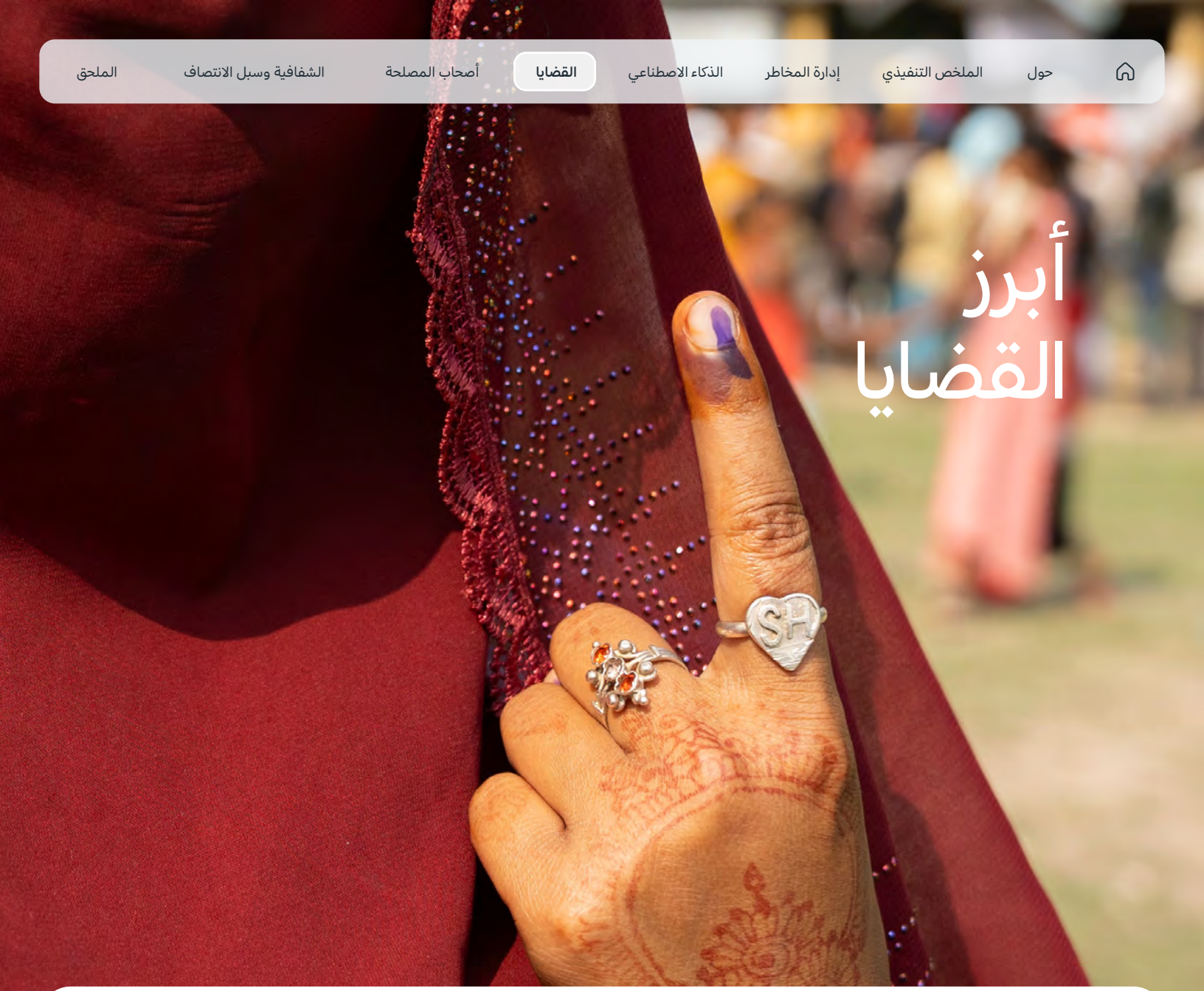
شاركنا النتائج المستخلصة من منتديات المجتمع التي عُقدت في الولايات المتحدة والبرازيل وألمانيا وإسبانيا لاستكشاف المبادى الموجِّهة للبرامج التلقائية للدردشة القائمة على الذكاء الاصطناعي التوليدي، وذلك بالشراكة مع مختبر الديمقراطية التداولية التابع لجامعة ستانفورد.

عقدنا سلسلة من ورش عمل Open Loop المصمَّمة لمعالجة التعقيدات والاستفادة من الفرص التي يتيحها الذكاء الاصطناعي مفتوح المصدر. جمعت هذه الورش صانعي السياسات وقادة المجال والأكاديميين وممثلي المجتمع المدني من مختلف أنحاء العالم، للعمل بشكل تعاوني على صياغة سياسات فعّالة ومسؤولة للذكاء الاصطناعي.

وبالتوازي مع المنتدى الثالث عشر للأمم المتحدة المعني بالأنشطة التجارية وحقوق الإنسان في جنيف، قمنا بتطوير وقيادة محاكاة تفاعلية متعددة أصحاب المصلحة حول العناية الواجبة بحقوق الإنسان في منتجات الذكاء الاصطناعي التوليدي، حيث شاركنا نهجنا وأسهمنا في بناء فهم متبادل أعمق للمخاطر والتحديات.

ومع استمرارنا في الابتكار في مجال الذكاء الاصطناعي، نؤكد التزامنا الدائم بتفاعل أصحاب المصلحة والتشاور معهم على نحوٍ شامل عالميًا وقوي وفعّال.

قراءة المزيد



أبرز القضايا

2024: عام الانتخابات

كان عام 2024 هو أكبر عام انتخابي في التاريخ. أجرت أكثر من 70 دولة، تمثل أكثر من نصف سكان العالم، انتخابات وطنية، وكان نحو 2 مليار شخص مؤهلين للتصويت.

تعترف شركة Meta بأهمية تمكين الحقوق الأساسية مثل حرية التعبير والتصويت والمشاركة في الشؤون العامة. وكان عملنا في مجال الانتخابات محور تركيز رئيسي طوال العام. لقد استعددنا لحجم الانتخابات وانتشارها وتواترها، وعملنا على التخفيف من المخاطر المرتبطة بها على المستخدمين، بما في ذلك المخاطر المحتملة الناجمة عن الاستخدام المتزايد للذكاء الاصطناعي.

نستعرض جهودنا في عام 2024 في الصفحات التالية، مع تقديم ملخصات توضيحية من الانتخابات في الاتحاد الأوروبي، والهند، والمكسيك، والولايات المتحدة.

الاستعداد للانتخابات على نطاق واسع

لقد قامت Meta بتطوير نهجها الأساسي تجاه الانتخابات على مدى السنوات القليلة الماضية. نطبّق هذا النهج في جميع البلدان التي تُستخدم فيها خدماتنا، مع تكييف استراتيجيتنا بما يتناسب مع الاحتياجات والمخاطر المحلية. كجزء من استعداداتنا لانتخابات 2024، كان هناك فريق مخصص مسؤول عن الجهود المشتركة بين الأقسام، ويضم خبراء من فرق الاستخبارات وعلوم البيانات والمنتجات والهندسة والبحث والعمليات والمحتوى وحقوق الإنسان والسياسات العامة والشؤون القانونية. على مدار العام، سعينا إلى تمكين قدرة الأشخاص على التعبير عن أنفسهم والتصويت والترشح للانتخابات.

وشمل نهجنا الجهود المتعلقة بإدارة مخاطر الذكاء الاصطناعي وتطبيق سياساتنا المتعلقة بالتدخل في الإدلاء بالأصوات، وتعطيل الشبكات العدائية، وتوفير شفافية الإعلانات السياسية، وربط الناخبين بالمعلومات الموثوقة. كما قيّمنا تغطية اللغات المناسبة لأدوات التصنيف والمراجعين البشريين في البلدان التي تجري فيها الانتخابات، لدعم جهودنا في اتخاذ الإجراءات ضد المحتوى المخالف للسياسات. ومن أبرز ما قمنا به:

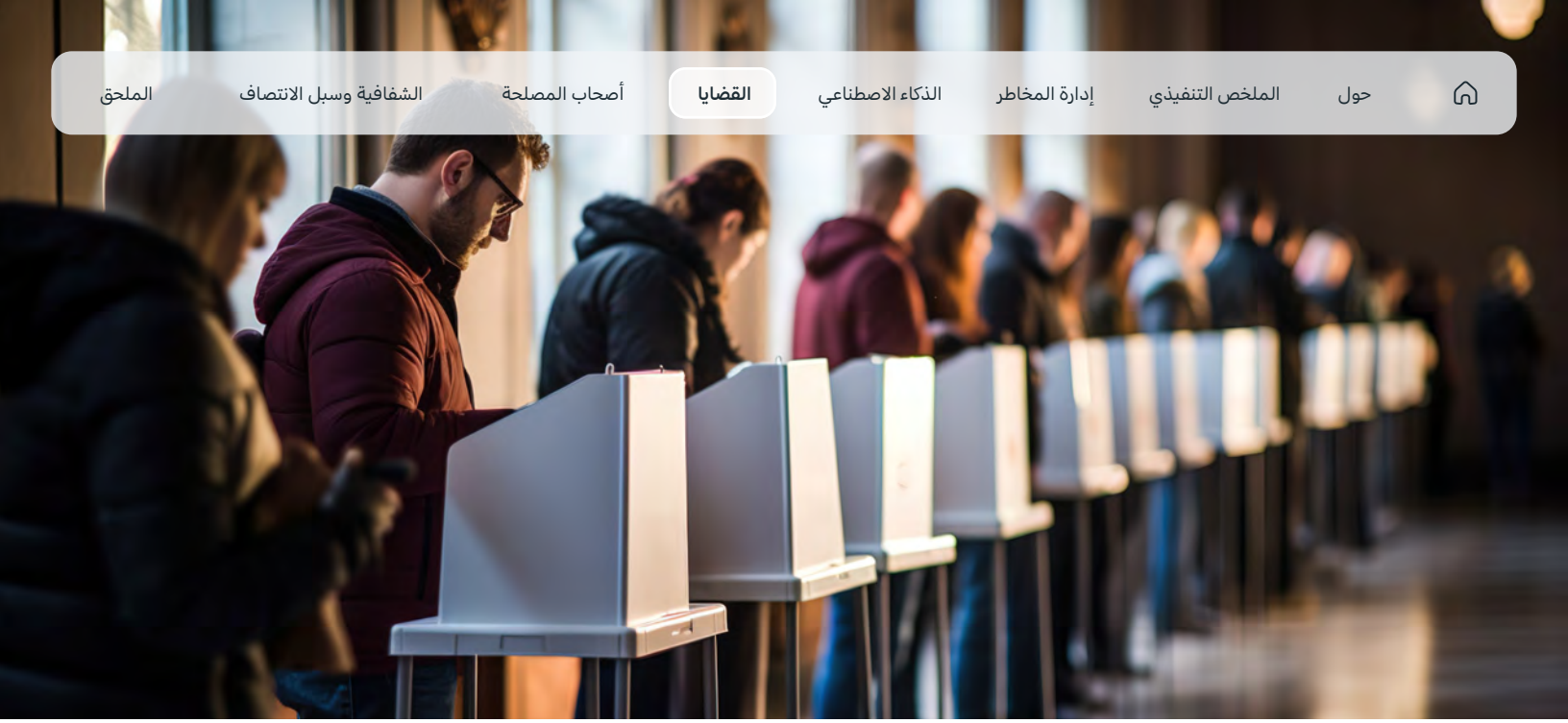
إدارة مخاطر تأثير الذكاء الاصطناعي

في بداية العام، كان الكثير من الأشخاص قلقين بشأن المخاطر التي قد يفرضها الذكاء الاصطناعي التوليدي على نزاهة الانتخابات، بما في ذلك مخاطر انتشار عمليات التزييف العميق وحملات التضليل المدعومة بالذكاء الاصطناعي. لقد استعددنا لذلك وراقبنا عن كثب التهديدات العدائية واحتمالية اضطراب الانتخابات نتيجة الذكاء الاصطناعي. من خلال ما رصدناه عبر خدماتنا، بدا أن هذه المخاطر لم تتحقق بشكل كبير، وأي تأثير محتمل كان متواضعًا ومحدود النطاق. على سبيل المثال، خلال فترة الانتخابات في مجموعة من الانتخابات الرئيسية، مثلت التقييمات المتعلقة بمحتوى الذكاء الاصطناعي المرتبط بالانتخابات والسياسة والموضوعات الاجتماعية أقل من 1% من إجمالي المعلومات المضللة التي تم التحقق من صحتها. يبدو أن سياساتنا وإجراءاتنا الحالية كانت كافية للحد من المخاطر المرتبطة بمحتوى الذكاء الاصطناعي التوليدي.

على مدار العام، ركزت جهودنا على إيقاف عمليات التأثير الخبيثة والاستفادة من الشراكات العالمية للمساعدة في الحفاظ على نزاهة الانتخابات.

لقد رصدنا عن كثب إمكانية إساءة استخدام الذكاء الاصطناعي التوليدي من خلال حملات منسقة تستخدم حسابات زائفة. وقد تبيّن لنا أن هذه الحملات حققت فقط زيادات طفيفة في الإنتاجية وتوليد المحتوى باستخدام الذكاء الاصطناعي التوليدي. ولم تُعيق هذه الزيادات الطفيفة قدرتنا على تعطيل هذه العمليات التأثيرية، لأننا نركّز عند التحقيق وإيقاف هذه الحملات على السلوكيات نفسها وليس على المحتوى الذي يتم نشره، سواء تم إنشاؤه بالذكاء الاصطناعي أم لا.

كما أننا تعاونّا مع جهات أخرى في المجال لمكافحة التهديدات المحتملة الناجمة عن استخدام الذكاء الاصطناعي التوليدي. على سبيل المثال، في فبراير 2024، وقعنا اتفاق الذكاء الاصطناعي للانتخابات إلى جانب العشرات من قادة المجال الآخرين، متعهدين بالمساعدة في منع المحتوى المخادع الذي يتم إنشاؤه بالذكاء الاصطناعي من التأثير على الانتخابات العالمية لعام 2024. يتم وصف أمثلة على المبادرات الخاصة بكل دولة المتعلقة بالذكاء الاصطناعي في الصفحات التالية.



جهود أخرى لضمان نزاهة الانتخابات

بالإضافة إلى التخفيف من مخاطر تأثير الذكاء الاصطناعي المحتملة على الانتخابات، سعينا أيضًا إلى تمكين الناخبين، ومنع التدخل الأجنبي، وتعزيز سلامة المرشحين، وبناء الشراكات، والمساعدة في ضمان الشفافية للمعلنين.

تمكين الناخبين

يُعد الوصول إلى معلومات موثوقة والاستخدام المسؤول للمنصات الإلكترونية أمرًا بالغ الأهمية خلال الانتخابات. في العديد من الدول، قدّمنا للأشخاص معلومات عن الناخبين وتذكيرات بيوم الانتخابات من خلال الإشعارات داخل التطبيقات على فيسبوك وInstagram. مكنت هذه الميزات الأشخاص من الوصول إلى معلومات رسمية وموثوقة من سلطات الانتخابات حول كيفية ومكان وموعد التصويت في يوم الانتخابات. على سبيل المثال، في الانتخابات المحلية في البرازيل، تفاعل الأشخاص مع هذه الإشعارات حوالي 9.7 مليون مرة عبر فيسبوك وInstagram. وقد شاهد أكثر من 63 مليون شخص على فيسبوك و118 مليون شخص على Instagram ملصق تسجيل الناخبين الذي يوجّههم إلى معلومات موثوقة حول الانتخابات والتصويت.

منع التدخل الأجنبي

قامت فرق الأمن لدينا بالتحقيق وإيقاف شبكات من الحسابات والصفحات والمجموعات غير الحقيقية والمنسقة. بالإضافة إلى ذلك، قدرنا أن آلية الكشف الآلية عن الحسابات الزائفة لدينا تمنع يوميًا إنشاء ملايين الحسابات الزائفة. وقد أوقفت فرقنا حوالي 20 عملية تأثير سرية حول العالم، بما في ذلك في الشرق الأوسط، وآسيا، وأوروبا، والولايات المتحدة. على سبيل المثال، في مولدوفا، أزلنا شبكة تستهدف الجماهير الناطقة بالروسية كجزء من تحقيقنا في سلوك زائف منظم مشتبه به في المنطقة.

سلامة المرشحين

قدمت Meta أيضًا حماية معززة ضد الاختراق وانتحال الشخصية والمضايقة لحسابات المسؤولين المنتخبين والمرشحين وفرق عملهم. عقدنا عدة تدريبات للمرشحين حول السلامة، موضحين فيها الإرشادات المتاحة للتعامل مع المضايقة على منصاتنا، ونشرنا محتوى تعليميًا ليكون متاحًا على نطاق واسع لجميع المشاركين في الانتخابات.

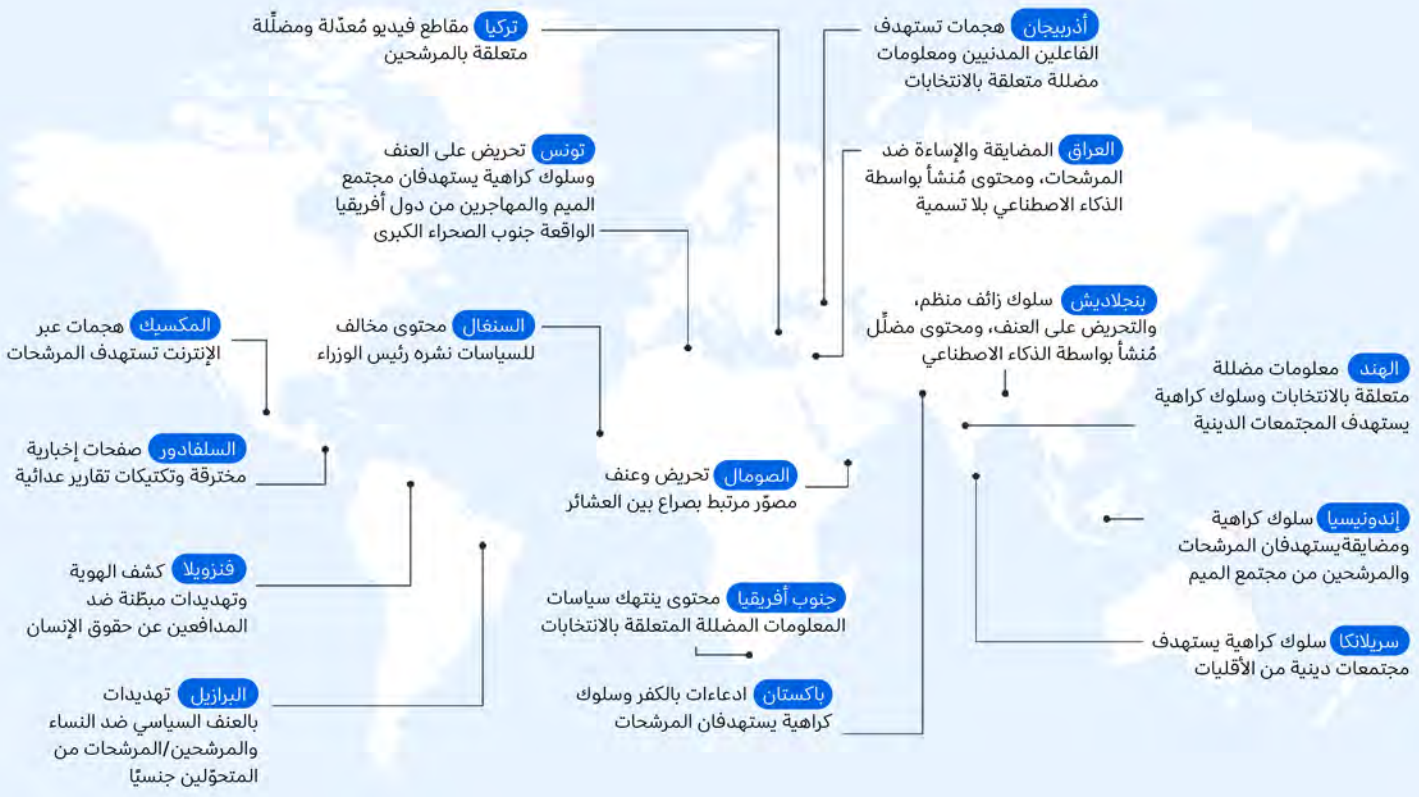
التواصل والشراكات

قمنا بإجراء عمليات تواصل وإقامة قنوات تواصل مع السلطات الحكومية ووكالات تطبيق القانون لتمكينها من الإبلاغ عن المحتوى الذي قد ينتهك معايير مجتمعنا أو القانون المحلي. كما تعاونّا مع مجموعات المجتمع المدني، وجهات تدقيق الحقائق، وشركات التكنولوجيا الأخرى لمساعدتنا في تحديد التهديدات الناشئة ووقف انتشار المعلومات الزائفة.

شفافية الإعلانات

واصلنا تقديم شفافية رائدة في المجال للإعلانات المتعلقة بالقضايا الاجتماعية والانتخابات والسياسة. في معظم الأسواق التي نقدم فيها هذه الإعلانات، يمر المعلنون عبر عملية التصريح ويجب عليهم تضمين إخلاء مسؤولية "مدفوع بواسطة" على محتواهم لتشغيل إعلاناتهم. قد يتضمن إخلاء المسؤولية معلومات عن المؤسسة أو الفرد المسؤول عن الإعلان، رغم أن المتطلبات قد تختلف حسب البلد. ثم تُخزن الإعلانات في مكتبة الإعلانات المتاحة للجمهور. في عام 2024، أضفنا متطلبًا يلزم المعلنين بالإفصاح عند استخدامهم للذكاء الاصطناعي أو تقنيات رقمية أخرى لإنشاء أو تعديل إعلان متعلق بقضية أو انتخابات أو سياسة في حالات محددة.

الشركاء الموثوق بهم دعموا جهود نزاهة الانتخابات في 25 دولة في عام 2024.



الاستعداد للانتخابات عالية المخاطر

اعتبرنا بعض الانتخابات عالية المخاطر، مما استلزم استعدادات إضافية، وموارد إضافية، وعملاً مخصصًا. نظرنا في عوامل مثل نوع الانتخابات، وحجم الدولة مقارنةً بقاعدتنا من المستخدمين، ومخاطر العنف السياسي، واستهداف الفئات الضعيفة، وقدرتنا التشغيلية. وشملت الجهود الإضافية إقامة عمليات مراقبة مخصصة وتدابير مؤقتة للاستجابة للمخاطر يمكن تصميمها وتطبيقها عبر الدول واللغات.

أقمنا عددًا من مراكز عمليات الانتخابات حول العالم لمراقبة الأحداث والتفاعل السريع مع القضايا التي تنشأ، بما في ذلك في الانتخابات عالية المخاطر. يمكنك العثور على مزيد من التفاصيل على الإنترنت حول جهودنا في البرازيل، وفرنسا، والهند، وإندونيسيا، والمكسيك، و باكستان، وجنوب أفريقيا، والمملكة المتحدة، والولايات المتحدة، والبرلمان الأوروبي.

أمثلة من الانتخابات الوطنية

تساعد الأمثلة الموجزة لأربع دول في الصفحات التالية على توضيح كيف سعينا لإدارة مخاطر الانتخابات في عام 2024. في كل سياق، بدأنا الاستعدادات قبل عام على الأقل.

الولايات المتحدة

في إطار الاستعداد للانتخابات الأمريكية، ركزت جهودنا على مساعدة الأشخاص في الوصول إلى معلومات موثوقة عن الناخبين، ومعالجة التدخل الأجنبي، وضمان شفافية المعلنين.

معلومات الناخبين

خلال الانتخابات العامة الأمريكية لعام 2024، تلقت التذكيرات في أعلى الموجز على فيسبوك و Instagram أكثر من مليار ظهور. وشملت هذه التذكيرات معلومات حول تسجيل الناخبين، والتصويت بالبريد، والتصويت المبكر حضوريًا، والتصويت في يوم الانتخابات. وقد نقر الأشخاص على هذه التذكيرات أكثر من 20 مليون مرة لزيارة المواقع الرسمية للحكومة للحصول على مزيد من المعلومات.

التدخل الأجنبي

استعددنا لمواجهة التدخل الأجنبي عبر الإنترنت في الانتخابات، من خلال توسيع جهودنا المستمرة في إنفاذ السياسات ضد وسائل الإعلام الخاضعة للدولة الروسية، واستمررنا في تعطيل واحدة من أكبر وأطول حملات التأثير السرية، المسماة "دوبلجانجر" (القَرين). تم إيقاف الغالبية العظمى من محاولات "دوبلجانجر" (القَرين) لاستهداف الولايات المتحدة في أكتوبر ونوفمبر بشكل استباقي قبل أن يشاهد أي شخص محتواها.

فترة حظر الإعلانات

خلال الأسبوع الأخير من حملة الانتخابات، حظرنا الإعلانات السياسية والانتخابية وإعلانات القضايا الاجتماعية الجديدة، وهي ممارسة استمررنا عليها منذ عام 2020. ظل الأساس المنطقي لفترة الحظر هذه كما هو في السنوات السابقة، في الأيام الأخيرة من الانتخابات، فنحن ندرك أنه قد لا يكون هناك وقت كافٍ للطعن على المطالبات الجديدة المقدمة في الإعلانات.

المكسيك

كان عام 2024 أكبر عام انتخابي في تاريخ المكسيك، حيث ترشح نحو 90000 مرشح لأكثر من 20000 منصب عام. كما بلغت مستويات العنف خلال الحملات الانتخابية أعلى مستوياتها على الإطلاق. قُتل ما لا يقل عن 37 مرشحًا، وسُجّلت أكثر من 828 هجومًا غير قاتل. شارك عدد أكبر من النساء في الترشح مقارنة بأي دورة انتخابية أخرى في تاريخ المكسيك، وتعرضت المرشحات لمعدلات عالية من العنف والقتل القائم على الجنس.

كانت جهودنا مشابهة لتلك المبذولة في بيئات عالية المخاطر الأخرى، واستفدنا من خبراء Meta على الأرض. قمنا بإزالة مستويات أعلى من المحتوى المخالف مقارنة بالمعتاد قبل وأثناء الانتخابات. وشمل المحتوى المخالف التأثير على الناخبين، وبيع الأصوات، والمحتوى الذي يحض على الكراهية، والتهديدات بالمضايقات والعنف القائم على الجنس ضد المرشحات على منصتي فيسبوك وInstagram.

ولتقليل فرص الاضطراب وخطر الضرر على أرض الواقع، ركزت جهودنا على سلامة المرشحين، وتوفير معلومات سهلة الاستخدام للناخبين، وتعزيز المعرفة الإعلامية.

سلامة المرشحين

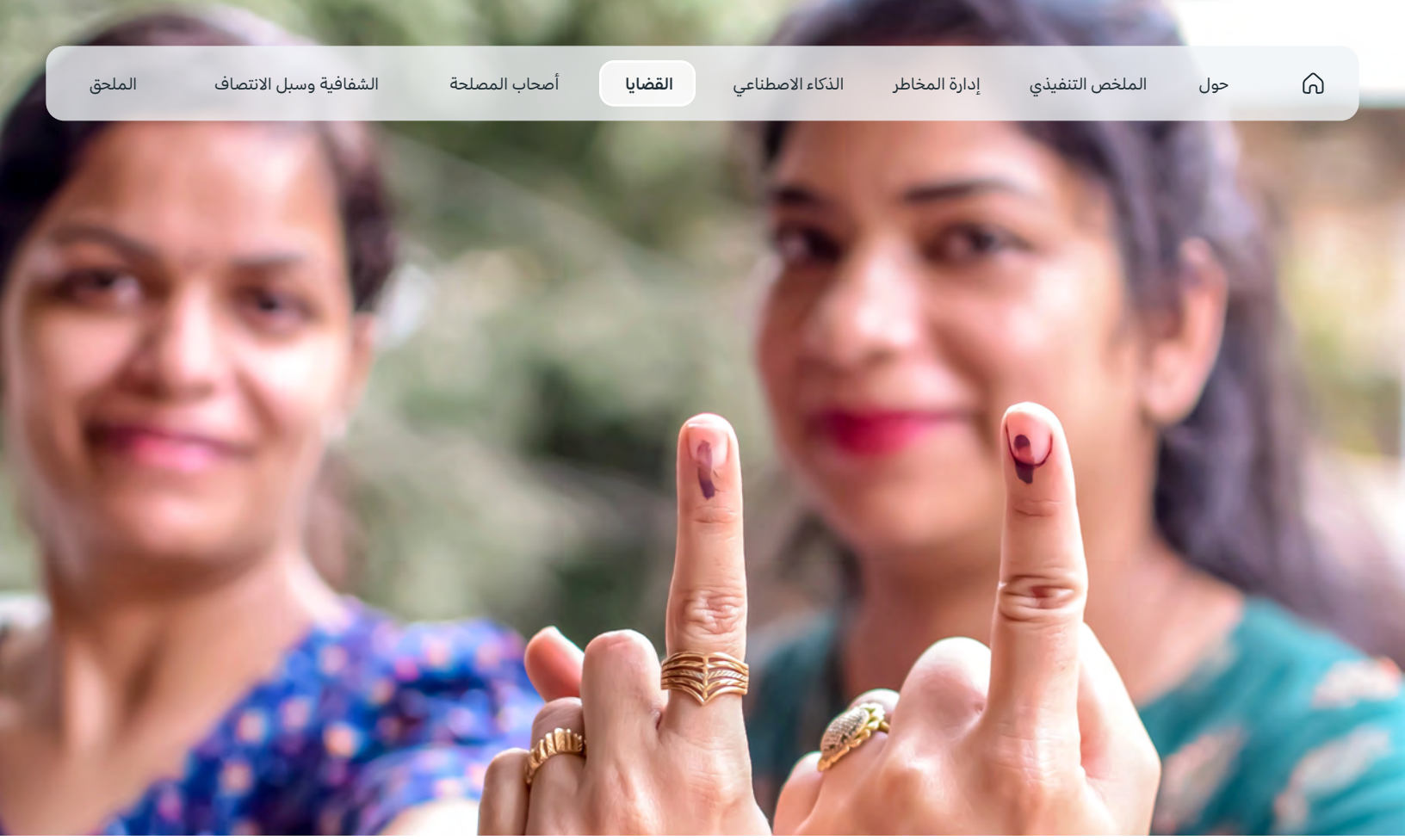
سجلنا أكثر من 3000 مرشح، بما في ذلك جميع المرشحين على المستوى الفيدرالي ومرشحي منصب الحاكم، في برنامج التحقق المتبادل للمساعدة في منع الأخطاء في تطبيق السياسات و/أو تطبيق الحماية المتقدمة على حساباتهم. وشمل ذلك مراقبة التهديدات المحتملة للاختراق. قمنا بتطوير حملة "صوّت ضد العنف،" التعليمية بالتعاون مع المنظمات غير الهادفة للربح ومجموعات الإعلام، بهدف ردع العنف القائم على الجنس على الإنترنت. وصلت الحملة إلى أكثر من 1.2 مليون شخص على منصاتنا، وتم تعزيز تأثيرها عبر قنوات أخرى. أرسلت السلطات طلبات إزالة عند اكتشاف العنف أو التهديدات بالعنف ضد المرشحين.

معلومات الناخبين

بالشراكة مع المعهد الوطني للانتخابات (INE)، أطلقنا برنامجًا تلقائيًا للدردشة "Inés" على واتساب لمساعدة الناخبين. أجاب البرنامج التلقائي للدردشة عن أسئلة حول العملية الانتخابية، مثل مكان وكيفية التصويت، وكيفية الحصول على بطاقات الناخبين، وإجراءات التصويت للمكسيكيين المقيمين في الخارج. في يوم الانتخابات، أرسلنا تذكيرات على فيسبوك وInstagram وأطلقنا ملصقات على كلا التطبيقين لتشجيع التصويت.

المعرفة الإعلامية

للمساعدة في منع انتشار المعلومات الزائفة، أطلقنا حملة "Soy Digital" ("التفكير الرقمي") بالتعاون مع INE وMovilizatorio، وهي منظمة مجتمع مدني. قدمت الحملة وحدات تعليمية وموارد يسهل الوصول إليها لبناء مهارات المواطنة الرقمية والوعي المعلوماتي، بما في ذلك كيفية البقاء آمنًا على الإنترنت. وصلت الحملة إلى أكثر من 15 مليون شخص. كما قمنا بتدريب 300 قائد على مستوى الدوائر الانتخابية، الذين بدورهم قاموا بتدريب آلاف موظفي الاقتراع على المعرفة الإعلامية.



الهند

بدأت Meta الاستعداد للانتخابات العامة الهندية لعام 2024 قبل 18 شهرًا. وكان التركيز على ضمان نزاهة المنصة وتعزيز توعية الناخبين. كان نهجنا مرنًا وقادرًا على دعم فترة انتخابات تمتد 60 يومًا، تم خلالها الإدلاء بأكثر من 640 مليون صوت. وشملت استعداداتنا ما يلي:

توعية الناخبين

أطلقت إشعارات تنبيه بشأن التصويت من صفحة لجنة الانتخابات في الهند على فيسبوك ووصلت إلى 145 مليون شخص. كما استخدمت لجنة الانتخابات الهندية واجهة API الخاصة بواتساب لإجراء حملات تذكير بالتصويت، وصلت إلى حوالي 400 مليون شخص.

ضمان نزاهة المنصة

اتخذنا إجراءات لمنع إساءة استخدام منصاتنا. عمل مراجعو المحتوى على محتوى فيسبوك وInstagram وThreads بأكثر من 20 لغة هندية بالإضافة إلى الإنجليزية. قمنا بإزالة الحسابات الزائفة والالتزام بالتعهدات الواردة في مدونة الأخلاقيات الطوعية التي انضممنا إليها، إلى جانب شركات وسائل التواصل الاجتماعي الأخرى، في عام 2019.

مكافحة المعلومات المضللة

أطلقنا خط مساعدة مخصص لتدقيق الحقائق على واتساب بالتعاون مع تحالف مكافحة المعلومات المضللة (MCA) لمكافحة المعلومات المضللة التي تم إنشاؤها بواسطة الذكاء الاصطناعي. أطلقنا خط مساعدة واتساب مع MCA، الذي أنشأ وحدة تحليل عمليات التزييف العميق الأولى من نوعها في العالم لتقييم أي محتوى صوتي أو فيديو يشتبه الأشخاص في كونه مزيفًا. كما قمنا بتدريب مئات من وكالات تطبيق القانون بالتعاون مع MCA لمكافحة عمليات التزييف العميق.

انتخابات البرلمان الأوروبي

استندت استعدادات Meta لانتخابات البرلمان الأوروبي إلى الدروس الرئيسية المستفادة من الانتخابات السابقة حول العالم، بالإضافة إلى الإطار التنظيمي المنصوص عليه في قانون الخدمات الرقمية والتزاماتنا ضمن مدونة الممارسات الخاصة بالاتحاد الأوروبي بشأن المعلومات المضللة.

ركزت إجراءاتنا الخاصة بالاتحاد الأوروبي على ما يلي:

تعزيز معلومات التصويت والمشاركة المدنية

قدمنا معلومات موثوقة حول الانتخابات ووجّهنا الأشخاص إلى معلومات حول العملية الانتخابية من خلال "وحدات معلومات الناخبين" و"معلومات يوم الانتخابات" داخل التطبيق. تفاعل الأشخاص مع هذه الإشعارات أكثر من 41 مليون مرة على فيسبوك وأكثر من 58 مليون مرة على Instagram.

التصدي لعمليات التأثير

ركزت جهودنا لإيقاف السلوك الزائف المنظم على التهديدات المرتبطة تحديدًا بانتخابات البرلمان الأوروبي. قمنا بتعطيل عدة شبكات استهدفت الاتحاد الأوروبي، بما في ذلك تنفيذ الإجراءات عدة مرات على الشبكة ذات الأصل الروسي المعروفة باسم "دوبلجانجر" (القَرين).

مكافحة المعلومات المضللة

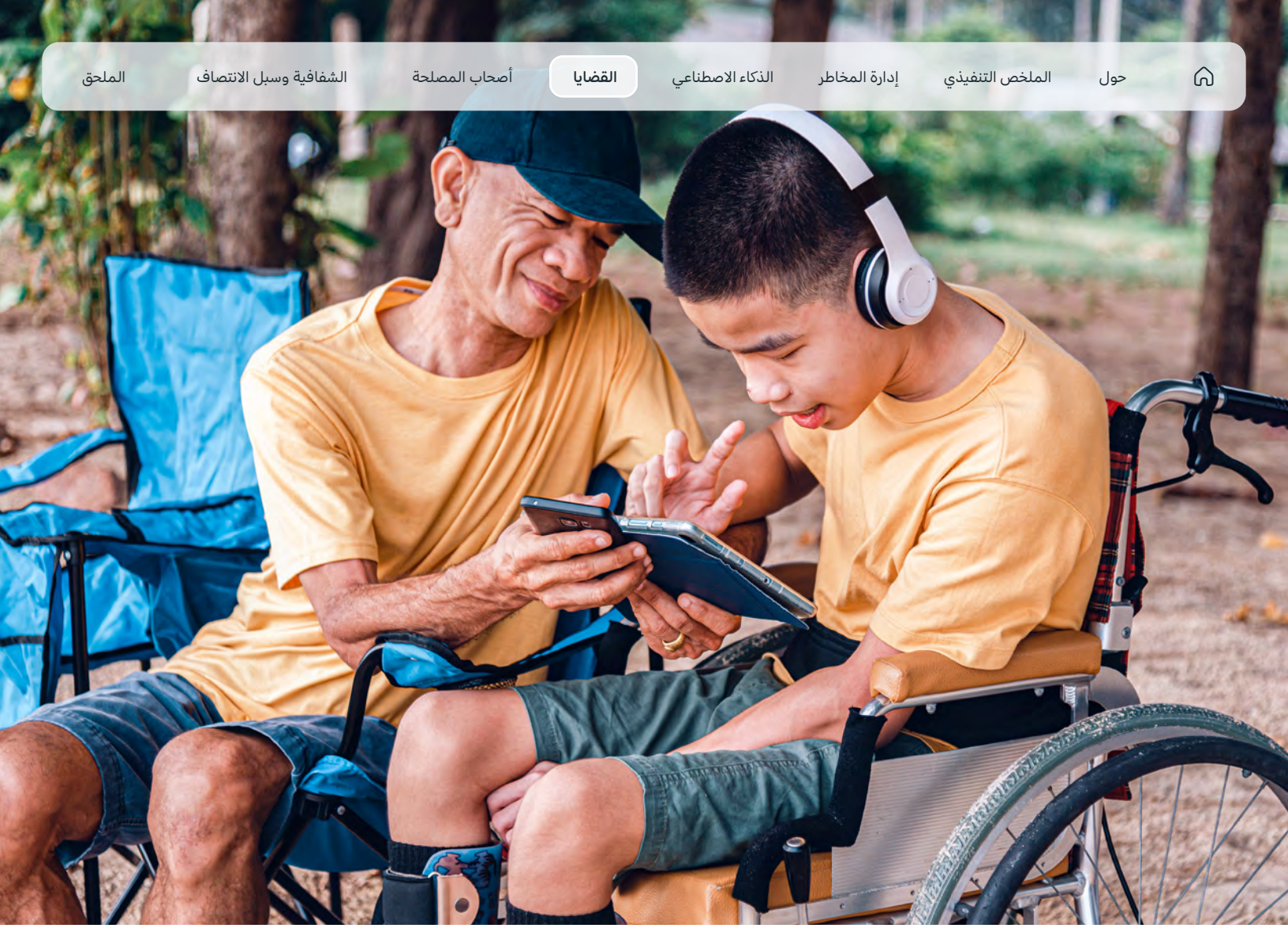
تعاوننا مع شبكة معايير تدقيق الحقائق الأوروبية للمساعدة في مكافحة الوسائط التي يتم إنشاؤها بواسطة الذكاء الاصطناعي أو تعديلها رقميًا، وكذلك لإطلاق حملة خاصة بالمعرفة الإعلامية لزيادة وعي الجمهور بالمخاطر المرتبطة بذلك.

مواجهة المخاطر المرتبطة بسوء استخدام تقنيات الذكاء الاصطناعي التوليدي

نتيجة للسياسات والإجراءات التي وضعناها للتعامل مع محتوى الذكاء الاصطناعي التوليدي، تم وسم ما يقارب 6000 إعلان عن قضايا اجتماعية أو انتخابات أو سياسة، وأكثر من 5.7 مليون جزء من محتوى على فيسبوك وInstagram في الاتحاد الأوروبي بإخلاء مسؤولية متعلق بالذكاء الاصطناعي حول انتخابات البرلمان الأوروبي، ما وفر شفافية محسّنة.

لعرض المزيد من التفاصيل، راجع مركز الشفافية.

الانتقال إلى مركز الشفافية



سلامة الأطفال والشباب

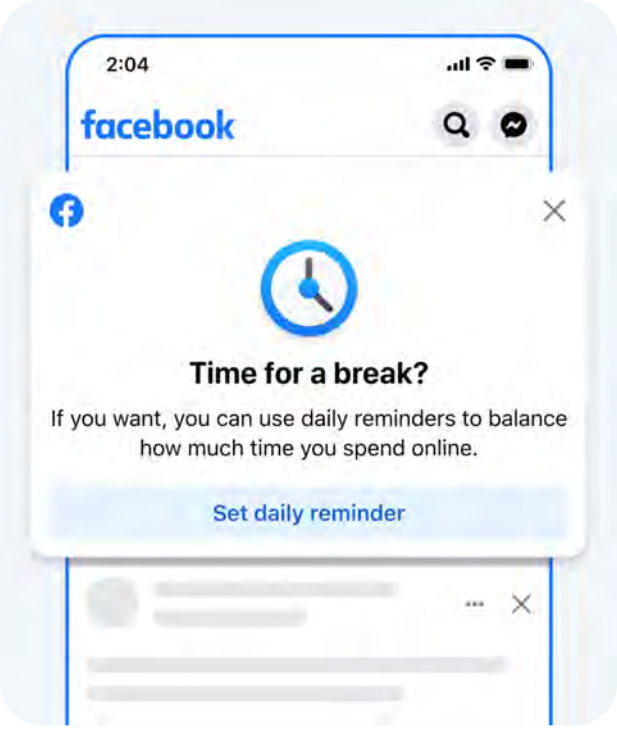
تُعد سلامة الأطفال على الإنترنت أولوية قصوى لدى Meta. نقدّم وسائل حماية مدمجة، بالإضافة إلى أدوات للمراهقين والوالدين لمساعدتهم على البقاء آمنين على تطبيقاتنا وخدماتنا.

وسائل حماية مدمجة للمراهقين

تتطلب حماية المراهقين على الإنترنت تعاونًا بين العديد من أصحاب المصلحة حول العالم، بما في ذلك الوالدين، وخبراء الأطفال، والأكاديميين، ونظراء المجال، والحكومات، والمجتمع المدني، وغيرهم. نحن ملتزمون بمساعدة المراهقين على الحفاظ على سلامتهم مع توفير مساحة لهم للتعبير عن أنفسهم والوصول إلى المعلومات تحت إشراف وتوجيه من الوالدين.

لقد طورنا على مدى سنوات أكثر من 50 أداة وموارد لدعم المراهقين ووالديهم والأوصياء عليهم، وقضينا أكثر من عقد في تطوير سياسات وتقنيات للتعامل مع المحتوى والسلوكيات المخالفة لقواعدنا.

في عام ٢٠٢٤، قمنا بتحديث سياساتنا وتصميم منتجاتنا لإنشاء تجربة فريدة وأكثر وضوحًا للمراهقين. تساعد هذه التحديثات المراهقين على رؤية المحتوى المناسب لأعمارهم استنادًا إلى إطار مصالح الطفل الفضلى. أضفنا إلى وسائل الحماية الموجودة على Instagram حسابات المراهقين على Instagram في الولايات المتحدة والمملكة المتحدة وكندا وأستراليا، مع خطط للتوسيع عالميًا لاحقًا. تأتي الحسابات المعاد تصميمها للمراهقين مع وسائل حماية مدمجة لتحديد من يمكنه التواصل مع المراهقين والمحتوى الذي يشاهدونه، بالإضافة إلى طرق لمساعدتهم في إدارة الوقت الذي يقضونه على التطبيق. توفر التغييرات أيضًا طرقًا جديدة للمراهقين لاستكشاف اهتماماتهم مع توجيه من الوالدين. تتماشى تجربة حسابات المراهقين الجديدة على Instagram مع الإرشادات الخبيرة ومبدأ القدرات المتطورة للطفل المنصوص عليه في اتفاقية الأمم المتحدة لحقوق الطفل.



لقد طورنا وأطلقنا عالميًا لوحة إشراف أبوي حيث يمكن للوالدين والأوصياء الذين يستخدمون أدوات الإشراف لدينا رؤية وإدارة حسابات أطفالهم في مكان واحد. يتيح ذلك للوالدين والأوصياء تحديد الضوابط عبر حساباتهم الخاصة لرؤية أي اتصالات غير مرغوب فيها أو محتوى غير مناسب وإدارتها، وتحديد حدود للوقت الذي يقضيه الأطفال أمام الشاشة.

كما أجرينا سلسلة من ورش العمل في الولايات المتحدة حول برنامج الاستخدام الذكي للشاشات لمساعدة الوالدين على إجراء محادثات مع أسرهم حول استخدام الأجهزة بأمان ومعرفة المزيد عن أدوات الإشراف الأبوي من Meta، باستخدام الحدود ووسائل الحماية التي تناسبهم بشكل أفضل.

"تساعد حسابات المراهقين الجديدة على Instagram من Meta الوالدين بشكل كبير على توجيه أبنائهم المراهقين دون المساس باستقلالية المراهقين الأكبر سنًا. تأتي الإعدادات الجديدة، جنبًا إلى جنب مع أدوات السلامة والخصوصية المحسّنة والنصائح، كخطوة كبيرة إلى الأمام."

- لاري ماجيد، الرئيس التنفيذي لمؤسسة ConnectSafely

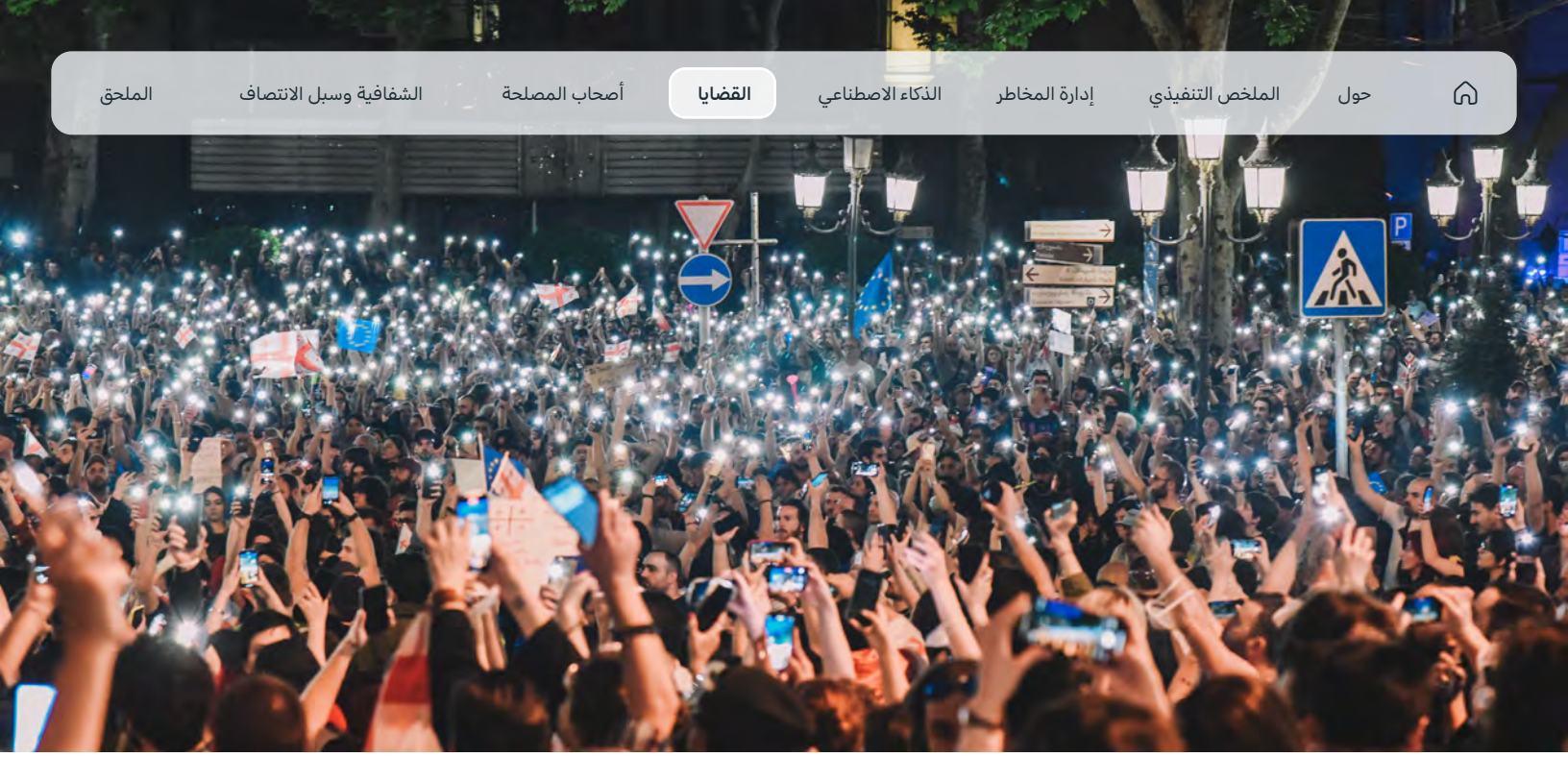
مكافحة الابتزاز الجنسي

تبقى مساعدة الأطفال على البقاء آمنين من المستخدمين الذين يحاولون إلحاق الضرر بهم أولوية قصوى بالنسبة لشركة Meta. في عام 2024، واصلنا العمل مع المركز الوطني للأطفال المفقودين والمستغلين (NCMEC) لتوسيع برنامج Take It Down ليشمل مزيدًا من البلدان واللغات، مما أتاح لعدد أكبر من المراهقين استعادة السيطرة على صورهم الحميمة. قمنا بتطوير أدوات جديدة للمساعدة في الحماية من الابتزاز الجنسي وجعل من الصعب على المحتالين والمجرمين المحتملين العثور على المراهقين والتفاعل معهم. كما أطلقنا حملة تعليمية، مستندة إلى خبرة NCMEC ومنظمة Thorn، لمساعدة المراهقين على التعرف على خدع الابتزاز الجنسي ومساعدة الآباء في دعم أبنائهم المراهقين لتجنّب هذه الخدع. توجّه الحملة المراهقين والآباء إلى نصائح خبراء، التي طوّرتها Thorn وتم تكييفها بواسطة Meta، لأي شخص يبحث عن دعم ومعلومات تتعلق بالابتزاز الجنسي.

نحن أيضًا أعضاء مؤسسون في Lantern، وهو برنامج تديره Tech Coalition يمكّننا وشركات التكنولوجيا الأخرى من مشاركة الإشارات المتعلقة بالحسابات والسلوكيات التي تنتهك سياسات حماية الأطفال الخاصة بنا. شاركنا إشارات خاصة بالابتزاز الجنسي إلى Lantern لتعزيز هذا التعاون المهم بين الجهات الفاعلة في المجال لمحاولة إيقاف خدع الابتزاز الجنسي عبر المنصات. في عام 2024، تضاعف عدد المشاركين في البرنامج، ليصل العدد الإجمالي للشركات المسجلة مع Lantern إلى 26.

انظر هنا للحصول على القائمة الكاملة لأدواتنا وميزاتها ومواردنا التي تساعد في دعم المراهقين والوالدين.

قراءة المزيد



كيفية الاستعداد للأزمات والاستجابة لها

نحن نستعد ونتعامل مع العديد من الأزمات، بما في ذلك النزاعات، والعنف داخل المجتمعات، والاضطرابات المدنية، والاحتجاجات الجماهيرية، والكوارث البيئية، فضلاً عن الهجمات الإرهابية وإطلاق النار في جميع أنحاء العالم. في عام 2024، بدأنا ونسقنا الاستجابات للأزمات في بنجلاديش، وجورجيا، وكينيا، وكاليدونيا الجديدة، ونيجيريا، وكوريا الجنوبية، والمملكة المتحدة وفنزويلا، من بين دول وأقاليم أخرى. واصلنا جهودنا للأزمات المحددة تحت بروتوكول سياسة الأزمات المتعلقة بالنزاعات في أوكرانيا، والسودان، والشرق الأوسط.

يعد بروتوكول سياسة الأزمات أداة رئيسية نستخدمها خلال فترات الأزمات. يوجه البروتوكول الاستخدام المعجّل للآليات للتقليل من الأضرار المحتملة في المجالات التالية:

- السياسة، مثل إصدار توجيهات إضافية للمراجعين. مثال على ذلك هو تقديم إرشادات حول الامتناع عن فرض عقوبات على بعض الانتهاكات المتعلقة بالمحتوى العنيف والصادم لتجنب معاقبة أو تقييد المستخدمين الذين يحاولون رفع الوعي بتأثيرات النزاع بشكل مفرط.
- المنتج، مثل تعديل تجربة المنتج. مثال على ذلك هو تغيير الإعدادات بحيث يمكن فقط للأصدقاء والعائلة التعليق على المنشورات.
- الأشخاص، بما في ذلك نقل الموارد للتركيز على قضايا محددة.

يساعدنا بروتوكول سياسة الأزمات على إجراء تقييم على أرض الواقع للمواقف التي قد تشكل مخاطر على المنصة. بمجرد تحديد الأزمة، نقوم بإجراء تقييم لتحديد المخاطر على المنصة وتحديد ما إذا كانت هناك حاجة إلى أي آليات إضافية. الأنواع المحددة من الاستجابات المتبعة تتماشى مع المخاطر الملاحظة. ويتم تحديدها استنادًا إلى التدخلات السابقة خلال الأزمات، ومبادى حقوق الإنسان، وقانون النزاع المسلح.

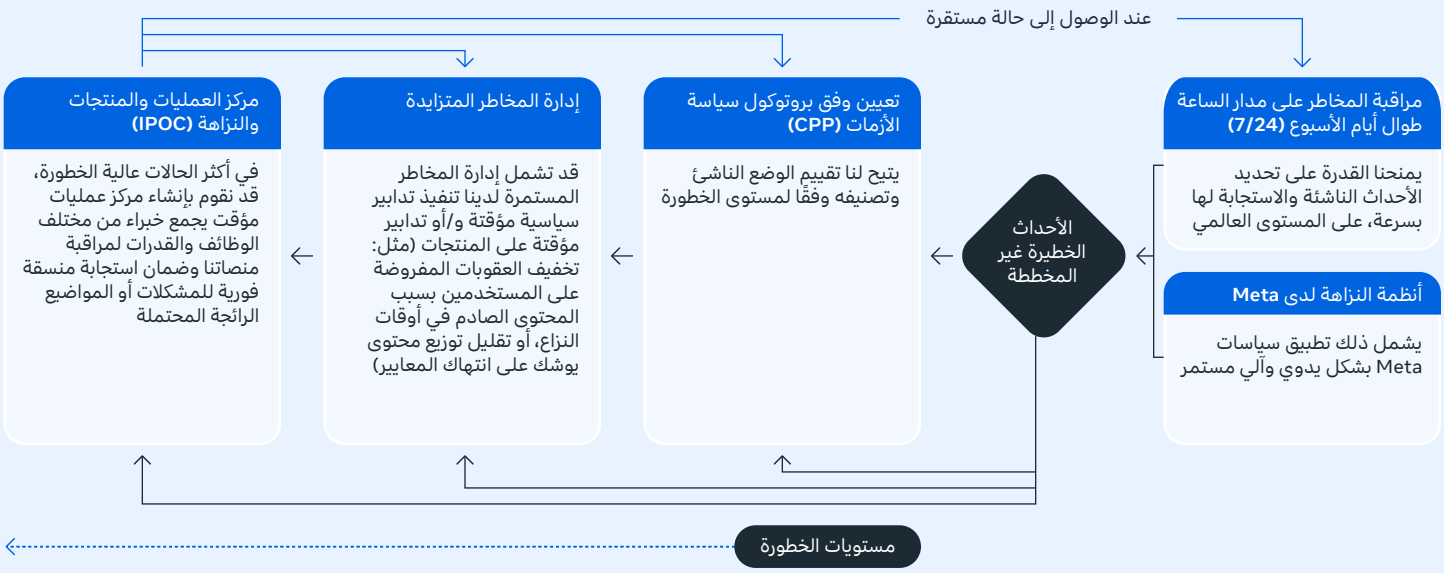
في الصفحة التالية، نوضح كيف نستعد للأزمات والنزاعات ونستجيب لها. ونقدم أمثلة على كيفية استخدام بروتوكول سياسة الأزمات والتنوع الجغرافي لجهودنا.

الاستعداد والاستجابة للأزمات والنزاعات[4]

يعد بروتوكول سياسة الأزمات وعملنا في الدول المعرضة للمخاطر أدوات رئيسية نستخدمها لمنع واكتشاف والتقليل من المخاطر. تقوم فرق المنتج والسياسات والعمليات لدينا بتقييم الديناميكيات المتغيرة على الأرض لتوجيه الاستجابات الفعّالة والمتناسبة.

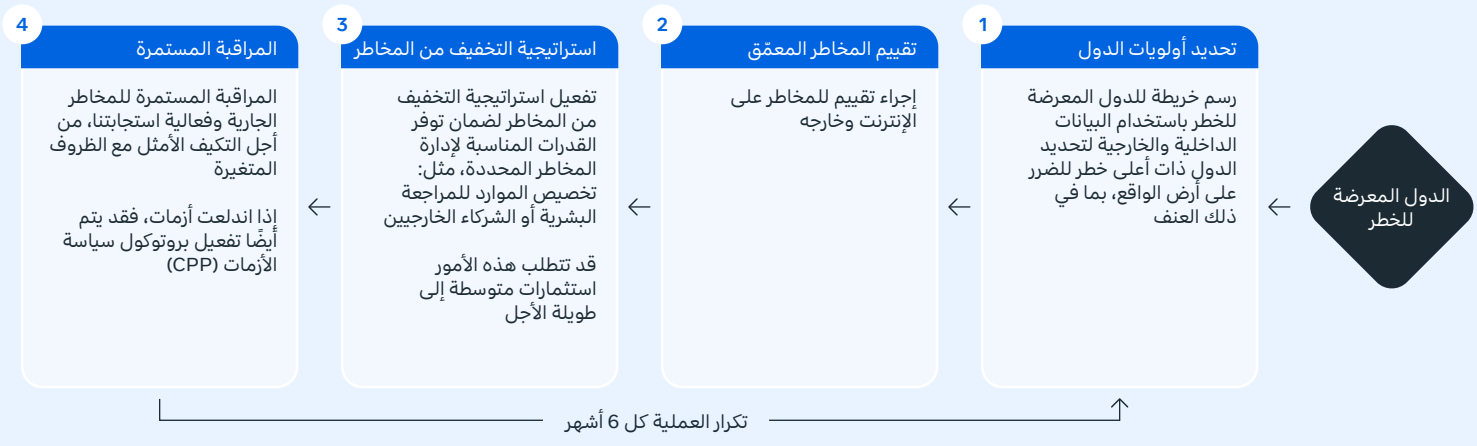
استجابة تفاعلية

كيف نضمن الاستجابة السريعة للأحداث الخطيرة غير المخطط لها؟



خطوات طويلة الأجل

ما التدابير طويلة الأجل التي نتخذها للحد من مخاطر النزاعات؟



[4] تشمل أعمالنا المتعلقة بالأزمات العديد من الحالات، بما في ذلك النزاعات، والعنف داخل المجتمعات، والاضطرابات المدنية، والاحتجاجات الجماهيرية، والكوارث البيئية، فضلاً عن الهجمات الإرهابية أو غيرها من الهجمات الإجرامية، في جميع أنحاء العالم.

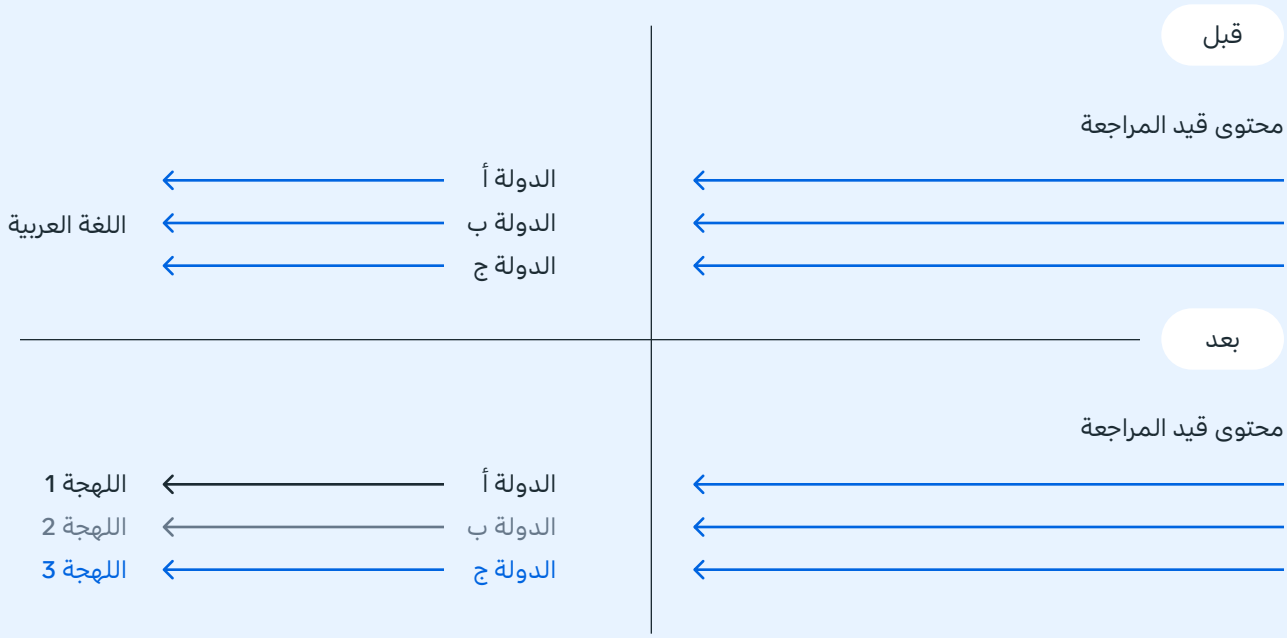
السودان

في عام 2024، تصاعد الصراع في السودان بين القوات المسلحة السودانية (SAF) وقوات الدعم السريع (RSF)، مما زاد من عدم استقرار البلاد والأزمة الإنسانية. ارتفع حجم المحتوى المخالف للسياسات، بما في ذلك العنف والتحريض، والإضرار المنسق، واستغلال البشر، والمنظمات والأفراد الخطيرين، مقارنة بما كان عليه قبل اندلاع الصراع، وظل مرتفعًا طوال العام.

لتقليل انتشار المحتوى المخالف للسياسات، بنينا على الإجراءات التي اتخذناها في 2023، مسترشدين ببروتوكول سياسة الأزمات الخاص بنا. مع استمرار الصراع، قمنا بتطبيق تدابير مؤقتة وتطوير حلول طويلة المدى لمعالجة مخاطر استمرار ارتفاع حجم المحتوى المخالف للسياسات.

إحدى هذه الحلول طويلة المدى كانت تصميم وبناء وإطلاق نظام يحدد اللهجات العربية المحددة ويعطي أولوية مراجعة المحتوى للمراجعين الأكثر قدرة على فهم الفروق اللغوية والسياق المحلي. كان النظام السابق يعتبر العربية لغة واحدة فقط ويوجه المحتوى إلى المشرفين القادرين على مراجعته. أما النظام الجديد، فيستطيع تحديد اللهجة العربية المستخدمة ويوجه المحتوى إلى المراجع الأكثر احتمالاً لفهمه. بالنسبة للسودان، أدى هذا التغيير إلى مراجعة حجم أكبر من المحتوى بدقة أعلى، وتقليل الأخطاء في تطبيق السياسات. استُند في هذا العمل إلى نتائج العناية الواجبة بحقوق الإنسان في إسرائيل وفلسطين، وبُني عليه.

التوجيه بناءً على الفروق اللغوية



خلال عام 2024، قام كلا طرفي الصراع بشكل متزايد بكشف هويات أسرى الحرب (POWs) على الإنترنت. كان كشف هوياتهم يزيد من مخاطر الأذى في العالم الواقعي ويُضعف حماية كرامة وسلامة أسرى الحرب وفقًا لاتفاقية جنيف المتعلقة بمعاملة أسرى الحرب. مسترشدين بتوصية من مجلس الإشراف صادرة في 2023 ومكررة في 2024 في قضية فيديو أسرى قوات الدعم السريع في السودان، أدركنا أيضًا أن بعض محتوى أسرى الحرب قد يكون في مصلحة الجمهور، على سبيل المثال رفع الوعي بشأن انتهاكات حقوق الإنسان المحتملة أو المساعدة في العثور على أسرى مفقودين. وبناءً على ذلك، قدمت Meta إرشادات لمراجعي المحتوى بشأن أسرى الحرب ضمن سياسة تنسيق الأنشطة الضارة والترويج للجريمة، بحيث يكونون أكثر قدرة على التعامل مع المحتوى المخالف المحتمل في المنطقة وعلى نطاق واسع.

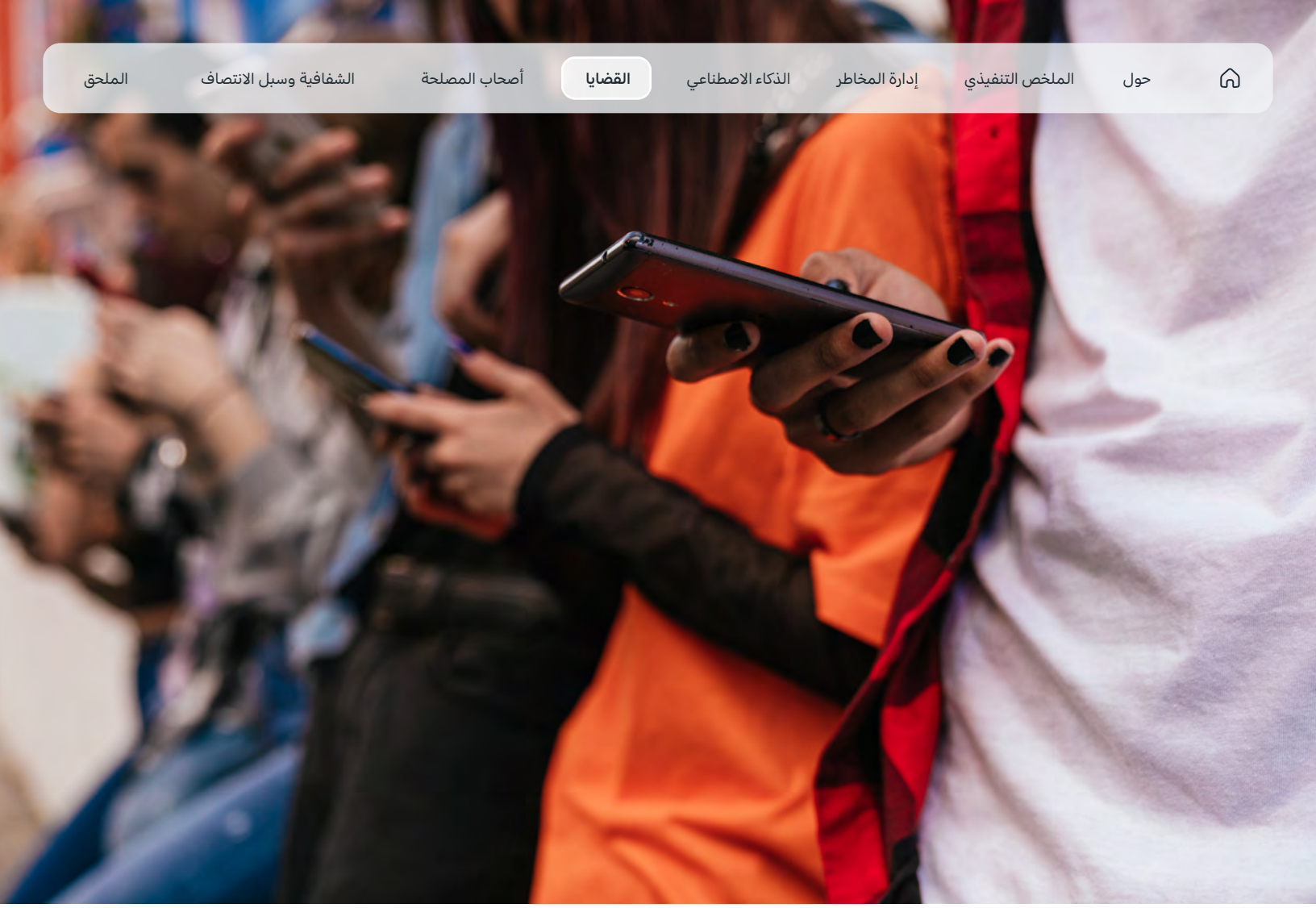
↖ قراءة تقرير مجلس الإشراف عن نصف السنة الأول من عام 2024

↖ عرض حالة مجلس الإشراف

لعب الشركاء الموثوق بهم دورًا حاسمًا من خلال تقديم رؤى مهمة حول التطورات المحلية والمحتوى المحتمل المخالف المتعلق بالصراع. ساعدت هذه الرؤى في تطبيق سياسات Meta ذات الصلة، بما في ذلك: السلوك الذي يحض على الكراهية والمضايقة والإساءة، واستغلال البشر، وتحديد الادعاءات المحتملة الضارة التي تمت مراجعتها مسبقًا ضمن سياسة المعلومات المضللة والأذى، وساهمت في النهاية في خلق بيئة أكثر أمانًا على الإنترنت. بموجب سياسة استغلال البشر، تمكنا من تحديد المخاطر المحتملة المتعلقة بصور الأطفال المجندين واستقطابهم، وإزالة هذا المحتوى، وتقليل معدل انتشاره.

قمنا بإجراء جلسات تدريبية مع المدافعين عن حقوق الإنسان، والصحفيين، والمنظمات الوطنية ومنظمات الشتات لمساعدتهم على تثقيف المستخدمين السودانيين، بما في ذلك المهاجرون واللاجئون. تركزت هذه الجلسات على سياسات المحتوى، والأمن الرقمي، وتعزيز حضورهم على منصات Meta.

تؤدي النزاعات المسلحة إلى نزوح جماعي للسكان، لذا ركزنا في السودان على تحديد استغلال البشر المحتمل، بما في ذلك تهريب البشر والاتجار بهم، والاستغلال الجنسي للنساء والفتيات، والزواج القسري. قمنا بحذف أكثر من 19100 منشور جماعي يقدم خدمات تهريب البشر، وأزلنا المحتوى الذي يمجد الزواج القسري. ظل التواصل مع الشتات استراتيجية مهمة لكشف اتجاهات المحتوى وتوفير الوضوح لمراقبة كل بلد. شمل ذلك أكثر من 30 تفاعلاً فريدًا للمساعدة في حماية المستخدمين على منصاتنا، بما في ذلك جهود تحديد كلمات وعبارات جديدة وناشئة مرتبطة بخطاب يحض على الكراهية وتهريب البشر. مكّننا هذه الرؤى من الكشف عن المحتوى المخالف لسياساتنا والاستجابة له بشكل أكثر فعالية.



الشرق الأوسط

ظلّ الصراع في الشرق الأوسط أولوية بالنسبة لـ Meta. في عام 2024، ركزنا على المخاطر الناجمة عن العنف المستمر في إسرائيل وقطاع غزة، خاصة مع توسع الحرب إلى المنطقة الأوسع وتورط أطراف إقليمية أخرى وتصعيدهم للصراع. عملنا على ضمان إمكانية استخدام منصاتنا للتعبير عن الرأي بحرية، مع السعي في الوقت نفسه لمنع انتشار المحتوى الذي يحض على الإرهاب أو العنف أو أي أضرار حقيقية أخرى.

لم يتغير نهجنا الأساسي مقارنة بعام 2023. وشمل هذا النهج الحفاظ على تصنيف هجوم 7 أكتوبر 2023 من قبل حركة حماس كـ "هجوم إرهابي" بموجب سياسة المنظمات الخطرة والأفراد الخطرين (DOI) والتعامل مع المحتوى المخالف وفق سياساتنا. كما أوقفنا التغييرات المؤقتة في المنتج التي تم إدخالها في 2023.

في أعقاب الهجمات الإرهابية في 7 أكتوبر مباشرةً، صنفت Meta العنف والصراع الناتج عند أعلى مستوى في بروتوكول سياسة الأزمات وطبقت إجراءات استجابة عاجلة للأزمات. شملت الإجراءات المتخذة فريقًا مخصصًا متعدد الوظائف يعمل على مدار الساعة، وطوال أيام الأسبوع، وتدابير مؤقتة في المنتج والسياسة. استندنا إلى مبادئ الأمم المتحدة التوجيهية بشأن الأعمال التجارية وحقوق الإنسان كأساس لسياسة حقوق الإنسان الخاصة بشركتنا، وإلى أعمال التحقق التي قمنا بها في 2022 لتوجيه نهجنا. يمكن الاطلاع على تفاصيل استجابتنا في تقرير حقوق الإنسان لعام 2023 ومنشورات غرفة الأخبار.

طوال عام 2024، واصلنا التفاعل مع مجموعة متنوعة من الجهات الحكومية ومنظمات المجتمع المدني وغيرهم في إسرائيل والدول العربية في الشرق الأوسط، وكذلك على الصعيد العالمي، لإظهار الشفافية وسرعة الاستجابة. كما استجبنا للعديد من حالات مجلس الإشراف.

بالإضافة إلى ذلك، واصلنا تنفيذ التوصيات الواردة في تقرير تقييم حقوق الإنسان 2022. قدمنا تقارير عن تقدم Meta للفترة من 30 يونيو 2023 إلى 30 يونيو 2024، بما في ذلك زيادة مواردنا للإشراف على المحتوى باللغة العبرية وتحديث على سياسة DOI للسماح بمزيد من النقاش الاجتماعي والسياسي في بعض الحالات. جاء هذا استجابةً للملاحظات التي أشارت إلى أن سياسة DOI غالبًا ما كانت تشمل محتوى مثل التقارير الإخبارية، والمناقشات المحايدة حول الأحداث الجارية، أو حتى الإدانة للجماعات الإرهابية وجماعات الكراهية. يظل المحتوى الذي يمدح أو يدعم المنظمات الخطرة أو الأفراد الخطيرين أو أفعالهم العنيفة أو مهامهم محظورًا.

بين 30 يونيو 2024 و30 يونيو 2025، أطلقنا نظامًا يكتشف المحتوى ويعطي أولوية للمشرفين الأكثر قدرة على فهم اللهجة العربية المحددة. كما قمنا بتحديث قناة تصعيد الشركاء الموثوق بهم، مما أدى إلى تحسين الاستجابة السريعة للتصعيدات. يمكن الاطلاع على تفاصيل تقدمنا في هذه الفترة في التحديث النهائي ديسمبر 2025: العناية الواجبة بحقوق الإنسان في إسرائيل وفلسطين.

التقدم الخاص بعام 2023 ↖ التقدم الخاص بعام 2024 ↖

بنجلاديش

ساعدتنا الاستعدادات لانتخابات يناير 2024 على التنبؤ بالعديد من المخاطر ومعالجتها خلال فترة إعداد التقارير. كان هدفنا حماية المستخدمين مع دعم قدرتهم على التصويت والتعبير عن آرائهم. مكنت هذه التحضيرات الانتخابية فريقنا من الاستجابة في منتصف العام لاحتجاجات الطلاب، والحملة العنيفة ضدهم، والتغيير اللاحق في الحكومة.

نظرًا لشدة الاضطرابات، قمنا بتفعيل بروتوكول سياسة الأزمات. حددنا بشكل استباقي المخاطر، بما في ذلك: خطاب يحض على الكراهية، والتحريض الموجه ضد الأقليات الدينية، والمعلومات المضللة، والسلوك الزائف المنظم. طبقنا تدابير تخفيف المخاطر، بما في ذلك: استخدام سياسة المواقع عالية الخطورة المؤقتة، والتعاون مع الشركاء الموثوق بهم وشبكة تدقيق الحقائق بواسطة جهات خارجية، وتطبيق حماية معززة على حسابات المدافعين عن حقوق الإنسان.

وشملت إجراءاتنا الأخرى:

- إعداد إشارات دقيقة للكشف عن زيادة المحتوى المخالف الذي يمكن التعامل معه في الوقت الفعلي، مثل العنف المصوّر والسلوك الذي يحض على الكراهية.
- تطبيق أدوات وتقنيات، بما في ذلك اكتشاف الذكاء الاصطناعي، لتحديد المحتوى المخالف والبحث عن الكلمات الرئيسية.
- تعيين مطالبات محتملة ضارة تمت مراجعتها مسبقًا ضمن سياسة المعلومات المضللة والإيذاء.

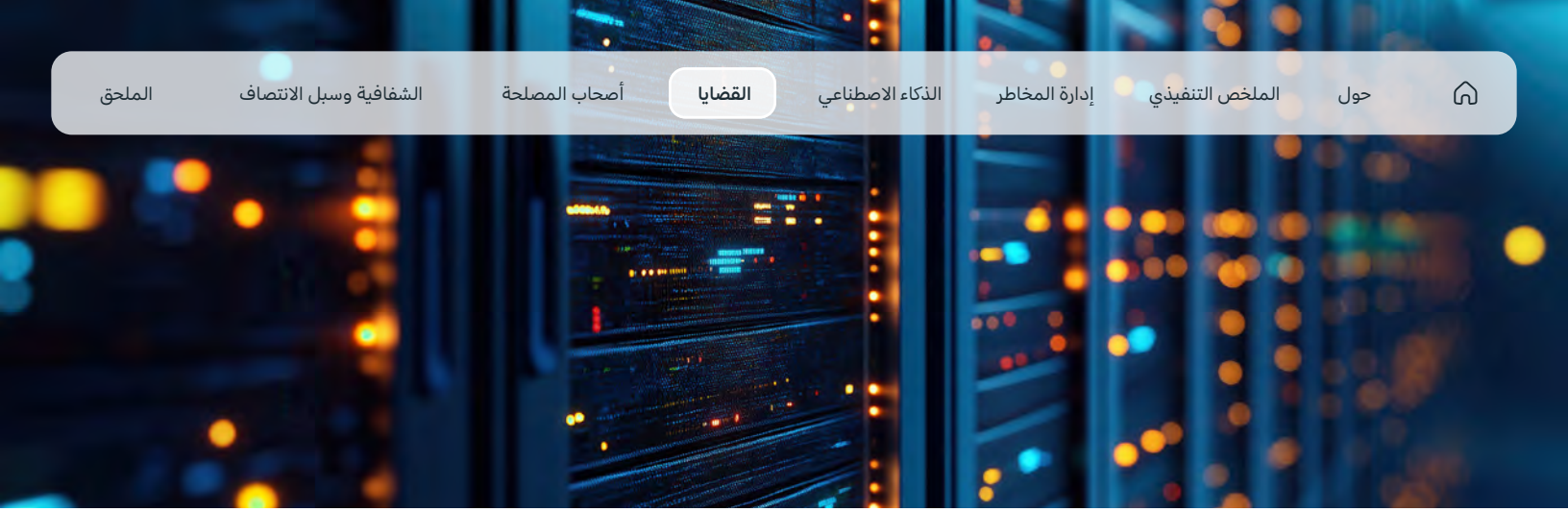
لم نمتثل لطلبات إزالة المحتوى من الحكومة المتعلقة بالاحتجاجات، إذا كان هذا المحتوى يتعارض مع المعايير الدولية لحقوق الإنسان. كان ذلك تماشيًا مع التزاماتنا كعضو في مبادرة الشبكة العالمية وسياسة حقوق الإنسان الخاصة بشركتنا.

جورجيا

طبقنا بروتوكول سياسة الأزمات مرتين في جورجيا خلال 2024. في مارس 2024، تم تفعيله لأول مرة بعد سلسلة من التظاهرات الجماهيرية المعارضة لمشروع قانون شفافية النفوذ الأجنبي. وأعدنا تفعيله في ديسمبر 2024 بعد الانتخابات الوطنية، عندما وقعت تظاهرات جماهيرية أخرى ترافقها تصاعد أعمال العنف من الشرطة وقوات الأمن الأخرى.

أتاح تطبيق بروتوكول سياسة الأزمات لفريقنا تعزيز جهود تخفيف المخاطر، والتعامل مع زيادة المحتوى المخالف، ومواجهة المخاطر المتزايدة للعنف الجسدي، والمساعدة في حماية المدافعين عن حقوق الإنسان. قمنا بمراجعة قائمة الإهانات، وهي الكلمات التي تُستخدم تاريخيًا لمهاجمة مجموعات معينة، لتحديد المحتوى الذي يحض على الكراهية وإدارته عبر منصاتنا. أزلنا الحسابات الزائفة المصممة للتلاعب بالرأي العام أو توزيع محتوى قد يكون ضارًا. كما قمنا بتعطيل شبكة السلوك الزائف المنظم (CIB) التي استهدفت جورجيا، إلى جانب حسابات غير أصيلة أخرى.

طوال فترات الأزمة، تعاوننا مع منظمات المجتمع المدني، وجهات تدقيق الحقائق، والشركاء الموثوق بهم، الذين ساعدونا على فهم الوضع المتطور وتسهيل تبادل المعلومات مع المجتمع المدني والمعارضة في جورجيا. قدّم الشركاء الموثوق بهم إشارات ورؤى حاسمة حول المحتوى المخالف الذي يستهدف مجموعات المعارضة. كما تواصلنا معهم لمساعدتهم على فهم أفضل لما يشكل محتوى مخالفًا وفقًا لمعايير مجتمعنا. بالإضافة إلى ذلك، تواصلنا مع شركاء المجتمع المدني لتحديد المدافعين عن حقوق الإنسان المعرضين للخطر لضمان توفير حماية معززة لحساباتهم.

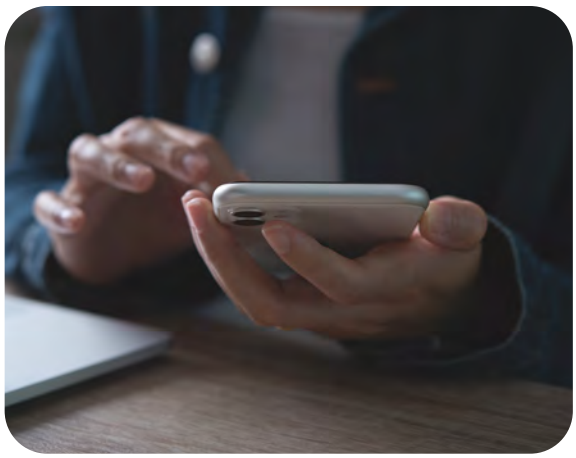


الأمن السيبراني

تُعد سياساتنا الأمنية بالغة الأهمية لتمكين حقوق المستخدمين في حرية التعبير، والوصول إلى المعلومات، والخصوصية، وغيرها من الحقوق. واصلنا العمل على مستوى الشركة بأكملها لتحديد التهديدات العدائية التي تستهدف منصاتنا والتصدي لها، بما في ذلك عمليات التأثير، والتجسس السيبراني، والمراقبة، وعمليات الاحتيال والخداع. ويُعد تعطيل الشبكات العدائية التي تنخرط في أنشطة ضارة عنصرًا أساسيًا في عملنا الأمني.

في عام 2024، قمنا بإزالة 20 شبكة ذات سلوك زائف منظم (CIB) في الشرق الأوسط وآسيا وأوروبا والولايات المتحدة لانتهاكها سياسة السلوك الزائف المنظم لدينا. وهذه الشبكات كانت تعمل على التلاعب بالنقاش العام لتحقيق أهداف استراتيجية، من خلال استخدام حسابات زائفة أو أساليب مضللة. نراقب محاولات الشبكات التي سبق إزالتها لإعادة التشكّل على منصاتنا ونتخذ إجراءات لإزالتها، وننشر معلومات علنًا عبر تقارير التهديدات الخاصة بنا، كما نعمل على إدماج الدروس المستفادة من التحقيقات في أنظمة الرصد لدينا وتصميم منتجاتنا، بما يعزز قدرتها على الصمود.

واصلنا كذلك رصد وإزالة شبكات CIB التي استهدفت و/أو انتحلت صفة جماعات عرقية أو دينية محددة. وفي عام 2024، كان أحد الأمثلة من بين العديد من الحالات شبكة نشأت في بنجلاديش، أُزيلت بسبب سلوك زائف منظم. وقد استهدفت جمهورًا محليًا باستخدام حسابات زائفة لنشر المحتوى وإدارة الصفحات. وانتحلت هذه الشبكة صفة جهات إخبارية وهمية واستخدمت أسماء مؤسسات إعلامية قائمة لنشر محتوى معادٍ للحزب الوطني البنجلاديشي وداعم للحزب الحاكم. وقد ارتبطت هذه العملية بأفراد على صلة بحزب رابطة عوامي ومنظمة غير هادفة للربح في بنجلاديش.



وشمل مثال آخر شبكة نشأت في الصين استهدفت المجتمع السيخي عالميًا، مستخدمة حسابات مخترقة وزائفة لانتحال صفة سيخيين والترويج لحركة ناشطة وهمية تُدعى عملية K، دعت إلى تنظيم احتجاجات مؤيدة للسيخ، بما في ذلك في نيوزيلندا وأستراليا. واستخدمت هذه العملية صورًا ومنشورات تم إنشاؤها بواسطة الذكاء الاصطناعي، باللغتين الإنجليزية والهندية، حول فيضانات في إقليم البنجاب، والمجتمع السيخي حول العالم، وحركة استقلال خالستان، واغتيال هارديب سينغ نيجار، وانتقادات

قراءة المزيد

للحكومة الهندية.

وفي إطار جهودنا لإنفاذ السياسات ضد شركات برامج التجسس، قمنا بتعطيل وإزالة أنشطة شركة Paragon Solutions، وهي مورّد لبرامج التجسس، كانت تستهدف عددًا من مستخدمي واتساب، بمن فيهم صحفيون وأفراد من المجتمع المدني. وتواصلنا مع مستخدمي واتساب الذين يُحتمل تأثرهم، ووفّرنا لهم موارد حول كيفية حماية أنفسهم. كما قدّمنا لهم معلومات عن The Citizen Lab في جامعة تورنتو، الذي يوفّر موارد إضافية لأفراد المجتمع المدني. وفي عام 2024، كنّا من الموقّعين المؤسسين على مذكرة Pall Mall Memorandum، وهي مبادرة متعددة الجنسيات تهدف إلى كبح سوء استخدام برامج التجسس.

وفي ديسمبر 2024، قضى قاضٍ فيدرالي أمريكي بمسؤولية NSO Group عن انتهاك قوانين الولاية والقوانين الفيدرالية، وعن خرق شروط خدمة واتساب. وكانت هذه المرة الأولى التي تُحمَّل فيها شركة برامج تجسس المسؤولية بموجب القانون الأمريكي. وقد رفعت Meta وواتساب هذه الدعوى في عام 2019 ضد شركة NSO Group، التي قامت بالوصول غير المصرح به إلى خوادم واتساب من أجل تثبيت برنامج التجسس Pegasus على الأجهزة المحمولة لأكثر من 1400 مستخدم لواتساب، من بينهم صحفيون، ونشطاء في مجال حقوق الإنسان، ومعارضون سياسيون، وغيرهم.

20

شبكة CIB تمت إزالتها

التفاعل مع أصحاب المصلحة

يساعد التفاعل الاستباقي والمنهجي مع مجتمع مستخدمينا العالمي في تشكيل سياسات Meta، ويُعد عنصرًا محوريًا في إدارة مخاطر حقوق الإنسان لدينا.

وخلال عام 2024، تواصلنا مع نطاق واسع من أصحاب المصلحة، شمل أفرادًا من المجتمع المدني، وأكاديميين، ومراكز فكر، وخبراء في حقوق الإنسان، وجهات تنظيمية. وشملت القضايا السياسية الرئيسية نهجنا تجاه الذكاء الاصطناعي المسؤول (AI) ونزاهة الانتخابات، بالإضافة إلى إشارات التعيين للمنظمات والأفراد الخطرين، والأحداث العنيفة.

فعلى سبيل المثال، ولتقييم مدى ملاءمة سياستنا المتعلقة باستخدام كلمة "صهيوني" (Zionist)، أجرينا مشاورات مع 145 جهة من أصحاب المصلحة من المجتمع المدني والأوساط الأكاديمية على مستوى العالم. وضمت قائمة المشاركين علماء سياسيين ومؤرخين وعلماء قانون وجماعات حقوق رقمية ومدنية، ودعاة حرية تعبير، وخبراء حقوق إنسان. كما تفاعلنا مع أصحاب مصلحة آخرين، من بينهم منظمات غير حكومية ضمن برنامج الشركاء الموثوق بهم لدينا، إضافة إلى مجموعة واسعة من مجتمعات الشتات التي تمثل وجهات نظر مختلفة.

وفي عام 2024، أنشأنا مجموعات عمل معنية بالصوت وحرية التعبير بالتعاون مع منظمات مجتمع مدني محلية في منطقتي دول أفريقيا الواقعة جنوب الصحراء الكبرى والشرق الأوسط وشمال أفريقيا، لفهم شواغلهم إزاء المقترحات التشريعية في المملكة العربية السعودية، والأردن، ونيجيريا، والسنغال، وغيرها من الدول. وخلال هذه الجلسات، ناقشنا سبل حماية الوصول إلى منصاتنا مع التعامل في الوقت نفسه مع قيود المحتوى المفروضة بموجب القوانين المحلية، وبما يتوافق مع التزاماتنا ضمن مبادرة الشبكة العالمية في دعم حرية التعبير وخصوصية المستخدمين.

كما أطلقنا بشكل تجريبي برنامج لجان حقوق الإنسان، شاركت فيه مؤسسات وطنية لحقوق الإنسان من إثيوبيا، وغانا، وكينيا، ونيجيريا، وجنوب أفريقيا، وركّز على كيفية تعامل Meta مع المحتوى الضار المحتمل وتنظيم المحتوى عبر الإنترنت.

إضافة إلى ذلك، عقدنا ورش عمل للاستجابة للنزاعات في إثيوبيا، وفلسطين، والصومال، والسودان، وتونس. وقمنا بتدريب مدافعين عن حقوق الإنسان وصحفيين من دول تشهد انتخابات، مع تزويدهم بالأدوات اللازمة لحماية حضورهم الرقمي.

ومن خلال برنامج Open Loop India وأعمال Open Loop Sprint، تعاونّا مع شركات، وصنّاع سياسات، وخبراء في الذكاء الاصطناعي لإنتاج رؤى حول دور إشراك أصحاب المصلحة عبر دورة حياة الذكاء الاصطناعي وسلسلة القيمة الخاصة به.

نهجنا في تفاعل أصحاب المصلحة

دمج مجموعة واسعة من وجهات النظر والخبرات: استخلاص رؤى مهمة وإشراك خبراء متخصصين في جميع المناطق، بما يضمن تنوع الرؤى العالمية ومراعاة الخصوصيات والسياقات المحلية.

توفير الشفافية: مناقشة التحديات ومجالات التحسين مع المساهمين الخارجيين.

إنشاء حلقة ملاحظات: إظهار كيفية تطور سياساتنا وتحسنها بمرور الوقت.

بناء الثقة: تعزيز مشروعية سياساتنا وآليات الإنفاذ المتبعة لدينا.

464

مساهم من **34** دولة قدموا المساهمات في **6** عمليات دفق عمل لمنتديات السياسة.

121

مساهم قدموا المساهمات في أعمال تطوير أخرى في Meta.

أكثر من 100

مساهم انخرطوا في إحاطات متعلقة بالانتخابات، مع إنتاج **7** نشرات إخبارية للانتخابات.

أكثر من 290

صحفي ومدافع عن حقوق الإنسان وناشطًا تم تدريبهم.

دورة تطوير السياسات في Meta

مراجعة مستمرة

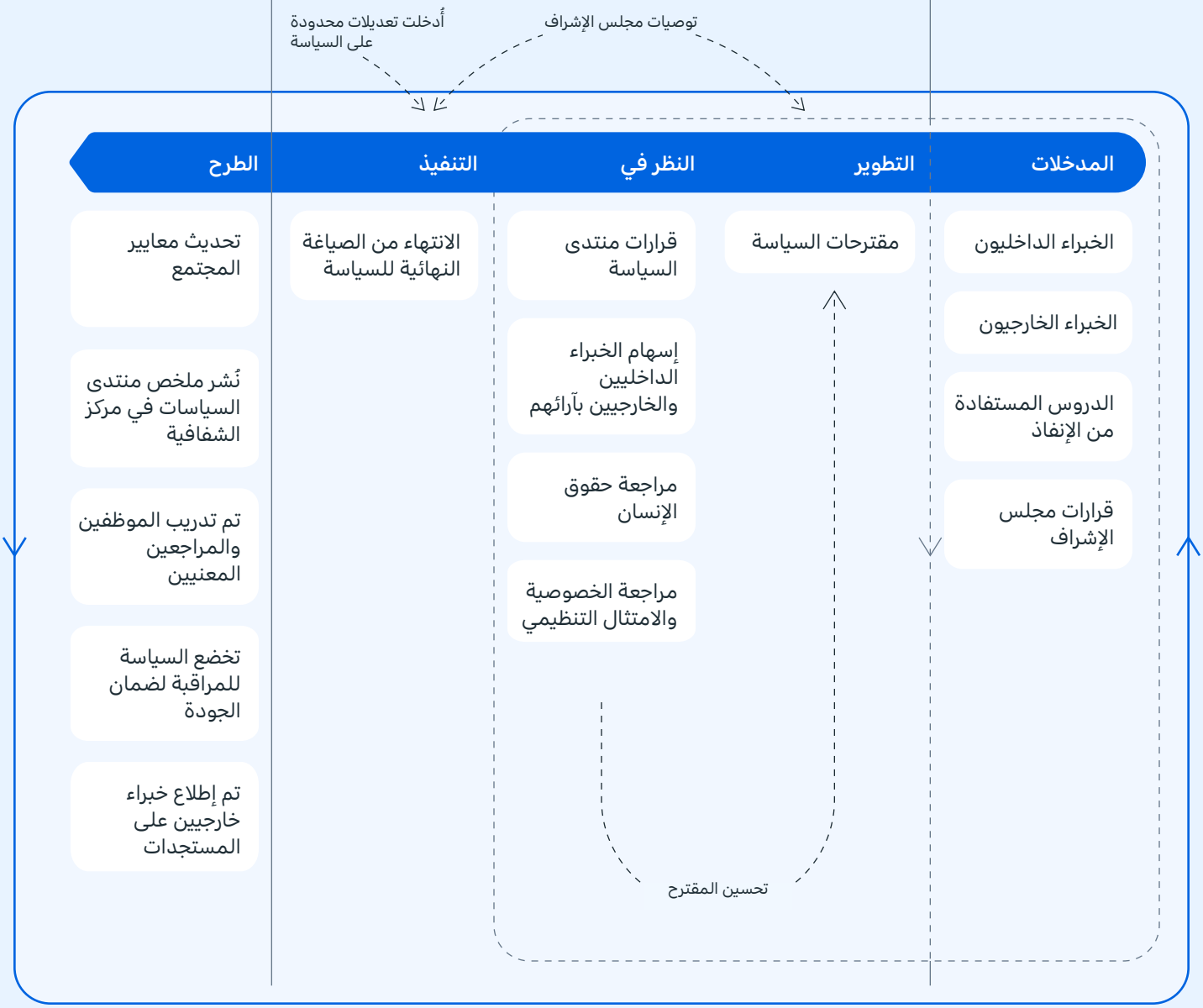
نراجع سياساتنا بشكل مستمر استنادًا إلى مدخلات واردة من مجموعة متنوعة من المصادر.

التطوير

تمر المقترحات بعملية تطوير دقيقة وصارمة لضمان اتساقها مع المبادئ المعتمدة، وقابليتها للتنفيذ، وسهولة شرحها وتوضيحها.

الطرح

يتم تحديث أنظمة الإنفاذ ودخول السياسات حيّز التطبيق على خدماتنا.



منتديات السياسة

نسعى إلى صياغة سياسات تحترم حقوق الإنسان وتتبنّى تنوّع وجهات النظر، بما يتيح سماع الآراء والمعتقدات المختلفة وعكسها ضمن سياساتنا. ويُعد منتدى السياسات في Meta اجتماعًا دوريًا يناقش فيه خبراء متخصصون التغييرات المحتملة على معايير المجتمع ومعايير الإعلانات. وتشمل هذه الاجتماعات اقتراح سياسات جديدة أو تعديل سياسات قائمة، وذلك وفق عملية تطوير سياسات تتضمن تفاعلاً موسعًا مع أصحاب مصلحة عالميين، إلى جانب مراجعة أبحاث خارجية وداخلية.

وخلال عام 2024، عقدنا ستة منتديات سياسات تناولت الموضوعات التالية:

1. استخدام مصطلح "صهيوني" كمصطلح بديل يُستخدم للدلالة على سلوك يحض على الكراهية
2. الأحداث العنيفة المخالفة للمعايير
3. المحتوى التجاري الذي قد ينطوي على مخاطر صحية أو متعلقة بالسلامة
4. إزالة الصور الحساسة
5. المحتوى المرتبط باضطرابات الأكل
6. عبارات التعزية المتعلقة بأفراد مصنّفين كجهات أو أشخاص خطرين

منتديات المجتمع

ترتكز منتديات مجتمع Meta على الحوكمة التداولية، وقد صُمِّمت للاستفادة من مدخلات الجمهور في القضايا التي تنطوي على توازنات متضاربة ولا توجد لها إجابات واضحة. ويهدف نهجنا إلى تمكين الأصوات من خارج الشركة من أن يكون لها دور أكبر في عملية اتخاذ القرار، كما يتيح لنا فهم كيفية تطور الرأي العام مستقبلاً.

في عام 2024، عقدت Meta منتدى مجتمعيًا، بالشراكة مع Stanford's Deliberative Democracy Lab، ركّز على المبادى التي يرغب المستخدمون في أن تقوم عليها عملية تطوير وكلاء الذكاء الاصطناعي. وشارك في المنتدى نحو 1000 شخص من الهند ونيجيريا والمملكة العربية السعودية وجنوب أفريقيا وتركيا. ويتوافر تقرير مفصّل هنا.

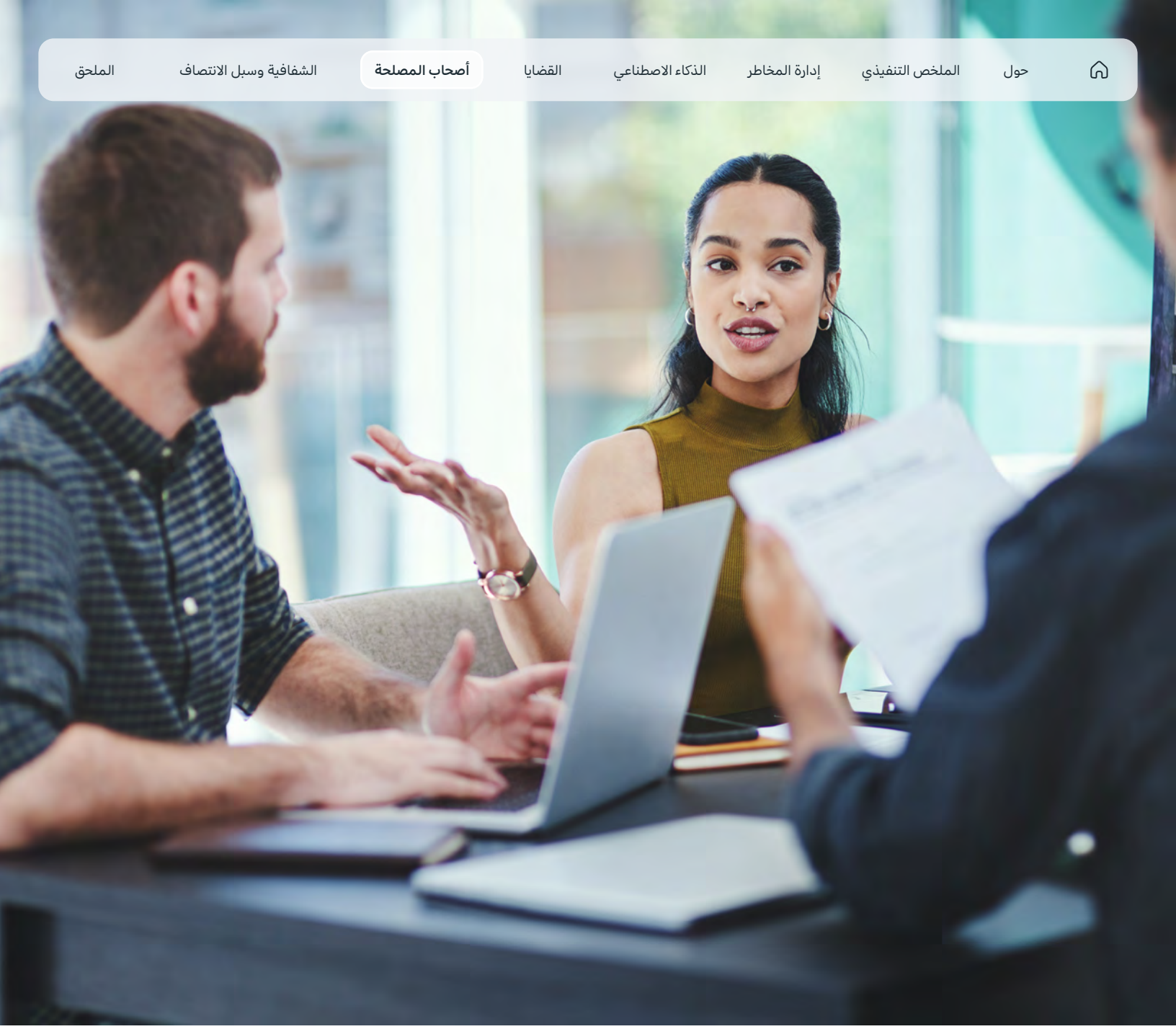
وكجزء من المنتدى، أُتيح للمشاركين الاستماع مباشرةً إلى خبراء مختصين، والتداول فيما بينهم، وتقديم ملاحظات قيّمة إلى Meta. وقد مكّن الأسلوب التداولي المشاركين من التعامل مع أوجه التوتر الكامنة في تقديم تجارب مخصّصة، من خلال الموازنة بين قيمة التخصيص والمفاضلات المصاحبة له، مثل جمع البيانات وتخزينها.

نهجنا في ضوابط المستخدم والتجارب المخصّصة

وقد أسهمت النتائج في توجيه نهجنا المتعلق بضوابط المستخدم والتجارب المخصّصة مع وكلاء الذكاء الاصطناعي. وشملت هذه النتائج ما يلي:

- أيّد المشاركون أن يقوم وكلاء الذكاء الاصطناعي بتذكّر محادثاتهم السابقة من أجل تخصيص التجربة، لا سيما عند توفر الشفافية وضوابط المستخدم.
- أبدى المشاركون دعمًا أكبر لوكلاء ذكاء اصطناعي مكيّفين ثقافيًا أو إقليميًا مقارنة بالوكلاء الموحّدين.
- أيّد المشاركون وكلاء ذكاء اصطناعي ذوي طابع شبيه بالبشر، قادرين على الاستجابة للإشارات العاطفية.

وبالإضافة إلى ذلك، أطلقنا مشروعًا تجريبيًا لإشراك الجمهور في تحديد ما الذي يسهم في إنشاء نموذج ذكاء اصطناعي ذي صلة ثقافية، وتطوير مجموعات بيانات للتفضيلات استنادًا إلى هذه الملاحظات، وإتاحة هذه البيانات كمصدر مفتوح لاستخدام المطورين. وستتمثل النتيجة في مجموعة بيانات متاحة بسهولة، تُسهم في جعل نموذج اللغة الكبير Llama أكثر ملاءمة وفائدة في سياقات ثقافية مختلفة.



الشركاء الموثوق بهم

واصلنا التواصل مع الشركاء الموثوق بهم لتحديد المواضيع الرائجة وفهم أفضل لتأثير المحتوى والسلوك عبر الإنترنت على المجتمعات المحلية واستكشاف سبل تعزيز قنوات التصعيد الخاصة بالمجتمع المدني.

ويُعدّ الشركاء الموثوق بهم حلفاء مهمين في تحديد الانتهاكات عالية الخطورة لمعايير المجتمع لدينا، وكانوا مفيدين على نحو خاص خلال عام 2024، وهو عام اتسم بكثرة الانتخابات. وقد قدموا رؤى مهمة وحددوا محتوى ضارًا في دول شهدت مستويات مرتفعة من الاضطرابات. من بينها بنجلاديش والبرازيل وكوت ديفوار وجمهورية الكونغو الديمقراطية وفرنسا واليونان والهند وإندونيسيا وكينيا وكردستان-العراق والمكسيك ونيجيريا وباكستان والسنغال وجنوب أفريقيا وسوريا وفنزويلا، من بين دول ومناطق أخرى.

وفي عام 2024، أزلنا أكثر من 100000 جزء من المحتوى المخالف للسياسات نتيجة برنامج الشركاء الموثوق بهم.

كما قدم الشركاء الموثوق بهم رؤى حول اتجاهات المحتوى المتعلق بالانتخابات لدعم جهود النزاهة، والمساعدة في اكتشاف المحتوى المخالف وإزالته، وتحديد المستخدمين ذوي المخاطر العالية على منصاتنا ممن يحتاجون إلى وسائل حماية إضافية. وكان الشركاء الموثوق بهم فعّالين في رصد الارتفاعات المفاجئة في الخطاب العدائي الموجّه إلى المجتمعات المهمّشة، وكذلك الاعتداءات على الصحفيين والمدافعين عن حقوق الإنسان، إضافةً إلى إساءة استخدام المحتوى المُنشأ بالذكاء الاصطناعي.

ولإدارة مخاطر السلوك القائم على الكراهية، قمنا بإزالة إهانات مُعيّنة. وعملنا مع الشركاء الموثوق بهم لفهم السياق الذي تُستخدم فيه هذه الإهانات بشكل أفضل، بما يتيح لنا تطبيق سياساتنا بدقة أعلى.

كما تشاورنا مع أكثر من 40 من الشركاء الموثوق بهم في 20 دولة لإثراء عمليات تطوير السياسات والمنتجات المتعلقة بإزالة الصور الحساسة، واستغلال البشر، وإشارات تصنيف المنظمات والأفراد الخطرين، واستخدام مصطلح "صهيوني" كمصطلح بديل للسلوك القائم على الكراهية، وبرامج تلقائية للدردشة المعتمدة على الذكاء الاصطناعي، وغيرها.

واستجابةً لتوصية مجلس الإشراف، قيّمت Meta مدى توقيت وفعالية الاستجابات للمحتوى الذي يتم الإبلاغ عنه عبر برنامج الشركاء الموثوق بهم. وخلال فترة تمتد لعامين من الربع الثاني لعام 2022 إلى الربع الرابع لعام 2024، حققت Meta تحسينات كبيرة في زمن الاستجابة للمحتوى المُبلّغ عنه عبر هذا البرنامج.

وقد أسهم الاستثمار في التدريب، وتبسيط أنظمة الإنفاذ، وتطوير أدوات جديدة في تحسين حجم البلاغات وكفاءة المراجعة خلال عام 2024.

النتائج العالمية

على المستوى العالمي، تلقّى برنامج الشركاء الموثوق بهم أكثر من **11800** بلاغ عن محتوى في الربع الثاني من عام 2022، وارتفع هذا العدد إلى أكثر من **49200** بلاغ في الربع الثاني من عام 2024، بما يعكس **زيادة بمقدار** أربعة أضعاف.

النمو العالمي لقناة الشركاء الموثوق بهم خلال عامين

الربع الثاني لعام 2022 – الربع الرابع لعام 2024

+4 أضعاف
أجزاء المحتوى المبلغ عنها من خلال برنامج الشركاء الموثوق بهم

+12 نقطة مئوية
نسبة الحالات التي تم حلها خلال 5 أيام من التصعيد

+15%
تحسن الكفاءة في متوسط زمن الاستجابة بالأيام

+15 ضعفًا
محتوى تم الإبلاغ عنه لمزيد من المراجعة السياسية

يمكن الاطلاع على أمثلة لتأثير الشركاء الموثوق بهم في باكستان وسوريا وفنزويلا في الصفحات التالية.

دراسة حالة

إعداد التقارير المستندة إلى رؤى من سوريا

في أعقاب انهيار نظام الأسد في ديسمبر 2024، لعب الشركاء الموثوق بهم دورًا محوريًا في الإبلاغ عن التطورات المحلية وتحليلها، من خلال تقديم رؤى حول اتجاهات المحتوى المحلي وتصعيد الانتهاكات عالية الخطورة.

وقد أسهمت تقارير الشركاء الموثوق بهم، اعتمادًا على خبراتهم المحلية، في دعم جهود الاستجابة للأزمات لدى Meta، ومكّنتنا من تطبيق سياساتنا والتخفيف من المخاطر بوتيرة أكثر سرعة وكفاءة. وأثار الشركاء الموثوق بهم مخاوف تتعلق بمخاطر كشف الهوية والادعاءات بالانتماء إلى النظام المُطاح به، والتي استهدفت أقليات عرقية ودينية، من بينها العلويون والمسيحيون والأكراد، كما نبهوا إلى تصاعد نفوذ فصائل متطرفة مختلفة داخل الجيش السوري السابق. وقد دعمت هذه الرؤى جهودنا للحد من مخاطر المنظمات والأفراد الخطرين على منصاتنا، وكذلك مخاطر الاعتداءات الجسدية القائمة على الخصائص الشخصية.

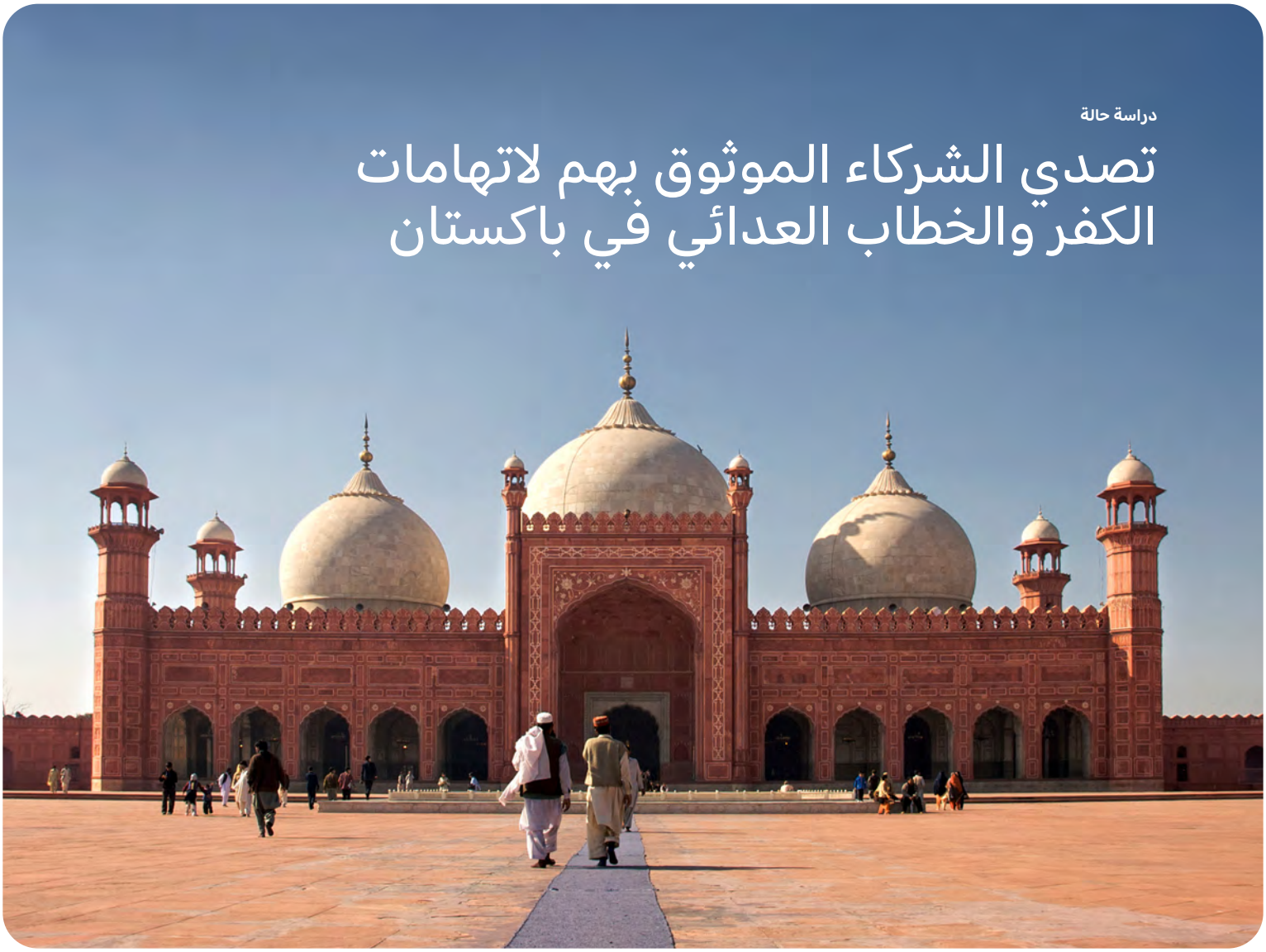
دراسة حالة

تخفيف حدة المخاطر التي تواجه الفاعلين المدنيين في فنزويلا

قبيل الانتخابات التي جرت في فنزويلا في 28 يوليو 2024، عملنا مع الشركاء الموثوق بهم لدينا، وبنينا علاقات جديدة مع منظمات المجتمع المدني، استعدادًا للمخاطر المرتبطة بالانتخابات، ولتعزيز الإبلاغ عن المحتوى الذي قد ينتهك السياسات.

وفي فترة ما بعد الانتخابات، اندلعت احتجاجات أعقبها ردّ قمعي من الحكومة. شمل اعتقالات جماعية وعمليات توقيف موجّهة ضد معارضين سياسيين. وقد قدّم الشركاء الموثوق بهم رؤى بالغة الأهمية حول التطورات على أرض الواقع. وأبلغوا عن محتوى ضار، بما في ذلك تهديدات مبطّنة وكشف هويات المتظاهرين ومؤيدي المعارضة، وهو ما عرّضهم لخطر الاحتجاز التعسفي والضرر البدني. كما أبلغ الشركاء الموثوق بهم عن هجمات استهدفت حسابات فاعلين مدنيين، مثل الصحفيين، وأعضاء المعارضة، والمدافعين عن حقوق الإنسان، وغيرهم.

وقد أسهمت هذه الرؤى في تحسين الرصد الاستباقي، ومكّنتنا من تفعيل الحماية المتقدمة لهذه الحسابات، ومنع أخطاء الإنفاذ من خلال تطبيق آلية التحقق الشامل. وساعدت هذه الإجراءات في دعم العمل الصحفي والمشاركة المدنية في بيئة قمعية.



دراسة حالة

تصدي الشركاء الموثوق بهم لاتهامات الكفر والخطاب العدائي في باكستان

في باكستان، اضطلع الشركاء الموثوق بهم بدور مهم في تنبيهنا إلى محتوى قد يكون ضارًا وموجّهًا ضد المجتمعات المهمّشة، بما في ذلك الأقليات الدينية وأقليات النوع الاجتماعي.

وخلال فترة الانتخابات في فبراير 2024، أبلغ الشركاء الموثوق بهم عن محتوى مرتبط بالانتخابات، مع تركيز خاص على الخطاب العدائي الموجّه إلى المرشحين السياسيين، وعلى اتهامات الكفر التي قد ترقى إلى مستوى التحريض. وفي باكستان، قد تؤدي اتهامات الكفر إلى إجراءات قانونية وأعمال عنف جسدي.

وبناءً على هذه البلاغات، تمكّنا من إزالة محتوى متعلق باتهامات الكفر بموجب سياسة تنسيق الأنشطة الضارة والترويج لارتكاب الجرائم.

كما قدّم الشركاء الموثوق بهم إشارات ورؤى خلال لحظات حرجة أخرى، مثل اندلاع أعمال عنف طائفي. وقد مكّننا عملهم من الاستجابة بسرعة، وإزالة المحتوى المخالف على منصاتنا، وتعزيز قدرات الرصد والإنفاذ.

إشراك أصحاب المصلحة في باكستان

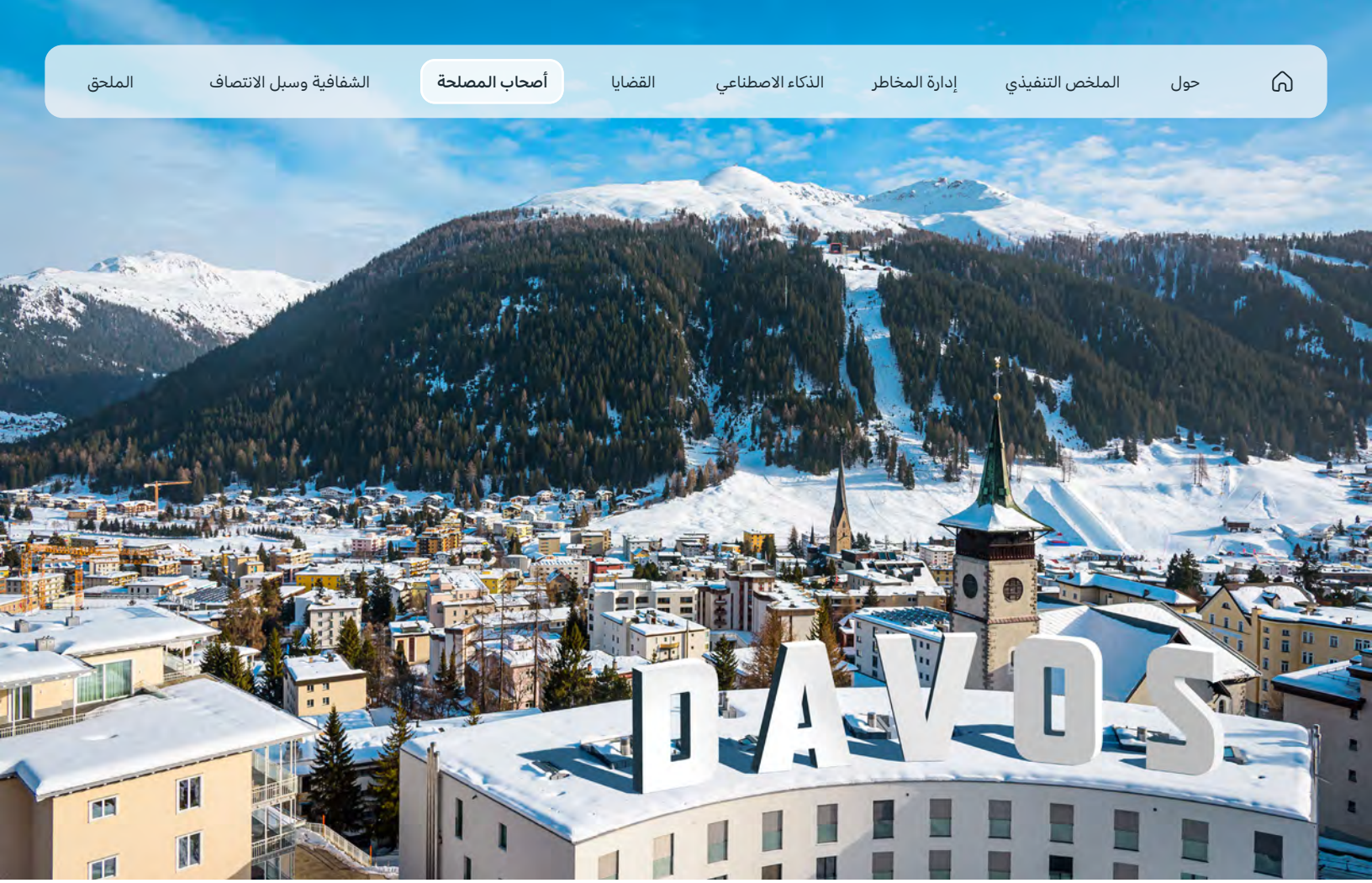
أجرت Meta سلسلة من اللقاءات في باكستان مع عدد من أصحاب المصلحة الحكوميين وغير الحكوميين، في إطار إجراءات العناية الواجبة بحقوق الإنسان، مع التركيز على تحقيق التوازن بين السلامة وحرية التعبير. وشملت أبرز النقاط ما يلي:

مائدة مستديرة حول سلامة الشباب على الإنترنت، عُقدت بالشراكة مع وزارة حقوق الإنسان، والمفوضية الوطنية لحقوق الطفل، والمفوضية الوطنية لحقوق الإنسان، ومؤسسة الحقوق الرقمية (Digital Rights Foundation). وناقشنا خلالها حسابات المراهقين، وإطلاق بوابة Take It Down باللغة الأردية لمستخدمي Meta في باكستان.

التفاعل مع مجموعة متنوعة من المدافعين عن حقوق الإنسان لجمع رؤاهم حول اضطرابات الإنترنت، واستكشاف سبل محتملة للتعاون في مجالات المناصرة. وقد قدموا رؤى قيّمة حول تأثير ما يُعرف بـ "جدار الحماية" وسياسة إضافة الشبكات الخاصة الافتراضية (VPN) إلى القائمة البيضاء من قبل الحكومة.

مائدة مستديرة مع منظمات المجتمع المدني لتقديم شرح معمّق لالتزامات Meta الأساسية في مجال حقوق الإنسان، ولأعمال فريق حقوق الإنسان بالشركة. وشمل ذلك مناقشة سبل التعامل مع الأوضاع الحساسة دون اللجوء إلى قطع الإنترنت أو تقييد عمل منصات التواصل الاجتماعي، بما في ذلك مجموعة تطبيقاتنا.

وفي جميع اللقاءات، أشار جميع المشاركين، في كل الفعاليات، إلى تأثير استهداف المستخدمين باتهامات عدائية وكيدية تتعلق بالكفر. وقد شعروا بالاطمئنان عندما اطّلعوا على سياسة Meta الخاصة بمخاطر الكشف عن الهوية وعلى الجهود المستمرة التي تبذلها الشركة للحفاظ على سلامة المستهدفين.



المنظمات الدولية

في عام 2024، تفاوضت الدول الأعضاء في الأمم المتحدة واعتمدت الميثاق الرقمي العالمي (GDC)، وهو إطار شامل للحوكمة العالمية للتقنيات الرقمية والذكاء الاصطناعي. عملنا مع الدول الأعضاء في الأمم المتحدة، ووكالاتها، وتحالفات القطاع الصناعي، من أجل استكمال النص النهائي للميثاق الرقمي العالمي. وقد ركّزت جهودنا على دعم حرية التعبير، مع الإسهام في بناء مستقبل رقمي أكثر أمانًا وشمولاً وانفتاحًا للجميع.

كما واصلت Meta الانخراط مع منظومة الأمم المتحدة بطرق متعددة على مدار العام. وشمل ذلك المساهمة في مبادرة اليونيسف المعنية بالتقنيات الرقمية وحقوق الطفل ورفاهيته، دعمًا لإجراءات العناية الواجبة في قطاع التكنولوجيا، وكذلك التعاون مع اليونيسكو في مجال حوكمة المنصات الرقمية ومكافحة المعلومات المضللة. وقدّمنا أيضًا دعمًا لليونيسكو من خلال واجهة ترجمة مبنية على نموذج الذكاء الاصطناعي Meta No Language Left Behind (NLLB)، بهدف توفير ترجمة عالية الجودة لما يقرب من 200 لغة. وشملت هذه اللغات لغات مهمّشة مثل الأسترية، واللوغندية، والماورية، والسواحيلية، والأردية، بما يسهم في تعزيز التنوع اللغوي وتوسيع نطاق الوصول إلى المعلومات.

وواصلت Meta العمل بشكل وثيق مع مكتب مفوضية الأمم المتّحدة السامية لحقوق الإنسان (OHCHR). حيث عقدنا اجتماعات منتظمة مع موظفي المفوضية، وشاركنا بفاعلية في مشروع B-Tech، الذي يوفر إرشادات وموارد مرجعية لتطبيق المبادئ التوجيهية للأمم المتحدة بشأن الأعمال التجارية وحقوق الإنسان في قطاع التكنولوجيا، إلى جانب المشاركة في مجتمع الممارسة التابع له، وهو مساحة للحوار السري بين شركات التكنولوجيا. كما شاركنا بفعالية في المناقشات الجارية حول معايير الذكاء الاصطناعي وحقوق الإنسان. بالإضافة إلى ذلك، شاركنا في منتدى الأمم المتحدة للأعمال التجارية وحقوق الإنسان لعام 2024، وقدّمنا مداخلات في جلسات تناولت موضوعي "خطاب يحض على الكراهية عبر الإنترنت" و"حماية حرية الصحافة".

كما انخرطت Meta في مناقشات السياسات على هامش قمة المستقبل والدورة التاسعة والسبعين للجمعية العامة للأمم المتحدة. حيث شملت الموضوعات دور الذكاء الاصطناعي في الحوكمة العالمية، وتمكين صنّاع المحتوى الرقمي، والابتكار الاقتصادي القائم على الجاليات في المهجر، وتأثير قوانين الجرائم الإلكترونية وتنظيم المحتوى على حرية التعبير. وشاركنا أيضًا في مناقشات تتعلق بحماية المدافعين عن حقوق الإنسان، وباستخدام وسائل التواصل الاجتماعي لتوفير معلومات منقذة للحياة في الأزمات الإنسانية.

وعلاوة على ذلك، تشاورنا مع الإجراءات الخاصة لمجلس حقوق الإنسان التابع للأمم المتحدة، وهم خبراء مستقلون في مجال حقوق الإنسان، بمن فيهم المقررون الخاصون المعنيون بحرية الرأي والتعبير، وبالمدافعين عن حقوق الإنسان، وغيرهم.

وطوال العام، تعاونت Meta مع مجموعة G7، ومجموعة G20، واليونيسكو، ومنظمة التعاون الاقتصادي والتنمية (OECD) في مسارات عمل تتعلق بشمولية الذكاء الاصطناعي وحوكمته. كما شاركنا في حوارات مع الحكومات حول أهمية نزاهة المعلومات. وواصلنا مشاركتنا في التحالف العالمي للسلامة الرقمية التابع للمنتدى الاقتصادي العالمي، والذي أسفر عن نشر تقرير "رحلة التدخل (The Intervention Journey)". خارطة طريق لتدابير فعّالة للسلامة الرقمية.

وإضافة إلى ذلك، شاركنا بفاعلية وانخرطنا مع أصحاب المصلحة في عدد من المنتديات متعددة أصحاب المصلحة، بما في ذلك قمة القضاء على الكراهية (Eradicate Hate Summit)، ومنتدى حرية الإنترنت في أفريقيا (FIFAfrica)، والمنتدى العالمي للإنترنت لمكافحة الإرهاب، ومنتدى حوكمة الإنترنت (IGF)، و RightsCon، ومؤتمر التكنولوجيا ضد الاتجار بالبشر، ولجنة الأمم المتحدة المعنية بوضع المرأة.

ومن خلال عضويتنا في مبادرة الشبكة العالمية ومشاركتنا في شراكة الثقة والسلامة الرقمية، حضرت Meta "منتدى إشراك أصحاب المصلحة الأوروبي بشأن الحقوق والمخاطر"، والذي أسهم في إثراء تقييماتنا للمخاطر المنهجية بموجب قانون الخدمات الرقمية.

الشفافية وسبل الانتصاف

يساعدنا مجلس الإشراف، بوصفه هيئة مستقلة، على معالجة بعض أكثر الأسئلة تعقيدًا المتعلقة بحرية التعبير على الإنترنت: ما الذي يجب إزالته، وما الذي يجب تركه، ولماذا. يراجع المجلس القضايا التي تحيلها Meta إليه، أو التي يطعن عليها أفراد على فيسبوك أو Instagram أو Threads ممن يعترضون على قرارات الإشراف على المحتوى، ويصدر قرارات ملزمة بشأن ما إذا كان ينبغي إزالة المحتوى أو تركه. كما يقدّم مجلس الإشراف توصيات لتحسين ممارسات الإشراف على المحتوى، ويصدر آراءً استشارية بشأن سياسات المحتوى عند الطلب.

أين يمكن العثور على معلومات حول أثر مجلس الإشراف

في عام 2024، انتقلنا من وتيرة تقارير ربع سنوية إلى تقارير نصف سنوية بشأن القضايا التي تحيلها Meta إلى مجلس الإشراف، وكذلك التحديثات المتعلقة بتقدّمنا في تنفيذ توصياته. وبالإضافة إلى ذلك، أطلقنا صفحة مركز الشفافية التي تتّبع أثر توصيات مجلس الإشراف. ويأتي ذلك إلى جانب صفحة توصيات مجلس الإشراف، حيث نعرض التوصيات المرتبطة بكل قضية، ومستوى التزامنا بها، وحالة تنفيذها.

الإجراءات المتعلقة بتوصيات مجلس الإشراف في عام 2024

توصيات مجلس الإشراف الصادرة

48

(66 في عام 2023)

تحت تقييم Meta و/أو قيد التنفيذ[5]

70

(69 في عام 2023)

توصيات تم تنفيذها[5]

41

(61 في عام 2023)

في عام 2024، نظر مجلس الإشراف في قضايا تتعلق بإنفاذ سياسات المحتوى في ضوء الإطار الدولي لحقوق الإنسان، بما في ذلك حرية التعبير، والحق في الصحة، والحق في المساواة وعدم التمييز، وغيرها. وفيما يلي بعض الأمثلة التوضيحية لإجراءاتنا استجابةً لقرارات مجلس الإشراف في عام 2024. تتوفر التفاصيل في تقارير Meta نصف السنوية بشأن مجلس الإشراف.

شملت الأمثلة على قرارات مجلس الإشراف في 2024 ما يلي:

ألغى مجلس الإشراف قرارات Meta بإزالة ثلاثة منشورات على فيسبوك تتضمن لقطات من الهجوم الإرهابي في موسكو في مارس 2024، وطالب باستعادة المحتوى مع وضع شاشات تحذيرية تحمل وسم "تمييز كمزعج". وخلص المجلس إلى أنه، رغم مخالفة المنشورات لسياسة Meta بشأن عرض لحظة الهجمات المصنّفة على ضحايا ظاهرين، فإن إزالتها لم تكن متسقة مع مسؤوليات الشركة في مجال حقوق الإنسان.

أيّد مجلس الإشراف قرار Meta بإزالة فيديو لسياسي باكستاني يلقي خطابًا مرفقًا بنص يدّعي أن السياسي "يتجاوز كل حدود الإيمان"، ويستخدم كلمة «كفر» للإيحاء بالكفر، وذلك نظرًا إلى مخاطر الضرر على أرض الواقع.

[5] بعض التقييمات و/أو عمليات التنفيذ قيد التقدم أو التوصيات التي تم تنفيذها بالكامل تتضمن توصيات من أعوام سابقة (راجع تقريرنا لحقوق الإنسان لسنة 2023 لمزيد من التفاصيل).

شملت الأمثلة على إجراءات اتُّخذت تنفيذًا لتوصيات مجلس الإشراف ما يلي:

عقب سلسلة من التوصيات (على سبيل المثال، المتعلقة بالذكاء الاصطناعي هنا)، أجرينا تغييرات على كيفية تعاملنا مع المحتوى الذي تم إنشاؤه بواسطة الذكاء الاصطناعي، شملت تحديث التسميات والسياسات، مثل سياسة المعلومات المضللة.

وبناءً على توصيات مجلس الإشراف بشأن سياسة المحتوى، عدّلت Meta سياسة المنظمات والأفراد الخطرين للسماح باستخدام مصطلح "شهيد" بجميع اللغات التي تستخدم هذا المصطلح، إلا في الحالات التي يكون فيها المحتوى مصحوبًا بإشارات إلى العنف أو مخالفًا لسياساتنا الأخرى (مثل تمجيد منظمات أو أفراد مصنّفين كخطرين).

في عام 2024، جرّب مجلس الإشراف أيضًا جدولاً زمنيًا أقصر للفصل في القضايا العاجلة. فعلى سبيل المثال، بعد الانتخابات الرئاسية في فنزويلا في يوليو وما أعقبها من أعمال عنف، أحلنا قطعتين من المحتوى تتعلقان بـ "Colectivos" إلى مراجعة مُعجّلة. ويُستخدم مصطلح "Colectivos" كمسمّى جامع لوصف عصابات مسلّحة غير نظامية أو مجموعات شبه عسكرية وثيقة الصلة بالحكومة. وقد صُدِرَت القرارات في هذه القضايا ضمن جدول زمني مُسرّع بلغ 14 يومًا.

كما تعاونّا مع مجلس الإشراف للتواصل مع الجهات التنظيمية ومنظمات المجتمع المدني، بما في ذلك في مناطق أفريقيا وأمريكا اللاتينية والشرق الأوسط وتركيا، لرفع مستوى الوعي بولاية مجلس الإشراف وآلية اختيار القضايا.

الملحق

كيفية حوكمة حقوق الإنسان والتعامل معها في Meta

يتولى خبراؤنا في مجال حقوق الإنسان توجيه عملية تنفيذ سياسة حقوق الإنسان الخاصة بشركتنا، التي يشرف عليها رئيس الشؤون العالمية (حاليًا رئيس الشؤون العالمية للشركة) ومدير الشؤون القانونية.

وتتضمن مهام خبراء حقوق الإنسان تعزيز دمج السياسة في السياسات والبرامج والخدمات القائمة والناشئة، وبذل العناية الواجبة، ودعم تدريب الموظفين على السياسة. تعمل السياسة بدورها على التوجيه نحو تصميم منتجات تحترم الحقوق، والاستجابة للأزمات الناشئة، والعمل بسرعة ومرونة لترسيخ حقوق الإنسان على نطاق واسع.

تُلزمنا سياسة حقوق الإنسان الخاصة بشركتنا برفع تقارير دورية إلى مجلس الإدارة حول القضايا الرئيسية المتعلقة بحقوق الإنسان. في عام 2024، قدّم مدير حقوق الإنسان إحاطة إلى لجنة المراجعة والإشراف على المخاطر في المجلس.

كما أطلقت Meta في عام 2024 محور مخاطر حقوق الإنسان ضمن برنامج إدارة مخاطر الجهة الخارجية للشركة. ويُظهر هذا الإجراء التزامنا بالتحسين المستمر لإدارة مخاطر حقوق الإنسان والسعي نحو تفاعلات مع الجهات الخارجية تكون مسؤولة وتحترم حقوق الإنسان.

تدريب موظفي Meta في مجال حقوق الإنسان

إن الكيفية التي نبني بها في Meta لا تقل أهمية عن الشيء الذي نبنيه. ويسلط تدريب حقوق الإنسان لدينا الضوء على التأثيرات الواقعية المحتملة والفعلية لخدماتنا وسياساتنا وقراراتنا التجارية على حقوق الإنسان. ويهدف التدريب إلى ترسيخ عقلية حقوق الإنسان في أعمالنا اليومية، وتشجيع احترام حقوق الإنسان بما يحقق صالح جميع الأشخاص الذين يستخدمون خدماتنا.

أطلقنا تدريب أكبر من Meta: حقوق الإنسان (Bigger than Meta: Human Rights) في عام 2022، واستمر العمل به طوال عام 2024. كما يدعم تدريب الخصوصية أهداف تدريب حقوق الإنسان من خلال التركيز على تطوير قدرتنا الجماعية على حماية الأفراد، بما يشمل، على وجه الخصوص، الفئات المهمشة، من الأضرار الناجمة عن معالجة بيانات الأشخاص.

روابط إلى تقارير مرجعية

تقرير الممارسات التجارية المسؤولة لعام 2025

تقرير الاستدامة لعام 2025

تقرير حقوق الإنسان لعام 2023، تقرير حقوق الإنسان لعام 2022، تقرير حقوق الإنسان لعام 2021

تقرير مناهضة الرق والاتجار بالبشر لعام 2024

تقرير سياسة معادن مناطق الصراع لعام 2024

تقارير الشفافية الخاصة بـ Meta

التقارير التنظيمية وتقارير الشفافية الأخرى

تقييمات التأثيرات الواقعة على حقوق الإنسان المنشورة سابقًا: التشفير التام بين الطرفين، الفلبين، وميانمار، وإندونيسيا، وكمبوديا، والهند، وسريلانكا، وإسرائيل وفلسطين

Meta