

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）回答における主な項目の比較

※総務省において要約したもの。詳細は各社の回答原本を参照。

参考資料4

社名		ヤフー株式会社	Facebook Japan株式会社		Google合同会社	LINE株式会社	Twitter Japan株式会社	
1-①	対象サービス名	Yahoo!知恵袋、Yahoo!ニュース(コメント欄のみ)、Yahoo!ファイナンス	Facebook	Instagram	YouTube, Google 検索, Google マップ	LINE タイムライン、LINE Openchat、LINE LIVE、LINE BLOG、livedoor Blog	Twitter	
1-②	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】	2020年度の任意の1か月における書き込み総数 ■知恵袋 投稿数:約440万件 ■ヤフーニュース 総投稿数:約1,300万件 ■ファイナンス掲示板 総投稿数:約210万件	月間アクティブアカウント数:2,600万人(2019年7月)	月間アクティブアカウント数:3,300万(2019年3月)	回答なし	国内MAU 8,600万人(2020年9月末時点)	回答なし
		【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)	月間アクティブアカウント数:10億(2018年6月)	YouTube: 毎月20億人以上のログインユーザー。毎日10億時間以上の動画視聴 Google 検索: 毎年何兆回もの検索回数 Google マップ: 10 億人以上	主要4か国(日本・台湾・タイ・インドネシア)MAU 1億6,700万人(2020年9月末時点)	回答なし
		【(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年第四半期:グローバル)
2-①	誹謗中傷等に関する申告や削除要請の件数 (i)一般ユーザー、(ii)司法機関・行政機関	【日本の数値】	・各サービスにおいて、特定の一か月の数値を算出。 ・また、サービスごとに利用している申告フォームの仕様や運用等が異なるため、集計できる数値に差異がある。 (i)(ii)合計値: ■知恵袋 「違反連絡リスト>不適切な投稿」で申告された案件数:約6.6万件 ■ヤフーニュース 「不快」を理由とする違反申告数:不明(取得不可) ※総違反申告数:約16万件 ■ファイナンス掲示板 不快(アダルト・グロ・罵詈雑言)での違反申告数:約1.6万件	サービス全ての合計、誹謗中傷以外を含む (i)一般ユーザー 非公開 (ii)司法機関・行政機関 86件(2019年6月~2020年5月) (行政機関および法執行機関からの情報開示請求)	(i)一般ユーザー [検索]: 名誉毀損問題の法的理由のためにコンテンツを報告するためにGoogle検索ウェブフォームを通じて提出されたURL:約7000件(2019年) (ii)司法機関・行政機関 2020年1月から6月までに日本政府(政府機関及び裁判所命令)から受けた名誉毀損に関する依頼は57件(アイテム数595件)。政府から受けた依頼全体の85%。	(i)一般ユーザー [タイムライン] 919件、[Openchat] 2,115件、[LIVE] 312件、[BLOG] 27件、[livedoor Blog] 2,479件 ※タイムライン/Openchat/LIVEについては、迷惑行為全般に関する件数、BLOG/livedoor Blogについては、プライバシー権・名誉権侵害に関する件数。 (ii)司法機関・行政機関 [タイムライン] [Openchat] [LIVE] [BLOG] 0件 [livedoor Blog] 19件	(i)一般ユーザー 回答なし (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	
		【グローバルの数値】					(i)一般ユーザー(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 (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	
2-②	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)一般ユーザー、(ii)司法機関・行政機関	【日本の数値】	(i)(ii)合計値: ■知恵袋 「違反連絡リスト>不適切な投稿」で申告された案件の削除数:約1.9万件 ■ヤフーニュース 違反申告されて「不快」を理由に削除された件数:約1.8万件 ※誹謗中傷等に関する申告件数が取得できないことから、削除された理由が「不快となっている」ものについて集計。 ■ファイナンス掲示板 違反申告されて「不快(アダルト・グロ・罵詈雑言)」で削除された件数:約7千件	(誹謗中傷以外を含む) (i)一般ユーザー 0件(2020年1月~6月) (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) (ii)司法機関・行政機関 0件(2020年1月~6月) (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) 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)	(i)一般ユーザー [Youtube]:コミュニティガイドライン違反(ヘイトスピーチ、嫌がらせやネットいじめ以外も含む)で削除された動画の数:49,180本(2020年10月から12月まで) ※ユーザーからの報告や機械によってフラグされたものを含む(参考)グローバルでは、削除された約1,000万本の動画のうち、ハラスメントやネットいじめが原因で削除された動画はそのうちの約0.8%。 [検索]:法律に基づく削除に関する問題を報告するためのウェブフォームを通じて一般ユーザー(もしくはその法的代理人)により名誉毀損を報告された推定7000件のURLのうち、法的理由で約1700件のURLを検索結果から非表示化(2019年)。非表示化した報告:約24%。非表示化しなかった報告:約55%。 Googleから報告者に追加情報の提供を求める必要があった報告:約20%。 ※Googleが追加情報を要求する理由としては、権利者が明確に特定されていない可能性がある、要求の法的根拠が明確に特定されていない可能性がある、などが挙げられる。 (ii)司法機関・行政機関 449件を削除(google全体、名誉毀損以外の請求などに関するものも含む。2020年1月から6月)	(i)一般ユーザー [タイムライン] 8件、[Openchat] 22件、[LIVE] 26件、[BLOG] 24件、[livedoor Blog] 約1,600件 (ii)司法機関・行政機関 [タイムライン] [Openchat] [LIVE] [BLOG] なし [livedoor Blog] 14件	(i)一般ユーザー 回答なし (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	
		【グローバルの数値】	(i) 630万件(2020年10月~12月) (コミュニティ規定に違反したコンテンツの削除件数)	(i)500万件(2020年10月~12月) (コミュニティガイドラインに違反したコンテンツの削除件数)			(i)一般ユーザー(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 (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 利用規約違反によるコンテンツ削除(その他の法的要請):25,110	

社名	ヤフー株式会社	Facebook Japan株式会社	Google合同会社	LINE株式会社	Twitter Japan株式会社
2-② 削除割合 (i)一般ユーザ、(ii)司法機関・行政機関	【日本の数値】 (i)(ii)合計値: ■知恵袋 誹謗中傷等に関する申告数における削除率: 29.3% ■ヤフーニュース 総違反申告数に対する全ての申告削除率: 17.5% ※誹謗中傷に限定されない ■ファイナンス掲示板 不快を理由とする違反申告数における削除率: 44.5%	回答なし	回答なし	回答なし	(i)一般ユーザ 回答なし (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】	回答なし	回答なし	回答なし	(i)一般ユーザ 回答なし (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】	回答なし	回答なし	(参考回答) リクエストに応じずコンテンツを削除しない理由は状況によって様々。政府やユーザーが求める削除対象が明確でない(たとえばリクエストにURLが記載されていない)場合や、コンテンツがすでに作成者によって削除されている場合などは、リクエストに応じることはできない。上記情報で一部は削除割合を示しているが、削除割合にとらわれることは全体像を理解する上で必ずしも得策ではないことに留意が必要。	
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】 回答なし	回答なし	回答なし	回答なし	回答なし
	【グローバルの数値】 回答なし	回答なし	回答なし	回答なし	回答なし
	【(質問に答えられない場合)参考となる数値】 申告窓口やバトロールの手法(人かAIか)、事象の軽重等により異なるため、一概に回答することは難しいが、AIや機械学習を用いた監視体制であれば投稿後、速やかに削除可否の判断が可能である。 また、人によるバトロールであっても、権利侵害性に疑義がない明確な投稿であれば一箇日中に判断できるものもあれば、判断に悩ましい投稿については、一担当者のみの判断によらず、複数人での協議や、法的判断を行っている専門部署へのエスカレーション、投稿者への意見照会などを経て最終的な判断がなされるため、相当の期間をもって慎重に判断がなされる。	【表示頻度(prevalence)】 ※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。 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。	【平均所要時間】に基づく評価は、いくつかの理由から問題があると考えている。①プラットフォームが展開した施策の有効性を評価するための最も効果的な指標ではない可能性がある。②プラットフォームに対して、迅速かつ慎重に作業するよりも、対処にかかる時間を早めることを奨励することとなる可能性がある。③平均的な所要時間を測定することは、例えば平均所要時間の中央値のようなデータと比較するなど、不用意にデータを歪めることを誘導しかねない。	(i)一般ユーザ 数日~10日前後 (ii)司法機関・行政機関 1週間程度~数ヶ月前後	回答なし
2-③ ②以外の誹謗中傷等に関する削除件数(AIを用いた自動検知機能の活用等)	【日本の数値】 ・各サービスにおいて、特定の一か月間のバトロール数値を算出。 ・また、サービスごとに利用しているバトロールツールの仕様、運用、基準等が異なるため、集計できる数値に差異がある。 ・効率的なバトロールとなるよう、重点的に対応すべきバトロール対象、運用方法については柔軟に見直しを実施。本件数値については、集計対象期間によっては必ずしも同じ結果となることを示すものではない。 ■数値 ■知恵袋 約3万2千件 ※違反申告を除く、低品質対応・積極巡回による不快削除件数 ■ヤフーニュース 約5万8千件 違反申告以外で不快で削除された件数 AIによる投稿削除数: 約25万件 ■ファイナンス掲示板 約5万4千件 ※違反申告以外での不快(アダルト・グロ・罵詈雑言などで)削除された件数	回答なし	回答なし	回答なし	(i)タイムライン] 78,962件、[Openchat] 78,770件、[LIVE] 3,780件、[BLOG] 6,358件 ※自動検知したうち一定の要件に合致した件数。
	【グローバルの数値】	48.8%(2020年10月~12月) 「How much did we find before users reported it」(事前対応率)	80.0%(2020年10月~12月) 「How much did we find before users reported it」(事前対応率)	930万本以上の動画をコミュニティガイドライン違反を理由に削除(2020年10月から12月)。これらの動画の94%に当たる8,800,082件は、人間ではなく機械によって最初にフラグ化。人工知能によって検出された動画のうち、35.9%は一度も視聴をされる前に、35.9%は1~10回以下の視聴回数のうちに削除。9億件以上のコメントを削除(2020年10月から12月)。削除されたコメントの99.6%が自動的に検出。	回答なし
	【(質問に答えられない場合)参考となる数値】	回答なし	回答なし		・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。 ・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)の組み合わせで明らかにされたもの。

社名		ヤフー株式会社	Facebook Japan株式会社	Google合同会社	LINE株式会社	Twitter Japan株式会社	
2-⑤	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】	不正な申告や不正な削除要請への対応は、ハトール部隊や専門部署により人の目により確認が行われているため、不正か否かにかかわらず、一件一件、内容を確認している。	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ない。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る。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しる。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る。	規定に違反するコンテンツについて、削除する前に多くの人に視聴されない、または一切視聴されることがない方が望ましいため、YouTubeを含む様々なプロダクトで、ポリシーに違反する可能性のあるコンテンツの検出を支援する自動化されたシステムを導入している。 またYouTubeは、コンテンツを迅速に削除するための最善の方法は、新たな性質の問題が大きくなる前に予測することと考えている。YouTubeのIntelligent Desk チームでは、ニュース、ソーシャルメディア、ユーザーの報告を監視して不適切なコンテンツに関する新しい傾向を検出し、大きな問題に発展する前に対処できるようチームの体制を整えている。	回答なし	
		【日本の数値】	何をもち「不正な申告や削除要請」であると判断するのが難しく、数値は計測していない。	回答なし	回答なし	[livedoor Blog] 857件 ※主に違法性がないと思われる投稿等への差別的な申告の件数。	回答なし
		【グローバルの数値】	回答なし	回答なし	回答なし	回答なし	回答なし
		【(質問に答えられない場合)参考となる数値】	回答なし	回答なし	回答なし	回答なし	回答なし
3-③	対応決定時における通知の内容、理由の記載の程度	問い合わせ方法により異なるが、プロバイダ責任制限法に基づく削除請求を受領した場合には、対応決定後、申立て者に対し回答書の発送を行う。回答書面においては、一般的な理由の記載はあるが、理由の詳細については回答していない。※個々のガイドラインに抵触する旨の説明は容易であるが、ガイドラインに抵触しないことを説明するのは非常に難しい。	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している。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る(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可能。 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。その利用者は、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く。 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ひすることがある。 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。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る。	削除要請に関するお知らせをユーザーにする方法は様々である。例えば、名誉毀損を理由に検索結果からあるURLの削除を依頼したユーザーには、受領確認としてケース番号が記載されたメールが送信される。ユーザーが削除リクエストの際にURLの管理者に連絡を試みたかどうかの確認が取れなかった場合、メールには管理者への連絡が推奨される理由や、キャッシュされたURLのコピーの削除を迅速に行う方法、特定のGoogleポリシーの削除フォームに関する情報も含まれる。削除するかどうかの決定の結果についても、メールで伝えられる。濫用を防ぐために結果に至った理由には詳細には伝えられないが、Googleとしては削除できない場合であってもその他取り得る手段などを伝えている。	削除可否等について通知。	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。	
		削除時に投稿者に対し通知がなされるか否かは、サービスにより異なる。 例えば、知恵袋においては、削除された質問・回答については、削除がなされてから2週間、My知恵袋のページにおいて削除理由を確認することができる。	回答なし	回答なし	回答なし	回答なし	回答なし
3-④	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	【日本の数値】	全サービス共通の数値ではあるが、ハトール部隊には数百名在籍し、24時間365日、シフトを組んで対応にあっている。プロバイダ責任制限法に基づく法的な判断をする部署については、7.8名程度で対応。そのほか、社外弁護士複数名に諮問して対応にあっている。	回答なし	回答なし	各サービスにつき、10名程度	回答なし
		【グローバルの数値】	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存在。過去数年の間に、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る。 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。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可能。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る。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る。	当社のチームは、製品の専門家、エンジニア、弁護士、データサイエンティスト、元法執行機関職員など、多様な専門性を有している。彼らは、専門家のネットワークと密に連携をとりながら仕事をしている。 現在ではGoogle全体で10,000人を超えるスタッフが、1日24時間、週7日、1年365日体制で、Googleのプラットフォーム全体でコンテンツの確認と削除に取り組んでいる。	回答なし	回答なし	
3-⑤	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	【日本の数値】	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る。 いくつかのレビューサイトの所在地は公表しているが、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ない。 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ある。最も重要なのは、言語と文化的背景。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る。	具体的人数などは非公開。 Googleでは、グローバルポリシーとコミュニティガイドラインに基づき、各国の文化的背景やトレンドを理解した上で実施。例えばYouTubeでは各国のコンテキストやトレンドも把握しながら対策を行っており、日本を含めた各国のコンテキストを理解できるスタッフも存在し、適切な体制(日本語のネイティブのレビュアー等)を整えている。	日本国内に拠点が有り、日本における責任者を置いている。	回答なし	
		【グローバルの数値】	回答なし	回答なし	回答なし	回答なし	
3-⑥	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	【(質問に答えられない場合)参考となる数値】	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。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る。	Googleのサービス全体で、Googleの利用規約やサービス固有の追加規約またはポリシーに違反していると思われる、または違反していると報告されたものに対して措置が取られることがある。ユーザーは再審査請求をすることで、その決定に不服を申し立てることが可能。 例えば、YouTubeでは再審査請求を認めている。動画が削除されたクリエイターには通知が発行され、再審査を請求するための手順が記載されたリンクも提供される。クリエイターが再審査請求を提出すると、その請求は審査担当者に送られたのち、判定が維持または破棄される。この再審査は、動画の削除について最初の判断を行った審査担当者とは異なる担当者によって行われる。クリエイターには、結果を知らせるフォローアップメールが送られる。 また例えば、Googleマップにおいて、自分の投稿の多くが削除された場合や、ローカルガイドプログラムから削除された場合は、このフォームから再審査請求することが可能。	[タイムライン][Openchat][LIVE] ユーザーからの異議申し立てがあった場合は、その内容に応じて適宜対応。 [BLOG] [livedoor Blog] モニタリングの削除基準に該当しないものについては、プロバイダ責任制限法に則って対応。	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可能。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から行うことが可能。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。	
		削除等への苦情や問い合わせについては、ご意見・ご要望フォームを通してCS窓口にて受付を行い、回答に当たり法的な判断が必要なものであれば、専門部署へエスカレーションが行われ、回答指示を行う。 今後は、後述の社内労働者会議において、AIによる対策の拡大などが進んだ場合における誤削除されたユーザーの救済が重要である旨の提言がなされたことを踏まえ、削除されたユーザーからの専用窓口を設置するとともに、削除理由の開示フローなどについて検討を進める。	回答なし	回答なし	回答なし	回答なし	回答なし
4-①	発信者情報開示請求を受けた件数 (i)裁判上の開示請求件数、(ii)裁判外の開示請求件数	【日本の数値】	(i)裁判上の開示請求件数: 30件 (ii)裁判外の開示請求件数: 55件	回答なし	回答なし	(i)裁判上の開示請求件数 6件(Openchat 1件、livedoor Blog 5件) (ii)裁判外の開示請求件数 27件(タイムライン 1件、Openchat 1件、BLOG 1件、livedoor Blog 24件)	回答なし
		【グローバルの数値】	回答なし	回答なし	回答なし	回答なし	回答なし
		【(質問に答えられない場合)参考となる数値】	回答なし	回答なし	回答なし	回答なし	回答なし
4-②	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】	(i)裁判上の開示件数: 13件(うち一部開示が4件) (ii)裁判外の開示件数: 5件(うち一部開示が4件)	回答なし	回答なし	(i)裁判上の開示件数 5件(Openchat 1件、livedoor Blog 4件) (ii)裁判外の開示件数 6件(BLOG 1件、livedoor Blog 5件)	回答なし
		【グローバルの数値】	回答なし	回答なし	回答なし	回答なし	回答なし
		【(質問に答えられない場合)参考となる数値】	回答なし	回答なし	回答なし	刑事(民事は含まない)87%	回答なし

社名	ヤフー株式会社	Facebook Japan株式会社	Google合同会社	LINE株式会社	Twitter Japan株式会社	
5-② その他の対策(警告表示・ミュート機能等)	<p>■知恵袋 ・不適切投稿の非表示ルールについて「利用規約」や「利用のルール」に書かれている禁止事項に該当する回答は、専門的なパトロールチームによる投稿の削除のみならず、アダルトや低品質な投稿については、不適切な投稿として機械的に判定し、質問詳細ページで非表示となる仕様が存する。</p> <p>また、当該投稿については、知恵袋Topページの新着一覧やランキングにおける非掲載も実施。</p> <p>・いたずらや中傷などの不快な投稿を繰り返す利用者をブラックリストに登録可能。</p> <p>■ニュースコメント 特定のユーザーのコメントを閲覧したくない場合には、ユーザー単位で非表示に設定することが可能。また、記事を読む際、コメント自体を表示させたくない場合には、コメント欄を非表示に設定することも可能。</p> <p>■フィナンス掲示板 不快な投稿を繰り返すなど、特定の利用者の投稿を見たくない場合は無視リストに登録すると、非表示にすることが可能。</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。 (2)コメントの非表示・削除機能。 (3)利用者をブロックする機能。 (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、報告が可能。</p> <p>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可能。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。</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、防止することが可能。</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。</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。</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。</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。</p>	<p>一度に交流したくない複数の人をブロックしたり、ポジティブなコメントをトップにピン留めしてポジティブな交流を促したり、タグ付けや投稿へのコメント、メッセージを許可する人を選択可能。</p> <p>「制限」機能により、あなたのアカウントを不要なやりとりから保護し、フォロワーがこれらを見ないように保護することが可能。</p> <p>特定の人をブロックしたり、特定の人からのコメントをブロックすることが可能。</p>	<p>例えば YouTube 上では、コミュニティガイドライン違反に近いものの実際には違反していないため、削除されない場合がある(=「ボーターラインのコンテンツ」)。こうしたコンテンツは、米国のYouTubeで視聴されているコンテンツの1%にも満たない数。YouTube のおすすめの機能において、ボーターライン上のコンテンツと有害な誤情報に関しても、ユーザーの目に触れる機会をできるだけ減らす取り組みを実施。</p>	<p>・LINEでは友だちをブロックすることができるため、誹謗中傷を投げかける友だちがいなくなった場合、それ以降のメッセージを受け取らない設定が可能。</p> <p>・見知らぬ人との出会いが思いがけぬトラブルに発展することがあることから、18歳未満はLINE ID検索ができないよう、機能制限が設けられている。</p> <p>・LINEでは、文部科学省や長野県、大津市などといった行政府と協力していじめを始めとする子ども向けの悩み相談窓口を用意。</p> <p>・誹謗中傷やデマなどの有害コンテンツ有害コンテンツへの対応方針を発表し、パブリックコメントを募集(2020年8月3日～同月31日)し、その結果を公表。</p> <p>・一般財団法人全国SNSカウンセリング協議会、一般社団法人全国心理業連合会と連携し、インターネット上の誹謗中傷による精神的ダメージを受けた方を対象とした無料のLINE相談窓口「SNS誹謗中傷等の心のケア相談」を開設。</p> <p>・ガイドライン違反の可能性があるテキストを投稿する前に、ユーザに対して警告を通知する機能の開発を進めている。2021年上半年にタイムラインに実装し、2021年下半年にOpenchat等の他サービスへ実装する予定。</p>	<p>・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る。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る。</p> <p>・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可能。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。</p> <p>・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る。ツイートを非公開にすることで、Twitterをより安全に活用可能。</p> <p>・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、管理をさらに強化可能。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可能。</p> <p>・セーフサーチ: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。</p> <p>・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。設定で調整することが可能。</p> <p>・自分のツイートに返信ができる範囲を選択: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つの選択肢から選択可能。</p>
5-③ 取組の効果分析	<p>各種取り組みの効果分析を行いながら対策の改善を検討している。</p> <p>外部へ公開している事例については、以下のリンクを参照。</p> <p>■ニュースコメント Yahoo!ニュースにおけるAIを活用した投稿時注意メッセージの掲出効果について (参照)https://about.yahoo.co.jp/pr/release/2021/02/16a/ (以下、抜粋)</p> <p>不適切なコメントを繰り返し投稿するアカウントが13.5%減少 ヤフー株式会社提供日本最大級のインターネットニュース配信サービス「Yahoo!ニュース」が、2020年7月からAIを活用して取り組んでいる投稿時の注意メッセージ掲出の効果により、「Yahoo!ニュース コメント」(以下、コメント欄)に不適切なコメントを繰り返し投稿しているアカウントが、掲出直後と比較して13.5%減少しました。</p>	<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、ポリシーに違反するコンテンツの防止と対策の詳細を記載。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た。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た。そのうち97.1%は利用者が報告を行う前に検出。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。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た。</p>	<p>四半期ごとにコミュニティ規定施行レポートを発行し、ポリシーに違反するコンテンツの防止と対策の詳細を記載。前回のCSER(2020年11月)によると、2020年10月から12月までの期間に、いじめや嫌がらせに関するポリシーに違反したInstagram上のコンテンツ500万件に対して対策を講じた。ヘイトスピーチのコンテンツに関しては、660万件に対して対策を講じた。そのうち95.1%は利用者が報告を行う前に検出。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。2年前には65万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44.6%がAIを介してユーザーが報告を行う前に識別された。</p>	<p>YouTube上のボーターライン上のコンテンツなどが、登録をしていないチャンネルからおすすめて視聴する動画の再生時間を70%削減(米国)。</p> <p>2017年6月にYouTube上の過激なコンテンツを検出するために機械学習技術を導入。より正確に悪質なコンテンツを分類できる機械を訓練するために、200万本以上のコンテンツをレビュー。自動化されたプログラムの品質と有効性を継続的に向上させるための取り組みを下記の通り進めてきた。</p> <p>● 2017年第1四半期には、暴力的過激主義ポリシーに違反して削除された動画の8%は、削除時の再生回数が10回未満。</p> <p>● 2018年第1四半期には、この数字は50%以上に達した。</p> <p>● 2019年第4四半期には、暴力的過激主義ポリシーに違反したとして削除されたアップロードされた動画の約90%が、再生回数が10回に満たないうちに削除された。</p> <p>● 2020年第4四半期には、ハラスメントや嫌がらせなどのその他のコミュニティガイドライン違反も含む全体でYouTubeは9億件以上のコメントを削除した。削除されたコメントの99.6%が自動的に検出された。</p>	<p>・長野県 いじめ・自殺相談では、2017年9月10日から23日だけで、長野県内の中高校生390人から計547件の相談が寄せられ、前年度1年間の電話相談259件を大きく上回った。</p> <p>・文部科学省「SNS等を活用した相談体制の構築事業」では、都道府県19、指定都市8、市町村3の合計30自治体で、11,039件の相談が寄せられた。</p> <p>・「SNS誹謗中傷等の心のケア相談」では、告知を開始した2020年8月4日から相談終了の同月19日までの14日間で、164件の相談があった。</p>	<p>・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。</p> <p>・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。</p> <p>・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。</p> <p>・簡単に新しくなった報告プロセスによって、2.5倍以上の個人情報削除。(2019年4月時点)</p>
6-③ 透明性レポートの公開	<p>有識者会議における提言書に基づき、2020年度の透明性レポートを本年夏頃を目処に作成予定。</p>	<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。 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/</p>	<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。 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/</p>	<p>日本語で公開 https://transparencyreport.google.com/?hl=ja</p>	<p>LINE Transparency Report (https://linecorp.com/ja/security/transparency/top)にて以下を公開: ・違反投稿への対応:当社におけるコンテンツモニタリングの仕組みと対応実績のレポート(全世界対象) ・ユーザからの削除申告への対応:名誉毀損やプライバシー侵害について当社が受領した削除申告の対応実績レポート(日本のみ対象) なお、本レポートは日本語で公開。</p>	<p>・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。 ・日本語訳も提供。最新版(2020年上半年)は2021年2月現在翻訳作業中。</p>