

誹謗中傷等への対策状況 ヒアリングシート（2021年2月25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
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】 月間アクティブアカウント数：2,600万人（2019年7月）
	【グローバルの数値】 月間アクティブアカウント数：28億人（2020年6月）
	【（質問に答えられない場合）参考となる数値】

<p>③ (i) 誹謗中傷等に関して禁止とする情報や行為</p>	<p>私たちのコミュニティ規定は下記のとおり幅広いテーマを扱っており、その多くは、私たちのサービスにおけるいじめや嫌がらせから人々を守るものです。</p> <p>アカウントの健全性と実名の使用：利用者のアイデンティティの真正性は、Facebook コミュニティの根幹をなすものです。実名の使用により、人々がお互いに対して、また Facebook に対して、有意義なあり方で責任をもつコミュニティ作りが促進されるものと Facebook は考えています。また、自分らしさを表現する多種多様な方法を Facebook のグローバルコミュニティ全体で許容することを目指しつつも、同時に、なりすましや身元の虚偽表示を防止する必要があります。</p> <p>ヘイトスピーチ：Facebook では、利用者が自身の属性を理由に攻撃を受けていると感じることがなければ、より自由な発言と交流が可能になると信じています。そのため、Facebook ではヘイトスピーチを許容しません。ヘイトスピーチは他者を脅迫したり排除したりする環境を生み出し、場合によっては現実世界での暴力行為を促すおそれがあります。Facebook が考えるヘイトスピーチとは、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、保護特性と呼ばれるものを理由に人々を直接攻撃することです。攻撃としてとらえるものには、暴力的な発言や人間性を否定するような発言、有害な固定観念、他者を見下した発言、侮辱的または嫌悪的な表現、過小評価する表現、暴言、他者を排除したり差別したりするような呼びかけが含まれます。</p> <p>いじめや嫌がらせ：いじめや嫌がらせは、あらゆる場所で、さまざまな形で起こります。例えば、個人情報を公開すると脅迫すること、脅迫的なメッセージを送信すること、悪意をもって一方的に連絡することもいじめや嫌がらせにあたります。Facebook はこのような行為を許容しません。利用者が安心して交流し、尊重されていると感じることができなくなるからです。Facebook は、いじめや嫌がらせが未成年者に与える精神的影響について認識しています。そのため弊社のポリシーでは、13 歳から 18 歳までの利用者に対する保護を強化しています。</p> <p>Facebook では、いじめや嫌がらせについて公人・著名人と一般の個人を区別して規定しています。これは、Facebook において議論や意見交換は認められる行為であり、報道で取り上げられたり社会的に注目を集めたりする人については、批判的な意見も含めた議論がされるためです。公人・著名人</p>
----------------------------------	--

		<p>に関しては、攻撃が重大な場合、また本人が投稿やコメントに直接タグ付けされた場合は削除します。一般の個人に対しては、保護の範囲が広がります。</p> <p>成人に対する性的搾取：性的な暴力や搾取に対する議論や注意喚起を行える場所であることが大切だと考えています。利用者が安全な環境で対話できる場とするために、被害者が自らの体験を共有できる場として Facebook を利用することは認められていますが、性的暴力や暴行、性的搾取の描写、またはこれらに関する脅しや助長などを含むコンテンツは削除の対象となります。また、同意のない性的行為の助長を避けるために、同意のない第三者との性的行為の表示、擁護、または手配を含むコンテンツも削除されます。「リベンジポルノ」や「公開許可のない性的描写がある画像」と呼ばれる画像をシェア、脅迫またはシェアする意図を表明したり、提供、要求したりすることに対しても、私たちは一切の寛容さを持っていません。</p> <p>詳しくは、https://www.facebook.com/communitystandards/をご覧ください。</p> <p>また、10代の若者、保護者、いじめやその他の紛争に関する問題のサポートを求める教育者のためのリソースである「いじめ防止ハブ」(https://www.facebook.com/safety/bullying)も設置されています。</p>
	(ii) 削除等の対応	<p>違反行為に応じて、Facebook は違反コンテンツを削除したり、センシティブなコンテンツ（例えば、行為を非難する文脈で共有された場合は物理的ないじめの画像など）に警告画面を適用したり、深刻な違反行為を行ったアカウントを停止したりします。</p>
<p>2. 削除等の対応</p>		

<p>①</p>	<p>誹謗中傷等に関する申告や削除要請の件数</p> <p>(i) 一般ユーザ、(ii) 司法機関・行政機関</p>	<p>【日本の数値】（当社サービス全て、誹謗中傷以外を含む）</p> <p>(i) 非公開</p> <p>(ii) 86 件（2019 年 6 月～2020 年 5 月）</p> <p>詳細は以下を参照ください。</p> <p>https://transparency.facebook.com/government-data-requests/country/JP</p> <p>（行政機関および法執行機関からの情報開示請求）</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>②</p>	<p>誹謗中傷等に関する申告や削除要請に対する削除件数</p> <p>(i) 一般ユーザ、(ii) 司法機関・行政機関</p>	<p>【日本の数値】（誹謗中傷以外を含む）</p> <p>(i) 0 件（2020 年 1 月～6 月）</p> <p>https://transparency.facebook.com/content-restrictions/country/JP</p> <p>（利用者の申告に基づき、コミュニティ規定に違反していないコンテンツで、日本の法令違反だとみなされたコンテンツの制限）</p> <p>(ii) 0 件（2020 年 1 月～6 月）</p> <p>https://transparency.facebook.com/content-restrictions/country/JP</p> <p>（司法機関の命令に基づき、コミュニティ規定に違反していないコンテンツで、日本の法令違反だとみなされたコンテンツの制限）</p> <p>全情報開示請求のうち 53%で何らかのデータを提供（2020 年 1 月～6 月）</p> <p>詳細は以下を参照ください。</p> <p>https://transparency.facebook.com/government-data-requests/jan-jun-2020</p> <p>（行政機関および法執行機関からの情報開示請求、当社サービス全て）</p>
----------	---	--

		<p>【グローバルの数値】</p> <p>(i) 630 万件 (2020 年 10 月～12 月)</p> <p>詳細は以下を参照ください。</p> <p>https://transparency.facebook.com/community-standards-enforcement#bullying-and-harassment</p> <p>(コミュニティ規定に違反したコンテンツの削除件数)</p>
		<p>【(質問に答えられない場合) 参考となる数値】</p>
	<p>削除割合</p> <p>(i) 一般ユーザ、(ii) 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>

		【（質問に答えられない場合）参考となる数値】
削除までの平均時間 (i)一般ユーザ、(ii) 司法機関・行政機関		【日本の数値】
		【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>Facebook は、削除される前の違反コンテンツによりもたらされた影響を評価する際には、削除されるまでの時間よりも閲覧された回数に着目することに意味があると考えています。</p> <p>インターネット上でコンテンツによりもたらされる害は、それが見られてしまうことによって生じます。インターネットの性質上、コンテンツが見られる回数は均等ではありません。非常に短い期間で多数の利用者の目に触れることになり、多くの流通を得ることとなるコンテンツはそれほどありません。一方で、他の多くのコンテンツはインターネット上に長い間ありこそすれ、一部の利用者の間でのみ見られることとなります。有害なコンテンツの取締りを理解するために使用するあらゆる手段は、この点を考慮に入れなければなりません。</p>

		<p>このような理由から、私たちは「表示頻度 (prevalence)」という指標が重要であると考えています。つまり、当社の基準に違反するコンテンツが実際にどれくらいの頻度で見られているか、Facebook 上で見られたコンテンツの総量との相対的な関係に着目しています。</p> <p>そのため、違反の種類によっては、表示頻度という指標を計算しています。表示頻度指標は、違反コンテンツの総閲覧回数 (分子) を、すべてのコンテンツの総閲覧回数 (分母) に占める割合として推定します。</p> <p>例えば、2020 年第 4 四半期のヘイトスピーチの表示頻度は 0.07%~0.08%、つまり、コンテンツの 1 万ビューごとにヘイトスピーチのビューが 7~8 回あったこととなります。</p>
③	②以外の誹謗中傷等に関する主体的な削除件数 (AI を用いた自動検知機能の活用等)	<p>【日本の数値】</p> <p>【グローバルの数値】</p> <p>48.8% (2020 年 10 月~12 月)</p> <p>詳細は以下を参照ください。 https://transparency.facebook.com/community-standards-enforcement#bullying-andharassment</p> <p>「How much did we find before users reported it」 (事前対応率) を参照ください。</p>

		【（質問に答えられない場合）参考となる数値】
④	③についての削除の方法・仕組み（AIを用いた自動検知機能の活用等）	<p>Facebook は、ヘイトスピーチや暴力的なコンテンツ、嫌がらせやいじめなどの不適切なコンテンツに対して、人工知能（AI）の技術や訓練を受けた専任チームの取り組みを活用して対処しています。</p> <p>Facebook は、利用者からの報告と AI を利用して、コミュニティ規定に違反する可能性のある投稿や写真、その他のコンテンツを特定しています。これらのレポートは、日本語を含む 50 以上の言語に対応している当社のコミュニティオペレーションチームが 24 時間 365 日体制でレビューしています。</p> <p>いじめや嫌がらせは性質上、非常に個人的なものであるため、多くの場合、特定したり削除したりする前に、その行動への利用者からの報告が重要になります。つまり、テクノロジーを使っていじめや嫌がらせを積極的に検出することは、他の種類の違反行為よりも困難である可能性があります。私たちは、この問題に取り組み、コミュニティを守るために、積極的な検知技術への投資を続けています。</p>
⑤	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】</p> <p>システムの悪用を防ぐための技術を導入しており、大量の報告がコミュニティ規定の実施方法に影響を与えることはありません。</p> <p>報告されたレポートは、Facebook と Instagram の両方のプラットフォームを担当するさまざまなチームによって審査されます。利用者からの重複報告を認識する技術を使用しており、1,000 人があるコンテンツを報告しても、1,000 人が同じコンテンツをレビューすることがないようにしています。</p>

		<p>また、AI は、コンテンツレビュアーがレビューするケースに優先順位をつけて、最も有害で時間的な問題のあるコンテンツを最初に処理できるようにしています。</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>3. 削除要請や苦情に関する受付態勢・プロセス</p>		

①	一般ユーザからの申告・削除要請への受付窓口・受付態勢、対応プロセス	<p>Facebook では、Facebook 上のコンテンツを簡単に報告できるようにしています。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を報告するために、報告ツールを使用することを奨励しています。Facebook の利用者は、各コンテンツに表示されるオプションを含め、さまざまな方法でコミュニティ規定違反を報告することができます。</p> <p>「報告」の方法については、以下を参照してください。</p> <p>https://www.facebook.com/help/1380418588640631/</p> <p>利用者が問題のあるコンテンツを「報告」すると、Facebook はそのコンテンツを審査し、コミュニティ規定に準拠していないコンテンツをすべて削除します。Facebook が問題の原因となった人物と連絡を取った場合、利用者名などの個人情報が開示されたり、通知されたりすることはありません。</p> <p>利用者が Facebook のコミュニティ規定に違反していることを「報告」した場合、利用者はサポート受信箱で報告の状況を確認することができます。報告した利用者だけがサポート受信箱を見ることができるようになります。</p>
②	申立時における申立者や発信者への通知の有無	<p>有り</p> <p>申告を受け取った日時、申告の理由の確認、内容についてのレビューの結果を通知しています。</p>
③	対応決定時における通知の内容、理由の記載の程度	<p>Facebook では、虐待や嫌がらせであると報告されたコンテンツを評価する際には、常に、その報告をした人にサポート受信箱でその人の言語で最新情報を提供することで、明確なコミュニケーションをとっています。報告者には、当社が報告を評価したことだけでなく、当社がいつ、どのような措置をとったかがわかります。（報告されたコンテンツに対して措置をとったか、とらなかったかの理由、報告者が好まないコンテンツや当社のコミュニティ規定に反する可能性のあるコンテンツへの対</p>

		<p>処方法の提案を含む) 場合によっては、報告者は Facebook の決定の見直しを要求することもできます。</p> <p>また、Facebook は、報告の対象となった投稿を行った利用者が、実際にコミュニティ規定に反するコンテンツを投稿したと判断した場合はいつでも、報告された利用者に通知します。その利用者には、当社が禁止しているコンテンツや行動の種類に関するより明確な情報とともに、当社がそのコンテンツの一部を削除したという通知が届きます。</p> <p>場合によっては、(アダルトヌード等の画像といった) 当社のコミュニティ規定に反する可能性のある、利用者が過去に投稿した類似のコンテンツを削除するよう、確認を利用者にお願いすることがあります。</p> <p>これは、プラットフォーム上でどのようなコンテンツや行動が許容されるかについて人々を啓発するための取り組みの一環であり、これ以上の違反は、利用者が特定のアカウント機能へのアクセスを失ったり、アカウントへのアクセスを完全に失ったりする可能性があることを説明するためのものです。また、自分が管理しているグループやページ、イベントを削除した際には、その旨をお知らせしています。</p> <p>報告についての詳細は下記 URL をご覧ください。</p> <p>https://www.facebook.com/safety/tools/safety</p> <p>https://www.facebook.com/help/181495968648557/?ref=sc</p>
④	一般ユーザからの申告や削除要請に対応する	【日本の数値】

<p>部署・チームの規模・人数</p>	<p>【グローバルの数値】</p> <p>コミュニティ運営チームには、日本語を含む 50 以上の言語で Facebook や Instagram のコンテンツをレビューするコンテンツレビュアーがおり、世界中のサイトでグローバルコミュニティにサービスを提供しています。</p> <p>過去数年の間に、Facebook は安全、セキュリティ、コンテンツレビューに特化したチームを 35,000 人以上に拡充し、そのうち約 15,000 人をコンテンツレビュアーに充てています。</p> <p>これにより、あらゆる時間帯をカバーし、日本語を含む 50 以上の言語に対応し、グローバルに展開することが可能になりました。言語能力が鍵となり、24 時間体制でコンテンツをレビューすることができます。例えば、日本の真夜中に誰かが日本語の投稿を報告した場合、日本語を話すレビュアーが常に存在し、迅速なレビューのためにレポートを転送することができます。しかし、実際には言語サポートを必要としないレポート（ヌードなど）もあり、言語に関係なく世界中のレビュアーに割り振られます。言語能力に加えて、各投稿の文脈を理解しているかどうかを確認し、有害かどうか、ポリシーに違反していないかどうかを判断するために、文化を理解している人員の確保に努めています。</p>
	<p>【（質問に答えられない場合）参考となる数値】</p>

⑤	④の部署・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>当社の審査は、ドイツ、アイルランド、ラトビア、スペイン、リスボン、米国、フィリピンなどの国の世界 20 以上のサイトに拠点を置いています。</p> <p>いくつかのレビューサイトの所在地は公表していますが、安全性とセキュリティ上の理由から、多くのサイトやそこで働く人々についての具体的な詳細は公開していません。</p> <p>私たちの人の目による審査の担当者は、様々なバックグラウンドを持った人たちです。最も重要なのは、言語と文化的背景です。当社のチームは、日本語を含む当社のコミュニティで最も広く使われている言語をカバーし、50 以上の言語で当社のポリシーに照らし合わせてコンテンツを審査しています。</p>
⑥	削除等への苦情や問い合わせに対する苦情受付態勢及び苦情処理プロセス	<p>Facebook では、いじめや嫌がらせ、ヘイトスピーチなどのコミュニティ規定違反で削除された決定に異議申し立てする機会を提供しています。コンテンツが誤って削除されたと判断された場合、削除されたコンテンツは復元されます。</p> <p>詳しくは以下のサイトをご参照ください。</p> <p>https://about.fb.com/ja/news/2018/04/comprehensive-community-standards/</p> <p>https://transparency.facebook.com/community-standards-enforcement/guide</p>
⑦	⑥の苦情・問い合わせ件数	<p>【日本の数値】</p> <hr/> <p>【グローバルの数値】</p>

		<p>申し立ての件数とそれに対応した件数については以下のサイトをご参照ください。</p> <p>https://transparency.facebook.com/community-standards-enforcement#bullying-and-harassment</p>
		【（質問に答えられない場合）参考となる数値】
4. 発信者情報開示への対応		
① <u>発信者情報開示請求を受けた件数</u>		【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】
		【（質問に答えられない場合）参考となる数値】
② 開示を行った件数		【日本の数値】
		【グローバルの数値】

	(i) 裁判上の開示件数、(ii) 裁判外の開示件数	【（質問に答えられない場合）参考となる数値】
5. その他の取組		
①	普及啓発	<p>私たちは、第三者の専門家と共に開発された様々なリソースを利用者が利用できるようにしています。例えば、「Facebook を安全に利用するために」のウェブサイトでは 60 以上の言語でリソースを用意しています。</p> <p>https://www.facebook.com/safety/</p> <p>さらに Facebook 社は、アジア太平洋地域の専門家と協力して、いじめ防止やオンラインでの嫌がらせなど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究会の協力のもと、デジタルリテラシーに関する授業を提供しています。</p> <p>https://wethinkdigital.fb.com/jp/ja-jp/</p>

<p>② その他の対策（警告表示・ミュート機能等）</p>		<p>Facebook には以下のような機能があります。</p> <p>(1) 自分のプロフィールや投稿を誰が見たり共有したりできるかを細かく設定できる機能。 https://www.facebook.com/help/1297502253597210</p> <p>(2) コメントの非表示・削除機能。 https://www.facebook.com/help/297845860255949</p> <p>(3) 利用者をブロックする機能。 https://www.facebook.com/help/185897171460026</p> <p>(4) 利用者は、公開したくない写真が無断で共有されている場合、報告することもできます。詳しい方法は以下をご参照ください。 https://www.facebook.com/safety/withoutmyconsent/tools。</p> <p>また、利用者はプラットフォーム上で他の人と友達になったり、フォローを解除したりすることができ、Facebook 上のプロフィール、ページ、グループ、投稿、コメント、その他のコンテンツを簡単に報告できるようにしています。また、専門家や利用者からのフィードバックに基づいて、プラットフォーム上での嫌がらせを防止するために、他の業界をリードするツールにも投資しています。これらの機能には以下のようなものがあります。</p> <p>ブロックした人が新しいアカウントを作成したり、自分が管理している別のアカウントから連絡を取ろうとしたりしたときに、友達リクエストやメッセージなどの望まないやり取りを自動で検知し、防止することができます。</p>
-------------------------------	--	---

		<p>送信者をブロックすることなく、Messenger に届いたメッセージを見ることがないように、自動的に受信トレイの外に移動するオプションを提供します。</p> <p>Facebook ページ上の不適切な言葉や任意の単語をブロックするフィルタを提供します。</p> <p>ページやグループの管理者にアカウントを禁止したり削除したりする機能を提供します。</p> <p>誰が Facebook Live にコメントできるかを制御するオプションを提供します。</p>
③	取組の効果分析	<p>四半期ごとにコミュニティ規定施行レポート (CSER) を発行しています。このレポートには、当社のポリシーに違反するコンテンツの防止と対策の詳細が記載されています。</p> <p>前回の CSER (2021 年 2 月) によると、2020 年 10 月から 12 月までの期間に、いじめや嫌がらせのポリシーに違反していた Facebook 上の 630 万件のコンテンツに対して対策を講じました。いじめや嫌がらせはその性質上、非常に個人的なものであるため、多くの場合、私たちが特定したり削除したりする前に、その行動を報告してくれる人が必要です。つまり、テクノロジーを使っていじめや嫌がらせを積極的に検出することは、他の種類の違反行為よりも困難である可能性があるということです。私たちは、この問題に取り組み、コミュニティを保護するために、積極的な検出技術への投資を続けています。</p> <p>また、ヘイトスピーチのコンテンツに関しては、Facebook 上の 2,690 万件のコンテンツに対処しました。そのうち 97.1% は利用者が報告を行う前に検出しました。これは、ヘイトスピーチの文脈に大きく依存する性質を考えると、長年の課題となっていたヘイトスピーチ対策の取り組みを大幅に改善したものです。2 年前には 340 万件のヘイトスピーチコンテンツを削除しましたが、そのうち 60.7% が AI を介して利用者が報告を行う前に識別されました。この大幅な改善は、当社のサービスにおけるヘイトスピーチの削減に向けた投資と進歩の証です。</p>

<https://transparency.facebook.com/community-standards-enforcement>

6. 情報分析・公開

① 誹謗中傷等の流通状況に関する分析・調査

コンテンツの傾向に応じて定期的にポリシーを更新しています。例えば、昨年新型コロナウイルス感染症に関連したコンテンツの増加に対応して、流行病の呼びかけから人々を守るために、いじめや嫌がらせに関するポリシーを更新しました。Facebook社は長い間、暴行の威嚇などを行う投稿を削除してきましたが、7月にFacebook社はまた、そうした投稿を認識したときには、そのアカウントを無効にすることを発表しました。最近行われた他のポリシーの更新については、以下を参照してください。

<https://www.facebook.com/communitystandards/recentupdates/bullying/>

また、当社のポリシーやツールについて、利用者や安全の専門家からのフィードバックを定期的に得ています。例えば、安全諮問委員会による会合を四半期ごとに開いています。また他には、今年はこのコミュニティに役立つツールやリソースについて理解を深めるために、ラウンドテーブルを2回開催しました。ラウンドテーブルの後、International Bullying Prevention Association (IBPA) と連携し、世界中のアクティビストに対するヘイトスピーチ、いじめ、嫌がらせの発生を減らすために活動している個人や組織向けに、いじめ防止のベストプラクティスをより充実させ、周知するためのツールやリソースを開発しました。

②	研究者への情報提供、 利用条件	Facebook 安全諮問委員会や 500 人以上の安全専門家で構成されるグローバル安全ネットワークなど、外部の専門家と協力して、ポリシーや実施の改善に取り組んでいます。
③	透明性レポートの公開 日本語で閲覧可能か	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ました。 https://about.fb.com/ja/news/2020/08/community-standards-enforcement-report-aug-2020/ 。 https://about.fb.com/ja/news/2020/11/community-standards-enforcement-report-nov-2020/
④	意見・補足	