

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	Twitter	
1. サービス概要とポリシー		
①	対象サービス名	Twitter
②	加入者数、月間アクティブユーザー数又は書き込み数	【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】 収益化可能な日時平均アクティブユーザー数（mDAU）：1億9,200万（2020年第4四半期；グローバル）
③	(i) 誹謗中傷等に関して禁止とする情報や行為	<p>【Twitter ルール】</p> <p>Twitterの目的は、公共の場における会話に寄与することです。暴力、嫌がらせ、およびその他の類似行為は、利用者の自己表現を抑制し、その結果、世界中で行われる公共の会話の価値を下げることとなります。Twitterでは、すべての利用者が自由に、安心して公共の会話に参加できるよう、Twitterルール（https://help.twitter.com/ja/rules-and-policies）を設けています。その代表的なものを以下に説明致します。</p> <p>攻撃的な行為/嫌がらせ</p> <ul style="list-style-type: none"> 個人または特定の集団が深刻な損害を被ることを願うこと 不適切な性的誘いかけ 他人を苦しめるまたはおびえさせる目的で侮辱すること 個人または特定の集団にいやがらせをするように他人をそそのかす、または呼びかけること

		<p>暴力</p> <ul style="list-style-type: none"> • 特定の個人または集団に対して、死に至らしめる、または深刻な身体的危害を負わせる意図を表明すること • 個人または集団に向けた暴力をほのめかす脅迫 • 暴力を賛美すること <p>ヘイト行為</p> <ul style="list-style-type: none"> • 人種、民族、出身地、社会的地位、性的指向、性別、性同一性、信仰している宗教、年齢、障害、深刻な疾患を理由とした他者への暴力行為、直接的な攻撃行為、脅迫行為を助長する投稿 <p>センシティブな画像/動画</p> <ul style="list-style-type: none"> • ライブ放送、プロフィール画像、またはヘッダー画像として、過度にグロテスクな、暴力を共有する、または成人向けコンテンツを含む画像/動画を投稿すること <p>合意のない裸体の描写</p> <ul style="list-style-type: none"> • 本人の同意を得ずに撮影または配布された、私的な画像や動画の投稿や共有 <p>個人情報</p> <ul style="list-style-type: none"> • 他の利用者の個人情報を、明確な許可を受けずに公開または投稿すること • 個人情報を公開すると脅迫する行為、または他者にこれを促す行為 <p>なりすまし</p> <ul style="list-style-type: none"> • 混同や誤解を招きかねない形で特定の他人、ブランド、または組織になりすますこと
--	--	---

		<p>児童の性的搾取</p> <ul style="list-style-type: none"> 児童の性的搾取に該当するコンテンツやこれを助長する行為
	(ii) 削除等の対応	<p>Twitter ルールに違反する行為があった場合、強制的対応を行います。強制的対応は、特定のコンテンツ（個人のツイートやダイレクトメッセージなど）またはアカウントに基づいて行われます。いくつかの対応を組み合わせる場合もあります。他にも、特定の国において、権限のある機関から適切な適用範囲を定めた有効な要請を受けた場合に、措置を実施することがあります。Twitter が行う場合のある強制的対応の例を以下に示します。</p> <p>ツイートレベルの対応</p> <ul style="list-style-type: none"> ツイートの表示を制限する ツイートの削除を求める 削除されるまで違反ツイートを非表示にする 公共の利益にかなう場合の例外措置 <p>ダイレクトメッセージレベルの対応</p> <ul style="list-style-type: none"> 報告された違反者と報告者のアカウントの会話を停止する ダイレクトメッセージを非表示にする <p>アカウントレベルの対応</p> <ul style="list-style-type: none"> メディアまたはプロフィールの編集を求める アカウントを読み取り専用モードに切り替える アカウント所有権を認証する 永久凍結する <p>(参考 : https://help.twitter.com/ja/rules-and-policies/enforcement-options)</p>

		<p>【特定の国におけるコンテンツの表示制限】 Twitter はあらゆる場所でサービスを提供できるよう継続的に取り組んでいるため、権限のある機関から適切な内容の有効な請求を受け取った場合は、特定の国において、必要に応じて特定のコンテンツを表示制限することがあります。このような表示制限が実施されるのは、特定の司法権が有効な法的請求を行った場合や、コンテンツが現地法に違反していることが判明した場合に限られます。</p> <p>(参考 : https://help.twitter.com/ja/rules-and-policies/tweet-withheld-by-country)</p>
2. 削除等の対応		
①	<p>誹謗中傷等に関する申告や削除要請の件数 (i) 一般ユーザ、(ii) 司法機関・行政機関</p>	<p>【日本の数値】 (ii) 司法機関・行政機関 (2020 年上半期 ; 誹謗中傷等に限定しない削除請求) 裁判所命令 : 21 その他の法的要請 : 19, 903 (参考 : https://transparency.twitter.com/en/reports/removal-requests.html#2020-jan-jun)</p> <p>【グローバルの数値】 (i) 一般ユーザ (2020 年上半期) 暴力をほのめかす脅迫 : 2, 203, 411 センシティブな内容のメディア : 3, 365, 868 個人情報 : 1, 115, 655 合意のない裸体の描写 : 609, 842 なりすまし : 1, 572, 672 ヘイト 行為 : 6, 055, 642 児童の性的搾取 : 92, 485 攻撃的な行為、嫌がらせ : 5, 138, 663 (参考 : https://transparency.twitter.com/en/reports/rules-enforcement.html#2020-jan-jun)</p>

		<p>(ii) 司法機関・行政機関（2020 年上半期；誹謗中傷等に限定しない削除請求） 裁判所命令：1,894 その他の法的要請：40,326</p> <p>【（質問に答えられない場合）参考となる数値】</p>
②	<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 (i) 一般ユーザ、(ii) 司法機関・行政機関</p>	<p>【日本の数値】</p> <p>(ii) 司法機関・行政機関（2020 年上半期；誹謗中傷等に限定しない削除請求） 表示制限されたアカウント：0 表示制限されたツイート：181 利用規約違反によるコンテンツ削除：7,148</p> <p>【グローバルの数値】</p> <p>(i) 一般ユーザ（2020 年上半期） 暴力をほのめかす脅迫：23,835 センシティブな内容のメディア：167,697 個人情報：25,756 合意のない裸体の描写：9,227 なりすまし：131,136 ヘイト 行為：635,415 児童の性的搾取：444,781 攻撃的な行為、嫌がらせ：398,057</p> <p>(ii) 司法機関・行政機関（2020 年上半期；誹謗中傷等に限定しない削除請求） 表示制限されたアカウント：82 表示制限されたツイート：3,069 利用規約違反によるコンテンツ削除（裁判所命令）：407</p>

		利用規約違反によるコンテンツ削除（その他法的要請）：25,110
		【（質問に答えられない場合）参考となる数値】
削除割合 (i)一般ユーザ、(ii)司法機関・行政機関	【日本の数値】	(ii)司法機関・行政機関（2020年上半期；誹謗中傷等に限定しない削除請求） 裁判所命令：95% その他の法的要請：32%
	【グローバルの数値】	(ii)司法機関・行政機関（2020年上半期；誹謗中傷等に限定しない削除請求） 裁判所命令：19% その他の法的要請：31.7%
		【（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】	
		【（質問に答えられない場合）参考となる数値】
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）	【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】

		<ul style="list-style-type: none"> 2019年4月時点で、強制執行措置の取られた攻撃的なコンテンツの38%は、利用者からの報告ではなく、テクノロジーを駆使しTwitterの専門のチームによって積極的に特定されたものでした。 2020年上半期に児童の性的搾取に関連する違反を理由に凍結されたアカウントのうち、91%が、テクノロジー（PhotoDNA、社内の独自ツールなど）の組み合わせで明るみに出たものでした。
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）	社内の独自ツール、PhotoDNA など
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	【具体的な方法・仕組み】
		【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】
3. 削除要請や苦情に関する受付態勢・プロセス		
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	Twitter ルールまたは利用規約に違反しているツイート、リスト、およびダイレクトメッセージを見つけた場合は、そのツイート、リスト、およびダイレクトメッセージを報告できます。報告対象となるのは、攻撃的もしくは有害なコンテンツ、スパム、なりすまし、著作権または商標権の侵害を含むツイート、リスト、およびダイレクトメッセージです。違反の内容によっては、個別のツイート、リスト、またはプロフィールから直接報告できません（スパム、攻撃的または有害なコンテンツ、不適切な広告、自傷行為、なりすましなど）。

		<p>Twitter では人による調査とテクノロジーを組み合わせる Twitter ルールの徹底に取り組んでいます。</p> <p>Twitter は 24 時間 365 日報告を調査・回答しています。また複数の言語に対応しています。</p> <p>報告を送信すると、報告を受け付けたことを通知する Twitter からの確認メッセージが表示されます。Twitter は報告対象のアカウント、ツイート、リスト、またはダイレクトメッセージを調査します。当該アカウント、ツイート、またはダイレクトメッセージがポリシーに違反していると判断された場合、Twitter は警告やアカウントの永久凍結など、違反の内容に応じた対応を行います。報告したユーザーからより詳しい説明が必要な場合や、Twitter が当該アカウント、ツイート、リスト、またはダイレクトメッセージに対する対応を行う場合は、その旨をユーザーに通知します。</p> <p>報告したツイートの元のコンテンツは、報告済みであることを示すお知らせに置き換えられます。必要に応じ、クリックしてツイートを表示することができます。</p> <p>(参考 : https://help.twitter.com/ja/rules-and-policies/twitter-report-violation、 https://help.twitter.com/ja/safety-and-security/report-a-tweet)</p>
②	申立時における申立者や発信者への通知の有無	有 (日本語)
③	対応決定時における通知の内容、理由の記載の程度	有 (日本語) 対応内容および該当するルール/ポリシーを記載。異議申し立てについても案内。
④		【日本の数値】

	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	【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】
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	
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	違反者は、Twitter の判断が間違っていると思う場合は、異議を申し立てることができます。異議申し立ては、Twitter プラットフォームのインターフェースまたはヘルプセンター（ https://help.twitter.com/forms/general?subtopic=suspended ）から行うことができます。Twitter は異議申し立てを受けても対応が妥当であると判断した場合、アカウントが違反しているポリシーについて説明することにより異議に対応します。
⑦	⑥の苦情・問い合わせ件数	【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】
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、	【グローバルの数値】
	(ii) 裁判外の開示請求件数	【（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 (i) 裁判上の開示件数、(ii) 裁判外の開示件数	【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】
5. その他の取組		
①	普及啓発	<ul style="list-style-type: none"> ヘルプセンター (https://help.twitter.com/ja) からの情報発信 Twitter 公式アカウントやブログ記事による、新機能や既存機能、Twitter の上手な使い方に関する情報提供 UNESCO と合同で作成したメディア情報リテラシーのハンドブック「Twitter を活用した教育と学習」 ソーシャルメディア利用環境整備機構 (SMAJ) による啓発活動への参加 安心ネットづくり促進協議会 (安心協) による啓発活動への参加 悩み相談を行う NPO への無償広告枠提供等を通じた支援、等
②	その他の対策（警告表示・ミュート機能等）	<ul style="list-style-type: none"> ミュート：他のアカウントをミュートすると、そのアカウントのツイートはタイムラインに表示されなくなります。また、見たくない特定のキーワードやハッシュタグを設定することでタイムラインに表示されなくなります。 ブロック：アカウントをブロックすると、そのアカウントとのやり取りを制限できます。かかわりたくないアカウントとのやり取りを避けるために効果的です。 ツイートの公開または非公開設定：ツイートを非公開にすると、ツイートがフォロワーだけに表示され、他の人は見られなくなります。ツイートを非公開にすることで、Twitter をより安全に活用できます。誰かが自分をフォローしようとするたびに、そのリクエストを許可するか拒否するか選ぶことになります。

		<ul style="list-style-type: none"> • 通知フィルター：受け取る通知をフィルタリングすることで、管理をさらに強化できます。クオリティフィルター、ミュート、詳細フィルターを使ってフィルタリングができます。 • セーフサーチ：セーフサーチ機能は、初期設定でセンシティブな内容が含まれている可能性があるツイートや、ブロックまたはミュートされているアカウントからのツイートを検索結果のページから除外します。 • センシティブなメディア：Twitterの初期設定では、センシティブな内容が含まれている可能性があるメディアには警告文が表示されます。こちらは設定で調整することができます。 • 自分のツイートに返信ができる範囲を選択できる：ツイートする前に、1) 全員が返信（Twitterでのデフォルト設定）、2) フォローしている人だけが返信、または、3) @アカウントで指定した人だけが返信、といった3つの選択肢から選べます。
③	取組の効果分析	<ul style="list-style-type: none"> • フォローしていないアカウントからの攻撃的な行為による報告は16%減少しました。 • 新たにアプリ内で異議申し立てが可能になったことにより、対応時間がこれまでより60%減少しました。 • 報告を受領後24時間以内に凍結した悪質なアカウントの件数が、前年同期比で約3倍に増加しました。 • 簡単で新しくなった報告プロセスによって、2.5倍以上の個人情報を削除しました。（2019年4月時点）
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	
②	研究者への情報提供、利用条件	無償の学術研究プロダクトトラックを提供しています。

		(参考 : https://blog.twitter.com/developer/ja_jp/topics/tools/2021/enabling-the-future-of-academic-research-with-the-twitter-api.html)
③	透明性レポートの公開 日本語で閲覧可能か	<ul style="list-style-type: none"> 2012年7月以来、年2回更新される「Twitter 透明性センター」(https://transparency.twitter.com/ja.html)を通じて、Twitter に寄せられる法的請求、知的財産に関する異議申し立て、Twitter ルールに基づく強制的対応、プラットフォームの操作の状況等を公表しています。 日本語訳も提供しています。最新版(2020年上半期)は2021年2月現在翻訳作業中。
④	意見・補足	<ul style="list-style-type: none"> ルールに違反するコンテンツ、特に個人情報、脅迫、その他の攻撃的なツイートを行う利用者について、他の利用者から報告されるより先に特定できるよう、Twitter の技術面を引き続き引き続き改良していきます。 公共の会話の場である Twitter の健全性と高め、利用者の皆様に安心してご利用いただくことが私たちの最優先事項です。