

Meta Platforms, Inc.の回答

1 総務省要請に関する取組について	
Q1-1(闇バイトの募集活動に係る投稿に対する利用規約等に基づく対応関係) 総務省要請の記1について、以下の3点を御教示下さい。 (1)利用規約等における記載内容 闇バイトの募集活動に係る投稿への対応に関する、利用規約、ポリシー、社内基準等(以下「利用規約等」という。)における記載内容(職業安定法の解釈に関する記載を含む。)	弊社には、コミュニティ規定の一環として、「詐欺、スキャン、欺瞞的行為ポリシー」があり、これにより禁止されるものとして、求人詐欺・スキャンも含まれます。 https://transparency.meta.com/ja/policies/community-standards/fraud-scams/。 このポリシーでは、以下のような求人詐欺・スキャンを含むコンテンツが対象となります： ● 仕事の内容が不明確または漠然とし、短時間で簡単に大金を稼ぐことができるという「すぐに富を得る」機会を約束するもの。 ● 求人情報がなく、単に求人があることを示すだけのもの。 ● 職種からして在宅勤務ができないにもかかわらず、在宅勤務を可能とするもの。 ● 給与を前払いすることを約束するもの。 ● 就職を保証するもの。 ● 仕事を開始する前に前払いを要求するもの。 また、弊社のコミュニティ規定は、労働搾取や強制的な犯罪を含む、さまざまな形態の人身売買のために人を募集したり、助長したり、搾取したりするようなコンテンツ、活動、交流を禁止しています。 私たちのポリシーを施行するために、私たちは人工知能と人間のレビューを組み合わせて使用しています。私たちは、人工知能と機械学習を活用した積極的な検出技術に多大な投資を行い、サービス上の有害なコンテンツや行動を特定して阻止しています。私たちの検出努力は、ポリシーに違反するコンテンツを特定するために進化し続けており、一部の場合には誰もが見る前に削除します。メタでは、約4万人の人が安全性とセキュリティに従事しています。私たちのコンテンツレビューチームは24時間365日コンテンツをレビューしており、運営している市場の言語や文化的背景を持つ人材を採用することに努めています。また、私たちは詐欺を事前に特定してブロックするためにシステムを常に改善し適応させる専門チームも持っています。

	この分野は敵対的な空間であり、悪意のあるグループは検出を回避するために常に手法を進化させています。弊社は継続的な投資を行っているものの、機械も人間もミスをするため、見逃したり誤って削除したりする例があることは承知しています。 そのため、弊社はコミュニティ規定に違反していると思われるコンテンツやアカウントを見つけた場合は誰からでも報告してほしいと思っています。これらに基づき、弊社が調査して措置を講じることができます。 弊社の詐欺の実行行為者に対する取り組みは多層的であり、利用者への情報提供・教育すること、セキュリティツールの更新すること、検出活動をスケールさせること、執行の継続的な改善すること、狡猾な悪意のある組織に対する専門家による調査などを行っているとともに、利用者の安全に焦点を当てたNGOや法執行機関との協力も行っています。
(2) 闇バイトの募集活動に係る投稿に関する流通状況 自らが提供するSNS等において流通する闇バイトの募集活動に係る投稿に関する流通状況(可能な限り、定量的な回答を御願いたします。)	下記(3)参照。
(3) 削除対応等の実施状況 自らが提供するSNS等において流通する闇バイトの募集活動に係る投稿に関する、利用規約等を踏まえた迅速な削除対応等(削除、表示制限、ラベルの付与、閲覧制限、アカウント停止等いわゆるコンテンツモデレーションをいう。以下同じ。)の実施状況	2024年の第2四半期と第3四半期に、日本におけるFacebookとInstagram上で削除したコンテンツ及びアカウントに関しては以下のとおりです。これらは、詐欺・スキャン、欺瞞的行為または容認できないビジネス慣行にに関する弊社のポリシーに違反したものです。 ● 利用者投稿:(事前対応率) ○ 第2四半期 約16,000件(95%) ○ 第3四半期 約18,000件(96%) ● アカウント: ○ 第2四半期 約2,000件(FB) ○ 第3四半期 約3,000件(FB) ○ 第2四半期 約11,000件(IG) ○ 第3四半期 約16,000件(IG)

(可能な限り、定量的な回答を御願いたします。)	● 広告コンテンツ: ○ 第3四半期 約326,000件(82%) ○ 第3四半期 約391,000件(89%) ● 広告アカウント: ○ 第2四半期 約18,000件 ○ 第3四半期 約21,000件 2024年には、ミャンマー、ラオス、カンボジア、アラブ首長国連邦、フィリピンにあるスキーム・センターに関連する200万以上のアカウントを削除しました。 https://about.fb.com/news/2024/11/cracking-down-organized-crime-scam-centers/ 弊社は、「pig butchering」スキームの背後にある犯罪組織を追跡しています。このスキームでは、メッセージング、デーティングアプリ、ソーシャルメディア、その他のアプリを通じて世界中の人々を標的にして、聞こえの良い誘いで「投資」するよう説得します。 弊社は、法執行機関や同業他社とも協力して、知見を共有し、悪意のある詐欺活動を崩壊させるために協力しています。 たとえば、最近、弊社のチームは、カンボジアに新しく設立された詐欺施設から日本語と中国語で人々を標的していると思われる悪意のある活動を崩壊させました。弊社は、詐欺実行行為者がコンテンツを生成し、翻訳するためにツールを使用したことを検出し、これを防いだOpenAIから弊社に共有された情報に基づいて、この活動の調査を開始しました。また、弊社が崩壊させたスキーム・センターに関する情報はこれを共有し、インターネット全体でさらなる調査を可能にしました。
Q1ー2(SNS等のアカウント開設時における本人確認手法の厳格化関係) 総務省要請の記2について、以下の2点を御教示下さい。 (1)本人確認の実施状況等 自らが提供するSNS等のアカウント開設時における本人確認の実施状況、及び、	利用者アカウントを作成する場合、弊社はメールアドレスまたは電話番号、生年月日、性別の登録を要求します。弊社は、二段階認証を設定してセキュリティを向上させることを強く推奨しています。認証のために、以下の3つのセキュリティ方法を提供しています: - 対応するデバイスでセキュリティキーをタップする。 - サードパーティーの認証アプリからログインコードを受け取る。 - 携帯電話からのSMSコードを受け取る https://ja-jp.facebook.com/help/401566786855239?helpref=faq_content https://ja-jp.facebook.com/help/401566786855239?helpref=faq_content 利用者は、以下の場合に追加の検証ステップを完了する必要がある場合があります。

本人確認手法の厳格化を含む措置の検討状況。	● 1) 特定のプロダクトにアクセスする場合。 ● 2) アカウントが詐欺行為に関連している可能性がある判断した場合、アカウント所有者は自身が偽のアカウントを運営したり、自分を誤って表現したりしていないことを示すためにいくつかのアクションを完了する必要があります。これらのチェックに失敗した場合、または弊社の審査員がポリシーの違反を決定した場合、アカウントを削除します。 昨年半ば以降、新しい広告主は、広告を公開する前に広告アカウントに関連付けられた確認済みの電話番号を持っている必要がある場合があります。必要なステップでは、アカウント管理者が弊社によってSMS、音声、またはWhatsApp経由で送信されたランダムに生成されたコードを確認することで電話番号を確認するように求められます。 テクノロジーやサイバーセキュリティの観点からは、本人確認強化はオンライン詐欺防止の万能薬ではありません。オンライン・プラットフォームを利用する多くの個人が、正式な身分証明書を入手できないか、それを共有したくないことには正当な理由があることは知られています。例えば、家庭内虐待に直面している人、トランスジェンダーのコミュニティの人たちなどが挙げられます。 強制的に本人確認を行うことは、運用上実現不可能なだけではありません。同時に、最終的にはプラットフォームが保有する個人情報の量を増やすことになることを意味し、深刻なプライバシー上の懸念をもたらします。このように、これらの要件は、データ最小化というプライバシー法の基本原則に反するものです。
(2) 犯罪実行者の特定に資するその他の取組 その他、犯罪実行者の特定に資する取組。	弊社は、悪意のある行為者と対立する空間で戦う必要があります。悪意のある行為者は常に弊社によるポリシーの執行を回避しようとしており、プラットフォーム上でネットワークを再構築します。弊社は、Strategic Network Disruption(戦略的ネットワーク・ディスラプション)と呼ばれる重要なアプローチを使用して、悪意のある行為者を排除します。弊社はAIを使用して、悪意のある行為者たちがどのようにプラットフォーム上で行動し、どこにいるのかを検出し、それらの情報を学習して、対象をプラットフォームから一気に排除します。 詐欺の実行行為者は、人々を詐欺できるならどのようなプラットフォームでも利用しようとします。また、ポリシーの執行を回避するためには、常に、我々の対応に適応していきます。しかし、弊社は、他人を欺くまたは搾取することを目的としたコンテンツに対して、これらを探知した場合には削除します。弊社は検出技術への投資を続け、法執行機関と情報を共有して詐欺実行行為者を訴追できるように協力しますし、法執行機関から受け取った情報は詐欺に対処するのに役立てていきます。 弊社は警察当局(警察庁)との連絡を担当する専門チームを設けており、密接に協力して、プラットフォーム上で利用者が安全に過ごせるように支援しています。特に、このチームはLaw Enforcement

	Online Records System(法執行機関オンライン記録システム)を運営しており、警察は、捜査活動に資する利用者に係る記録の提供を求めることができます。 https://transparency.meta.com/reports/government-data-requests/country/JP/
Q1ー3(捜査機関等からのSNS等に対する照会への回答の円滑化関係) 総務省要請の記3について、以下の2点を御教示下さい。 なお、捜査機関等への対応の円滑化に当たり、その前提として通信ログの保存状況を把握する必要がある点を踏まえて回答ください。 (1)捜査機関等からの対応状況 捜査機関等からの照会に対して、(i) 照会から開示に至るまでの標準対応期間、(ii) 対応件数の推移、(iii)円滑に回答できる体制整備の実施及び検討状況・課題。	世界中の政府からユーザーデータの要求を受けることがあります。弊社は、より多くの情報を共有する取り組みの一環として、定期的に政府からの要求に関して、レポートを公表しています。このレポートでは、これらの要求の性質、範囲、そしてこれらの要求を処理するための厳格なポリシーとプロセスについての情報を提供しています。 たとえば、2024年1月から6月までの間に、弊社は日本の政府機関から491件の要求を受けました: 緊急要求485件、法律に基づく要求6件。また、利用者アカウントの保存要求68件も受けました。 https://transparency.meta.com/reports/government-data-requests/country/JP/ なお、上記専門チームは、定期的に現地国における警察当局とやりとりしています。
(2)捜査機関等への対応の円滑化に資する通信ログの扱い 捜査機関等への対応の円滑化に資する通信ログについて、(i)通信ログの保存状	提出資料参照

況(保存している通信ログの種類、保存期間含む。)、(ii) 通信ログの保存期間算出の根拠、(iii)保存期間の延長の可否やその理由等、その他の課題。	
Q1-4 (SNS等の利用者に対する注意喚起・周知活動関係) 総務省要請の記4について、自らが提供しているSNS等のユーザーに対する闇バイト等の重大な犯罪に加担する危険性についての注意喚起・周知活動の実施状況を御教示下さい。	2025年3月、日本における詐欺的な求人募集に利用者が応えてしまうことを防ぐため、Facebook、Instagram、Threadsで検索への介入機能を開始しました。警察庁との連携のもと、労働搾取に関連するキーワードを特定しました。検索への介入機能では、以下の内容が導入されています： - 労働搾取につながる可能性の高い、詐欺的な求人広告に関連するキーワードを利用者が入力すると、警告メッセージが表示される - 利用者が検索結果の表示を許可するまで、その表示をブロックします。 - 被害者となり得る利用者を、支援やリソースを提供できる組織につなげます。 これらの介入による影響を継続的に監視し、そのパフォーマンスに関するデータを収集していきます。 さらに、弊社は日本における闇バイトを含むさまざまな詐欺を抑止するために、複数のプロダクト上の工夫を行っています： - メッセンジャーにおいて、スキャン・セーフティ通知(疑わしいと判断したアカウントがメッセージ送信要求を送信した場合、警告通知を表示) - メッセンジャーにおいてなりすましセーフティ通知(Facebookの友人の名前とプロフィール写真と似ていけれども、つながっていない利用者が、連絡したきた際に警告通知を表示) - インスタグラムでの不審なフォローのセーフティ通知(個人または事業をなりすましているか、潜在的に詐欺的な活動を行っている可能性があるアカウントがフォローをリクエストした場合、警告を促す通知を表示) - セーフブラウジング(E2E暗号化されたメッセージ内の不審なリンクをクリックした場合にポップアップを表示) また、オンライン上で人々を安全に保つためには、デジタル・リテラシーの育成が重要な要素となります。弊社は、この分野に継続的に投資しており、2020年以降、国内で40,000人以上の中高生が「みんなのデジタル教室」の授業を受講しました。さらに、関連省庁等と協力し、FacebookおよびInstagram上の大規模な詐欺対策キャンペーンなども行っています。 特に闇バイトについては、警察庁と連携して、加害者の手口と警察の相談窓口に関する認識を高めるための啓発キャンペーンを実施しました。詐欺対策および闇バイト防止の啓発キャンペーンは、2024年

	7月から2025年2月までの間に関係団体との協力で実施され、国内で1500万人以上にリーチし、6000万回以上のインプレッションを獲得しました。 また、日本語版の詐欺対策ポータルを公開しています。このポータルを通じて、利用者と企業が詐欺を識別し、回避することができるように情報提供をしています。弊社のサイバーセキュリティの最新の進歩や資料、そして詐欺の報告方法、アカウントへのアクセス問題、知的財産権保護および不正使用のためのクイックリンクが含まれています。https://about.meta.com/jp/actions/safety/anti-scam/
2 その他の違法情報等への対応について	
<u>1)違法情報への対応について</u> Q2-1(違法情報への対応における利用規約等における扱い) 違法情報(他人の権利を侵害する情報を除く。以下同じ。)への対応に関して、利用規約等において、①どのような種類・性質の情報や行為を禁止としており、	一般的に、裁判所の判断なくしてコンテンツが違法であるかどうかを評価することは非常に困難です。世界中の裁判所でも、違法なコンテンツを特定するという問題に苦慮しています。 弊社は、違法なコンテンツに対処するための包括的なアプローチを取っています。これには、コミュニティ規定に照らしてのレビュー、法的レビュー、および人権への影響の評価が含まれます。 コミュニティ規定によるレビュー 違法なコンテンツに対する第一線の防御は、コミュニティ規定に照らしての判断です。これらの規定は、プラットフォーム上でどのようなタイプのコンテンツが許容され、利用者からどのような行動が期待されるかを概説しています。コンテンツが潜在的に違法であるとして報告された場合、これらの規定に対して徹底的なレビューが行われます。モデレーターとAIツールが協力してコンテンツを評価し、弊社のポリシーに違反しているかどうかを判断します。コンテンツがコミュニティ規定に違反していると判定された場合、プラットフォームから削除されます。 https://transparency.meta.com/policies/community-standards/ 法的レビュー コンテンツがコミュニティ規定に違反していない場合、現地法に違反しているかどうかを判断するために法的レビューが行われます。このレビューは、さまざまな管轄区域の法律に精通した経験豊富な弁護士チームによって実施されます。彼らは、言論の自由、ヘイトスピーチ、知的財産権などの要素を考慮しながら、適用可能な法律と規制に対してコンテンツを評価します。コンテンツが現地法に違反していると

	判定された場合、プラットフォームから削除するか、関連する管轄区域でのアクセスを制限する措置を講じます。 人権レビュー コンテンツが複雑な人権問題を引き起こす場合、人権レビューを実施します。このレビューは、国連が示している「ビジネスと人権に関する指導原則」を含む、国際人権基準に基づいて行われます。専門家チームは、表現の自由の権利、差別を受けない権利、プライバシーの権利などの要素を考慮しながら、これらの原則に照らしてコンテンツを評価します。このレビューの目的は、弊社の行動が国際人権基準に沿っていることを確認しながら、国籍、民族、背景に関係なく、すべての個人の権利を尊重することです。 弊社は、日本の利用者に対して、権利や利益を侵害するコンテンツを報告するためのいくつかのチャネルを提供しています。また、利用者自身が自分の体験を管理し、サービス上で安全を保つことができるように、強力なコントロールを移譲しています。 • https://www.facebook.com/help/ipreporting/report/copyright • https://www.facebook.com/help/contact/516343134409068 • https://www.facebook.com/help/contact/430253071144967 さらに、弊社は、現地法に違反するコンテンツの削除要求を行うために、規制当局とのレポート・チャンネルを構築しています。これらの規制要求に応じた弊社の対応は、透明性センターで公開されています。 https://transparency.meta.com/reports/content-restrictions/country/JP/
②どのような削除対応等を行うことと定めているのか御教示下さい(行為や情報の類型ごとに取り得る対応が異なる場合、そのことが分かるようにお示し下さい。)	上記参照
Q2-2(違法情報の削除対応等の取組)	上記参照

違法情報の削除対応等の取組について、次の5点を御教示下さい。 (1)利用者からの申告対応 貴社のSNS等のサービス上で流通する違法情報について、利用者からの申告を受け付けているか。受け付けている場合、受付窓口・受付体制をどのように定めているか。受付窓口を定めている場合、どのような方法(ウェブフォームによる申請、書面による申請等)による申告を可能としているか。また、申告の受付のためどのような体制を整備しているか。当該受付窓口・受付体制について公表しているか。	
(2)行政機関等からの申告対応 貴社のSNS等のサービス上で流通する違法情報について、捜査機関、その他の行政機関、インターネット・ホットラインセンター等の第三者機関からの申告や要請を受け付けているか。受け付けている場合、(1)とは別に受付窓口・受付体制を定めているか。受付窓口を定めている場合、どのような方法(ウェブフォームによる申請、書面による申請等)による申告や要請を可能としているか。また、申告や要請の受付のためどのような体制を整備	弊社は、法務省、厚生労働省、金融庁などの規制当局とのレポーティング・チャネルを構築しています。 弊社は、規制当局に以下の情報を提供するよう依頼しています：(a)削除すべきコンテンツの正確なURL、(b)削除請求の理由(例：Metaのコミュニティ規定の違反または現地個別法の違反)、(c)裁判所命令などの証拠など。

しているか。当該受付窓口・受付体制について公表しているか。	
(3) 違法情報への対応等 貴社のSNS等のサービス上で流通する違法情報について、(1)や(2)の申告や要請によらず、AIを活用するなど主体的に検知や削除対応等を実施しているか。実施している場合、どのような方法・体制によって検知・削除対応等を実施しているか(検知を担当するチームの有無、機械的な検索等)。	上記、特に2(1)を踏まえ、弊社のテクノロジーは、誰かが報告する前に、違反するコンテンツの大部分を事前に検出して削除しています。 例えば、弊社は、FacebookとInstagramで毎日何百万もの違反する投稿やアカウントを削除しています。これらのほとんどは自動的に行われ、テクノロジーが裏側で動作して違反するコンテンツを削除しています—多くの場合、誰かがそれを見る前に行っています。時には、弊社のテクノロジーは、潜在的に違反するおそれのあるコンテンツを検出しますが、それはレビューチームに回送し、必要な対応をします。 こうしたプロセスは決して終わりません。悪意のある人々は弊社のテクノロジーを回避しようとするので、改善し続ける必要があります。 https://transparency.meta.com/enforcement/detecting-violations/
(4) (1)～(3)における体制整備の状況 (1)から(3)の各場合について、以下の各事項。 ① 利用規約等に反しているかを審査するプロセス・体制(各プロセスにおいて、人間が審査しているのか、AIなどの活用により機械的に審査しているか等も明示願います)。 ② 申告・要請や検知がなされてから、削除対応等を行うまでの期間(標準処理期間の設定の状況、実際に削除対応等に要した期間等)。	① 上記のように、弊社はコミュニティ規定を執行するために、機械と人間の両方によるレビューを使用しています。 ② 弊社は、ターンアラウンドタイムを検出および施行システムの効率の有意義な尺度とは見なしていません。代わりに、自動化されたシステムを改善することへ投資しており、これらのシステムは報告される前にポリシー違反のコンテンツを見つけて削除することを目指しています。弊社がどのようにして積極的なレートを計算するかについての詳細はこちらをご覧ください。 https://transparency.meta.com/en-gb/policies/improving/proactive-rate-metric/ 2024年の第2四半期と第3四半期には、詐欺・スキャンム、および欺瞞的行為に関するポリシーに違反する利用者投稿コンテンツの95%以上が、利用者が報告する前に弊社によって見つけて削除されたものでした。 ③ 通知は次のように送信されます: (コミュニティ規定違反の場合)

③ 申告・要請をした者及び発信をした者に対する削除対応等の実施の判断に係る通知の有無及びその内容。 ④ 削除対応等の実施の判断に対する不服申立制度の有無及びその内容。	あなたのコンテンツが削除されるべきではないと考え、監視委員会に異議申し立てたい場合は、まずレビューを要求するプロセスを経る必要がある旨を伝えます。 (現地法違反の場合) 規制当局が主張した理由を削除の理由として伝えます。 ④ 利用者がコンテンツの削除決定に同意しない場合、コンテンツの再審査を要求することができます。 https://www.facebook.com/help/2090856331203011?helpref=faq_content https://help.instagram.com/280908123309761?helpref=faq_content
(5)その他の体制整備の状況 ① 違法情報の削除対応等の実施の判断基準及び意思決定プロセス。	上記2(1)のとおり。
② 自ら提供するSNS等において流通する違法情報への削除対応等について、日本語で対応が可能な人員として常用している人数。また、違法情報の削除対応等の審査・判断に際し、日本の法令等に精通した人材がいる場合、当該人材の知識経験(弁護士等)及び人数。	提出資料参照
<u>2) 違法・有害情報への共通する対応について</u>	ほとんどの違反について、繰り返しの警告や制限にもかかわらず、コミュニティ規定に違反するコンテンツを投稿し続ける場合、弊社はアカウントを無効化します。

Q2-3 違法・有害情報の流通・拡散を抑止するための方策について、次の5点を御教示下さい。 (1) 繰り返し行為に対する取組 繰り返し違法・有害情報を発信・拡散する行為を抑止するための取組(貴社のサービス 設計・機能上の工夫等)。	一定回数の違反を繰り返した場合には、30日間のコンテンツ作成制限を受けることがあります。または、違反の重さと頻度に応じて、アカウントを削除する場合があります。場合によっては、違反が非常に深刻であるため、1回の発生だけでアカウントを無効化する場合があります。例えば、児童性的搾取コンテンツや人身売買・労働搾取コンテンツを投稿した場合などです。 https://transparency.meta.com/ja-jp/enforcement/taking-action/restricting-accounts/
(2) その他の態様・場面に着目した取組 その他の違法・有害情報の流通の態様・場面に着目した取組(1回限りの捨てアカウントに対する貴社のサービス設計・機能上の工夫、災害時等の特定の場面に着目した取組等)。	コミュニティ規定に違反するコンテンツを投稿した場合、弊社はこれを削除し、その投稿を行ったFacebookまたはInstagramアカウントに1ストライクをカウントします。ストライクをカウントするかどうかは、コンテンツの重性、共有された状況、および投稿された時期によって決まります。 このコンテンツをFacebook上で管理しているページまたはグループに投稿した場合、そのページまたはグループにもストライクがカウントされる場合があります。グループを管理している場合、承認した違反もそのグループに対するストライクとしてカウントされる場合があります。場合によっては、違反が非常に深刻であるため、例えば児童性的搾取コンテンツを投稿した場合など、1回のストライクだけでFacebook上のアカウント、ページ、またはグループ、またはInstagram上のアカウントを無効化する場合があります。

(3) 社会的影響の評価

諸外国における制度(欧州のデジタルサービス法等。以下同じ。)の適用を受けている場合において、当該制度に基づき、自らが提供するSNS等において流通する違法・有害情報の社会的影響について、社会的影響の評価をどのように実施しているか。また、最新の評価結果はどのような内容か。

弊社では、違法および有害なコンテンツの社会的影響を評価し、軽減するためのアプローチを採っていますが、このアプローチはすべての国に一貫したものであり、以下のように、違法および有害な情報を取り扱うだけでなく、次のように多面的な取り組みを示しています。

堅牢なポリシー：弊社のポリシーは、ヘイトスピーチや児童搾取など、違法性を帯びる可能性のある幅広い危害にも対処するように設計されています。弊社は、コミュニティ規定に違反するコンテンツを検出し、これを削除します。これらの規定は、プラットフォーム上で許可されていないものを明確に概説しています。このポリシーは、オンラインの脅威の進化に応じて定期的に更新されます。

検出と執行：弊社は、テクノロジー・ドリブンのソリューション、人間によるレビュー、そして利用者の報告を組み合わせて、ポリシーに違反するコンテンツを検出し、削除しています。弊社の検出システムは、潜在的な違反も迅速かつ正確に特定するように設計されており、訓練されたレビューチームがフラグ付きのコンテンツを慎重に評価して、規定に合致しているかどうかを判断します。

違法であると主張されているけれども、必ずしもコミュニティ規定に違反していないコンテンツについては、政府のレポーティング・チャンネルを構築して、法的枠組みへの遵守を確保します。政府および認可された組織・団体は、削除要求をすることができます。これらの要求は、有効性(validity)、法的根拠(legal basis)、および国際基準への遵守について審査されます。法的に求められる場合、その管轄区域内でのコンテンツへのアクセスを制限しながら、意思決定プロセスの透明性を確保します。

教育とパートナーシップ：弊社は、業界およびNPO/NGOらと密接に協力して、新たな脅威について情報を収集し、それらに対処するための効果的な戦略を開発しています。

ユーザーツールとコントロール：弊社は、利用者が設定をカスタマイズし、不審な活動を報告し、望まないインタラクションをブロックできるようにするためのツールとコントロールを移譲しています。

弊社はまた、違法および有害なコンテンツのモデレーションに関する詳細な知見を提供する透明性レポートを定期的に公開しています。これらのレポートは、コンテンツの削除数、政府の要求、およびプラットフォーム全体での執行措置を強調しています。弊社の透明性センターは、この情報を統合し、コンテンツポリシーとその執行に関する包括的なビューを提供します。

<https://transparency.meta.com/>

継続的な改善への取り組みの一環として、弊社は違法および有害なコンテンツに関連するトレンドを評価するための定期的なリスクアセスメントを行っています。先ごろ、他国で法律に準拠して実施されたリスク評価では、リスクの状況に影響を与えるような新たな傾向は確認されませんでした。しかし、生成型AI技術がもたらすような新たなリスクに対しては、引き続き警戒と適応を続けていきます。

(4)社会的影響の評価結果を踏まえた軽減措置の内容 諸外国における制度の適用を受けている場合において、当該制度に基づき、自らが提供するSNS等において流通する違法・有害情報の社会的影響について、上記(3)の最新の評価結果を踏まえた軽減措置としてどのような措置を講ずることとしているか。	上記2(3)参照
(5)(3)、(4)の制度に係る評価 諸外国における制度の適用を受けている場合において、(3)の社会的影響の評価及び(4)の軽減措置の制度について、どのように評価をしているか。	海外の規制、例えばEUの規制を踏まえて検討する場合には、デジタルサービス法(DSA)がその目標を達成するかどうか、またどのように達成するかについてまだ完全な絵が描けていないことを認識することが重要です。DSAは誕生から間もない萌芽期にあり、何らの検証も受けていません。プラットフォームと欧州連合の規制当局はすべてこのプロセスから学習の途上であり、特に違法および有害なコンテンツに対処する上で、DSAが意図通りに機能するかどうか判断するには、時期尚早です。実際、ある国の既存の法律を他の管轄区域に焼き直そうとしても、うまく機能しません。しかし、いくつかの国際標準化機関はコンテンツモデレーションの基準を確立するために取り組んでいます。これらのリソースを活用することで、国家の優先事項に沿ったルールを作成しながらも、プラットフォームが国境を越えたグローバルなインターネット上で運営できるようにし、その国のインターネットユーザーのユーザーエクスペリエンスを断片化させないようにすることができます。 DSAはプラットフォームに対して重大な義務を課していますが、DSA以前に導入された規制モデルは、DSAを導入するとどのような課題に直面することになるのか、深い洞察を与えています。 例えば、Computer and Communication Industry Association(CCIA)は、ドイツのネットワーク強化法(NetzDG)(DSAで採用されている特定の条項)に関する研究Germany's Network Enforcement Act(Netzwerkdurchsetzungsgesetz or NetzDG)において、以下のことを指摘しています。 1. 政府の命令には実効性が乏しいこと: 1. 追加的削除件数は微小:NetzDGによって、4つの主要なソーシャルメディアサイト(これらのサイトは年間数兆件の利用者投稿コンテンツを扱っている)において、5,138件のコンテンツ削除が行われましたが(2022年)、前年度からの増分は微小でした。

	2. 高い第一種の誤り(false positive): 上記に加えて、84.4%の報告がfalse positiveであり、残りの報告の98.8%はすでにプラットフォームのポリシー・基準によって対処されていました。 2. 高額なコンプライアンス費用: 1. 大幅なコンプライアンスコストの増大:コンプライアンス体制を敷くには、4つのソーシャルメディアサイトで計441人のスタッフを必要とし、増加した一件の削除のために1,741ドルから5,116ドルを要した計算になります。 2. 経済的影響:ドイツ経済へのより広範なコストは、毎年22.25百万ドルと推定されており、これは増加した一件の削除のために約4,336ドルを要したと計算になります。 3. 競争抑制的な側面: 1. 小規模サイトに課せられることとなったアンバランスな負の影響:高額なコンプライアンスコストは、小規模のソーシャルメディアサイトやスタートアップにとって参入障壁となり、競争を阻害するものでした。スタートアップは、リソースが少なく、大規模で確立されたプラットフォームに比べてコンプライアンス要件の影響を受けやすくなります。 4. 有害なコンテンツへの限定的な影響: 1. 有害なコンテンツの減少は微減にとどまる:NetzDGにより削除された利用者投稿コンテンツの量は極めて少なく、毎日数十億件の投稿を扱うプラットフォームでは6ヶ月ごとに2,000件未満の投稿が削除されたに過ぎませんでした。 これらの不確実性を考えれば、定期的なレポートを通じた継続的な透明性が、プラットフォームが執行アプローチを改良し、リスク軽減戦略を向上させながら、説明責任とユーザーの安全を確保するための最善の方法であると言えます。
--	---