

誹謗中傷等への対策状況 ヒアリングシート (2022年3月7日)

社名	Facebook Japan 株式会社	
1. サービス概要とポリシー		
①	対象サービス名	Facebook/Instagram
②	加入者数、月間アクティブユーザ数又は書き込み数	【日本の数値】 Facebook 月間アクティブ利用者数 : 2,600 万人 (2019 年 3 月) Instagram 月間アクティブアカウント数 : 3,300 万 (2019 年 3 月)
		【グローバルの数値】 Meta が提供するプラットフォーム利用者数 : 月間アクティブ利用者数 35 億 9,000 万人 (2021 年 12 月時点)
		【(質問に答えられない場合) 参考となる数値】

③

(i) 誹謗中傷等に関して禁止とする情報や行為

Facebook コミュニティ規定及び Instagram コミュニティガイドラインでは、それぞれ Facebook 及び Instagram で認められる行為と認められない行為の概要を示しています。
弊社のコミュニティ規定は、ヘイトスピーチ、いじめや嫌がらせ、自殺や自傷行為、成人の性的搾取など、さまざまな種類の有害なコンテンツに適用されます。また、コミュニティ規定は日本語で公開されています。
Facebook コミュニティ規定について、下記 URL をご参照ください。

<https://www.facebook.com/communitystandards/introduction>

Instagram コミュニティガイドラインについて、下記 URL をご参照ください。

https://help.instagram.com/477434105621119/?helpref=uf_share

弊社のいじめや嫌がらせに関するポリシーに違反する行為の例としては、以下のようなものがあります（ただし、これらに限定されません）。

- 一方的に望まれない形で又はセクシャルハラスメントの形で誰かに繰り返し接触する行為
- 特定の人物やグループの自傷行為や自殺を求める行為
- 性的行為に関する侮蔑的な言葉（淫売、尻軽など）を用いた攻撃
- 個人の私用の電話番号、住所または電子メールアドレスを公表するとの脅迫
- 性的行為を行う意思の表明または性的行為を他人に呼びかける発言（例：レイプの脅し）
- 著しく性的なコメント
- 加工された軽蔑的な性的画像や性的描画のシェア
- いじめや嫌がらせを望む発言や、いじめや嫌がらせを行う意思の表明

詳細は、下記 URL をご参照ください。

<https://transparency.fb.com/ja-jp/policies/community-standards/bullying-harassment>

弊社のヘイトスピーチに関するポリシーに違反する行為の例としては、以下のようなものがあります（ただし、これらに限定されません）。

保護の対象となる特性（人種、民族、国籍、障がい、宗教、社会階級、性的指向、性別、ジェンダーアイデンティティ、重度の病気などや移民ステータス）に基づき、個人またはグループを標的とした、

- 暴力的またはそれを支援する、記述または視覚的な形による発言
- 人間性を否定するような発言や画像

詳細は、下記 URL をご参照ください。

<https://transparency.fb.com/ja-jp/policies/community-standards/hate-speech/>

	<p>(ii) (i)で挙げた行為に対して取り得る対応</p>	<p>弊社がFacebookのコミュニティ規定やInstagramのコミュニティガイドラインに違反するコンテンツがあると判断した時点で、弊社はそれを削除します。その際にはコンテンツを共有した利用者には、削除した理由を理解してもらい、今後違反するコンテンツを投稿しないようにする方法を理解してもらうために通知します。この情報は、利用者のサポート受信箱 (www.facebook.com/support)でも確認できます。Instagramにおいても同様の確認ができます。また、弊社の判断が何か間違っていると思われる場合には、その措置を不服として訴えることができるプロセスも用意されています。</p> <p>弊社は、投稿するコンテンツに対して利用者に責任を持ってもらうために、違反回数を数えてストライクシステムを運用しています。違反しているポリシー、弊社プラットフォームにおける違反の履歴、ストライク（弊社プラットフォームで定めているポリシーに抵触あるいは違反した）回数に応じて、アカウントを制限または停止することもあります。</p> <p>ストライク回数のカウント、アカウントの制限及び停止等についての詳細については下記URLをご参照ください。 https://transparency.fb.com/ja-jp/enforcement/taking-action/</p>
<p>④</p>	<p>ポリシー等の見直し状況及び外部レビューの有無とそのタイミング</p>	<p>弊社のコミュニティ規定は生きた文書であり、弊社のポリシーもオンラインにおける人々の行動の変化に合わせて進化しています。例えば、かつては無害だった言葉が害のある言葉になることはありますし、その逆もあり得ます。また、特定のグループが急に攻撃の対象になることもあり得ます。</p> <p>弊社のコンテンツポリシーチームは、2週間毎にポリシーフォーラムという会議を開き、コミュニティ規定及び広告ポリシーについて、調整の必要がないか、議論します。議論は、セーフティチーム、サイバーセキュリティチーム、カウンターテロリズムの専門家、グローバルオペレーションチーム、プロダクトマネジャー、リサーチャー、公共政策担当、法務担当、人権とダイバーシティチームなど、社内の幅広い専門家を集めて行われ、時にはジャーナリストやアカデミアの方にオブザーバーとして参加していただくこともあります。弊社の透明性と説明責任へのコミットメントの一環として、ポリシーフォーラムの議事録は公開されています。詳細については下記URLをご参照ください。 https://transparency.fb.com/ja-jp/policies/improving/deciding-to-change-standards/ https://transparency.fb.com/ja-jp/policies/improving/policy-forum-minutes/</p> <p>外部専門家との連携 弊社のコンテンツポリシー/ステークホルダーエンゲージメントチームは社外ステークホルダーと定期的に連携し、彼らの世界を網羅する知識と助言をポリシー策定プロセスに取り入れてポリシーをより強固なものにするために尽力しています。</p>

さらに、弊社のポリシーが与える影響を深く理解するため、NGO や学術機関などの専門家との関係を築いたり、世界中のさまざまな市民社会団体と連携したりしています。

ステークホルダーとの連携への弊社の取り組みについては下記 URL をご参照ください。

<https://transparency.fb.com/ja-jp/policies/improving/input-from-external-stakeholders/>

これに加えて、2020 年、弊社が行う最も重要で困難なコンテンツに関する決定について、独立したチェックを行うために、監督委員会が設立されました。監督委員会は、コンテンツに関する決定が Facebook や Instagram のポリシーや価値観、さらには国際的な人権規範の枠組みの中で表現の自由を守るというコミットメントと一致しているかどうかを審査します。そして、監督委員会は、これらの原則及び利用者と社会への影響に基づいて決定を下します。これらの決定には拘束力があり、監督委員会の会則で、弊社は、監督委員会の決定の実施が法律に違反するおそれがない限り、実施しなければならないこととされています。また実際に弊社は、監督委員会の決定を全て、その公表後速やかに実施しています。

<https://transparency.fb.com/oversight/>

<https://about.fb.com/ja/news/2019/07/global-feedback-on-oversight-board/>

<https://about.fb.com/ja/news/2019/12/oversight-board-update/>

<https://about.fb.com/ja/news/2020/02/facebooks-oversight-board/>

<https://about.fb.com/ja/news/2020/05/welcoming-the-oversight-board/>

2. 削除申告等の受付・審査のプロセス

①

(i) 一般ユーザからの
申告・削除要請への
受付窓口・受付態勢

弊社は、コミュニティからの報告、弊社のコンテンツ・モデレーション・チームによる審査、および技術を組み合わせて、コミュニティ規定を施行しています。

コミュニティからの報告については、

- Facebook では、ページ、グループ、プロフィール、投稿、写真、ビデオ、コメント、広告など、すべてのコンテンツを報告することができます。また、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は、誰でも弊社に報告することができます。
- Instagram でも同様に、アプリ内のフィード投稿、ビデオ、ストーリーズ、リール、コメントなど、あらゆるコンテンツを簡単に報告することができます。

また弊社は、世界各地の市民団体とのパートナーとしてのネットワークを構築しており、彼らは専用のチャンネルを通じて弊社に連絡を取り、新たな問題を警告したり、弊社のチームが知らないような重要な情報を提供してくれます。

詳細については下記 URL をご参照ください。

<https://www.facebook.com/help/1380418588640631/>

<https://www.facebook.com/help/263149623790594/>

<https://help.instagram.com/2922067214679225/>

現在、Facebook と Instagram の違反コンテンツの特定には、主に人工知能（AI）が利用されています。弊社の技術は、コンテンツが弊社のコミュニティ規定に違反していると認め、（多くの場合、誰かの目に触れる前に）自動的に削除します。また、弊社のプラットフォームで審査するコンテンツの量に鑑み、AI は、弊社のコンテンツモデレーターが審査する案件の優先順位付けにも役立っており、最も有害で一刻を争うコンテンツを優先的に審査しています。優先順位付けは、重大性、バイラリティ、違反の可能性など、いくつかの要素に基づいて行われます。

弊社は、15,000 人を超える世界各国の審査担当者を備えています。審査担当者は、さまざまなバックグラウンドを持っています。そして弊社として人材の採用及び最適化において最も重要であると考えているのが、言語と文化的背景です。

Meta の審査チームはグローバル体制の下、24 時間年中無休でコンテンツを審査しています。ドイツ、アイルランド、ラトビア、スペイン、ポルトガル、フィリピン、米国など、弊社が全世界に展開している 20 以上の拠点でコンテンツの審査を行うこれらのチームは全体で日本語を含む数十種類に及ぶ言語に対応しています。

審査チームについての詳細は下記 URL をご参照ください。

<https://transparency.fb.com/ja-jp/enforcement/detecting-violations/how-review-teams-work/>

(ii) 主体的な誹謗中傷等の情報の検知態勢

Meta では独自のテクノロジーを利用して違反コンテンツを積極的に検出し、利用者によって報告される前にその大半を削除しています。エンジニア、データサイエンティスト、審査チームはこのテクノロジーを更新してさらに洗練されたものにするために協力して取り組んでいます。こうした中、審査チームはこのテクノロジーを活用して、コンテンツの審査優先順位の決定にも役立てています。

弊社が Facebook と Instagram で削除している違反投稿・違反アカウントの数は 1 日あたり数百万件に達しています。これらの大半は、背後にあるテクノロジーにより自動的に実行され、しかもそのほとんどが利用者に表示される前に処理されています。ときには、違反の可能性のあるコンテンツが検出された後、確認と措置の実施のために審査チームに転送されることもあります。

		<p>違反コンテンツに対する取り組みに終わりはありません。弊社のテクノロジーによる監視の目を巧妙に逃れようとする行為は尽きることがないため、弊社は絶え間なくテクノロジーを改良し続ける必要があります。テクノロジーによる違反検出のしくみ等についての詳細は下記 URL をご参照ください。 https://transparency.fb.com/ja-jp/enforcement/</p>
②	<p>申立時における申立者への通知の有無</p>	<p>利用者は、行った報告について、サポート受信箱 (at www.facebook.com/support) でその報告のステータスを確認できます。 同様に、Instagram の場合、Instagram アプリの[サポートリクエスト]から報告のステータスを確認できます。利用者のアカウントが日本語に設定されている場合、すべての通知やレポートは日本語で提供されます。 https://www.facebook.com/help/338745752851127 https://help.instagram.com/484700649522086</p>
③	<p>申立てに対するポリシー違反等の審査プロセス、案件ごとの優先度の決定方法</p>	<p>弊社は、世界で 30 億人以上の人々が国や文化、言語を超えて自由に自己表現できるサービスを提供しています。しかし、人々が安心して自由に自分を表現できる環境を整えるためには、コミュニティの安全性、プライバシー、尊厳、信頼性を維持する必要があります。</p> <p>そのため、弊社はコミュニティ規定を設けています。これは、どのコンテンツを公開し、どのコンテンツを削除するかを決定するための一貫した枠組みを提供するものです。</p> <p>多くの場合、コミュニティ規定の言葉は、意図的に、特定的かつニュアンスを含む表現になっています。これは、表現の自由（弊社はこれを基本的人権と考えています）を可能にする一方で、最も有害なタイプのコンテンツを削除することで、コミュニティの安全と福祉を確保するためです。</p> <p>コミュニティ規定をどのように実施し、何を掲載すべきか、何を削除すべきかをどのように決定するかについての詳細は、下記 URL をご覧ください。 https://about.fb.com/news/2018/04/comprehensive-community-standards/ https://about.fb.com/news/2019/03/inside-feed-vanity-fair-policy-team/ https://about.fb.com/news/2018/08/hard-questions-free-expression/ https://about.fb.com/news/2018/04/community-standards-examples/</p> <p>コミュニティ規定の施行には、コミュニティからの報告、コンテンツモデレーションチームによる審査、および技術を組み合わせて使用しています。</p>

コミュニティからの報告については、

- Facebook では、ページ、グループ、プロフィール、投稿、写真、ビデオ、コメント、広告など、すべてのコンテンツを報告することができます。また、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は、誰でも弊社に報告することができます。
- Instagram でも同様に、アプリ内のフィード投稿、ビデオ、ストーリーズ、リール、コメントなど、あらゆるコンテンツを簡単に報告することができます。

また弊社は、世界各地の市民団体とパートナーとしてのネットワークを構築しており、彼らは専用のチャンネルを通じて弊社に連絡を取り、新たな問題を知らせたり、弊社のチームが知らないような重要な情報を提供してくれます。

毎週、数百万件のレポートを処理していますが、大半のレポートは 24 時間以内にレビューされています。

人間によるレビュー

- 15,000 人以上のコンテンツモデレーターがいます。このチームはグローバルに活動しており、日本語を含む数十の言語で 24 時間 365 日コンテンツをレビューしています。
- 弊社のコンテンツモデレーターはさまざまなバックグラウンドを持っていますが、弊社が人材の採用及び最適化において最も重要と考えているのは、言語と文化的背景です。
- (人間の) コンテンツモデレーターは、公正かつ正確にコミュニティ規定を実施するために不可欠であり、特にコンテンツを取り巻く文脈が重要な場合には、その重要性が増します。
例えば、ヘイトスピーチなどです。弊社のシステムは、ヘイトスピーチとしてよく使われる特定の言葉を認識できませんが、それを使う人の意図は必ずしも認識できません。そのため、審査チームがこのコンテンツをレビューします。

テクノロジーを使って有害なコンテンツを主体的に識別

- Meta では主に人工知能 (AI) を利用して、Facebook や Instagram 上の違反コンテンツを特定しています。弊社のテクノロジーは、コンテンツの一部が Facebook コミュニティ規定や Instagram のコミュニティガイドラインに違反していることを確信して、(多くの場合、誰かの目に触れる前に) 自動的に削除することができます。
- また、AI は、弊社のプラットフォームで審査するコンテンツの量が多いため、弊社のコンテンツモデレーターが審査する案件の優先順位付けにも役立っており、最も有害で一刻を争うコンテンツを優先的に審査しています。優先順位付けは、以下のようないくつかの要素に基づいて行います。
 - **バイラリティ** 違反する可能性のあるコンテンツがすぐにシェアされている場合、シェアや閲覧がないコンテンツよりも優先的に審査されます。
 - **重大性** 現実の被害に関連するコンテンツは、他のカテゴリーよりも優先的に審査されます。

		<p>○ 違反の可能性 弊社のポリシーに違反した他のコンテンツに類似したシグナルを持つコンテンツは、そのようなシグナルを持たないコンテンツよりも優先的に審査されます。</p> <p>- 弊社は、四半期ごとに発行されるコミュニティ規定施行レポート（Community Standards Enforcement Report）で、AI を使った進捗状況を公開しています。 https://transparency.facebook.com/community-standards-enforcement</p>
④	<p>一般ユーザからの申告や削除要請に対応する部署・チームの規模・人数 及び部署・チームに関する日本国内の拠点の有無、日本における責任者の有無</p>	<p>【日本の数値】</p> <p>【グローバルの数値】 Meta の審査チームは、数多くの職務の一環としてコンテンツ審査を担当する弊社の社員と、弊社のパートナーに従事するコンテンツ審査担当者によって構成されています。Meta が提供するプラットフォームのコミュニティの多様性の豊かさを反映して、審査チームには退役軍人や法律専門家に加え、児童の安全、ヘイトスピーチ、対テロ対策などの各種ポリシー分野に精通する規定施行の専門家など、さまざまな背景と職歴を持つ人たちが集結しています。 また、弊社はさまざまな企業から参画している 15,000 人以上の審査担当者と協力しています。Meta の審査チームはグローバル体制の下、24 時間年中無休でコンテンツを審査しています。ドイツ、アイルランド、ラトビア、スペイン、ポルトガルフィリピン、米国など、弊社が全世界に展開している 20 以上の拠点でコンテンツの審査を行うこれらのチームは全体で日本語を含む数十種類に及ぶ言語に対応しています。</p> <p>https://transparency.fb.com/ja-jp/enforcement/detecting-violations/people-behind-our-review-teams/</p> <p>【（質問に答えられない場合）参考となる数値】</p>
⑤	<p>(i) 対応決定時の申立者への通知の有無とその理由説明内容</p>	<p>利用者が Facebook や Instagram で報告した場合、その報告の状況を Facebook ではサポート受信箱や Instagram ではサポートリクエストから確認することができます。</p> <p>報告を受け取った日時、報告の理由の確認、内容についてのレビューの結果を通知しています。 弊社は、コンテンツがコミュニティ規定に反すると判断した場合、コンテンツを削除します。また、そのコンテンツが弊社のポリシーに違反していない場合には、その旨をお知らせします。その際、コンテンツの掲載を継続するという弊社の判断にご同意いただけない場合には、通常、再審査をリクエストする機会を提供しています。</p>

		<p>利用者が再審査をリクエストされた場合、コミュニティ運営の専門チームがコンテンツを再度審査します。間違いが見つかった場合には、利用者にお知らせし、報告されたコンテンツは非表示または削除されます。最も有害な種類のコンテンツ（児童搾取画像など）の一部については、再審査を申請することはできません。再審査のリクエストについては下記 URL をご参照ください。</p> <p>www.facebook.com/help/134552198624586/?helpref=uf_share</p> <p>利用者のアカウントが日本語に設定されている場合、すべての通知やレポートは日本語で提供されます。また、下記 URL にて利用者にとってポリシーの施行がどのようなものとなるかの例をご覧ください。</p> <p>https://transparency.fb.com/policies/community-standards/bullying-harassment/#user-experiences</p>
	<p>(ii) 対応決定時の対象となる情報の発信者への通知の有無とその理由説明内容及び異議申立て方法の案内の有無</p>	<p>Facebook のコミュニティ規定や Instagram のコミュニティガイドラインに違反するコンテンツがあると判断した時点で、弊社はそれを削除します。その際には コンテンツを共有した利用者には、削除した理由を理解してもらい、今後違反するコンテンツを投稿しないようにする方法を理解してもらうために通知します。この情報は、その人のサポート受信箱でも確認できます。また、弊社が何か間違っていると思われる場合には、その措置を不服として訴えることができるプロセスも用意されています。</p> <p>コンテンツの削除の流れについては下記 URL をご参照ください。</p> <p>https://transparency.fb.com/enforcement/taking-action/taking-down-violating-content/#takedown-experience</p>
<p>⑥</p>	<p>削除等への異議申立てや問い合わせに対する受付態勢及び処理プロセス</p>	<p>Meta がコンテンツに対して措置を講じると決定した時点では、その判断が誤っている場合もあります。自分のアカウントやコンテンツに対して行われた措置について、弊社の判断が誤っていると考えられる場合、ほとんどの状況において、異議申し立てを行うことができます。</p> <p>利用者が Meta における決定に異議申し立てを行うには、自らのコンテンツが削除または警告付きでぼかしが入れられたことについて弊社からの通知を受けた後、「審査のリクエスト」を選択します。</p> <p>異議申し立てがなされた場合、弊社の専門家であるコミュニティ・オペレーション・チームがそのコンテンツを再度審査し、弊社の判断が正しかったかどうかを確認します。</p> <p>弊社の判断が誤っていた場合には、利用者にもその旨をお知らせし、投稿、写真、ビデオは復元されます。</p> <p>また、弊社が措置を講じたコンテンツだけでなく、報告されたにもかかわらず対処されなかったコンテンツについても、異議申し立ての機会を提供し始めています。後者の異議申し立てについては、コミュニティ規定施行レポートには含まれません。</p> <p>児童の搾取に関連する画像など、重大な安全上の懸念がある違反については、異議を申し立てることができません。</p>

		<p>https://www.facebook.com/help/2090856331203011 異議申し立ての手続については下記 URL をご参照ください。 https://transparency.fb.com/en-gb/policies/improving/appealed-content-metric/ コンテンツが Facebook や Instagram から削除され、異議申し立ての手続を経ている場合は、監督委員会に申し立てることでその決定に異議を唱えることができます。 詳しくは下記 URL をご参照ください。 https://oversightboard.com/news/833880990682078-the-oversight-board-is-now-accepting-cases/</p>
⑦	⑥の異議申立・問い合わせ件数	<p>【日本の数値】</p>
		<p>【グローバルの数値】 最新の件数については下記 URL をご参照ください。 https://transparency.fb.com/data/community-standards-enforcement/bullying-and-harassment/facebook/ https://transparency.fb.com/data/community-standards-enforcement/hate-speech/facebook/ https://transparency.fb.com/data/community-standards-enforcement/bullying-and-harassment/instagram/ https://transparency.fb.com/data/community-standards-enforcement/hate-speech/instagram/</p>
		<p>【（質問に答えられない場合）参考となる数値】</p>
⑧	不正な申告や削除要請への対策の方法・仕組み、対応件数	<p>【具体的な方法・仕組み】 一度に大量の報告があったとしてもコミュニティ規定の施行には影響しません。弊社ではシステムの濫用を阻止する技術的対策が取られています。1 件でもコミュニティ規定に違反するコンテンツに関する報告があれば弊社は対応します。 また、アカウントやコンテンツの過剰な報告は、弊社のプラットフォームの濫用とみなされ、弊社のコミュニティ規定および利用規約に違反する可能性があります。弊社は、このような行為を制限し、また、このような行為の過度の違反を検知する技術を有しています。 利用者のネットワークが組織的にこの機能を悪用し、弊社の制限や執行を逃れようとしていることが判明した場合、弊社はそれらの利用者に対して措置を講じます。</p>
		<p>【日本の数値】</p>

		【グローバルの数値】
		【（質問に答えられない場合）参考となる数値】
3. 削除等の対応の実施状況及び機能上の工夫		
①	誹謗中傷等に関する 申告や削除要請の件数 (1)一般ユーザ、(2) 司法機関・行政機関	【日本の数値】 【グローバルの数値】 【（質問に答えられない場合）参考となる数値】 誹謗中傷を含むとして弊社が措置をとったコンテンツ数はこちらに記載されています。 https://transparency.fb.com/data/community-standards-enforcement/bullying-and-harassment/facebook/
②	(i) 誹謗中傷等に関する 申告や削除要請に 対する削除件数 (1)一般ユーザ、(2) 司法機関・行政機関	弊社では、Facebook と Instagram におけるポリシーの施行状況を公的かつ透明性をもって追跡し、Facebook と Instagram を安全で包括的なものにするための継続的な取り組みを示すために、コミュニティ規定施行レポートを四半期ごとに発行しています。 【日本の数値】 (2) Facebook や Instagram 上のコンテンツが現地国の法律に違反していると弊社に報告されたものの、弊社のコミュニティ規定には違反していない場合、弊社はそのコンテンツが違法であると主張されている当該国でのコンテンツの利用を制限することがあります。 詳細については、以下の URL をご確認ください。 https://transparency.fb.com/data/content-restrictions/country/JP/

	<p>【グローバルの数値】 (1) 【Facebook】 2021年第2四半期から2021年第3四半期、いじめや嫌がらせに関して措置を講じたコンテンツは790万件から920万件に増加しました。これは事前検出テクノロジーの言語の対象を拡大したことによるものです。2021年第3四半期での表示頻度は0.14%～0.15%でした。なお、今回いじめや嫌がらせに関する表示頻度を公開するのは初めての試みです。 2021年第2四半期から2021年第3四半期、ヘイトスピーチに関して措置を講じたコンテンツは3,150万件から2,230万件に減少しました。これは、弊社の事前検出テクノロジーを調整し、アクション実施の精度を向上したことによるものです。2021年第3四半期のヘイトスピーチに関する表示頻度は0.03%で、2021年第2四半期から減少しました。 【Instagram】 2021年第2四半期から2021年第3四半期、いじめや嫌がらせに関して措置を講じたコンテンツは450万件から780万件に増加しました。これは事前検出テクノロジーの言語の対象を拡大したことによるものです。2021年第3四半期での表示頻度は0.05%～0.06%でした。なお、今回いじめや嫌がらせに関する表示頻度を公開するのは初めての試みです。 2021年第2四半期から2021年第3四半期、ヘイトスピーチに関して措置を講じたコンテンツは980万件から600万件に減少し、第2四半期以前の水準に戻っています。2021年第3四半期での表示頻度は0.02%でした。なお、今回ヘイトスピーチに関する表示頻度を公開するのは初めての試みです。</p>
	<p>【（質問に答えられない場合）参考となる数値】</p>
<p>(ii) 削除以外の対応（非表示化やラベルの付与等）を行った件数 (1) 一般ユーザ、(2) 司法機関・行政機関</p>	<p>【日本の数値】</p>
	<p>【グローバルの数値】</p>
	<p>【（質問に答えられない場合）参考となる数値】</p>
	<p>【日本の数値】</p>

	(iii) 対応を行わなかった件数 (1) 一般ユーザ、(2) 司法機関・行政機関	【グローバルの数値】
		【（質問に答えられない場合）参考となる数値】
	(iv) 削除までの平均時間及び/又は削除対象の閲覧数の抑制程度 (1) 一般ユーザ、(2) 司法機関・行政機関	【日本の数値】
		【グローバルの数値】 違反コンテンツへの対応時間は、コンテンツの種類や重大性、悪質性、違反の可能性によって、数分から数時間、あるいは数日に及ぶことがあるため、明確な数字を示すことは困難です。 弊社の目標は、最も有害なコンテンツにできるだけ早くたどり着き、報告してくださった利用者に対応することです。 重要なのは、AI 技術の発展と、時系列レビューからランク付け（または優先順位付け）レビューへの移行により、コンテンツがいつ Facebook でシェアされたか、利用者から報告されたか、または弊社の技術で検知されたかに関わらず、コンテンツに対する判断ができるようになったことです。
③	(i) 主体的な削除件数（AI を用いた自動検知機能の活用等）	【日本の数値】
		【グローバルの数値】
		【（質問に答えられない場合）参考となる数値】
	(ii) 主体的な削除以外の対応（非表示化やラベルの付与等）を行った件数	【日本の数値】
		【グローバルの数値】
		【（質問に答えられない場合）参考となる数値】
	(iii) 主体的な対応までの平均時間及び/又は対象情報の閲覧数の抑制程度	【日本の数値】
		【グローバルの数値】

		【（質問に答えられない場合）参考となる数値】
④	<p>その他、誹謗中傷等個人の権利侵害を防止・低減するためのサービス上の工夫 （特に、いわゆる炎上への対策に資するもの）</p>	<p>弊社は2021年10月、弊社のアプリの利用者の保護を強化するために、いじめや嫌がらせに関するポリシーの更新を発表しました。</p> <ul style="list-style-type: none"> - これは、個人のプロフィールや投稿へのダイレクトメッセージやコメントが、望ましくない形で繰り返し大量に送られてくるなど、集団的な嫌がらせや脅迫から人々を守るためのポリシーです。 - また、公人を攻撃するより有害なコンテンツを削除するとともに、人権擁護活動家やジャーナリストなど、不本意ながら有名になってしまった公人の保護を強化します。 <p>Facebookにおいては、コメントがいじめや嫌がらせに利用されることがあることを受け、投稿のオプションメニューから複数のコメントを一度に非表示にしたり削除したりする方法を導入しました。また、自分のFacebookプロフィールやFacebookページのコメントに特定の言葉が表示されないようにすることもできます。 https://www.facebook.com/help/1182883832161405 https://www.facebook.com/help/131671940241729</p> <p>望まない注目を浴びることはストレスになりますし、いじめや嫌がらせを報告することに抵抗を感じる人もいます。また、いじめや嫌がらせは、被害者の目の届かないところで起こっている場合もあります。友人や家族がいじめや嫌がらせを受けているのを見かけたら、気になる投稿の上にあるメニューから、その人に代わって通報することもできます。</p> <p>さらに、以下のような機能も備えています。</p> <ul style="list-style-type: none"> - ブロックした人が新しいアカウントを設定したり、自分が管理している別のアカウントから利用者に連絡を取ろうとしたときに、友達リクエストやメッセージなどの迷惑な連絡を積極的に認識し、防止する手助けをする <p>また、Instagramは2019年10月、望まないやりとりからアカウントを保護するための「制限」機能を日本を含む全世界で提供開始したことを発表しました。</p> <p>オンライン上のいじめは複雑な問題です。オンライン上でいじめを経験する若い世代は、被害を報告したり相手をブロックしたりすることをためらう傾向があることがわかっています。「制限」機能は、Instagram上で相手がどのような態度をとっているのか把握しながらも、相手に知られることなく望まないやり取りからアカウントを守ることができます。 https://about.fb.com/ja/news/2019/10/restrict/</p> <p>さらに、Instagramは2021年8月、利用者を誹謗中傷から守る以下の新機能を導入することを発表しました。</p> <ul style="list-style-type: none"> - コメントやダイレクトメッセージ（DM）を送るアカウントを一時的に制限し、それ以外のアカウントから送られたメッセージを非表示にすることで、急激に注目を集めた利用者が望まないやりとりを防ぐことができる「抑制」機能 - 誹謗中傷を含むDMを分類できる「非表示ワード」機能を導入

弊社には、誰もが安心して Instagram を利用できるようにする責任があります。ヘイトスピーチやいじめは許容しません。また、そもそもそのような体験をする必要がないよう、利用者を守りたいと考えています。そのため、専門家やコミュニティからのフィードバックに常に耳を傾け、Instagram における体験を利用者自身が管理しやすくしたり、望まないやりとりから身を守るための機能を継続して開発しております。

Meta のコンテンツ配信に関するガイドラインでは、ニュースフィードでの配信が抑制されるコンテンツタイプについて説明しています。Meta はこのガイドラインの策定に伴い、問題のあるコンテンツや低品質のコンテンツをどのように定義し扱うかについて引き続き透明性を確保していきます。

ニュースフィードで問題のあるコンテンツの表示を減らすために Meta が行う施行は、「利用者の直接的なフィードバックへの対応」、「パブリッシャーによる高品質コンテンツへの投資の奨励」、「安全なコミュニティの構築」という価値観に対する取り組みに根ざしています。

例えば、「安全なコミュニティの構築」について、コンテンツの中には、その意図の有無にかかわらず、コミュニティにとって問題となるものがあります。Meta は、このようなコンテンツ（コミュニティ規定のボーダーライン上のコンテンツやコミュニティ規定に違反している可能性が高いコンテンツなど）が、そのコンテンツを積極的に求めている人以外の目に届きにくくなるようにしています。

4. 発信者情報開示への対応		
①	発信者情報開示請求を受けた件数	【日本の数値】
	(i) 裁判上の開示請求件数、(ii) 裁判外の開示請求件数	【グローバルの数値】
		<p>【（質問に答えられない場合）参考となる数値】</p> <p>日本における政府からのデータ開示要求への弊社の対応については、透明性レポートに開示されていますのでご覧ください。</p> <p>https://transparency.fb.com/data/government-data-requests/country/JP/</p>
②	開示を行った件数	【日本の数値】
	(i) 裁判上の開示件数、(ii) 裁判外の開示件数	【グローバルの数値】

		<p>【（質問に答えられない場合）参考となる数値】</p> <p>日本における政府からのデータ開示要求への弊社の対応については、透明性レポートに開示されていますのでご覧ください。</p> <p>https://transparency.fb.com/data/government-data-requests/country/JP/</p>
5. 情報分析・公開		
①	透明性レポートの公開 日本語で閲覧可能か	<p>コミュニティ規定施行レポートを日本語に翻訳し公表しています。</p> <p>https://about.fb.com/ja/news/2020/11/community-standards-enforcement-report-nov-2020/ https://about.fb.com/ja/news/2021/02/community-standards-enforcement-report-q4-2020/ https://about.fb.com/ja/news/2021/05/community-standards-enforcement-report-q1-2021/ https://about.fb.com/ja/news/2021/08/community-standards-enforcement-report-q2-2021/</p>
②	誹謗中傷等の取組の 効果分析	<p>Facebook と Instagram におけるポリシーの施行状況を公的かつ透明性をもって追跡し、Facebook と Instagram を安全で包括的なものにするための継続的な取り組みを示すために、コミュニティ規定施行レポートを四半期ごとに発行しています。</p>
③	誹謗中傷等の流通状況に関する分析・調査及びユーザに与える影響の分析・調査	<p>弊社は、改善すべき点を積極的に特定するために調査に投資していますが、弊社の調査は社内での検討に役立つように設計されており、文書は調査の背景と限界を理解している人々のために作成され、使用されるものであるため、これらのすべてを外部で共有することはありません。</p>
④	研究者への情報提供、利用条件	<p>弊社は、独立した研究者とプライバシーが保護されたデータを共有することで、Meta が与える社会への影響を研究することに尽力しています。</p> <p>弊社は、プラットフォームを利用する人々のプライバシーを維持しつつ、何十億ものデータポイントを学術関係者と共有するという複雑な問題のバランスを取っています。</p> <p>例えば、弊社は、研究者向けに以下のような様々なデータセットを提供しています。</p>

<https://research.facebook.com/data/>

The Open Research and Transparency program では、学術関係者や独立した研究者に、選挙、民主主義、幸福に焦点を当てて、Meta が世の中に及ぼす影響を研究するのに必要なツールとデータ（例えば、Ad Targeting Transparency Data Sets と URL Shares Data Set など）を提供しています。

CrowdTangle は、公開設定されているページ、公開設定されているグループ、検証済みプロフィール、および Instagram の公開アカウントにおけるエンゲージメントを測定し分析できるようにするものです。これにより、公開されているコンテンツを簡単にフォロー、分析、レポートすることができ、ソーシャルメディア全体で何が起きているのかについて理解できるようになります。CrowdTangle は、プライバシーに配慮した公開データや製品を提供することで、外部の組織が Facebook や Instagram で起きていることをより透明に把握するための最良の手段となっています。

Data for Good プログラムでは、Facebook プラットフォーム上のプライバシー保護されたデータから構築された地図、調査、洞察を提供するほか、衛星画像や国勢調査データなど、商業的にも公的にも利用可能なソースを使って開発されたツールを提供しています。

広告ライブラリ (Ad Library) は、Instagram を含む Meta のアプリやサービスで現在表示されているすべての広告を包括的に検索可能なコレクションとして提供し、広告の透明性を提供しています。

<https://www.facebook.com/ads/library/>

最近では、大規模なデータの分析を必要とする独自の研究を支援するために、FORT Researcher Platform を立ち上げました。FORT Researcher Platform は、管理された環境で機密情報へのアクセスを可能にするスケーラブルなプラットフォームとして開発されたものです。

<https://research.facebook.com/blog/2022/1/introducing-the-researcher-platform-empowering-independent-research-analyzing-large-scale-data-from-meta/>

6. その他の取組

①

普及啓発

【みんなのデジタル教室】

弊社は、アジア太平洋地域の専門家と協力して、Facebook や Instagram といったオンライン上での嫌がらせやいじめなどに対処し、責任あるデジタル市民によるグローバルコミュニティを構築するためのリソースを提供するプログラム「みんなのデジタル教室」を立ち上げました。日本では特定非営利活動法人企業教育研

研究会の協力のもと、デジタルリテラシーに関する出張授業を国内の中学校・高等学校などで行っています。、2021年12月末までに13,000名以上の学生が授業を受講しており、90%以上が“授業を受けて、インターネットやアプリ、SNSへの関心が高まった”と回答しています。

<https://wethinkdigital.fb.com/jp/ja-jp/>

https://about.fb.com/ja/news/2020/12/we_think_digital/

【保護者のためのInstagramガイド】

保護者の方は未成年のお子様がおオンラインでどんなことをしているのかを気にかけて、お子様がソーシャルメディアを安全に利用できるようにサポートするにはどうすればよいか考えていらっしゃるかと存じます。そこでInstagramでは『保護者のためのInstagramガイド』を作成しました。このガイドは、認定NPO法人カタリバ、特定非営利活動法人キッズドア、特定非営利活動法人ストップいじめ！ナビ、認定NPO法人3keys、認定NPO法人育て上げネット、TDU・雫穿大学といった子供の安全や若者、子育てを支援する団体と専門家の協力を得て作成したものです。

このガイドでは、Instagramでプライバシーや他の利用者との交流、利用時間、セキュリティを管理する方法について取り上げています。また、Instagramの基本やツールについても説明しています。保護者が未成年のお子様とInstagramについてオープンに話し合うための方法に関するアドバイスもご覧いただけます。

<https://about.fb.com/ja/news/2021/03/parentguide/>

<https://about.instagram.com/ja-jp/community/parents>

<https://help.instagram.com/154475974694511>

【#インスタ ANZEN カイギ】

若年層の利用者の皆様と一緒に安心安全なInstagramの使い方について考えるプロジェクト「#インスタ ANZEN カイギ」をUUUM株式会社と発足しました。クリエイター自身も、多くのフォロワーとInstagram上で接する中で、安全な使い方について悩む時があるところ、クリエイターと彼らのファンである利用者が、Instagramの安心安全な使い方について一緒に考える場を提供しています。

<https://about.fb.com/ja/news/2020/03/instaanzenkaigi>

【SAVE YOUR LOVE 好きっていいね。】

自分の“LOVE（好き）”を楽しむために、安心安全にInstagramを利用するための機能について知ってもらい、SNSの使い方について考えるきっかけにしてほしいという思いを込め、デジタル広告・渋谷駅構内広告の展開、キャンペーンステッカーの配布、Instagramでのライブ配信などを実施しました。

相手に知られずに、コメントやDMを制限することができる「アカウント制限」や投稿にコメントできる人や、コメントの内容を管理するための「コメント制限」などの機能を紹介しました。

		<p>https://about.fb.com/ja/news/2021/12/save_your_love/</p> <p>【リールで知ろう！インスタ ANZEN ヒント】 Instagram では、若年層の利用者にアプリの安心安全な使い方について考えてもらう機会を作るためセーファーインターネットデー（Safer Internet Day：2022年2月8日）に合わせ、「リールで知ろう！インスタ ANZEN ヒント」を実施しました。2020年3月から継続しているプロジェクト「#インスタ ANZEN カイギ」の一環として行い、国内のクリエイター 5名とコラボレーションし、リール動画 5本を共同制作しました。”嫌がらせだと感じるコメントがきた”、“ついアプリを長時間使ってしまう” など、若年層の利用者に身近な悩みであり、クリエイター自身も体験したことがある問題に対して、知っておきたい Instagram の機能を紹介しています。 いじめや嫌がらせコメントを防ぐ非表示ワード機能や、相手をブロックせずに DM などのやりとりを制限できる機能、アプリの利用時間を管理できる機能など、クリエイター自身も実際に活用している機能を気軽に見られる短尺動画（リール）で紹介することで、彼らのファンである若年層の利用者に安心安全な Instagram の使い方について知ってもらう狙いです。 https://about.fb.com/ja/news/2022/02/instagram_creator_safety_reels/</p>
②	<p>誹謗中傷等以外にポリシーで禁止・制限されている情報や行為</p>	<p>Facebook コミュニティ規定及び Instagram コミュニティガイドラインでは、それぞれ Facebook 及び Instagram で認められる行為と認められない行為の概要を示しています。 これらは日本語で公開されています。 コミュニティ規定が掲げている項目は幅広いですが、例えば以下のようなものがあります。</p> <p>暴力と扇動、危険な人物および団体、危害を加えるための計画および犯罪の助長、制限されている商品およびサービス、不正行為および詐欺、自殺と自傷行為、児童に対する性的搾取、虐待、児童のヌード、成人に対する性的搾取、いじめと嫌がらせ、人的搾取、プライバシーの侵害、ヘイトスピーチ、暴力や過激な描写を含むコンテンツ、成人のヌードと性的行為、性的行為の勧誘、アカウントの安全性と実名の使用、スパム、サイバーセキュリティ、偽装行為、フェイクニュース、加工されたメディア、追悼アカウント、知的財産権</p> <p>詳しくは下記 URL をご参照ください。 https://transparency.fb.com/ja-jp/policies/</p>
③	意見・補足	